Institute for Fiscal Studies

OLS Estimation of the Intra-Household Distribution of Consumption

IFS Working Paper W19/19

Valérie Lechene
Krishna Pendakur
Alex Wolf

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

OLS Estimation of the Intra-Household Distribution of Consumption

Valérie Lechene, Krishna Pendakur, Alex Wolf*

July 2, 2019

Abstract

Individuals may be poor even if their household is not poor, because the intra-household distribution of resources may be unequal. We develop a model wherein the *resource share* of each person in a collective household---defined as their share of household consumption---may be estimated by simple linear regressions using off-the-shelf consumer expenditure micro-data. The model is a linear approximation of Dunbar, Lewbel and Pendakur (2013), whose nonlinear structural model can be computationally difficult. Our model allows for complex household types, including those with multiple adult men and/or women and single parent households. We also provide a simple linear pre-test to check for model identification.

Resource shares are obtained as nonlinear functions of estimated coefficients from OLS regressions. 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 always rejected. We also find evidence of large gender gaps in resource shares, and consequently in poverty rates, in a few countries.

Individuals have indifference maps; households don't, because households are economic environments in which individuals live. Individuals obtain utility from consumption and thus poverty is experienced by individuals not households. The measurement of consumption, and of the extent and composition of poverty, should therefore be done at the individual level. The measurement of gender gaps in consumption, which summarise intra-household gender

*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: Geoffrey Dunbar and Arthur Lewbel.

inequality, also requires the measurement of consumption at the individual level. Policy targeting of poor individuals, monitoring of movements in and out of poverty and evaluation of policies designed to reduce poverty would also be better if we measured poverty at the individual level.

The Millenium Development Goals include the promotion of gender equality and the empowerment of women. Achieving this goal requires measuring the consumption of women, as opposed to the consumption of households. The World Bank has set two goals for the world to achieve by 2030: to end extreme poverty by decreasing the percentage of people living on less than $1.90 a day to no more than 3%; and to promote shared prosperity by fostering the income growth of the bottom 40% for every country. A pre-requisite to ending extreme poverty is its measurement.

An ingredient to measuring the within-household distribution of consumption is the "resource share", defined as the fraction of total household consumption consumed by each member.[1] If, for example in a nuclear household, the woman has a smaller resource share than the man, then there is gender inequality in consumption. 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 consumption, and compares that to US$1.90 per day. A strategy that respected the idea that consumption lives at the individual level would instead assign each person their (resource) share of household consumption and compare that to US$1.90 per day.

The use of these types of methods is consequential, and may help us understand other phenomena. For example, Calvi (2019) estimates poverty at the individual level in India and finds that women have higher poverty rates than men, especially older women. Further, Calvi shows that this higher poverty rate among older women explains the quantitative finding of Anderson and Ray (2010) that 1.7 million woman over age 45 in India are "missing" from the the expected population.

Although intuitively compelling, this turns out to be hard to implement. Indeed, it is relatively straightforward to collect information on household expenditure, but not on individual consumption. For instance, we may know that the household bought a bottle of milk, but we may not know who drank the milk. Collecting information on the allocation of consumption inside the household is costly and prone to measurement error (e.g, via the Hawthorne effect). Furthermore, there are goods with different degrees of shareability inside

[1] The resource share is not the only an ingredient in measuring intra-household distribution of consumption. Knowledge of scale economies and sharing is also an important element.

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.

Building on the collective model developed by Chiappori (1992), work by Browning, Chiappori and Lewbel (2013: BCL) and Dunbar, Lewbel and Pendakur (2013: DLP) introduces 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 (BCL) or observed data on assignable goods (DLP), defined as goods consumed by a single known person in the household (e.g., men's clothing), to identify resources shares. But they require estimation 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 approximation to the structural model of Dunbar, Lewbel and Pendakur (2013) 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 consumption spent on a good against total household consumption.) In this model, resource shares are identified by the slopes of these Engel curves: if the slope of a person's assignable good Engel curve is larger (in absolute value), then their resource share is larger. We note that in our approximation of DLP, we approximate only level terms; the slope terms are specified exactly.

