

Lechene, Valérie; Pendakur, Krishna; Wolf, Alex

**Working Paper**

## OLS estimation of the intra-household distribution of expenditure

IFS Working Papers, No. W20/06

**Provided in Cooperation with:**

The Institute for Fiscal Studies (IFS), London

*Suggested Citation:* Lechene, Valérie; Pendakur, Krishna; Wolf, Alex (2020) : OLS estimation of the intra-household distribution of expenditure, IFS Working Papers, No. W20/06, Institute for Fiscal Studies (IFS), London,  
<https://doi.org/10.1920/wp.ifs.2020.620>

This Version is available at:

<https://hdl.handle.net/10419/223282>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# OLS Estimation of the Intra-Household Distribution of Expenditure

IFS Working Paper W20/06

Valérie Lechene  
Krishna Pendakur  
Alex Wolf

# OLS Estimation of the Intra-Household Distribution of Expenditure

Valérie Lechene, Krishna Pendakur, Alex Wolf\*

February 28, 2020

## Abstract

Individuals may be poor even if their household is not poor, because the intra-household distribution of resources may be unequal. Dunbar, Lewbel and Pendakur (2013) develop a model wherein the *resource share* of each person in a collective household---defined as their share of total household expenditure---may be estimated via nonlinear estimation of Engel curves, which in practice can be difficult. In this paper, we provide a linear representation of their model, so that resource shares are revealed by estimated coefficients from OLS regressions using off-the-shelf consumer expenditure micro-data. We also provide a simple linear pre-test to check for model identification.

We apply the model to data from 12 countries, and investigate resource shares, gender gaps in expenditure, and individual poverty. We find that equal sharing--the implicit assumption underlying household-level poverty calculations---is rejected. We also find evidence of large gender gaps in resource shares, and consequently in poverty rates, in a few countries.

## 1 Introduction

Most empirical measurement of poverty is done at the household level, in the sense that if a household has income or consumption below a threshold, we call all its members poor. But, this is a matter of convenience and data availability more than a matter of principle. There

---

\*The authors acknowledge the financial support of the following agencies: the Social Sciences and Humanities Research Council of Canada (Pendakur); Economic and Social Research Council (Lechene); the World Bank. We also thank our friends and colleagues who gave us feedback on this work: Orazio Attanasio, Andrew Chesher, Geoffrey Dunbar, Dave Freeman, Arthur Lewbel, Kevin Milligan and Kevin Schnepel.

are many aspects of well-being that depend critically on the intra-household distribution of income or consumption. For example, The Millenium Development Goals include the promotion of gender equality and the empowerment of women. Some of this goal has to do with women’s access to resources within households, and achieving this goal requires measuring the consumption of women, as opposed to the consumption of households.

Many important questions in labour, public and development economics hinge on the intra-household distribution. For example, it is well-understood that poverty in childhood has negative long-term consequences (see, e.g.: Heckman 2006; Campbell et al 2014), but much of this literature is underpinned by the assumption that if a child is in a low-income household, then they have low consumption. But, what if parents devote a greater fraction of resources to their children than to themselves? Questions regarding the measurement of poverty, its persistence and correlation with long-term outcomes cannot be answered properly if the measurements we use preclude within-household inequality.

We seek to measure how household expenditure is divided across the people inside households. Conceptually, this is a simple exercise: measure the expenditure of each person in a household, and compare them. But, this goal is frustrated by a lack of data on expenditure at the individual level. For instance, we may observe in the data that the household bought a bottle of milk, but we may not observe who drank it. Furthermore, there are goods with different degrees of shareability inside households, such as a common dwelling, and ascribing a value to the services from the use of these goods to each individual is not straightforward. Our model accomodates both the fact that people may have unequal access to resources and that some goods may be shared (to unknown degrees).

A useful tool to describe the within-household distribution of expenditure is the *resource share*, defined as the fraction of total household expenditure enjoyed by each member. If, in a given household, the women has a smaller resource share than the men, then there is gender inequality in expenditure. Further, in that case, there is the possibility that although the household may have enough resources to keep all members out of poverty, the woman may have such poor access to resources that she is nonetheless poor. The standard World Bank poverty measurement strategy assigns each household member their per-capita share of household expenditure, and compares that to US\$1.90 per day. A strategy that respects the idea that consumption lives at the individual level would instead assign each person their resource share of household expenditure and compare that to US\$1.90 per day.

These methods may help us understand other phenomena. For example, Calvi (2019) estimates resource shares, and poverty at the individual level, in India and finds that women—especially older women— have lower resource shares than men. This then implies that older Indian women have much higher poverty rates than previously thought. Calvi shows that

this higher poverty rate (driven by lower resource shares) among older women can explain the finding of Anderson and Ray (2010) that Indian women over the age of 45 have higher mortality rates than do Indian men (a phenomenon they call “missing women”).

Chiappori (1992) develops a model of quantity demand decisions by *collective households*, defined as households comprised of individual people who maximize utilities. Using this general framework, Browning, Chiappori and Lewbel (2013) and Dunbar, Lewbel and Pendakur (2013, hereafter DLP) introduce structural models that allow us to use off-the-shelf data, of the sort collected routinely by statistical agencies and the World Bank, to reveal the resource shares of individual household members. They use either observed price variation or observed data on *assignable goods*, defined as goods consumed by a single known person in the household (e.g., men’s clothing), to identify resources shares. But they require the use of complex nonlinear models that can be computationally difficult to estimate. The core of the computational difficulty lies in the fact that resource shares must be between 0 and 1, and they enter the model nonlinearly, implying that bounded nonlinear estimation is required.

In this paper, we provide a linear reframing of the structural model of DLP yielding a theory-consistent and easy-to-implement model, requiring only the estimation of linear Engel curves for assignable goods. (An “Engel curve” relates the fraction of total household expenditure spent on a good to total household expenditure.) In this model, the *levels* of Engel curves pick up a mixture of differences across household members’ resource shares and differences across household members’ preferences. But, the *slopes* of Engel curves with respect to the household budget are driven only by differences across household members’ resource shares. So, resource shares are identified by the slopes of Engel curves: if the slope of a person’s assignable good Engel curve is twice as large (in absolute value) than another person’s, then their resource share is twice as large.

Empirically, we run OLS regressions of observed household-level spending on assignable goods (as a fraction of total household expenditure) on observed household demographics, log total household expenditure and their interactions. Then, we compute estimated resource shares as functions of estimated regression coefficients.

We also extend the model of DLP to allow for complex household types, including those with multiple adult men and/or women and single parent households. Calvi (2019) and Dunbar, Lewbel and Pendakur (2019) allow for households with multiple men and women in their nonlinear models. But, our linear reformulation of DLP also makes it very easy to incorporate households with adults of just one gender (e.g., single-parent households) into the mix, and, to our knowledge, ours is the first paper to do so.

Finally, we provide a pre-test base on OLS regression that indicates whether the model is identified. Essentially, our pre-test is: do Engel curves respond to total household ex-

penditure? Because resource shares are identified via the relative slopes of Engel curves, if the slopes are zero, the model is not identified. Previous tests of this identifying restriction required nonlinear estimation methods; our linear recasting of DLP delivers this OLS-based pre-test.

We use the model to estimate resource shares and individual poverty rates (including women’s poverty and children’s poverty) with data from 12 countries, using household surveys from the World Bank LSMS data, and 1 other national survey from Bangladesh. We use person-level clothing expenditure as the assignable good. Clothing Engel curves pass the pre-test for 5 of 12 countries, so we estimate resource shares and person-level poverty for these countries.

We find that equal sharing—the implicit assumption underlying standard household-level poverty calculations—is falsified by the data, and that there are significant gender gaps in resource shares and poverty rates in some countries. For example, we find estimated women’s resource shares to be 5 and 8 percentage points lower than men’s in Bangladesh and Malawi, respectively. This results in women’s poverty rates that are 4 and 12 percentage points higher than men’s in Bangladesh and Malawi, respectively.

Our data from Bangladesh have both person-level clothing expenditure and person-level food expenditure (including implicit expenditure on home-produced food). We find that using food data to identify resource shares delivers estimates that are very similar to those generated from clothing data (this echoes the findings of Bargain, Lacroix and Tiberti 2018). Since food expenditure is larger than (and possibly better measured than) clothing expenditure, this suggests that statistical agencies and the World Bank should focus data gathering efforts on the collection of person-level food consumption. This view is strengthened by the fact that food Engel curves are broadly known to be downward-sloping and by the fact that we find statistically significantly sloped clothing Engel curves for only half of our countries.

Our work suggests that data collection efforts aimed at learning about poverty should collect data on household expenditure and consumption (as is already typically done), should additionally collect data on person-level expenditure on at least 1 assignable good, and should put effort into collecting assignable food consumption data. This applies to sample surveys like the LSMS and to field experiments where the object of interest is poverty or consumption.

In section (2), we review the theoretical foundations of our work, and discuss identification of resource shares in DLP’s nonlinear model. We then show how to re-express the model linearly, and how to recover resource shares using OLS regression. In section (3), we present the data, and in section (4), estimated resource shares, gender gaps and poverty rates. We finish with a brief discussion of the implications of our work.

## 2 Theory

Consider for the moment a world where all goods are non-shareable (we will come to the case where some goods are shareable shortly). Let *expenditure* on a good be the quantity of the good times its price.<sup>1</sup> Dream data to measure the expenditure of individuals within households would look like Table 1a. Here, we directly observe the expenditure on each good by the man, woman and child in a nuclear household with 1 child. The poverty line of \$1.90 per person per day defines a household-level poverty line of \$2080. Since this household has total expenditure of only \$1850, standard poverty measure measurement (which assumes equal division within the household) would call all members of this household as “poor”.

However, with the dream data here, we observe the (unequal) expenditure level and resource share of each person, and can compare individual expenditure levels to individual poverty thresholds. The column totals give individual total expenditure levels. The man’s total expenditure on all goods is \$800, so his resource share (fraction of total household expenditure on all goods) is 43% (equals 800/1850). The individual poverty threshold is \$1.90 per day, equalling \$694, so the man is not poor. However, the woman’s total expenditure is \$600, and her resource share is 32%. Since her expenditure falls below \$694, she is poor. Similarly, the child’s total expenditure is \$450 and their resource share is 24%, and s/he is poor. Thus, the dream data reveal within-household inequality in resource shares, and the fact the some members are poor while others aren’t.

