Identification of Random Resource Shares in Collective Households With an Application to Microcredit in Malawi.

Geoffrey R. Dunbar, Arthur Lewbel and Krishna Pendakur [1,2]
November 2013

Abstract

We propose methods to estimate resource shares of individuals in collective households that do not require restrictions on individual preferences, but rather rely on the existence of distribution factors. We provide theorems that show identification of the distribution of these resource shares. Thus, we can identify the conditional mean of resource shares given observable demographics and distribution factors, and we allow for and identify random variation in resource shares given these observables. We use our model to investigate the effects of credit on the within-household allocation consumption in Malawi. We find: that agricultural credit and microcredit may divert resources away from children; that large loans shift resources to men; and if loans have female signatories, then resources are diverted to women and children.

Very Preliminary, Please Do Not Cite without Permission

JEL classification: D13, D11, D12, C31, I32.

Keywords:: Collective Household Model, Cost of Children, Bargaining Power, Microcredit, Sharing Rule, Demand Systems, Engel Curves.

1 Introduction

The microcredit revolution in household finance in developing economies began as a group-lending initiative, the Grameen Bank, in Bangladesh in 1983. By 2010, almost 140 million households were receiving some type of microcredit worldwide, with the number of women receiving microcredit rising from 10.3 million in 1999 to over 110 million in 2010. The Grameen Bank and it's founder, Muhammed Yunus, were awarded the Nobel Peace Prize in 2006. It would not be unfair to claim that microcredit has become a keystone development initiative worldwide.

Despite the proliferation of microcredit, there is much that is still not known about its effects on participating households (see, e.g., Banerjee (2013) for a review of the literature on the effects of microcredit). Perhaps most prominently, we do not yet understand the effects of microcredit on the allocation of consumption within households. Certainly, there is reason to hope that the effects within households favour those members most vulnerable in developing societies. Many group-lending initiatives are targeted at women, based on evidence that female empowerment is positively associated with micro-credit. See, e.g., Ngo and Wahhaj (2012) and Ashraf, Karlan and Yin (2010). What is less clear is whether this empowerment is accompanied by an associated increase in women's share of household consumption, and the impact on children's shares.

There is also evidence that micro-credit loans increase total household consumption (Pitt and Khandker (1998), Kaboski and Townsend (2005)) although there is also evidence of no effect of microcredit on household consumption (Morduch (1998), Morduch and Roodman (2010), Crepon *et al.* (2011), Banerjee *et al.* (2013)). However, Kaboski and Townsend (2011) find that micro-credit programs in Thailand cost more overall than the benefits they provide, which suggests that understanding the effects of microcredit for inequality is crucially important in any assessment of its desirability as a mechanism to deliver aid.

In this paper we build on Dunbar, Lewbel and Pendakur (DLP 2013) to estimate the effect of micro-credit on the allocation of resources within households. The earlier literature on Chiappori (1988, 1992) type Pareto efficient collective household models, including Browning, Bourguignon, Chiappori, and Lechene (1994), Browning and Chiappori (1998), Vermeulen (2002), and Chiappori and Ekeland (2009), showed that each household member's resource share, defined as his or her share of total household expenditures, can not be identified from household-level demand data. However, what this earlier work shows could be identified is the impact of distribution factors on resource shares, where distribution factors are defined as variables that affect bargaining power or claims on resources within the household, but do not affect preferences for goods and services. [1].

Given these results, some additional information or assumptions are needed to identify resource shares.

[1]Other papers that make use of this sharing rule concept include Bourguignon and Chiappori (1994), Chiappori, Fortin and Lacroix (2002), and Blundell, Chiappori and Meghir (2005).

One direct approach, taken e.g. by Menon, Perali and Pendakur (2013), is to collect intrahousehold consumption data, though this method requires detailed data collection and suffers severe measurement problems in the allocation of shared goods. Another approach is taken by Cherchye, De Rock and Vermeulen (2012) and Cherchye, De Rock, Lewbel, and Vermeulen (2013) who, without imposing restrictions on preferences, identify bounds on resource shares.