Empirically, resource shares are obtained in two steps. First, we run OLS regressions of observed household-level spending on assignable goods (as a fraction of total household consumption) on observed household demographics and total household consumption. Then, we obtain resource shares as nonlinear functions of estimated slope coefficients from these regressions.

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. This is important, as household compositions are varied, and limiting the analysis of poverty to nuclear households risks missing many households. Finally, we provide a pre-test that indicates whether the model will work in the dataset at hand. Essentially, our pre-test is: are the Engel curves flat? If they are flat, then you cannot use these methods because the identifying variation comes from the slopes of Engel curves.

We use the model to estimate resource shares and individual poverty rates (including women's poverty and children's poverty) in 12 countries, using household surveys from the World Bank LSMS data, and 1 other national survey from Bangladesh. We use person-level clothing consumption as the assignable good, and use this to test the model and estimate resource shares where possible. We show that neglecting intra-household inequality severely biases poverty estimates in some countries. We find that equal sharing—the implicit assumption underlying standard household-level poverty calculations—is always rejected and that this implies higher rates of female and child poverty in the countries we study. Finally, we show that there are significant gender gaps in some countries, and that these contribute significantly to women's poverty.

Our data from Bangladesh also have person-level food consumption, which we use to estimate resource shares. We find that using food data to identify resource shares delivers estimates that are very similar to those generated from clothing data. Since food consumption is larger than (and possibly better measured than) clothing consumption, this suggests that statistical agencies and the World Bank should focus data gathering resources on the collection of person-level food consumption.

In section (1), we briefly review the theoretical foundations that underpin our work. We discuss the identification of resource shares using assignable goods; the model specification and finally we show how it can be interpreted as a linear approximation of DLP's model. We then present the off the shelf household survey data we use, and we focus on features of the data that inform our modeling choices in section (2). In this section, we highlight the diversity of household compositions and describe the information available in off the counter household surveys for the twelve countries we analyse. We present results in section (3) and we discuss individual poverty in section (3.4). We conclude with a brief discussion of the implications of our work for ongoing data collection efforts.

1 Theory

Dream data to measure the consumption of individuals within households would like Table 1a. Here, we directly observe the consumption of 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 consumption 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) consumption level of each person, and can compare individual consumption levels to individual poverty thresholds. The individual poverty threshold is \$1.90 per day, equalling \$694. We can observe that the man consumes \$800 which exceeds the poverty threshold, so he is not poor. However, the woman and child both consume less than the \$694 poverty threshold, so they are both poor.

Table 1a: Dream Data

	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

Table 1b: Real Data

	Man	Woman	Child	Total
Food				900
Clothing	50	75	25	150
Shelter				300
Other				500
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. Unfortunately, to our knowledge, this is the only case where individual-level consumption data for all consumption categories is collected.

Real-world consumption data tend to look more like Table 1b. In this type of data, we see *household-level* consumption for all the goods and services comprising total consumption, 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 Study (LSMS) of the World Bank. So, with real-world data, we face a missing data problem: we do not have full data on individual consumption; instead, we have data on just 1 or 2 commodities collected at the individual level.

BCL and DLP aim to address this missing data problem via modeling the allocation problem of the household, and "backing out" the resource shares from these incomplete data. In doing so, they also allow for further complications, such as the possibility that some goods are shared. In the context of Table 1a, this would mean that to achieve a shelter consumption of \$100 each for the man, woman and child, the household may only need to purchase \$100 (not \$300) worth of shelter. Or, suppose that "other" is motorcycle fuel, which is partly shared. E.g, it might be that the man rides the motorbike with the woman half the time, and with the child half the time. In this case, to achieve consumption of \$250 for the man and \$125 each for the woman and child, the household would only need to purchase \$250 (not \$500) worth of motorcycle fuel.

1.1 An Efficient Collective Household Model

BCL provide a 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. Here, we briefly sketch those models, and show a linear approximation of DLP that may be estimated by ordinary least squares. Further, we will extend the model to allow for households of complex composition (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.

Let $h = 1, ..., H$ 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 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 a type t person is called their *resource share*, denoted η_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 $\eta_h^t y_h / n_h^t$.[2]

Shadow prices for goods are the within-household prices of consumption. Let $\boldsymbol{p}$ denote the market price vector for goods and let $\widetilde{\boldsymbol{p}}$ 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 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.