Table 1a: Dream Data					Table 1b: Real Data				
	Man	Woman	Child	Total		Man	Woman	Child	Total
Food	400	300	200	900	Food				900
Clothing	50	75	25	150	Clothing	50	75	25	150
Shelter	100	100	100	300	Shelter				300
Other	250	125	125	500	Other				500
Total	800	600	450	1850	Total				1850

If data like those in Table 1a were widely available, poverty measurement at the person level would be straightforward. Cherchye, Demuynck, De Rock, and Vermeulen (2017) collected this type of data for the Netherlands, and used it to, among other things, estimate consumption inequality within households. Brown, Ravallion and van de Walle (2019) use this type of data to investigate individual-level poverty and food deprivation, and Bargain, Lacroix and Tiberti (2019) use this type of data to validate the modeling assumptions of col-

<sup>1</sup>In principle, we include home-produced quantities, so expenditures include the imputed value of home-production.

lective household models. To our knowledge, these are the only cases where individual-level expenditure data for all consumption categories is collected.

In this paper, we use the terms *shareable* and *non-shareable* to refer to the consumption technology of the household and to related scale economies in household consumption. A nonshareable good has the property that the quantities consumed by each person add up to the total quantity purchased by the household. For example, food is nonshareable because food eaten by one member cannot be eaten by another, so if two members eat 1 unit each, the household must buy 2 units. In contrast, a shareable good has the property that the quantities of each person add up to more than the quantity purchased by the household. For example, if two people ride a motorcycle together, they both consume a motorcycle ride, but the household only has to purchase gasoline for one motorcycle ride. If the two people ride together only part of time, then this good is partially shared. If they ride the motorcycle together all the time, then it is fully shared.<sup>2</sup>

A key feature is that shareable goods feel cheaper to people living in households than they do to people living alone, but non-shareable goods feel just as expensive to people living in households as they do to people living alone. Thus in Table 1a, if we assume that food is non-shareable, then the \$400 expenditure on food for the man buys the same quantity that it would if he were living alone. In contrast, if we assume that shelter were fully shareable, the \$100 expenditure on shelter for the man would buy a larger quantity of shelter than it would if he were living alone. In principle, that \$100 expenditure for the man in a three-person household could buy three times the quantity attained by a \$100 expenditure for a person living alone (see Appendix 1 for a full description of this).

Thus, the benefit of shareability of goods to people in households is that the shadow price of consumption of those goods in the household is lower than the market price of those goods. Regardless of how shareable different goods are, the resource share of each person has the same interpretation: it is the fraction of total household expenditure enjoyed by each person.

Real-world expenditure data tend to look more like Table 1b. In this type of data, we see *household-level* expenditure for all the goods and services comprising total expenditure, and we may see one or two goods at the person level (in this case clothing). Such data are widely available in rich countries, because they are collected by statistical agencies that estimate the rate of price inflation, and are increasingly common in developing countries, in part due to international research efforts like the 100+ datasets in the Living Standards Measurement

---

<sup>2</sup>*Public* goods are a particular type of shareable good. First, public goods are *fully shareable* in the sense that the household can attain a quantity level of  $q$  for each household member by spending  $pq$  on that good. (For a nonshareable good, it would have to spend  $Npq$ , where  $N$  is the number of members.) Second, for public goods, each member must consume *the same amount* (equal to  $q$ ).



Study (LSMS) of the World Bank. So, with real-world data, we face an incomplete data problem: we do not have full data on individual expenditure; instead, we have data on just 1 or 2 commodities collected at the individual level.

DLP aim to address this missing data problem via modeling the allocation problem of the household, and “backing out” the individual resource shares from these incomplete data. In particular, DLP uses information on individual-level spending on non-shareable assignable goods (clothing in the Tables) to infer the individual-level to total expenditure of each person (the bottom row of Table 1a). DLP *does not* infer the individual-level expenditure on any particular good, just the individual-level total expenditure on all goods. So, for example, the fact that the man has less clothing expenditure than the woman does not imply that he has less food expenditure than her. Consequently, the fact that the man has less clothing expenditure than the woman does not imply that he has less total expenditure than her. Instead, the link between individual assignable goods expenditure and individual-level total expenditure is driven by response of the former to the total household budget. If the man’s clothing expenditure responds more to the household budget than does the woman’s, then he has a larger claim on household resources than she does.

If no goods are shareable, then there are no scale economies in household consumption; if some goods are at least somewhat shareable, then there are scale economies in household consumption. The methods proposed by DLP allow for unknown degrees of shareability for each consumption good, and therefore for unknown and unrestricted scale economies in consumption. But, they require that at least one good is not shareable and is assignable (observed at the individual level). Their methods reveal resource shares whose interpretation does not depend on the degree of shareability of the non-assignable goods.

## 2.1 An Efficient Collective Household Model

Browning, Chiappori and Lewbel (2013: BCL) provide a very general efficient collective household model with scale economies in consumption, preference heterogeneity across people, and possibly unequal distributions of household resources. DLP take that model and impose sufficient restrictions on it to make it implementable with real-world data via nonlinear estimation of household-level Engel curves for assignable goods. We now briefly sketch those models, and then show a linear recasting of DLP that may be estimated by ordinary least squares. Further, we will extend the model to allow for non-nuclear households (e.g., multigenerational households and single-parent households).

*Collective households* are households comprised of a collection of individuals. The individuals have utility functions; households are just environments in which individuals live.

*Efficient* collective household models are those in which the individuals in the household are assumed to reach the (household) pareto frontier. Like in earlier results in general equilibrium theory, the assumption of pareto efficiency is very strong: it means that the household-level allocation problem is observationally equivalent to a decentralised, person-level, allocation problem. In this decentralised allocation, each household member demands a vector of consumption quantities given their preferences and a personal budget constraint, and the household purchases the sum of these demanded quantities (adjusted for shareability/economies of scale).

The model thus has us picture the household as a machine that makes budget constraints for its members. Each person’s budget constraint is characterised by a shadow budget and a shadow price vector. They are “shadow” budgets and prices because they govern each person’s consumption demands but they are not observed and do not equal the observed household budget or market prices. Welfare analysis of people in living in households considers these shadow budget constraints.

Let  $h = 1, \dots, H$  index households. Let  $t$  index the types of individuals, in our case,  $m$  for adult male,  $f$  for adult female and  $c$  for children. Let the household  $h$  consist of  $N_h^t$  individuals of each type  $t$ , and let  $N_h = \sum_t N_h^t$  be the total number of individuals in household  $h$ . The types of individuals are in some sense defined by the data, as we will see below. Let  $y_h$  denote the observed household income (budget). Each type of person gets a shadow budget, and these shadow budgets must add up to the full household budget.

The share of the household budget allocated to persons of type  $t$  in household  $h$  is called their *resource share*, denoted  $\eta_h^t$ . Resource shares sum to 1 in each household  $h$  so that  $\sum_t \eta_h^t = 1$ . They may in general depend on household budgets, prices, household and individual characteristics (including so-called “distribution factors”). Most importantly, they can vary across the types of individuals in the household, for example, men’s and women’s resource shares are not assumed to be equal. But, within types, we assume that resources are distributed equally (if there is one person for each type, then this is not restrictive). For example, in a household with two children where the children’s resource share is  $\eta_h^c = 0.40$ , we have that 40 per cent of the household budget is allocated to children, with 20 per cent going to each child. In general, the total shadow budget of all the people of a given type  $t$  in a household  $h$  is  $\eta_h^t y_h$ , and the shadow budget of each person of that type is  $\eta_h^t y_h / N_h^t$ .<sup>3</sup>

Shadow prices for goods are the within-household prices of consumption. Let  $\mathbf{p}$  denote the market price vector for goods and let  $\tilde{\mathbf{p}}$  denote the shadow price vector of goods. Shadow prices must be the same for all household members. If they were not the same, then there

---

<sup>3</sup>DLP allowed for multiple children. Here, like Dunbar, Lewbel and Pendakur (2019), we extend that notation to allow for multiple members of any type.

would be gains from trade across household members, a violation of the assumption of efficiency.

Shadow prices differ from market prices. Shadow prices are weakly lower than market prices because some goods may be shareable. The more shareable is the good, the lower is its shadow price of consumption within the household. Shadow prices may in principle be as low as  $p/N_h$  for a good that is so shareable that each household member can consume the purchased quantity. For goods that are not shared, the shadow price equals the market price. If shadow prices are less than market prices, then there are scale economies in household consumption, because individuals face lower prices within households than when they live alone. (Holding the budget constant, lower prices means higher utility.)

Resource shares and shadow prices are interesting because they completely characterise the budget constraint of each person in the household. To the extent that budget constraints provide welfare measures, resource shares and shadow prices allow us to peer inside the household to consider the distribution of welfare therein. See BCL and Pendakur (2018) for a discussion of how to use resource shares and shadow prices in welfare analysis.

In this work, we wish to identify resource shares from household-level consumption choice data, but we will not try to identify shadow price vectors (which reveal scale economies in consumption due to sharing).<sup>4</sup> Resource shares are interesting even without knowledge of shadow prices. First, resource shares provide a measure of consumption within the household: higher resource shares mean higher consumption. Second, they speak to inequality within the household: if resource shares are very unequal, then there is a lot of inequality within the household. Third, resource shares may respond to policy variables in the context of poverty reduction. If we can find policy variables that shift resource shares upwards for disadvantaged individuals, then their poverty rates may decrease.

Define an *assignable good* as one where we can observe which household member consumes the good. We focus on nonshareable assignable goods. Such goods are very useful for identification of household models (see, e.g., Chiappori and Ekeland 2009).<sup>5</sup>

In our context, the available data on assignable goods will define our typology of individuals. In many data sources, assignable spending on clothing is available for adult men, adult women and children, but it could be recorded by gender for children as well as for adults, in which case there would be 4 types of individuals, adults or children and males or females. It is for this reason that we specify the model in terms of types of individuals rather than in terms of individuals themselves. Of course, if we had an assignable good for each individual,

---

<sup>4</sup>Our methodology estimates resource shares at a given price vector, without knowledge of relative prices. Since we don't observe market prices, we cannot estimate shadow prices. Other methodologies use observed price variation to identify shadow prices and thus scale economies (e.g., BCL and Pendakur 2018).

<sup>5</sup>Chiappori and Ekeland refer to nonshareable assignable goods as *private* assignable goods.

we could estimate the resource share of each rather than estimating the resource shares by type.