A third method is to completely identify resource shares from household level data by imposing empirically supportable restrictions on preferences and on resource shares. Papers that use this method include Lewbel (2003), Lise and Seitz (2011), Browning, Chiappori and Lewbel (2007), Lewbel and Pendakur (2008), Couprie, Peluso and Trannoy (2010), Bargain and Donni (2009,2012), and DLP. In particular, DLP show identification of each household member's resource share, including children's shares, using just household-level Engel curve data (no price variation is observed) by imposing testable restrictions on the sharing rule function and on shape of preferences.

Our approach is closest to that of DLP, who use Browning, Chiappori and Lewbel's (2013 BCL) general collective household model, applied to an assignable good for each household member, and the restriction that each person's resource share is invariant to household expenditure. With an additional restriction on individual preferences, They show identification of each person's resource share, using just household expenditure data without price variation.

We generalize DLP in three important ways. First, we completely identify resource shares without imposing DLP's restrictions on the shape of preferences, by exploiting observable distribution factors in a new way. Second, we allow resource shares to vary randomly across households, equivalent to the existence of unobserved distribution factors. Chiappori and Kim (2013) also allow for unobserved variation in resource shares across households, though unlike us they do not completely identify the resource shares. Third, we allow unobserved preference heterogeneity (random utility parameters) to affect the demands of individual household members. While an enormous literature exists on randomness in demand functions, the only other paper we know of that identifies unobserved preference heterogeneity in collective household models is Matzkin and Pérez-Estrada (2011).

Our identification theorems allow for both observed and unobserved distribution factors. Unobserved distribution factors cause random variation in resource shares conditional on observables. In collective household models resource shares multiply expenditures in structural demand equations, so unobserved distribution factors affect demand functions in a way that is analogous to measurement error in expenditures. Consequently, ignoring unobserved distribution factors generally induces attenuation bias in measured slopes with respect to expenditure. This attenuation bias causes bias in estimated structural parameters, including resource shares. Our estimators that allow for random resource shares essentially correct for this bias.

We apply our model to identify and estimate the effects of microcredit on the allocation on resources within households using data from two survey waves of a household survey in Malawi. The Malawi data include extensive consumption and demographic data, which are useful for the estimation of our model, and also include information on any credit obtained by households, including the name of the source of that credit. Based on the credit sources, we divide credit providers into 3 categories: formal micro-credit programs such as those initiated by NGOs (non governmental organizations); agricultural credit programs, and; a final category that includes store credit, money lenders and banks. We also observe which household member received the loan. We can thus compare the effects of credit origination on the consumption allocations within households. Our main empirical results are that:

1. Loans given to women increase the resource shares of women and children at the expense of men;

2. Large loans divert resources to men;

3. Agricultural credit diverts resources from children to men, and;

4. Microcredit loans may divert resources away from children.

Our first three findings appear qualitatively consistent with the existing empirical evidence of the effects of microcredit. Our estimates suggest that loans received by women shift the resource shares away from men by roughly 10 percentage points and that women gain 6.5 percentage points while children gain the remaining 3.5 percentage points. These estimates would seem consistent with a positive association between female empowerment and credit. We also find that large loans divert resource shares to men, although our estimates are not sufficiently precise to determine whether it is mothers or children who lose. This result is qualitatively consistent with evidence of loan ”pipelining” of microcredit (where women loan recipients hand over loan proceeds to men) documented in Goetz and Gupta (1996) and Ligon (2002). That agricultural credit shifts resource shares away from children to men is similar in spirit to arguments regarding household labor supply and efficiency in the presence of credit constraints, *e.g.* Morduch (1999), Baland and Robinson (2001) and Shimamura and Lastarria-Cornhiel (2009). One notable difference is that we quantify empirically a shift in the resource allocation of households from the provision of agricultural credit rather than a correlation with a labor market response.

Our last finding, also with the first and the third, is illustrative of the importance of explicitly modelling children as economic agents within households, as we do. Our preferred estimates imply that loans from microcredit providers – such as international NGOs and Malawian microcredit providers such as CUMO, OIBM, FINCA, Mardef and Pride – reduce the resource share allocated to children by almost 3 percentage points. We note that if mothers are the loan recipient this negative effect appears to be undone by the positive benefit to children from having mothers receive the loan. However, overall, we find no evidence that microcredit has a positive effect on the resource share allocated to children regardless of which parent receives the loan. We stress that this does not rule out a welfare benefit for children because microcredit may increase the overall expenditure of the household.