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). [3] 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 a *private good* to be a good that is not shared, and an *assignable good* as one where we can observe which household member consumes the good. Private assignable goods are very useful for identification (see, e.g., Chiappori and Ekeland 2009).

In our context, the available data on assignable goods will define our typology of individ-

[2]Dunbar, Lewbel and Pendakur (2013) allowed for multiple children. Here, we extend that notation to allow for multiple members of any type.

[3]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 that use observed price variation can speak to scale economies (e.g., BCL, described below).

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

BCL show that, given the model described above, household demands are related to individual demands very simply. 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 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 commodity is the fraction of the overall budget (spent on all commodities) commanded by that commodity (Engel 1895).[4] 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 BCL on Engel curves for private assignable goods.

Let $\eta^t(y)$ be the resource share of person t when the household faces a fixed market price vector $\boldsymbol{p}$.[5] Assume that shadow prices are linear in market prices, with $\widetilde{\boldsymbol{p}}_h = \boldsymbol{A}_h \boldsymbol{p}$ for some diagonal matrix $\boldsymbol{A}_h$. BCL provide identification results for more general relationships, but this restriction substantially simplifies Engel curves.

Let a private 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 $\widetilde{\boldsymbol{p}}$ and a budget y. The price vector is held constant at the shadow price vector $\widetilde{\boldsymbol{p}}$.[6] Since the only demander for this good is person t, the household Engel curve for assignable good t, evaluated at the market

[4]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".

[5]The resource share may also depend on other covariates, including but not limited to those which affect preferences, but we suppress that dependence here.

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

price vector $\boldsymbol{p}$, is given by:

$$W^t(y) = \eta^t(y)\, w^t\left(\eta^t(y)\, y/n^t\right). \tag{1}$$

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 shares of the relevant people times their Engel curves (at shadow prices, held fixed).[7]

BCL 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 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 impose: a) resource shares do not depend on the household budget so that $\eta^t(y) = \eta^t$; b) that 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) that preferences are similar—but not identical—across people, such that $\beta^t = \beta$.[8] Substituting these assumptions into (1) gives

$$W^t(y) = \eta^t\left(\alpha^t + \beta\left(\ln y + \ln \eta^t - \ln n^t\right)\right) \tag{2}$$

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 relative consumption distribution will not change.[9] Surprisingly, there is some empirical support for this restric-

[7] For a longer exposition of the implications of assignable goods in the BCL model, and for a story of how BCL connects to DLP13, see Pendakur (2018). For more strategies to identify resource shares from Engel curve data, see Dunbar, Lewbel and Pendakur (2018).

[8] 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). DLP assume that: BCL holds; $\widetilde{\boldsymbol{p}} = Ap$; 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 as in the Almost Ideal demand system for identification. They (and we) use loglinear Engel curves to make estimation easier.

[9] 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

tion. 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 η^t sum to 1, the sum of this semi-elasticity across types is β. 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 either necessities or inferior goods (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 η^t multiplies β, and requires positive resource shares, due to the $\ln \eta^t$ term. 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 η^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 an approximation strategy that retains the spirit of our identification but 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.

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.

1.2 Linear DLP

We now present our model, a theory-consistent linearisation of DLP, extended to accommodate multiple household types and demographic characteristics. Rewrite equation (2) with a subscript h on all observed variables, and an additive error term ε_h^t, as follows