BCL show that, given the model described above, household demands are related to individual demands in an intuitive way. Given the sharing in the household, the household purchases enough of each commodity (adjusted for sharing) so as to give each individual in the household exactly what they would have purchased had they faced their individual shadow budget and the shadow price vector.

Much work on consumer demand estimation models the choice of *Engel curves*. The Engel curve of a good is the fraction of the overall budget (spent on all goods) commanded by that good (Engel 1857, 1895).<sup>6</sup> Engel curve functions hold prices constant at some vector, and evaluate the fraction of expenditure as a function of the total household budget (and possibly other demographic characteristics). DLP derive the implications of the BCL model on Engel curves for nonshareable assignable goods.

Let  $\eta^t(y)$  be the resource share of person  $t$  when the household faces a fixed market price vector  $\mathbf{p}$ .<sup>7</sup> Assume that shadow prices are linear in market prices, with  $\tilde{\mathbf{p}} = \mathbf{A}\mathbf{p}$  for some diagonal matrix  $\mathbf{A}$ . BCL provide identification results for more general relationships, but this restriction substantially simplifies Engel curves. The matrix  $\mathbf{A}$  governs the shareability of goods, and therefore shadow prices. Nonshareable goods have corresponding elements of  $\mathbf{A}$  equal 1, and their shadow prices equal their market prices. Shareable goods have corresponding elements of  $\mathbf{A}$  less than 1, and their shadow prices are lower than their market prices.

The matrix  $\mathbf{A}$  is completely unrestricted, other than that there is assumed to be one assignable non-shareable good. For that good, the corresponding element of  $\mathbf{A}$  equals 1. Importantly, this means that other than knowing that one good is observed, assignable and nonshareable, we don't need to know *a priori* which of the other goods are shareable or public, nor the degree of their shareability or publicness.

Let a nonshareable assignable good (e.g., clothing) be observed for each type of person in a collective household. Let  $w^t(y)$  be the *Engel curve function* for a person of type  $t$  for their assignable good. It gives the fraction of expenditure commanded by that good for a person of that type if they lived alone and faced the shadow price vector  $\tilde{\mathbf{p}}$  and a budget  $y$ .<sup>8</sup> Since the only demander for this good is person  $t$ , the household Engel curve,  $W^t(y)$ , for

---

<sup>6</sup>Engel curve functions are often called “budget share” functions, for obvious reasons. We use the phrase Engel curve rather than budget share so that it is not confused with “resource share”.

<sup>7</sup>The resource share may also depend on other covariates, including but not limited to those which affect preferences. Since the entire model can be conditioned on such variables, we suppress that dependence here.

<sup>8</sup>Let  $q^t(y, p)$  be the quantity demand function of type  $t$  for their assignable good. By definition, the household's expenditure on the assignable good for type  $t$  individuals, at market price  $p$  and income level  $y$ ,

assignable good  $t$ , evaluated at the market price vector  $\mathbf{p}$ , is given by:

$$W^t(y) = \eta^t(y) w^t(\eta^t(y) y / N^t). \quad (1)$$

This simple relationship says that the household's Engel curves (at market prices, held fixed) for the assignable goods for  $t = m, f, c$  are equal to the resource share of the relevant type times the unobserved Engel curve of a person of that type facing the shadow price vector and their shadow budget.<sup>9,10</sup>

Browning, Chiappori and Lewbel (2013) show that if we observed the functions  $w^t(y)$  and the functions  $W^t(y)$ , then the resource shares  $\eta^t(y)$  are identified. In general, this is possible if we observe Engel curves at many observed price vectors *and* assume that single individuals have the same preferences as individuals that live in collective households *and* that the Engel curves of single individuals are observable. In many settings, including most developing countries, at least one of these conditions is likely to be violated. For example, we do not observe children living alone and, in many countries, unmarried men and women live in collective households rather than on their own.

DLP provide sufficient restrictions on the model such that resource shares are identified from data on just Engel curve functions of collective households facing a single price vector. They do not assume that the Engel curves of single individuals are observed. They assume: a) resource shares do not depend on the household budget so that  $\eta^t(y) = \eta^t$ ; b) individual Engel curve functions are given by the Almost Ideal demand system of Deaton and Muellbauer (1980), so that  $w^t(y) = \alpha^t + \beta^t \ln y$ ; and c) preferences are similar—but not identical—across

---

$pq(y, p)$  is equal to  $n^t \tilde{p} q\left(\frac{\eta^t(y)y}{n^t}, \tilde{p}\right)$ , the household's expenditure on the same good at the shadow price and the resource share of type  $t$ .

<sup>9</sup>For a longer exposition of the implications of assignable goods in the model, and for a story of how this connects to DLP13, see Pendakur (2018). For more strategies to identify resource shares from Engel curve data, see Dunbar, Lewbel and Pendakur (2019).

<sup>10</sup>The model implies that, given the sharing in the household, the household purchases enough of each commodity so as to give each individual in the household exactly what they would have purchased given their shadow constraint. For private assignable goods, there is no sharing, so the household purchases a quantity equal to the quantity the individual would have purchased:  $Q^t(y) = q^t(\eta^t(y)y/N^t)$ , where  $Q^t$  and  $q^t$  are household and individual quantity demands. Dividing both sides by the  $y$  and multiplying by  $p$  yields household Engel curves  $W^t(y) = pQ^t(y)/y = pq^t(\eta^t(y)y/N^t)/y$ . Since,  $w^t(\eta^t(y)y/n^t) = N^t pq^t(\eta^t(y)y/N^t)/\eta^t y$ , we have  $\eta^t w^t(\eta^t(y)y/N^t)/N^t = pq^t(\eta^t(y)y/N^t)/y$ . Substituting in, and multiplying by  $N^t$  (to get the aggregate Engel curve for all persons of type  $t$ ) yields equation (1).

people, such that  $\beta^t = \beta$ .<sup>11</sup> Substituting these assumptions into (1) gives

$$W^t(y) = \eta^t \left( \alpha^t + \beta \left( \ln y + \ln \eta^t - \ln N^t \right) \right) \quad (2)$$

The assumption that resource shares do not depend on the household budget is strong. It implies, for example, that, all else equal, if a household gets richer the intra-household relative consumption distribution will not change.<sup>12</sup> Surprisingly, there is some empirical support for this restriction. Menon, Perali and Pendakur (2011) show that reported (stated preference) resource shares in Italian survey data do not vary much with household budgets. Cherchye, De Rock, Lewbel and Vermeulen (2015) use revealed preference methods to show that although resource shares do depend on variables like relative wages and education, they do not vary much with household budgets.

The intuition for identification of resource shares in the above model is as follows. The observable budget semi-elasticity of household-level Engel curves for assignable goods,  $\partial W^t(y)/\partial \ln y$ , is equal to  $\eta^t \beta$ . Since  $\eta^t$  sum to 1, the sum of this semi-elasticity across types is  $\beta$ . Consequently, the relative magnitude of budget semi-elasticities determines resource shares. If the household's response to an increase in the budget is larger for men's clothing than for women's clothing, it is because the men's resource share is larger. Note that that it is budget responses, not levels, that identify resource shares. If women's clothing Engel curves were higher than men's, but men's had the larger budget response, then men would have the higher resource share.

A key feature of DLPs model is that identification requires  $\beta \neq 0$ , because if  $\beta = 0$  clothing budget shares are homothetic and  $\partial W^t(y)/\partial \ln y = 0$ . In this case, because all slopes are zero, we cannot identify resource shares. In economic terms, this means that because we use budget responses to identify resource shares, the assignable goods must be

---

<sup>11</sup>DLP define a property called "similar across people" (SAP) as being satisfied if the Engel curves for assignable goods are given by  $w^t(y) = w^t(y/G^t) + g^t$  for some constants  $G^t$  and  $g^t$ . This condition is satisfied if preferences satisfy "shape-invariance" (see, e.g., Pendakur 1999 or Blundell, Chen and Kristensen 2014). It is also satisfied if cost functions satisfy "independence of base" (Lewbel 1989) or "equivalence-scale exactness" (Blackorby and Donaldson 1993). In addition to the assumptions of BCL, DLP assume that:  $\tilde{\mathbf{p}} = A\mathbf{p}$ ; resource shares do not depend on household budgets; and SAP holds. Given these assumptions, DLP show that resource shares are identified from the Engel curves of collective households at a single price vector. So, they do not require the Engel curves are log-linear for identification. When applied to the log-linear Engel curves, SAP implies  $\beta^t = \beta$ . We (and they) use log-linear Engel curves to make estimation easier, not to achieve identification.

<sup>12</sup>In fact, DLP require less than full independence, in two ways. First, they only require that resource shares are invariant to expenditure *over some range* of household expenditure. So, for example, if this invariance held only for the poorest households, we could still identify resource shares for the very poor, and consequently identify poverty at the individual level for this subpopulation. Second, the independence of resource shares from household expenditure is *conditional* on other observed covariates, which may include, for example, wealth.

either necessities (whose Engel curve declines with the budget) or luxuries (whose Engel curve increases with the budget).

The econometric model defined by equation (2) is nonlinear due to the fact that  $\eta^t$  multiplies  $\beta$ , and requires positive resource shares, due to the  $\ln \eta^t$  term. So, estimation requires nonlinear optimization subject to bounding restrictions on parameters. This combination can be tricky to estimate, because nonlinear optimizers may have numerical issues if they search through a space with a negative value of  $\eta^t$  (and thus undefined  $\ln \eta^t$ ). Further complications arise if one tries to condition the model on observed covariates, as we do below, because negative resource shares have to be avoided at all observed values of the covariates.

We now propose a reframing of DLP that uses ordinary least squares regression to identify resource shares. This strategy is easier to implement because negative resource shares don't cause the estimator to crash; instead, they are estimated, and provide evidence to the researcher that the model is not a good one for the data at hand.