Malawi is, in many respects, an ideal environment in which to assess the effects of microcredit as it is one of the poorest countries in the world with an average per capita income of less than $1 US per day. Microcredit is also reasonably prevalent in Malawi. In 2012 there were roughly 450,000 borrowers who borrowed an average of $110. The effect of microcredit programs in Malawi has been studied by Hazarika and Guha-Khasnobis (2008) who found, using self-reported credit limits and nutritional Z-scores, that mother's access to micro-credit appeared to improve the nutrition of girls under six years of age but had no effect on boys. One difficulty in interpreting their results is that it is unclear whether micro-credit provision simply has a level effect on consumption to which girls are more sensitive or whether the results point to an increase in the resource share allocated to girls by household. Our estimates here would suggest the former interpretation. Baird et al. (2013) use an experimental design involving cash transfers to female students to study the effects of educational interventions on measures of female empowerment in Malawi. They found that improving the educational opportunities of young women tended to improve their agency within the household. Our results likewise suggest that targetting financial support at women tends to improve their agency within the household.

2 Collective Households with Children

A key component of collective household models, going back to the earliest frameworks of Becker (1965, 1981) and Chiappori (1988, 1992), are resource shares. Resource shares are defined as the fraction of a household's total resources (spent on consumption goods) that are allocated to each household member. Each member's resource share may differ from those of other members. Resource shares, which are closely related to Pareto weights, are often interpreted as measures of the bargaining power of each household member, however, they are also determined by altruism, particularly the shares claimed by children.

Our model starts with the Pareto efficient collective household model of BCL. Unlike earlier collective household models, BCL does not require goods to be purely public or purely private, but instead permits goods to be shared using a consumption technology function. BCL fully identifies this consumption sharing technology, and resource shares, by substantially limiting the differences in preferences between individuals living alone (singles) vs living together (couples).

Lewbel and Pendakur (2008) modify BCL to permit identification of resource shares from data that does not contain price variation (Engel curve type data), by placing restrictions on how prices and the consumption technology function interact, and by imposing the constraint that resource shares not vary with total expenditures. Theoretical and empirical evidence supporting this identifying assumption will be provided later. DLP also assume resource shares do not vary with total expenditures in this model, but substantially relax the BCL restriction limiting differences in preferences, replacing them with some demand function shape restrictions. DLP also only requires observing and estimating household Engel curves on one private, assignable goods for each household member, rather than (as in the previous papers) all the goods the household purchases. Private goods are goods that are not shared, and assignability means that we can observe which household member consumes the good.

The generalizations in DLP permit identification of the resource shares of both adults and of children, where children are treated as having their own utility functions and welfare. This is in contrast to most of the collective household empirical literature, where children are modeled just as public goods in the adult's utility function. This identification of children's utility functions and associated resource shares is necessary to answer questions regarding the welfare of children in the household, separate from the welfare of the parents.

The present paper, like DLP, is a model with both adult's and children's utility functions, and a data environment based on observable assignable goods that does not require price variation. As discussed in the introduction, the present paper generalizes DLP in some important ways, including the identification of resource shares without imposing DLP's restrictions on the shape of preferences (by exploiting observable distribution factors in a new way) and by allowing resource shares to vary randomly across households, equivalent to the existence of unobserved distribution factors.

In all these models, Pareto efficiency and duality makes maximizing the household's objective function observationally equivalent to the following procedure. First, within the household resource shares are determined. There resource shares may depend on distribution factors, defined as variables that affect bargaining power or claims on resources within the household, but do not affect preferences for goods and services. A vector of shadow prices for goods faced by each household member is also determined, based either partly or entirely on the consumption technology function. An individual's resource share within the household, along with the shadow price vector, define a shadow budget constraint for that member. Each member then determines his or her own demand for each consumption good by maximizing their own utility function given their shadow budget. The household's demand functions for each good then equals the sum of each individual member's demand, taking into account the consumption technology function that accounts for the extent to which each good is shared..

The shadow budget constraint faced by individuals within households can be used to conduct consumer surplus exercises relating to individual well-being. One example of this is the construction of 'indifference scales', a tool BCL develop for comparing the welfare of individuals in a household to that of individuals living alone, analogous to an equivalence scale.

Resource shares for each individual may also be of interest even without knowledge of shadow prices. The resource share times the household expenditure level gives the extent of the individuals' budget constraint for consuming resources within the household, and is therefore an indicator of that individual's material well-being. For example, Lise and Seitz (2004) use estimated resource shares to construct national consumption inequality measures that account for inequality both within and across households.