$$W_h^t = a_h^t + b^t \ln y_h + \varepsilon_h^t \tag{3}$$

where

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

and

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

Let us approximate the a_h^t term with

$$a_h^t = a_0^t + a_n^t \ln n_h^t.$$

This approximation soaks up the troublesome $\ln \eta^t$ term into the parameter a_0^t, and so solves the problem of taking the log of a negative and crashing the estimator.[10] The model may be estimated by linear regression of the observed household-level assignable good expenditure share, W_h^t, on a constant, the number of members of type t, 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 n_m lny) (W_f n_f lny) (W_c n_c lny)
```

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

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

If $\beta = 0$, then the estimated value of the denominator may be close to 0, yielding "crazy" estimates of resource shares. Hence, $\beta \neq 0$ is an identifying restriction. Note that we have only used an approximation for the level term, a_h^t, and that this estimator for resource shares does not depend on that term. One could implement this estimator for, e.g., η^m in Stata via:

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

The model above does not include any conditioning variables, such as demographic prefer-

[10]One could substitute the restriction that $a_1^t = b_t^t$, but it is not necessary to do so.

ence shifters. Including them does not affect identification, but does require some additional notation. Let $\boldsymbol{z}$ be all variables that affect preferences, including the numbers of household members of each type $\boldsymbol{n} = \{n^t\}$. Let $\widetilde{\boldsymbol{z}}$ be the subvector $\boldsymbol{z}$ of that excludes n: $\boldsymbol{z} = [\boldsymbol{n} \quad \widetilde{\boldsymbol{z}}]$. Let $\widetilde{\boldsymbol{z}} = 0$ for some meaningful reference value of these characteristics. Assume that preference parameters α^t and β depend on $\boldsymbol{z}$. Let $\boldsymbol{d}$ be so-called *distribution factors* that affect resource shares but not preferences; note that $\boldsymbol{d}$ and $\widetilde{\boldsymbol{z}}$ can be empty. Let $\boldsymbol{m} = [\boldsymbol{z} \quad \boldsymbol{d}] = [\boldsymbol{n} \quad \widetilde{\boldsymbol{z}} \quad \boldsymbol{d}]$ be the vector of all variables affecting resource shares η^t. Substituting this into (2), and expanding out the terms, we have:

$$W^t(y, m) = \eta^t(\boldsymbol{m})\alpha^t(\boldsymbol{z}) + \eta^t(\boldsymbol{m})\beta(\boldsymbol{z}) \ln y + \eta^t(\boldsymbol{m})\beta(\boldsymbol{z}) \ln \eta^t(\boldsymbol{m}) - \eta^t(\boldsymbol{m})\beta(\boldsymbol{z}) \ln n^{\mathrm{t}}, \qquad (4)$$

for households with members of type t. (If the household has no members of type t, $W_h^t = 0$.) The nonlinear structural model (4) has been implemented by several researchers on data from several countries (e.g., DLP in Malawi; Bargain, Donni and Kwenda 2014 in Cote D'Ivoire; Calvi 2014 in India; De Vreyer and Lambert in Senegal).

As with equation (2), this model contains a term linear in the log of the resource share, $\ln \eta^t(\boldsymbol{m})$. If η^t is parameterised as a linear index (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. For example, they may try to evaluate the function in a region of the parameter space where $\eta^t(\boldsymbol{m})$ is negative, yielding a missing value for $\ln \eta^t(\boldsymbol{m})$. Further, the model is 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 $\eta^t(\boldsymbol{m})\beta(\boldsymbol{z}) \ln y = \eta^t(\boldsymbol{d}, \boldsymbol{z})\beta(\boldsymbol{z}) \ln y$ has quadratic interactions in $\boldsymbol{z}$ multiplying $\ln y$. These make it difficult to precisely identify the dependence of resource shares $\eta^t(\boldsymbol{n}, \boldsymbol{z})$ on $\boldsymbol{z}$, because $\boldsymbol{z}$ affects both η^t and β.

Rewrite equation (4) with a subscript h on all observed variables, and an additive error term ε_h^t, as follows:

$$W_h^t = a_h^t + b_h^t \ln y_h + \varepsilon_h^t, \qquad (5)$$

where

$$a_h^t = \eta^t(\boldsymbol{m}_h)\alpha^t(\boldsymbol{z}_h) + \eta^t(\boldsymbol{m}_h)\beta(\boldsymbol{z}_h) \ln \eta^t(\boldsymbol{m}_h) - \eta^t(\boldsymbol{m}_h)\beta(\boldsymbol{z}_h) \ln n_h^{\mathrm{t}}, \qquad (6)$$

and

$$b_h^t = \eta^t(\boldsymbol{m}_h)\beta(\boldsymbol{z}_h). \tag{7}$$

Let us approximate the a_h^t term with

$$a_h^t = a_0^t + \boldsymbol{a}_m^{t\prime}\boldsymbol{m}_h = a_0^t + \boldsymbol{a}_n^{t\prime}\boldsymbol{n}_h + \boldsymbol{a}_{\tilde{z}}^{t\prime}\tilde{\boldsymbol{z}}_h + \boldsymbol{a}_d^{t\prime}\boldsymbol{d}_h. \tag{8}$$