## 2.2 Linear DLP

We now present our model—a theory-consistent linear recasting of DLP. Consider first the case with no demographic covariates (the entire model can be written conditionally on covariates, which we do below). Rewrite equation (2) with a subscript  $h = 1, \dots, H$  indexing households, and an additive error term  $\varepsilon_h^t$ , as the following linear model:

$$W_h^t = a_h^t + b^t \ln y_h + \varepsilon_h^t \quad (3)$$

where

$$a_h^t = \eta^t \alpha^t + \eta^t \beta \ln \eta^t - \eta^t \beta \ln N_h^t,$$

and

$$b^t = \eta^t \beta. \quad (4)$$

Rewrite  $a_h^t$  as

$$a_h^t = a_0^t + a_N^t \ln N_h^t,$$

where

$$a_0^t = \eta^t \alpha^t + \eta^t \beta \ln \eta^t \quad \text{and} \quad a_N^t = -\eta^t \beta$$

Here, the troublesome  $\ln \eta^t$  term is absorbed into the parameter  $a_0^t$ , and so solves the problem of taking the log of a negative and crashing the estimator.<sup>13</sup> The model may be estimated by linear regression of the observed household-level assignable good expenditure share,  $W_h^t$ , on

---

<sup>13</sup>One could impose the restriction that  $a_N^t = -b^t$ , but it is not necessary to do so.

a constant, the log of the number of members of type  $t$ ,  $\ln N_h^t$ , and the log of the household budget,  $\ln y_h$ . E.g., for data on households with  $t = m, f, c$ , one could implement the linear seemingly unrelated regression system in Stata via:

```
sureg (W_m ln_N_m lny) (W_f ln_N_f lny) (W_c ln_N_c lny)
```

Rearranging equation (4), we have

$$\eta^t = b^t / \beta.$$

Denote the regression estimates as  $\hat{a}_0^t, \hat{a}_1^t$  and  $\hat{b}^t$ . Since resource shares sum to 1, we can use  $\sum_t \hat{b}^t$  as an estimate of  $\beta$ , which implies that an estimate of the resource share of type  $t$ ,  $\eta^t$ , is given by

$$\hat{\eta}^t = \hat{b}^t / \left( \sum_{t=1}^T \hat{b}^t \right).$$

Notice that the estimated resource share does not depend on the estimate of the level term  $\hat{a}_h^t$ . We identify resource shares in DLP solely via the relative slopes of Engel curves. One could implement this estimator for, e.g.,  $\eta^m$  in Stata via:

```
nlcom [w_m]lnx/([w_m]lnx+[w_f]lnx+[w_c]lnx)
```

In this model,  $\beta \neq 0$  is an identifying restriction. If  $\beta = 0$ , then the estimated value of the denominator may be close to 0, yielding “crazy” estimates of resource shares. We use this fact to form the basis of our pretest, below.

The model above does not include any conditioning variables, such as demographic preference shifters. Including them does not affect identification, but does require some additional notation. Let  $\mathbf{z}$  be all variables that affect preferences, including the numbers of household members of each type  $\mathbf{N} = \{N^t\}$ . Let  $\tilde{\mathbf{z}}$  be the subvector  $\mathbf{z}$  of that excludes  $\mathbf{N}$ :  $\mathbf{z} = [\mathbf{N} \quad \tilde{\mathbf{z}}]$ . Let  $\tilde{\mathbf{z}} = 0$  for some meaningful reference value of these characteristics. Assume that resource shares,  $\eta^t$ , and preference parameters,  $\alpha^t$  and  $\beta$ , depend on  $\mathbf{z}$ .<sup>14</sup> Note that  $\tilde{\mathbf{z}}$  can be empty. Substituting this into (2), and expanding out the terms, we have:

$$W^t(y, \mathbf{z}) = \eta^t(\mathbf{z})\alpha^t(\mathbf{z}) + \eta^t(\mathbf{z})\beta(\mathbf{z}) \ln y + \eta^t(\mathbf{z})\beta(\mathbf{z}) \ln \eta^t(\mathbf{z}) - \eta^t(\mathbf{z})\beta(\mathbf{z}) \ln N^t, \quad (5)$$

for households with members of type  $t$ . (If the household has no members of type  $t$ ,  $W_h^t = 0$ .) This nonlinear structural model (5) has been implemented by several researchers on data

---

<sup>14</sup>*Distribution factors* are variables that affect resource shares,  $\eta^t$ , but not preferences,  $\alpha^t$  and  $\beta$ . We can think of these as elements of  $\mathbf{z}$ , imposing the restriction that they do not affect  $\alpha^t$  and  $\beta$ . These restrictions do not affect the linear estimator for resource shares that we propose below. So, in this paper, we don't separately track distribution factors from other observed covariates that affect preferences and/or resource shares.



from several countries (e.g., DLP in Malawi; Bargain, Donni and Kwenda 2014 in Cote D'Ivoire; Calvi 2014 in India; De Vreder and Lambert 2016 in Senegal; Bargain, LaCroix and Tiberti 2018 in Bangladesh).<sup>15</sup>

As with equation (2), this model contains a term linear in the log of the resource share,  $\ln \eta^t(\mathbf{z})$ . If  $\eta^t$  is parameterised as a linear index (and especially if it contains an unbounded variable), then search algorithms trying to find the minimum/maximum of the sum of squares, likelihood function or GMM criterion function can stop before finding a solution. This is similar to the problem of the linear probability model giving predicted probabilities outside  $[0, 1]$ , but with the additional consequence that it may induce numerical problems in nonlinear solvers. For example, they may try to evaluate the function in a region of the parameter space where  $\eta^t(\mathbf{z})$  is negative, yielding a missing value for  $\ln \eta^t(\mathbf{z})$ .

The model may be difficult to implement if some households do not have at least one member of each type, because those types have a resource share of zero, and the log of the resource share enters the demand equations. An additional problem relative to equation (2) comes from the fact that the term  $\eta^t(\mathbf{z})\beta(\mathbf{z}) \ln y$  has quadratic interactions in  $\mathbf{z}$  multiplying  $\ln y$ . These make it difficult to precisely identify the dependence of resource shares  $\eta^t(\mathbf{z})$  on  $\mathbf{z}$ , because  $\mathbf{z}$  affects both  $\eta^t$  and  $\beta$ .

Rewrite equation (5) with a subscript  $h$  on all observed variables, and an additive error term  $\varepsilon_h^t$ , as the following linear model:

$$W_h^t = a_h^t + b_h^t \ln y_h + \varepsilon_h^t, \quad (6)$$

where

$$a_h^t = \eta^t(\mathbf{z}_h)\alpha^t(\mathbf{z}_h) + \eta^t(\mathbf{z}_h)\beta(\mathbf{z}_h) \ln \eta^t(\mathbf{z}_h) - \eta^t(\mathbf{z}_h)\beta(\mathbf{z}_h) \ln N_h^t, \quad (7)$$

and

$$b_h^t = \eta^t(\mathbf{z}_h)\beta(\mathbf{z}_h). \quad (8)$$

Here,  $a_h^t$  and  $b_h^t$  are functions of the vector of conditioning variables  $\mathbf{z}_h$ . Suppose that  $\eta^t$  and  $\beta$  are linear indices in  $\mathbf{z}_h$ . Then,  $a_h^t$  is a third-order function in  $\mathbf{z}_h$ , and  $b_h^t$  is quadratic in  $\mathbf{z}_h$ . Defining  $\mathbf{Z}_h$  as the list of level and interaction terms up to the third order in  $\mathbf{z}_h$ , OLS regression of  $W_h^t$  on a constant,  $\ln y_h$  and  $\mathbf{Z}_h \cdot \ln y_h$  would suffice.

Alternatively, both  $\eta^t$  and  $\beta$  could have unknown functional forms. In this case, one could let  $a_h^t$  and  $b_h^t$  be nonparametric functions of  $\mathbf{z}_h$ , and use standard semiparametric methods to estimate the model. One such approach would be to let  $a_h^t$  and  $b_h^t$  be a multivariate polynomials over  $\mathbf{z}_h$ , with the degree of the polynomials increasing with the sample size.

---

<sup>15</sup>Menon, Perali and Piccoli (2018) provide a different almost linear formulation of Engel curves that could be approximated similarly to our development below.

## 2.3 Approximation

Unfortunately, neither of these approaches is practical with a high-dimensional conditioning vector  $\mathbf{z}_h$ . For example, with a constant and 9 conditioning variables in  $\mathbf{z}_h$ , third-order interactions requires 444 regressors.<sup>16</sup> So, we recommend approximating the model. Approximate the  $a_h^t$  term with

$$a_h^t = a_0^t + \mathbf{a}_m^{t'} \mathbf{z}_h = a_0^t + \mathbf{a}_N^{t'} \mathbf{N}_h + \mathbf{a}_z^{t'} \tilde{\mathbf{z}}_h. \quad (9)$$

This approximation soaks up the troublesome  $\ln \eta^t$  term into the parameter vector  $a_0^t$ , and so solves the problem of taking the log of a negative and crashing the estimator.<sup>17</sup>

Similarly, we recommend approximating the slope term analogously to the level term as

$$b_h^t = b_0^t + \mathbf{b}_z^{t'} \mathbf{z}_h = b_0^t + \mathbf{b}_N^{t'} \mathbf{N}_h + \mathbf{b}_z^{t'} \tilde{\mathbf{z}}_h, \quad (10)$$

where  $\mathbf{b}_N^t$  and  $\mathbf{b}_z^t$  refer to the relevant subvectors of  $\mathbf{b}_m^t$ . From inspection of equation (8), it is easy to show that approximation for  $b_h^t$  is exact if  $\eta^t$  is linear in  $\mathbf{z}_h$  and  $\beta$  is independent of  $\mathbf{z}_h$  (that is, if  $\beta$  is a constant).

This approximate model may be estimated via OLS regression of  $W_h^t$  on a constant,  $\mathbf{z}_h$ ,  $\ln y$  and  $\mathbf{z}_h \cdot \ln y$ . The estimated coefficients on  $\ln y$  and  $\mathbf{z}_h \cdot \ln y$  are estimates of  $b_0^t$  and  $b_z^t$ , respectively. These may be used to construct and estimate  $\hat{b}_h^t$  of  $b_h^t$ .

Regardless of the specification of  $b_h^t$ , and regardless of whether not it is taken to be an approximation or exact (due to prior knowledge of the functional form  $\eta^t$  and  $\beta$ ), we can solve for resource shares. Since resource shares sum to 1, we can use  $\sum_t \hat{b}_h^t$  as an estimate of  $\beta(\mathbf{z}_h)$ , which implies that an estimate of the resource share of type  $t$  in a household with characteristics  $\mathbf{z}_h$  is given by

$$\hat{\eta}_h^t = \hat{\eta}^t(\mathbf{z}_h) = \hat{b}_h^t / \left( \sum_{t=1}^T \hat{b}_h^t \right). \quad (11)$$

Resource shares may then be computed via (11). From a practical standpoint, if the denominators in (11) had a lot of variation, or if they were close to zero, estimated resource shares might be somewhat wild. However, we can simplify the denominator by imposing the restrictions

$$\sum_t \mathbf{b}_z^t = \mathbf{0}. \quad (12)$$

---

<sup>16</sup> $10^3 = 1000$  triples, deleting permutations, is 222 unique combinations, times 2 for the intercept and slope.