In addition, because within-household shadow prices are the same for all household members, resource shares describe the relative consumption levels of each member. Consequently, they can be used to evaluate the relative welfare level of each household member, and as noted above are sometimes used as measures of the bargaining power of household members. BCL show a one to one relationship between resource shares and collective household model Pareto weights on individual utility, which are also used as measures of member bargaining power.

Identification of household shadow prices, individual resource shares, and individual preferences from household demand functions may proceed in many different ways. BCL show that if individual preferences of household members are observed (e.g., by observing individuals making choices while living alone), and if household budgets and demands are observed, then the resource shares of each individual and shadow prices are identified. The intuition here is that observed demands may be inverted through known utility functions to recover the unobserved shadow budget constraints.

DLP relax the BCL restriction that each person's preferences are known. Instead, they impose two other restrictions (in addition to the Lewbel and Pendakur 2008 restrictions used to avoid the need for price variation. First, DLP assume that the resource share of each person is independent of household expenditure. Second, they impose the restriction that either preferences are functionally related across people, or that preferences are functionally related across household sizes. The functional relationships that DLP exploit translate into shape restrictions on Engel curves across individuals.

In this paper, like DLP, we do not require the restriction that preferences are known and we impose the restriction that resource shares are invariant to total household expenditure. Unlike DLP, we do not impose demand function shape restrictions across individuals. Instead, we use distribution factors to provide sufficient variation in household behaviour to identify the resource shares of each person in the household. We then greatly extend the DLP and BCL models to also allow both for unobserved random variation in resources shares across individuals and households, and to allow for unobserved preference heterogeneity (i.e., random utility parameters) iin individual's demand functions for goods.

2.1 The Model

Let x be a household's total budget, and let p denote the $M-$vector of market prices for M commodities (goods and services) that the household buys. Our general identification theorems allow households to contain any integer number J of individuals, indexed by $j = 1, ..., J$. However, for ease of exposition, and to match our empirical application, assume households consist of an adult male indexed by $j = m$, an adult female indexed by $j = f$, and $k \geq 1$ children indexed by $j = c$. Let z denote observable attributes of individual household members like age, education, etc., that may affect their preferences. Note that k can be an element of z. Let η_j denote the *resource share*, defined as fraction of the household's total expenditures x consumed by person j. By Pareto efficiency no household resources are wasted, so resource shares η_j sum to one.

Let d denote a vector of distribution factors, defined as variables which may affect resource shares η_j but which do not affect individual preferences. Distribution factors are important variables in the collective household literature for three reasons. First, as discussed in the introduction, in general collective household models the response of the resource share with respect to a change in distribution factor may be identified from household-level demand behaviour. Second, distribution factors are closely related to individual's relative bargaining power within households, and so are important components of marriage markets and other literatures associated with household formation, stability, and function. Third, distribution factors, such as the local supply of education or the local availability of nursing, may be policy variables, allowing governments the ability to affect the within-household distribution of resources.

Resource shares determine shadow budgets, since each person's shadow budget is equal to $\eta_j x$. Resource shares have a one-to-one correspondence with Pareto-weights, defined as the marginal response of person j's utility in the overall household optimization problem. In general resource shares η_j may depend on prices p (see Samuelson 1953), on preference shifters z and distribution factors d, and may depend also on the household budget x.

In our identification theorems, we assume that resource shares are independent of the household budget. Lise and Seitz (2007), Lewbel and Pendakur (2008), Bargain and Donni (2009, 2012) and DLP all use this restriction in their identification results. Cherchye et al (2013) and Menon, Perali and Pendakur (2013) provide empirical support for this restriction using Dutch and Italian data, respectively. In addition, resource shares are only assumed to be independent of x after conditioning on p, z, and d, and both z and d could include variables closely related to x, such as wealth, income (which equals x plus savings), education, wages, etc.

Let s denote the $M-$vector of shadow prices faced by household members in determining their demand functions. Following BCL, we assume a Gorman (1976) type linear consumption technology without overheads. This makes shadow prices $s = A(z)p$ for some M by M matrix valued function A, meaning that shadow prices are linear in market prices, in a way that may depend on attributes z (including the number of children k). As discussed in BCL, rather than limiting some goods to be purely public and others purely private within the household, this model allows goods to possess varying degrees of publicness. Each household member j maximizes their own utility function subject to the budget constraint that their vector of consumed quantities, when priced at shadow prices s, equals their own budget $\eta_j x$.