This approximation soaks up the troublesome $\ln\eta^t$ term into the parameter vector $\boldsymbol{a}_m^{t\prime}$, and so solves the problem of taking the log of a negative and crashing the estimator.[11]

Suppose that η^t and β are linear indices in $\boldsymbol{m}_h$ and $\boldsymbol{z}_h$, respectively. Then, b_h^t is quadratic in $\boldsymbol{m}_h$. In this case, we don't need to approximate the slope term b_h^t, because it is exactly given by a quadratic. Therefore, OLS regression of W_h^t on a constant, $\boldsymbol{m}_h, \ln y, \boldsymbol{m}_h \ln y$ and $(\boldsymbol{m}_h\boldsymbol{m}_h')\ln y$ would correspond to an approximation of the level term with the exact specification of the $\ln y$ term. Alternatively, both η^t and β could have unknown functional forms. In this case, one could let b_h^t be a nonparametric function of $\boldsymbol{m}_h$, and use standard semiparametric methods to estimate the model. One such approach would be to let b_h^t be a multivariate polynomial over $\boldsymbol{m}_h$, with the degree of the polynomial increasing with the sample size.

In practice, it may be impractical to specify b_h^t as a high-order polynomial over $\boldsymbol{m}_h$. So, we recommend approximating the slope term analogously to the level term as

$$b_h^t = b_0^t + \boldsymbol{b}_m^{t\prime}\boldsymbol{m}_h = b_0^t + \boldsymbol{b}_n^{t\prime}\boldsymbol{n}_h + \boldsymbol{b}_{\tilde{z}}^{t\prime}\tilde{\boldsymbol{z}}_h + \boldsymbol{b}_d^{t\prime}\boldsymbol{d}_h, \tag{9}$$

where $\boldsymbol{b}_s^t$, $\boldsymbol{b}_{\tilde{z}}^t$ and $\boldsymbol{b}_d^t$ refer to the relevant subvectors of $\boldsymbol{b}_m^t$. We note that this approximation is exact if η^t is linear in $\boldsymbol{m}_h$ and β is independent of $\boldsymbol{z}_h$ (that is, if β is a constant).

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 η^t and β), 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 $\beta(\boldsymbol{z}_h)$, which implies that an estimate of the resource share of type t in a household with characteristics $\boldsymbol{m}_h$ is given by

$$\hat{\eta}_h^t = \hat{\eta}^t(\boldsymbol{m}_h) = \hat{b}_h^t / \left(\sum_{t=1}^{T} \hat{b}_h^t\right). \tag{10}$$

As before, if $\beta(\boldsymbol{z}_h) = 0$, 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

[11] One could additionally add $\ln\boldsymbol{n}_h$ to the linear index, and in that case, impose the restriction that $a_n^t = b_h^t$. But it is not necessary to do so, since this is an approximating term.

would have very high variance (be unstable). Consequently, it is valuable to have a pre-test to tell us whether or not these methods will work at all. Let the overall clothing 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

$$W_h = a_h + b_h \ln y_h + \varepsilon_h \tag{11}$$

and OLS regression of W_h on $1, \boldsymbol{m}_h, \ln y_h$ and $\boldsymbol{m}_h \ln y_h$ yields an estimate $\widehat{b}_h$ of $\beta(\boldsymbol{z}_h)$. 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 $E[b_h] = \widehat{b}_0 + \widehat{\boldsymbol{b}}'\overline{\boldsymbol{m}}_h$, where $\overline{\boldsymbol{m}}_h$ is the sample average of $\boldsymbol{m}_h$, as a test statistic. This is a test of the economic hypothesis that the overall clothing share Engel curve, evaluated at the mean value $\boldsymbol{m}_h$, is either a normal or inferior good. 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 $\widehat{b}_h = \widehat{b}_0 + \widehat{\boldsymbol{b}}'\boldsymbol{m}_h$ 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 (of course, other cutoffs could be used).

Resource shares may then be computed via (10). From a practical standpoint, if the denominators in (10) 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

$$\sum_t \boldsymbol{b}_{\tilde{z}}^t = \sum_t \boldsymbol{b}_d^t = \boldsymbol{0}. \tag{12}$$

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

$$\widehat{\eta}^t(\boldsymbol{m}_h) = \frac{b_h^t}{\sum_t (b_0^t + b_{n^m}^t n_h^m + b_{n^w}^t n_h^w + b_{n^c}^t n_h^c)}.$$

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

A final note on functional form is 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 $\boldsymbol{b}_{\tilde{z}}^t = \boldsymbol{b}_d^t = \boldsymbol{0}$ 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, $\eta^t(\boldsymbol{m}_h) = n^t / \sum_t n^t$.