<sup>17</sup>One could additionally add  $\ln \mathbf{n}_h$  to the linear index, and in that case, impose the restriction that  $a_{\ln n}^t = b_n^t$ . But it is not necessary to do so, since this is an approximating term.

implying that  $\sum_t b_h^t = \sum_t (b_0^t + b_{N^m}^t N_h^m + b_{N^w}^t N_h^w + b_{N^c}^t N_h^c)$ . Then, estimated resource shares are equal to

$$\hat{\eta}^t(\mathbf{z}_h) = \frac{b_h^t}{\sum_t (b_0^t + b_{N^m}^t N_h^m + b_{N^w}^t N_h^w + b_{N^c}^t N_h^c)}.$$

Here, we expect  $b_{N^t}^t$  to all have the same sign, and that the variation in the denominator would be tamped down.

We note that this functional form for resource shares allows for the possibility that the resource shares of person types equal their per-capita share household members. In particular, if  $\mathbf{b}_z^t = \mathbf{0}$  for all  $t$  and  $b_{N^{t'}}^t = 0$  for all  $t' \neq t$  and  $b_{N^t}^t = \kappa$  for all  $t$ , then we get per-capita resource shares,  $\eta^t(\mathbf{z}_h) = N^t \kappa / \sum_t N^t \kappa = N^t / \sum_t N^t$ .<sup>18</sup>

Our model may be estimated by equation-by-equation ordinary least squares (OLS), or, if the restrictions (12) are used, with seemingly unrelated regression (SUR).

Previous implementations of DLP's nonlinear model (Calvi 2019 and Dunbar, Lewbel and Pendakur 2019) that include complex household types did not include households that didn't have at least 1 adult man or 1 adult woman. The basic reason is that DLP's nonlinear model would have that such households have one type whose resource share is zero, and therefore a troublesome  $\ln \eta^t$  term for that type. In our linear recasting of DLP, this problem vanishes, because that term is absorbed into  $a_h^t$ .

Let *composition* be a variable indicating whether or not different types of people are present in a household. In our work below, we consider 4 compositions of types: households with men, women and children; households with men and children only; households with women and children only; and households with men and women only. A pooled estimator would simply interact composition with all the regressors in the model ( $\mathbf{z}$ ,  $\ln y$  and  $\mathbf{z} \cdot \ln y$ ); equivalently one could estimate the model separately for each composition. In our empirical work, we do the latter. That is, to compute resource shares for people living in households with men, women and children, we run regressions on observations with at least 1 man, 1 woman and 1 child in each household. To compute resource shares for people living in households with just women and children, run regressions on observations with no men, and at least 1 woman and 1 child. And analogously for the other 2 compositions.

## 2.4 Pre-Test

As noted above, if  $\beta(\mathbf{z}_h) = 0$  then resource shares are not identified. In this case, the estimated value of the denominator may be close to 0, and the resulting estimated resource shares would be unreliable. If it were the case in the limit, then inference is polluted by weak

---

<sup>18</sup>One could additionally restrict  $\sum_t \mathbf{b}_N^t = \mathbf{0}$ , implying that  $\hat{\eta}^t(\mathbf{z}_h) = b_h^t / (\sum_t b_0^t)$ . This further simplifies the denominator, but at the cost of not nesting the per-capita model.

identification problems (see Han and McCloskey 2019). Consequently, it is valuable to have a pre-test to tell us whether or not these methods will work at all. Previous papers (DLP; Dunbar, Lewbel and Pendakur 2019; Han and McCloskey 2019) have tested this identifying restriction, but their tests all involve estimating nonlinear models. Our linear recasting of DLP straightforwardly delivers an OLS-based test of whether or not this restriction that supports the identification of resource shares is supported by the data.

Let the *overall assignable budget share* of the household be given by  $W_h = \sum_t W_h^t$ , and let  $a_h = \sum_t a_h^t$ ,  $b_h = \sum_t b_h^t$  and  $\varepsilon_h = \sum_t \varepsilon_h^t$ . Then, our approximate model above implies

$$W_h = a_h + b_h \ln y_h + \varepsilon_h \quad (13)$$

and OLS regression of  $W_h$  on  $1, z_h, \ln y_h$  and  $z_h \ln y_h$  yields an estimate  $\hat{b}_h$  of  $\beta(z_h)$ . We propose that an easy and useful pre-test for the use of this methodology is to check whether or not overall clothing budget shares for households are statistically significantly upward or downward sloping.

Below, we use two results from our pre-test regression to consider whether our methods should be applied to the data at hand. First, we use  $E[b_h] = \hat{b}_0 + \hat{\mathbf{b}}' \bar{\mathbf{z}}_h$ , where  $\bar{\mathbf{z}}_h$  is the sample average of  $\mathbf{z}_h$ , as a test statistic. This is a test of the economic hypothesis that the overall clothing share Engel curve, evaluated at the mean value  $\mathbf{z}_h$ , is either a necessity or a luxury (is increasing or decreasing). If it is neither, then our strategy to estimate resource shares should not be used. Second, for every observation in the data, we test whether or not  $\hat{b}_h = \hat{b}_0 + \hat{\mathbf{b}}' \mathbf{z}_h$  is statistically significantly different from zero, and report the fraction of households for which it is statistically significant. Here, we think that a “large” fraction of households should have an estimated overall Engel curve that is either upward or downward sloping, where “large” is taken to be 75% of the sample (other cutoffs could be used).

### 3 Data

In most countries in the world, national statistical offices regularly collect household expenditure survey data. These data are used as input in national accounts, for the calculation of the GDP, to measure inflation, to analyse household spending patterns and behaviour, and to evaluate policy. Since the early 1980s, the World Bank has been providing assistance to national statistical offices in the design and implementation of household surveys through the Living Standards Measurement Study (LSMS). These data are standardised to some extent, and are the best tool available for cross country comparisons of poverty in low- and middle-income countries.

LSMS surveys exist for about 40 countries, and often several waves exist. There are in total 87 country-waves potentially available for the analysis of household consumption behaviour. We analyse the most recent waves from 12 countries for which LSMS data include clothing expenditure by type of individual (men, women and children), a measure of total expenditure for the household, and a minimal set of demographic variables (age, sex and education level of household members). We also include non-LSMS data from the Bangladesh Integrated Household Survey so that we can consider using food as the assignable good (see below).<sup>19,20</sup>

Table 2: Descriptive Statistics

Country	total H	single N	compositions	Our H	Nuclear H	budget	std dev
Albania	3599	239	mw, mwc	3279	612	11084	6477
Bangladesh	6434	219	mw, wc, mwc	6120	2122	6416	6268
Bulgaria	3018	801	mw, mwc	2099	412	13117	7954
Ethiopia	4717	503	mw, wc, mwc	3845	1481	3092	3645
Ghana	8687	1922	mw, mc, wc, mwc	6313	2195	5096	4835
Iraq	17513	288	mw, wc, mwc	14297	5487	26188	14287
Malawi	12271	1030	mw, wc, mwc	10873	5488	3189	3758
Nigeria	4600	349	mw, wc, mwc	3556	1013	6656	20322
Tajikistan	1503	54	mw, mwc	1275	192	10483	6250
Tanzania	3352	320	mw, wc, mwc	2677	1133	7219	5164
Timor Leste	4477	229	mw, wc, mwc	3788	1577	4954	4116
Uganda	3117	257	mw, wc, mwc	2468	1014	2462	2262
Bangladesh–Food	6434	219	mw, wc, mwc	5604	1916	6445	6287

Descriptive statistics for the sample of countries are in table 2. Altogether, these countries represent about 9% of the world population. Starting from the publicly available LSMS data (and the Bangladesh data), for each country, we exclude observations with missing data on clothing expenditures, total household expenditures or the age, sex and education level of household members. This yields sample sizes reported in column (2). There is a wide range of sample sizes after this initial cleaning, from 1,503 households in Tajikistan to 17,513 households in Iraq. Below we will pay attention to whether sample size matters to the feasibility of the method.

In column (3), we report the number of households which are composed of a single adult man or woman. Since these households only have one individual, there is no sharing of resources, and they are not used in the estimation of resource shares, but they are included in

<sup>19</sup>A variety of reasons makes the data from the other countries unusable. In some cases, no data on assignable goods is collected; in others, information on elements of non durable expenditure is missing.

<sup>20</sup>Code to go from publicly available online data to our working data files for each of the 12 countries, and code to estimate all tables, is available on request.

the subsequent poverty analysis. It is worth noting that there are few singles, and that most households contain more than one type of person, highlighting the importance of modeling the within-household allocation of resources.<sup>21</sup>

For the estimation of the resource shares, we use all household compositions apart from singles and we allow for any number of individuals of each type. The possible compositions are  $mw$ ,  $mw c$ ,  $wc$ , and  $mc$ . These indicate that individuals of the type  $m$  for men,  $w$  for women and  $c$  for children, are present in the households, but it does not indicate how many individuals of each type there are. We exclude households belonging to a composition for which there are less than 100 observations (since estimation is done separately for each composition). The compositions remaining in the sample after this selection are indicated in column (4), and “our H” (5) gives the total number of observations of these compositions. This latter column shows that we are able to exploit most of the data.

Column (6) shows the number of nuclear households in each country. In contrast to previous work, we are not limited to using only nuclear households. This shows that the selection to just nuclear households can be very restrictive indeed in some countries; nuclear households are less than 25% of all households in 6 of our 12 countries.

We then provide the mean and standard deviation in our sample (excluding singles) of the overall budget in (PPP) \$US 2010. In some countries in our data, the average household budget is close to the World Bank poverty line of \$US7.60 per day for a 4 person household (e.g., Ethiopia, Malawi and Uganda); in some countries, it is well above (e.g., Bangladesh, Iraq). In all countries, the standard deviation is of comparable order to the mean, which is desirable since identification rests on budget variation.