As in DLP, we base our identification and estimation of resource shares on the household's demand functions for private assignable goods. What makes a good be a private good in our model is that its within-household shadow price equals its market price, meaning that its row of $A(z)$ has a 1 on the main diagonal and 0 elsewhere.) This means that private goods are goods that are not jointly consumed, and so do not have any economies of scale in consumption. For example, food is private to the extent that any unit consumed by one person cannot also be eaten by another. A private good is defined to be assignable if it is consumed exclusively by one known household member. For example, qat or men's clothing could be private assignable goods for men. For any private good, assignability is an attribute of observability in the data, not of behaviour, e.g., rice could be assignable if we were able to observe how much of it was eaten by each household member. In our application we observe separate expenditures on men's, women's, and children's clothing, which we take to be private and assignable.

Our definition of a private assignable good is relatively strict, but we do not need to rule out all externalities. In particular, we can allow for externalities of private assignable goods onto the *utilities* of other household members. Further, we can allow for consumption externalities which are internalized by the household in the sense that the shadow price vector s includes a within-household linear Pigouvian tax that exactly offsets the externality. We also only need just one private assignable good for each household member, so all other goods can be shared (jointly consumed), and can be partly public and partly private.

Our identification theorems assume there is at least one private assignable good for each household member j, which for convenience we will denote as good j. In our empirical work we will (because of data limitations) treat all children within a household the same, so a single private assignable good (children's clothing), denoted good c, will be assigned to all the children in the household). If we had data on private assignable goods for each child separately, rather than for all the children together, we would then have been able to estimate a separate resource share for each child, instead of estimating a single share for all the children.

Let $g_j(x, p, z)$ be the Marshallian demand function of person j's private assignable good. This means that if a hypothetical individual having person j's preferences maximized his utility function subject to the standard linear budget constraint that his purchased bundle cost less than or equal to x at market prices p, then the quantity of good j that he would consume is $g_j(x, p, z)$. Since person j within the household chooses quantities based on shadow prices $s = A(z)p$ and budget $\eta_j x$, we can write his demand function as $h_j(\eta_j x, p, z) = p_j g_j(\eta_j x, A(z)p, z)$. Here the quantity g_j is multiplied by its market price p_j, so h_j is money spent buying the private good j. Unlike in BCL, these individual demand functions h_j are not assumed to be observable (and not identifiable from observing the behavior of single people living alone).

Let $X_j(x, p, z)$ be the amount of money spent by the household on buying person j's private assignable good. While the demand functions for goods that are not private and assignable are more complicated, it follows immediately from results in BCL[2] that the household demand functions for the private assignable goods have the simple forms

$$\begin{aligned} X_c\left(x,p,z,d\right) &= kh_c\left(\eta_c(x,p,z,d)x,p,z\right) \\ X_m\left(x,p,z,d\right) &= h_m\left(\eta_m(x,p,z,d)x,p,z\right) \\ X_f\left(x,p,z,d\right) &= h_f\left(\eta_f(x,p,z,d)x,p,z\right) \end{aligned} \tag{1}$$

where the resource share functions $\eta_j(x,p,z,d)$ can in general depend on the household's Pareto weights and on the utility functions of all the members of the household.

Putting aside for now potential random variation in resource shares and in demand functions across households, the functions on the left of equation (1) can be estimated by observing the purchases of the private assignable goods by households with various x, z when facing various p regimes. The goal is identifying functions on the right of equation (1), in particular the resource share functions. In our empirical application, we will additionally want to estimate this functions for a given price regime, without observing price variation.

The main obstacle to this identification is that there are too many functions subscripted by j. We have three observable X_j functions on the left, while on the right there are five distinct unobserved structural functions; three h_j and two η_j functions. There are only two distinct unknown η_j functions because the third is identified given the other two by the constraint that resource shares η_j sum to one.