This model may be estimated by ordinary least squares (OLS), or with seemingly unrelated regression (SUR). To deal with the issue that some households do not have all types of people, we recommend estimation of resource shares via separate regressions for households with different strata of types. For example, to compute resource shares for people living in households with men, women and children, 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. In our work below, we consider 4 strata 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.

2 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). [12][13]

[12]We discuss the impact of omitting relevant characteristics on the estimates in the appendix. This is not the usual omitted variable case, since here variables would be omitted from the slope, rather than as is usually the case, from the intercept.

[13]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.

Table 2: Descriptive Statistics

Country	total N	single N	compositions	Our N	Nuclear N	budget	std dev
Albania	3599	239	mw, mwc	3279	612	11084	6477
Bangladesh	6434	143	mw, wc, mwc	4288	1472	6460	6879
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	143	mw, wc, mwc	3929	1330	6511	6969

Descriptive statistics for the sample of countries are in table 1. Altogether, these countries represent nearly 9% of the world population. Starting from the publicly available LSMS data (and the Bangladesh data), for each country, we reject 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, reflecting both variation in the LSMS survey samples and differences in the quality of the data. 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 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.[14]

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*, *mwc*, *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 N" (5) gives the total number of observations of these compositions. This latter column shows that we are able to exploit most of the data. The samples for

[14]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.

Albania, Bulgaria and Tajikistan contain households with men and women, and any number of children, but no single parents. In Bangladesh, Ethiopia, Ghana, Iraq, Malawi, Nigeria, Tanzania, Timor Leste and Uganda, as well as mw and mwc, there are households with women and children, wc. Note that these can have any strictly positive number of women and of children. Finally, only in the Ghanaian data are there more than 100 households with men and children, so that all compositions present in the estimation sample.

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.[15]

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 $US1.90 per day (e.g., Ethiopia, Malawi and Uganda); in some countries, it is well above (e.g., Bangladesh, Iraq). For Bangladesh, we also have assignable expenditure data on food, which represents 30% of non durable expenditure on average.

3 Results

We estimate equation (5) under the restrictions (12) via seemingly unrelated regression in Stata. Our observed vector of demographic variables $\boldsymbol{z}_h$ is comprised of: the numbers of men, women and children ($\boldsymbol{n}_h$); 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. We do not include any distribution factors $\boldsymbol{d}_h$ in this empirical work, so $\boldsymbol{m} = [\boldsymbol{z}] = [\boldsymbol{n} \quad \widetilde{\boldsymbol{z}}]$. Descriptive statistics for these variables are available in an online appendix.

3.1 Pre-test

Evaluation 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 (totaled across household members), and the slope evaluated at average characteristics, along with a z-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.

[15] Calvi (2018) uses the Indian data to estimate DLP resource shares, on non nuclear as well as nuclear households. It is not clear how she specifies the model to account for non nuclear households.

Clothing is not a large budget share. Clothing represents between 1.7% and 7% of the budget (in Nigeria and Ethiopia respectively). 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.

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 evaluated at mean characteristics are not statistically significantly different from zero in Ghana, Tajikistan, Tanzania, Timor L'este and Uganda, suggesting that 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 budget share functions 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 (3929) on food than clothing (4288) because there is some non-response in the daily food diary data. Food budget shares are much larger than clothing budget shares—in Bangladesh, whereas clothing accounts for only 3 per cent of total household consumption, fully 56 per cent of household consumption is embodied in 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 z-test of -14 and 99 per cent of the sample with significant slopes. Since, as noted above, resource shares are identified off variation in the slopes of assignable good Engel curves, the large size and steep slope of the food Engel curve suggests that it is a good candidate for these 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.