The bottom row gives summary statistics for the data on assignable food in Bangladesh. It is different from the clothing data because different observations have valid assignable food data versus clothing data. In the analysis below, we will compare estimated resource shares from assignable food with those from assignable clothing.

## 4 Results

We estimate equation (6) under the restrictions (12) via seemingly unrelated regression in Stata. Our observed vector of demographic variables  $\mathbf{z}_h$  is comprised of: the numbers of men, women and children ( $\mathbf{n}_h$ ); the average ages of men, women and children; the minimum age of the children; the average education levels of the men and women; and, in some specifications, a dummy variable indicating that the household lives in an urban area.

---

<sup>21</sup>For households with, e.g., multiple men but no women or children, the underlying model could be collective but it could only be estimated if there were an observed assignable good for each of the men.

## 4.1 Pre-test

The statistical significance of the slope of the Engel curve for the sum of household assignable goods provides the pre-test for the applicability of the method. In Table 3, we give the mean and standard deviation of assignable goods budget shares (summed across household members), and the slope of the Engel curve evaluated at average characteristics, along with a  $t$ -test for its difference from zero. In the rightmost column, we give the fraction of observations whose estimated slope (conditional on their observed covariates) is statistically significantly different from zero.

Clothing is not a large budget share. Clothing represents between 1.7% and 7% of the budget. The standard deviation of clothing shares is high relative to the mean, so there is considerable dispersion in the distribution of clothing shares in each country.

country	sample N	budget share	std dev	slope at $\bar{z}$	$t$ -test of slope	% of sample significant
Albania	3279	0.0411	0.0416	0.0139	4.6	84
Bangladesh	6120	0.0394	0.0206	-0.0157	-21	100
Bulgaria	2099	0.0355	0.0398	0.0144	5.1	90
Ethiopia	3845	0.0716	0.0636	-0.0109	-3.5	65
Ghana	6313	0.0476	0.04	-0.0021	-1	62
Iraq	14297	0.07	0.0465	0.0209	14.8	99
Malawi	10873	0.0246	0.0361	0.0092	10	98
Nigeria	3556	0.0171	0.0235	-0.0017	-2	50
Tajikistan	1275	0.0578	0.0502	0.0075	1.8	5
Tanzania	2677	0.0436	0.0578	-0.0022	-1	12
Timor Leste	3788	0.0223	0.0205	-0.0025	-1.8	48
Uganda	2468	0.0545	0.0521	-0.0039	-1.2	5
Bangladesh–Food	5604	0.5679	0.1504	-0.1204	-17.5	100

Clothing is found to be a luxury in Albania, Bulgaria, Iraq, and Malawi and a necessity in Bangladesh, Ethiopia, Nigeria.

The slopes of the clothing Engel curves are not statistically significantly different from zero in Ghana, Tajikistan, Tanzania, Timor L’este and Uganda. Since the formula for resource shares uses this slope as a denominator, for these countries, the model may not be identified.

We also report the percentage of the sample for which the slope is significant. For our method to work, this needs to be high enough, so that we further eliminate Ethiopia and Nigeria because less than 75% of observations in those countries have predicted Engel curve slopes that are statistically significantly different from zero. This leaves us with 5 countries

which pass the pre-test, hence for which the model is identified and resource shares can be estimated.

For Bangladesh, we also have assignable data on food consumption. We have fewer observations (5604) on food than clothing (6120) because there is some non-response in the daily food diary data. Food budget shares are much larger than clothing budget shares: whereas clothing accounts for only 3 per cent of total household consumption, fully 56 per cent of household consumption is food.

A long history of demand analysis, dating back to Engel (1890), has shown that food is a necessity whose Engel curve is therefore downward sloping. The Bangladeshi data reflect this with a strongly declining food Engel curve, whose estimated slope with respect to the log of household expenditure is  $-0.12$ , with a  $t$ -test of  $-14$ , and 99 per cent of the sample with significant slopes. Food Engel curves are therefore different from clothing Engel curves in two important ways: 1) Food budget shares are large while clothing budget shares are small; and 2) Food Engel curves slope downwards while clothing Engel curves sometimes slope upwards, sometimes slope downwards, and are sometimes flat. Both of these differences suggest that food is a preferable candidate for our methods. Consequently, in our analysis below, we pay special attention to the difference—or lack thereof—between estimates of Bangladeshi resource shares based on clothing versus food Engel curves.

## 4.2 Resource shares

Estimated per-person resource shares,  $\eta_h^t/N_h^t$ , of men, women and children, are shown in Table 4, for the countries whose data pass our pre-tests. We report both the resource shares estimated at the mean of observed covariates,  $\bar{z}$ , and the mean of the resource shares evaluated at all  $z_h$ . For the former, we give the standard error and for the latter, the standard deviation.

In Albania, the estimated men’s and women’s per-person resource shares at the average  $z_h$  are 28 per cent and 24 per cent, respectively, with small standard errors, of 3 per cent. Because resource shares are nonlinear functions of estimated OLS regression coefficients, the estimate of resource shares at average  $z_h$  does not equal the average of estimated resource shares over all  $z_h$ . However, they are similar: the sample averages of the resource shares are 29 and 24 per cent, respectively, for men and women. Variation in estimated resource shares is driven by variation in observed covariates  $z_h$ . The standard deviation of these estimated resource shares are 44 and 36 per cent, indicating quite a lot of heterogeneity in resource shares driven by the sample variation in observed covariates.



Table 4: Predicted Resource Shares, Selected Countries

Country	sample size	Evaluated at $\bar{z}$			Evaluated at all $z_h$			$\eta$ outside [0,1]	per cap test Wald, df <i>p-value</i>
		men	women	children	men	women	children		
		est <i>std err</i>	est <i>std err</i>	est <i>std err</i>	mean <i>std dev</i>	mean <i>std dev</i>	mean <i>std dev</i>		
Albania	3279	0.2823 <i>0.0311</i>	0.2448 <i>0.032</i>	0.1372 <i>0.0304</i>	0.2911 <i>0.4448</i>	0.2416 <i>0.3564</i>	0.132 <i>0.1604</i>	0.0561	28, 30 <i>0.600</i>
Bangladesh	6120	0.3148 <i>0.0122</i>	0.2852 <i>0.0143</i>	0.1183 <i>0.0104</i>	0.3134 <i>0.1136</i>	0.2838 <i>0.1198</i>	0.1209 <i>0.0593</i>	0.0005	78, 38 <i>0.000</i>
Bulgaria	2099	0.3048 <i>0.0371</i>	0.3702 <i>0.0408</i>	0.1893 <i>0.0617</i>	0.2951 <i>0.1404</i>	0.382 <i>0.2182</i>	0.1771 <i>0.2078</i>	0.0734	36, 30 <i>0.204</i>
Iraq	14297	0.2685 <i>0.0092</i>	0.2362 <i>0.0114</i>	0.0413 <i>0.0059</i>	0.2678 <i>0.1336</i>	0.2354 <i>0.132</i>	0.042 <i>0.0698</i>	0.0124	103, 38 <i>0.000</i>
Malawi	10873	0.3116 <i>0.0281</i>	0.2738 <i>0.0302</i>	0.124 <i>0.0112</i>	0.3109 <i>0.2219</i>	0.2664 <i>0.1668</i>	0.1275 <i>0.0915</i>	0.0143	101, 38 <i>0.000</i>
Bangladesh Food	5604	0.3011 <i>0.035</i>	0.2393 <i>0.0149</i>	0.1731 <i>0.0104</i>	0.3061 <i>0.1161</i>	0.2329 <i>0.1114</i>	0.1751 <i>0.0795</i>	0.0328	135, 38 <i>0.000</i>

The rightmost column of table 4 gives the fraction of resource shares which fall outside of the [0,1] interval. The largest is Bangladesh (clothing) with 7%. The consequence of estimating our model on data where Engel curves are not very steep, that is, where the assignable good is neither very strongly a normal or inferior good would be large fraction of estimated shares outside [0, 1] and implausible estimates for resource shares.<sup>22</sup>

According to the point estimates, men get a larger share of household resources than women in all countries, except Bulgaria. Children get between 12% and 17% everywhere, except in Iraq where they get about 4% of resources each.

A standard resource share in current use by the World Bank and other agencies is the per-capita share of household members, that is,  $\eta_h^t = N_h^t / \sum_s N_h^t$ . This would assign each person their per-capita share of household consumption. Given our model, this obtains if  $b_z^t = \mathbf{0}$  for all  $t$  and  $b_{n^{t'}}^t = 0$  for all  $t' \neq t$  and  $b_{n^t}^t = \kappa$  for all  $t$ . The Wald test statistic for this hypothesis and its associated degrees of freedom are presented in the rightmost column of Table 4, with p-values in *italics* below.

Table 4 shows lots of inequality across household members, so it should not be surprising that the per-capita model is not supported by these estimates in most countries. The per-capita model is rejected in data from Iraq, Malawi and Bangladesh (for both clothing and food), but it is not rejected in Albania or Bulgaria. Notably, these latter two countries have the smallest samples by a factor of about 2. This suggests to us that rather large sample sizes are needed to estimate these models.<sup>23</sup>

<sup>22</sup>Estimates for all countries, even those where data do not pass the pretests, are available on request.

<sup>23</sup>We note that one can pool multiple waves of data for a given country, just by including a year dummy as

We will show below that the failure of the per-capita model implies the existence of both gender gaps in consumption and gendered poverty that has been missed by previous investigations.

### 4.3 Gender gaps

In Table 4, we see some evidence that women get smaller per-person resource shares than men. However, those estimates include all types of households, including those that don't have an adult man or those that don't have an adult women. To construct an estimated gender gap that refers strictly to within-household inequality, we present in Table 5 estimates on the subset of households that include both adult men and adult women. In the leftmost columns, we present the mean and standard deviation of estimated resource shares evaluated at all values of the covariates. In the right-hand columns, we present estimated resource shares, and their standard errors, for men and women evaluated at the average value of observed covariates. The difference between these two per-person resource shares is our gender-gap estimate, provided with standard errors, and 1, 2 or 3 stars to indicate statistical significance at the 10, 5 and 1 per cent level.