BCL overcome this identification problem by assuming that each h_j function equals the observable demand function of good j by single people of type j living alone. However, since single children are not observed living alone, this strategy is not feasible for a setting with children. Donni et (2012) argue that if just the adult demand functions h_m and h_f are observable from single's demands, that is sufficient to exactly identify the remaining three unknown functions, i.e., the children's demand function h_c and the two unknown resource share functions.

There are many reasons to think that preferences of family members differ from those of singles living alone, which casts doubt on these identifying assumptions if those differences are large. DLP therefore take a different approach that does not depend on the demand functions of singles. They instead impose shape restrictions on demand functions, assuming that they are either similar across people or similar across household types. Either assumption implies that a particular transformation of demand functions is invariant across people or across households, and applying that transformation to (1) yields a structure with only 1 unknown preference function and 2 unknown resource share functions, allowing for identification. To permit identification without price variation, DLP also assume the η_j functions do not depend on x.

In this paper, we provide a new identification theorem that requires neither the ability to directly observe the personal demand functions h_j (as from singles data) nor any shape restrictions on preferences. Instead,

[2] BCL did not consider children, but the extension of their model (though not of their identification strategy) to include children is straightforward.

we use distribution factors to provide sufficient variation in household behaviour to identify the resource shares of each person in the household. As discussed in the introduction, previous identification results relating to distribution factors showed that, given just household level demand functions, one can in general identify the changes in η_j that result from changes in d. We improve on this result by first showing that, when resource shares do not vary with x, we can identify resource shares themselves (not just their changes with d), provided we have distribution factors that can take on at least J values.

To provide some intuition for this identification, drop p and z for now, and impose the constraint that η_j not depend on x, giving private assignable demand functions in equation (1) for each person j of the form $X_j\left(x, d\right) = h_j\left(\eta_j(d)x\right)$. Taking the derivative of this observable demand function with respect to x, and evaluating the result at $x = 0$, identifies $\eta_j(d)h_j'\left(0\right)$ where h_j' is the derivative of the function h_j. This shows identification of the resource share $\eta_j(d)$ up to the unknown constant $h_j'\left(0\right)$. Then, for each value of d, the constraint that resource shares sum to one imposes one linear restriction on the unknown constants $h_j'\left(0\right)$. There are J such unknown functions, so given at least J values for d, we get enough equations to identify these constants and thereby identify the resource shares $\eta_j(d)$. This derivation is just intended to illustrate that identification is possible; our formal identification theorems do not entail actually evaluating demand functions at $x = 0$.

All of the above analyses can be interpreted as applying to a single household, given their observed demand behavior. In practice, we observe a cross section of many households. We therefore extend the above results by identifying the distributions of resource shares and of demands across households. Given a cross section of households, we can identify the conditional distribution (across households) of the vector of assignable good expenditures $(X_1, ...X_J)$, conditioning upon x, z, d, and (if there is price variation in the data) p. From this conditional distribution we show that it is possible to identify the distribution of resource shares $(\eta_1, ...\eta_J)$. So, e.g., two households that otherwise appear identical could have different allocations of resource shares, due to unobserved differences in bargaining power, altruism, etc. We can interpret these unobserved differences as unobserved distribution factors, so essentially by identifying the conditional distribution of resource shares across households, we are identifying the effects of both observed and unobserved distribution factors.

The next section describes these new results formally, proving nonparametric identification of the conditional distribution of resource shares. It is also possible, with some additional restrictions on preferences, to allow for unobserved preference heterogeneity (i.e., random utility parameters) across individuals and across households, and to semiparametrically identify the distribution of this unobserved preference heterogeneity along with the nonparametric distribution of resource shares. These additional semiparametric results are deferred to an Appendix.

2.2 Nonparametric Identification of Resource Shares and Their Distribution

ASSUMPTION A1: For every individual $j \in \{1, ..., J\}$ in the household there is a private, assignable good, which will denoted as good j, and the household's demand function for good j is given by $X_j = h_j(\eta_j x, p, z)$. The unknown functions $h_1, ..., h_J$ are differentiable and strictly increasing in their first element. Resource shares η_1,...,η_J are random variables having some unknown joint distribution across households, with $\sum_{j=1}^{J} \eta_j = 1$.

The assumptions that $X_j = h_j(\eta_j x, p, z)$ with $\sum_{j=1}^{J} \eta_j = 1$ follow immediately from assuming either BCL or other standard Pareto efficient collective household model, with goods j being private and assignable. Having h_j increasing in its first element just means that good j is a normal good, i.e., a good for which demand goes up when total expenditures goes up.