Table 3: Pre-Test						
country	sample N	budget share	std dev	slope at mean	z-test of slope	%age of sample significant
Albania	3279	0.0411	0.0416	0.0139	4.6	84
Bangladesh	4288	0.0407	0.0207	-0.0157	-21	99
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	3929	0.5621	0.1507	-0.1181	-14.6	99

3.2 Resource shares

Estimated per-person resource shares, η_h^t/n_h^t, of men, women and children, are shown in Table 4, for the countries where the resource shares are identified. We report both the resource shares estimated at the mean of observed covariates $\boldsymbol{m}_h$ and the mean of the resource shares evaluated at all $\boldsymbol{m}_h$. For the former, we give the standard error and for the latter, the standard deviation.

For example, in Albania, the estimated men's and women's per-capita resource shares at the average $\boldsymbol{m}_h$ are 28 per cent and 24 per cent, respectively, with small standard errors, of 0.03. Because resource shares are nonlinear functions of estimated OLS regression coefficients, the estimate of resource shares at average $\boldsymbol{m}_h$ does not equal the average of estimated resource shares over all $\boldsymbol{m}_h$. 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 $\boldsymbol{m}_h$. 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.

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 (see the Appendix, where we provide estimates for all countries, even those where the model is not identified).

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, $n_h^t / \sum_s n_h^t$. This would assign each person their per-capita share of household consumption. Table 4 shows lots of inequality across household members, so it should be surprising that the per-capita model is not supported by these estimates. Given our model, this obtains if $\boldsymbol{b}_{\tilde{z}}^t = \boldsymbol{b}_d^t = \mathbf{0}$ for all t and $b_{n^{t'}}^t = 0$ for all $t \neq t'$ and $b_{n^t}^t = \kappa$ for all t. This is a testable restriction, it is rejected at conventional levels for all 5 countries.

Table 4: Predicted Resource Shares, Selected Countries

Country	sample N	Evaluated at mean $\boldsymbol{m}_h$ men est *std err*	women est *std err*	children est *std err*	Evaluated at all $\boldsymbol{m}_h$ men mean *std dev*	women mean *std dev*	children mean *std dev*	fraction with an eta outside [0,1]
Albania	3279	0.2823	0.2448	0.1372	0.2911	0.2416	0.132	0.0561
		0.0311	*0.032*	*0.0304*	*0.4448*	*0.3564*	*0.1604*	
Bangladesh	4288	0.3096	0.2683	0.1303	0.3076	0.2676	0.1332	0.0005
		0.0132	*0.0166*	*0.0122*	*0.1137*	*0.1175*	*0.0642*	
Bulgaria	2099	0.3048	0.3702	0.1893	0.2951	0.382	0.1771	0.0734
		0.0371	*0.0408*	*0.0617*	*0.1404*	*0.2182*	*0.2078*	
Iraq	14297	0.2685	0.2362	0.0413	0.2678	0.2354	0.042	0.0124
		0.0092	*0.0114*	*0.0059*	*0.1336*	*0.132*	*0.0698*	
Malawi	10873	0.3116	0.2738	0.124	0.3109	0.2664	0.1275	0.0143
		0.0281	*0.0302*	*0.0112*	*0.2219*	*0.1668*	*0.0915*	
Bangladesh	3929	0.2922	0.2405	0.1722	0.304	0.2306	0.1718	0.0501
Food		*0.0153*	*0.0175*	*0.0126*	*0.112*	*0.1234*	*0.0761*	

3.3 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

		Evaluated at all z		Evaluated at mean z		Gender Gap	
	sample N	men	women	men	women		
		mean	mean	est	est	est	sig
		std dev	*std dev*	*std err*	*std err*	*std err*	
Albania	3279	0.3413	0.2808	0.2823	0.2448	0.0376	
		0.3501	*0.253*	*0.0311*	*0.032*	*0.0571*	
Bangladesh	3850	0.3419	0.2848	0.3096	0.2525	0.0571	***
		0.1136	*0.0951*	*0.0132*	*0.0127*	*0.0237*	
Bulgaria	2099	0.3171	0.4435	0.3048	0.3702	-0.0654	
		0.1501	*0.2116*	*0.0371*	*0.0408*	*0.0683*	
Iraq	14040	0.3346	0.2871	0.2685	0.2327	0.0358	***
		0.1387	*0.1217*	*0.0092*	*0.0086*	*0.0166*	
Malawi	9490	0.3633	0.2797	0.3116	0.2532	0.0584	
		0.225	*0.1404*	*0.0281*	*0.0291*	*0.0542*	
Bangladesh	3522	0.3281	0.2263	0.2922	0.2217	0.0704	***
Food		*0.1167*	*0.1145*	*0.0153*	*0.0144*	*0.0273*	

Here, we see that the evidence given in Table 4 that women have a greater share of household resources than men is not a statistically significant finding. Because the estimates of men's and women's resource shares covary, the estimated 6.5 percentage point gender 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. In fact, we only see a statistically significant gender gap in Bangladesh and Iraq.