Table 5, Estimated Resource Shares and Gender Gaps, Selected Countries  
Households with Both Men and Women Present

	sample N	Evaluated at all $\mathbf{z}_h$		Evaluated at $\bar{\mathbf{z}}$		Gender Gap at $\bar{\mathbf{z}}$	
		men mean <i>std dev</i>	women mean <i>std dev</i>	men est <i>std err</i>	women est <i>std err</i>	est <i>std err</i>	sig
Albania	3279	0.3413 <i>0.3501</i>	0.2808 <i>0.253</i>	0.2823 <i>0.0311</i>	0.2448 <i>0.032</i>	0.0376 <i>0.0571</i>	
Bangladesh	6120	0.3500 <i>0.1108</i>	0.3003 <i>0.0915</i>	0.3148 <i>0.0112</i>	0.2659 <i>0.0107</i>	0.0489 <i>0.0199</i>	**
Bulgaria	2099	0.3171 <i>0.1501</i>	0.4435 <i>0.2116</i>	0.3048 <i>0.0371</i>	0.3702 <i>0.0408</i>	-0.0654 <i>0.0683</i>	
Iraq	14040	0.3346 <i>0.1387</i>	0.2871 <i>0.1217</i>	0.2685 <i>0.0092</i>	0.2327 <i>0.0086</i>	0.0358 <i>0.0166</i>	***
Malawi	9490	0.3633 <i>0.225</i>	0.2797 <i>0.1404</i>	0.3116 <i>0.0281</i>	0.2532 <i>0.0291</i>	0.0584 <i>0.0542</i>	
Bangladesh	4958	0.3330 <i>0.1204</i>	0.2328 <i>0.1008</i>	0.3011 <i>0.0135</i>	0.2216 <i>0.0127</i>	0.0795 <i>0.0240</i>	***
Food							

Here, we see that the evidence given in Table 4 that women have a greater share of household resources than men in Bulgaria is not a statistically significant finding. Because the

---

an additional element of  $\mathbf{z}$ . To do this, the model should include the additional restriction that the function  $\beta$  is price-independent (as in Deaton and Muellbauer 1980, but not as in Muellbauer 1974,1975).

estimates of men’s and women’s resource shares covary, the estimated 6.5 percentage point gender gap has a large standard error of 6.8 percentage points, even though the estimated resource shares of men and women have standard errors of only around 4 percentage points. Consequently, the difference between them—the gender gap—is statistically indistinguishable from zero.

The point estimates of the gender gap in Albania and Malawi are positive (3.8 and 5.4 percentage points, respectively), but are statistically insignificantly different from zero. In fact, we only see a statistically significant gender gap in Bangladesh and Iraq, and both of these show larger resource shares for men. The Iraqi data suggest a gender gap of 3.6 percentage points. In the Bangladeshi data, the estimated gender gap from assignable clothing data is larger, about 5.7 percentage points, and from the assignable food data, about 7 percentage points. The similarity between the estimates coming from clothing data and food data is striking: they are within about 1/2 a standard error of each other.<sup>24</sup>

## 4.4 Individual poverty

A standard poverty line used by the World Bank and other international organizations concerned with poverty is US\$1.90 per person per day (using PPP adjusted values). In Table 6, we measure poverty using that measure, and using our measures of resource shares. The poverty line is taken to be US\$1.90\*365=US\$693.50. In the leftmost column, we compute for each person in the household,  $y_h/N_h$ , compare this to the poverty line, and report the poverty rate.<sup>25</sup>

In the middle three columns, we compute for each man, woman and child in the dataset,  $y_h\eta_h^t/N_h^t$ , compare this to the poverty line, and report the poverty rate. Like DLP, we use a poverty line 40% lower for children (equal to US\$1.14 per day). In the final column, we report the overall poverty rate, at the person-level and using our resource shares, for the entire sample. Note that for these estimates, we include single-member households, where  $N_h^t = \eta_h^t = 1$ , and households with just one type of person (e.g., a two-man household), where each of the  $N^t$  people is assigned  $y_h/N_h^t$ . We provide asymptotic standard errors, computed via the bootstrap.<sup>26</sup>

---

<sup>24</sup>Additionally, if we estimate the food and clothing models together, under the (nonlinear) restriction that resource shares are the same, we find that the estimate is similar, and the test of whether the resource shares under the two models are different indicates that they are not statistically significantly different.

<sup>25</sup>This standard estimate of the poverty rate does not account for scale economies consumption. The OECD uses an alternative estimate, wherein household income is divided by a number less than  $N_h$  (for example,  $\sqrt{N_h}$ ) to account for the fact that members of large households can access scale economies. In this paper, we do not estimate scale economies, and so do not address this issue. Tractable estimation of scale economies in household consumption remains a task for future research.

<sup>26</sup>We bootstrap the standard errors (rather than using the delta method) because poverty rates are a

Table 6, Estimated Poverty Rates, Selected Countries					
country	per-capita	men	women	children	all people
	est	est	est	est	est
	<i>std err</i>	<i>std err</i>	<i>std err</i>	<i>std err</i>	<i>std err</i>
Albania	0.003	0.055	0.037	0.075	0.053
	<i>0.001</i>	<i>0.043</i>	<i>0.041</i>	<i>0.074</i>	<i>0.031</i>
Bangladesh	0.109	0.019	0.063	0.298	0.135
	<i>0.004</i>	<i>0.009</i>	<i>0.019</i>	<i>0.043</i>	<i>0.016</i>
Bulgaria	0.003	0.024	0.027	0.223	0.054
	<i>0.001</i>	<i>0.039</i>	<i>0.032</i>	<i>0.120</i>	<i>0.031</i>
Iraq	0.000	0.000	0.002	0.146	0.063
	<i>0.000</i>	<i>0.001</i>	<i>0.001</i>	<i>0.070</i>	<i>0.030</i>
Malawi	0.629	0.469	0.585	0.727	0.626
	<i>0.004</i>	<i>0.025</i>	<i>0.035</i>	<i>0.030</i>	<i>0.009</i>
Bangladesh	0.103	0.037	0.149	0.098	0.098
Food	<i>0.005</i>	<i>0.018</i>	<i>0.031</i>	<i>0.025</i>	<i>0.011</i>

Some variation in poverty across gender results from the sorting of men and women into households of different income levels. As a consequence, we see higher women’s poverty than men’s poverty in all countries, even Bulgaria, where women on average get higher resource shares than men. However, the key message here is that the variation across types in resource shares that we observed in Tables 4 and 5 translates directly into variation in estimated poverty rates across types. The point-estimates of the gender gap in resource shares are largest for Bangladesh, Malawi and Iraq. In these countries, we see higher women’s poverty than men’s poverty. In Bangladesh, women are 4 percentage points more likely to be poor than men; in Malawi, they are 12 percentage points more likely to be poor than men.

## 5 Discussion

We provide evidence of substantial within-household consumption inequality. This suggests that current standard practice for poverty measurement in developing countries—asking whether or not per-capita household income falls below a threshold—can be misleading. This current practice ignores within-household inequality, and so mischaracterises poverty rates. For example, if a household has income slightly above the poverty line, then by the per-capita method we would call it non-poor, but even a small amount of within-household inequality will result in some members being poor. Further, within-household inequality may be biased against certain groups. Among the 5 countries for which we estimate resource shares, we see statistically significant gender gaps in resource shares that favour men over

---

discontinuous function of the estimated resource shares, which are themselves nonlinear functions of estimated OLS regression coefficients. For an alternative, see Wouterson and Ham (2013).

women in two countries and we see no statistically significant evidence of gender gaps that favour women. Further, these gaps in resource shares result in gender gaps in poverty rates.

If within-household inequality is real, and affects the incidence of poverty among men, women and children, then its accurate measurement is of paramount importance. Our work suggests that Statistical agencies, and the World Bank programs they work with, should focus more data gathering effort on assignable goods. There are two strategies available here. First, resources could be directed to gathering assignable person-level consumption flows for all categories of goods and services (aka: dream data in Table 1). With these data, we would not need a structural model such as ours to estimate resource shares—we could measure them directly. Second, resources could be directed to gathering assignable consumption flows for 1 or 2 categories of goods and services that can be measured well and which represent a large fraction of total household expenditure. With these data, we could estimate resource shares using our structural model (or any household model that bases identification on assignable goods).

Our estimates of resource shares, gender gaps and poverty rates for Bangladesh come from two different assignable goods. We use clothing, which is roughly 4 per cent of the household budget, and food, which is roughly 56 per cent of the household budget. Clothing has a venerable history as an assignable good used in this literature (e.g., survey of Donni and Molina 2018; Calvi 2019; etc). However, the use of clothing is due to its availability in public-use datasets, not to its superiority in other ways.

We find in our work that using food data as an assignable good to identify resource shares delivers estimates that are very similar to those generated from clothing data. But, food data have five advantages over clothing data. First, food is more plausibly assignable than is clothing. Clothing can be handed down from member to member, but the same food cannot be eaten by two members. Second, food consumption is typically measured in quantities (like grams of legumes), whereas clothing is measured in currency units of expenditure. So, there may be more unobserved quality heterogeneity in clothing than in food. Third, food budget shares are known to be downward sloping (e.g., Engel 1857, 1895), and therefore satisfy the identifying restriction of our model. Fourth, clothing is much more durable than food. Consequently, observed clothing expenditure may not equal clothing consumption, due to infrequency of purchase. Fifth, food shares are typically much larger than clothing shares. This is not a gain in terms of the model in any formal sense, but it does seem like a worthwhile auxiliary feature. All together, this suggests that statistical agencies and the World Bank should focus significant data gathering resources on the collection of person-level food consumption.

## 6 Conclusions

We show how to estimate the resource share of each person in a collective household via simple linear regressions of assignable goods Engel curves. This may be implemented with off-the-shelf consumer expenditure micro-data, such as that collected through the World Bank's Living Standards Measurement Study. We apply the model to data from 12 countries, and investigate resource shares, gender gaps and individual poverty. We find that equal sharing--the implicit assumption underlying household-level poverty calculations---is rejected. We also find evidence of large gender gaps in resource shares, and consequently in poverty rates, in a few countries.

## 7 References

Anderson, S. and Ray, D., 2010. Missing women: age and disease. *The Review of Economic Studies*, 77(4), pp.1262-1300.

Bargain, O., Donni, O. and Kwenda, P., 2014. Intrahousehold distribution and poverty: Evidence from Cote d'Ivoire. *Journal of Development Economics*, 107, pp.262-276.

Bargain, O., Lacroix, G. and Tiberti, L., 2018. Validating the collective model of household consumption using direct evidence on sharing. *Partnership for Economic Policy Working Paper*, (2018-06).

Blundell, R., Chen, X. and Kristensen, D., 2007. Semi-nonparametric IV estimation of shape-invariant Engel curves. *Econometrica*, 75(6), pp.1613-1669.

Brown, C. M. Ravallion and D. van de Walle. 2019. Are poor Individuals Mainly Found in Poor households?, forthcoming, *the Review of Economics and Statistics*

Browning, M., Chiappori, P.A. and Lewbel, A., 2013. Estimating consumption economies of scale, adult equivalence scales, and household bargaining power. *Review of Economic Studies*, 80(4), pp.1267-1303.

Bütikofer, A., 2011. Sibling sex composition and cost of children. Unpublished manuscript, Department of Economics, Norwegian School of Economics, Bergen, Norway.

Calvi, R., 2019. Why are older women missing in india? the age profile of bargaining power and poverty. forthcoming, *Journal of Political Economy*.

Calvi, R., Lewbel, A. and Tommasi, D., 2017. LATE with Mismeasured or Misspecified Treatment: An Application to Women's Empowerment in India.

Campbell, F., Conti, G., Heckman, J.J., Moon, S.H., Pinto, R., Pungello, E. and Pan, Y., 2014. Early childhood investments substantially boost adult health. *Science*, 343(6178), pp.1478-1485.

Cherchye, L., De Rock, B., Lewbel, A. and Vermeulen, F., 2015. Sharing rule identification for general collective consumption models. *Econometrica*, 83(5), pp.2001-2041.

Cherchye, L., Demuynck, T., De Rock, B. and Vermeulen, F., 2017. Household consumption when the marriage is stable. *American Economic Review*, 107(6), pp.1507-34.

Chiappori, P.A., 1992. Collective labor supply and welfare. *Journal of political Economy*, 100(3), pp.437-467.

Chiappori, P.A. and Ekeland, I., 2009. The Microeconomics of Efficient Group Behavior: Identification 1. *Econometrica*, 77(3), pp.763-799.

Chiappori, P.A. and Meghir, C., 2015. Intrahousehold inequality. In *Handbook of income distribution* (Vol. 2, pp. 1369-1418). Elsevier.

Deaton, A. and Muellbauer, J., 1980. *Economics and consumer behavior*. Cambridge university press.

Donni, O. and Molina, J.A., 2018. Household Collective Models: Three Decades of Theoretical Contributions and Empirical Evidence.

Doss, C., 2013. Intrahousehold bargaining and resource allocation in developing countries. *The World Bank Research Observer*, 28(1), pp.52-78.

Dunbar, G.R., Lewbel, A. and Pendakur, K., 2013. Children's resources in collective households: identification, estimation, and an application to child poverty in Malawi. *American Economic Review*, 103(1), pp.438-71.

Engel, E (1857), Die Productions- und Consumtionsverhältnisse des Königreichs Sachsen" in *Zeitschrift des Statistischen Bureaus des Königlich-Sächsischen, Ministerium des Innern*, No. 8 u. 9, 1-54. Reprinted as in Engel E. (1895) "Die Lebenskosten Belgischer Arbeiter Familien früher und jetzt," *Bulletin de l'institut international de statistique*, tome IX, premiere livraison, Rome.

Han, S. and McCloskey, A., 2019. Estimation and inference with a (nearly) singular Jacobian. *Quantitative Economics*, 10(3), pp.1019-1068.

Heckman, J.J., 2006. Skill formation and the economics of investing in disadvantaged children. *Science*, 312(5782), pp.1900-1902.

Menon, Martina, Krishna Pendakur and Federico Perali, 2012, "On the Expenditure-Dependence of Resource Shares", *Economics Letters* 117 (2012), pp. 739-742.

Menon, M., Perali, F. and Piccoli, L., 2018. Collective consumption: An application to the passive drinking effect. *Review of Economics of the Household*, 16(1), pp.143-169.

Pendakur, K. 1999. "Estimates and Tests of Base-Independent Equivalence Scales", *Journal of Econometrics* 88(1), pp 1-40.

Pendakur, K. 2018. "Welfare Analysis When People Are Different", *Canadian Journal of Economics*, May, 2018.

De Vreyer, P. and Lambert, S., 2016. Intra-household Inequalities and Poverty in Senegal. mimeo, Paris School of Economics.

Woutersen, T. and Ham, J.C., 2013. Calculating confidence intervals for continuous and discontinuous functions of parameters (No. CWP23/13). cemmap working paper.

## 8 Appendix 1: Dream Data with Scale Economies

Suppose that there are scale economies as in Browning, Chiappori and Lewbel (2013: BCL), with shadow prices that are linear in market prices  $\mathbf{p}$ , where shadow prices equal  $\mathbf{A}\mathbf{p}$  for some diagonal matrix  $\mathbf{A}$ . ( $\mathbf{A}$  can depend on observed household characteristics like household size, but we suppress that dependence here.) An element of  $\mathbf{A}$  says how shareable a good is. If it is 1, then the good is not shareable; if it is less than 1, the good is shareable. The essence of this model of scale economies is that if individuals demand the vector of quantities  $\mathbf{q}_j$ , the household can satisfy all these demands with a market purchase of the quantity vector  $\mathbf{A}\sum_j \mathbf{q}_j$ . For example, in a 3 member household, the value of the element of  $\mathbf{A}$  corresponding to a perfectly shareable good might be 1/3. This means that the household could deliver a quantity  $q$  to each of the 3 people in the household with a market purchase of only  $q$ .

Table A1a gives dream data about the quantities consumed by each person in a world where scale economies are governed by the model. In this world, we observe for each person in the household the quantity of the good that they personally got to consume. Normalize market prices to 1 for all goods, so that we can think of consumed quantities as measured in dollars. We additionally observe the total expenditure of the household on each good.

Table A1a: Dream Data: Quantities

	Quantities			Expend.
	Man	Woman	Child	Total
Food	400	300	200	900
Clothing	50	75	25	150
Shelter	300	300	300	300
Other	500	250	250	500
Total				1850

Table A1b: Dream Data, Expenditure

	Expenditure			
	Man	Woman	Child	Total
Food	400	300	200	900
Clothing	50	75	25	150
Shelter	100	100	100	300
Other	250	125	125	500
Total	800	600	450	1850

For nonshareable goods (food and clothing in this example), the total expenditure of the household is simply the sum of the individual quantity levels (prices are normalized to 1). However, for goods that are shared, this is not the case. In this example, shelter is considered to be a fully shared good. Here, we have that each member reported that they personally consumed \$300 worth of shelter. But, because shelter is fully shared, the household only had to purchase \$300 of housing to accomplish this. This means that the household purchased



only  $1/3$  of the total housing consumption of the 3 members. It is as if the household was able to scale its housing spending up by a factor of 3, and then each member bought housing as a private good out of this scaled purchase. Consequently, we identify the matrix  $\mathbf{A}$  from these data: the element of  $\mathbf{A}$  corresponding to shelter is  $1/3$ , because the household only needs to buy  $1/3$  of the total consumed quantities of all the members.

Goods do not have to be either fully shared or non-shareable in the BCL model; they can be partly shared. Suppose that “other” is transportation, and that transportation costs are for riding a motorcycle. The individual-level quantities in Table A1a are the individual-level numbers of km ridden and the household purchased quantity would be the total number of kilometers shown on the odometer. The sum of the former would exceed the latter, because sometimes people ride together. Suppose the man is the only member who knows how to drive a motorcycle. If the man rode 250km with the woman and 250km with the child, then their consumed quantities would be as in Table A1a, with 1000 person-km driven. But, the motorcycle would only have travelled 500km, so the household would have purchased only 500km. Here, the element of  $\mathbf{A}$  corresponding to transport (other) would be  $1/2$ .

In Table A1b, we turn individual-level quantities into individual level expenditures by multiplying quantity by price. Since market prices  $\mathbf{p}$  are normalized to 1, within household prices given by  $\mathbf{Ap}$ , this means we multiply by the diagonal matrix  $\mathbf{A}$ . Since nonshareable goods have an element of  $\mathbf{A}$  equal to 1, for the nonshareable goods of food and clothing, the rows of Table A1b are identical to the rows of Table A1a. The elements of  $\mathbf{A}$  for shelter and other, respectively, are  $1/3$  and  $1/2$ . So, for shelter, we multiply by  $1/3$  and for other, we multiply by  $1/2$ . This yields Table A1b which gives the expenditure of each person on each good. These can be summed down columns to yield the total expenditure of each person, and these person-level total expenditures add up to household-level total expenditure in the bottom right corner.

Scale economies in the BCL model are thus driven by the matrix  $\mathbf{A}$  which scales prices. We like scale economies because we like low prices. The value of scale economies is just the cost of living index corresponding to the difference between facing a price vector  $\mathbf{p}$  and facing a price vector  $\mathbf{Ap}$ . BCL show how to identify resource shares and the matrix  $\mathbf{A}$  from knowledge of individual demand vector functions for all goods and household demand vector functions from all goods (as functions of prices and budgets).

DLP do not attempt to identify the matrix  $\mathbf{A}$ . Instead, they show how to identify just the resource shares from knowledge of just household Engel curve functions (without price variation) for assignable goods, where the assignable goods are assumed to be non-shareable. The model of DLP does not make any assumptions about how shareable the non-assignable goods are. In terms of the matrix  $\mathbf{A}$ , DLP assume that the single element of  $\mathbf{A}$  corresponding

to the non-shareable assignable good equals 1, and make no assumptions about the other elements of  $A$ .

Although the model of DLP is not affected by whether or not scale economies are assumed to exist, the characteristics of the dream data are affected by this assumption. In particular, if we want to identify scale economies *as well as* resource shares directly from data, then such data must provide (at least) the individual-level experienced quantities of each good as well as household level expenditure on these goods.

The matrix  $A$  governs scale economies and is relevant to poverty calculations. The standard tool used to estimate poverty in developing countries is to compare per-capita income to a poverty threshold of US\$1.90 per day. The assumption on scale economies underlying this strategy is that there are no scale economies. If we took scale economies seriously in the measurement of poverty, we would scale up household consumption by the matrix  $A$  to give an estimate of the total consumption of all people in the household. If we then take within-household inequality seriously in the measurement of poverty, we would multiply this scaled household consumption by the resource share of each person, and compare this quantity to the poverty threshold of US\$1.90 per day. This paper deals with only the latter issue. Simple estimation tools to recover scale economy parameters in household models remain an important issue for future research.