The Iraqi data suggest a small gender gap of 3.6 percentage points. In the Bangladeshi data, the estimated gender gap from assignable clothing data is much 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.[16]

3.4 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

[16]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.

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.

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. Note that we do not adjust the poverty threshold for children (e.g., DLP use a poverty rate of $1.20 per day for children). In the final column, we report the 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 $\eta_h^t = 1$, and households with just one type of person (e.g., a two-man household), where each person is assigned y_h/n_h^t. We provide asymptotic standard errors, computed via the bootstrap.[17]

Table 6, Estimated Poverty Rates, Selected Countries

country	per-capita est *std err*	men est *std err*	women est *std err*	children est *std err*	all people est *std err*
Albania	0.003	0.055	0.037	0.075	0.053
	0.001	*0.043*	*0.041*	*0.074*	*0.031*
Bangladesh	0.111	0.022	0.086	0.224	0.117
	0.004	*0.009*	*0.019*	*0.043*	*0.016*
Bulgaria	0.003	0.024	0.027	0.223	0.054
	0.001	*0.039*	*0.032*	*0.120*	*0.031*
Iraq	0.000	0.000	0.002	0.146	0.063
	0.000	*0.001*	*0.001*	*0.070*	*0.030*
Malawi	0.629	0.469	0.585	0.727	0.626
	0.004	*0.025*	*0.035*	*0.030*	*0.009*
Bangladesh	0.104	0.036	0.173	0.093	0.104
Food	*0.005*	*0.018*	*0.031*	*0.025*	*0.011*

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 translate 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

[17] We bootstrap the standard errors because poverty rates are a nonlinear function of the estimated resource shares, which are themselves nonlinear functions of estimated OLS regression coefficients. We could have alternatively used the delta method.

than men's poverty. In Bangladesh, the women are 6 percentage points more like to be poor than men; in Malawi, they are 12 percentage points more likely to be poor than men.

4 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—is wrong. This current practice ignores within-household inequality, and so mischaracterises poverty levels. For example, if a household has income slightly above the poverty line, then by the per-capita method would call them 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 women in two countries. Further, these gender gaps in resource shares result in gender gaps in poverty.

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, we could direct resources 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, we could direct resources 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 of 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, this use is due to its availability in public-use datasets, not to its superiority in other ways.

We find in our work that using food as an assignable good to identify resource shares delivers estimates that are very similar to those generated from clothing data. But, food data have four 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 dollars 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, food shares are typically much larger than clothing shares. This is not gain in terms of the model in any formal sense, but it does seem like a worthwhile auxilliary 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.

5 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 always rejected. We also find evidence of large gender gaps in resource shares, and consequently in poverty rates, in a few countries.

6 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.

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

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.

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

Han, S. and McCloskey, A., 2016. Estimation and inference with a (nearly) singular Jacobian.

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

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.

7 Material for online appendix

- The countries we are include are Albania (population: about 3 million), Bangladesh (170m), Bulgaria (7m), Ethiopia (110m), Ghana (30m), Iraq (40m), Malawi (20m), Nigeria (200m), Tajikistan (10m), Tanzania (60m), Timor Leste (1.5m) and Uganda (45m).
- Descriptive statistics for the demographic variables included in the empirical analysis: the numbers of men, women and children (nh); the average ages of men, women and children; the minimum age of the children; the average education levels of the men and women; and a dummy variable indicating that the household lives in an urban area.