Distribution factors d are defined to be characteristics that affect η_j but not h_j. Previous collective household models assumed that each η_j is a deterministic function of d and other observed variables. In contrast, we assume that η_j varies randomly across households, or equivalently, that there exist unobserved distribution factors. This variation in each η_j induces variation in observed private assignable goods demands X_j. For now we are assuming the only source of random variation across households are the resource shares η_j, but later we will add additional random variation that could be due to preference heterogeneity or measurement errors in X_j.

Let $F_X(X_1, ...X_j \mid p, x, d, z)$ denote the joint distribution of expenditures on private assignable goods $X_1, ...X_j$, conditioning on p, x, d, and possibly a vector of other observable characteristics z, across all household of a given type.

ASSUMPTION A2: $F_X(X_1, ...X_J \mid p, x, d, z)$ is identified from data for all p, x, d, and z in some sets Φ_p, Φ_x, Φ_d, and Φ_z, respectively. The set Φ_x is an interval, and the set Φ_p is not empty.

Assumption A2 is the standard type of assumption used for identification theorems in econometrics, that is, it starts from assuming that a distribution of observable data can be uncovered. In practice Assumption A2 will hold, and F_X could be consistently estimated, by observing a random sample of households in different price and total expenditure regimes. The sets Φ_d, and Φ_z could be empty, corresponding to not observing any distribution factor d or other characteristics z. The set ϕ_p could just contain a single element, in which case we will have Engel curve data with no price variation. We are assuming to be able to see households in some (possibly arbitrarily small) range of possible total expenditure values x, given by the interval Φ_x.

ASSUMPTION A3: Assume that η_1,...,η_J, conditional on any $p \in \Phi_p$, $d \in \Phi_d$, $z \in \Phi_z$, and any $x \in \Phi_x$, is independent of x and is continuously distributed. Let $F_\eta(\eta_1, ..., \eta_J \mid p, d, z)$ denote the unknown joint distribution of η_1,...,η_J conditional on p, d, z.

Lewbel and Pendakur (2008) and Bargain and Donni (2009) make Assumption A3 to obtain identification in the case of deterministic rather than random resource shares. BCL and Lise and Seitz (2004) imposed this assumption on their empirical models (again having deterministic shares). DLP and Menon, Pendakur, and Perali (2012) provide both theoretical and empirical evidence supporting this assumption. Note that Assumption A3 only needs to hold after conditioning on observables z that can include demographic characteristics and observable distribution factors. One way to interpret Assumption A3 is to assume there exist unobservable distribution factors, including at least one that is continuously distributed, and that the distribution of unobservable distribution across households does not depend on x, after conditioning on p and z.

THEOREM 1: Let Assumptions A1, A2, and A3 hold. Then, for some unknown functions $c_1(p, z), ..., c_J(p, z)$ the joint distribution of $\eta_1 c_1(p, z), ..., \eta_J c_J(p, z)$ conditional on p, d, z is identified for all $p \in \Phi_p$, $d \in \Phi_d$ and $z \in \Phi_z$.

A well known nonidentification result in the collective household literature (see Chiappori and Ekelund 2009 for a current general version) is that without restrictions on preferences, the levels of (deterministic) resource shares cannot be identified. Instead, only changes in resource shares with respect to observed distribution factors can be identified. This is equivalent to saying that, if each η_J were a deterministic function of observables, then only $\eta_1 c_1(p, z), ..., \eta_J c_J(p, z)$ could be identified for unknown functions $c_j(p, z)$. Theorem 1 provides a substantial generalization of this result to stochastic resource shares, since it says that when η_j are random, the entire joint distribution of resource shares (and hence the effects of all unobserved distribution factors) is identified up to the same unknown deterministic functions $c_j(p, z)$.

One way to identify the $c_j(p, z)$ functions, and hence identify the entire joint distribution function of the resource shares, would be to impose restrictions on preferences across individuals, like the assumption that individual household member's preferences are known as in BCL, or the SAP or SAT (stable acrosss people or stable across types) restrictions in DLP. As an alternative to restrictions on preferences, we consider the following mild restriction on resource shares.

ASSUMPTION A4: Assume Φ_d contains at least J elements, which without loss of generality will be denoted $d_1, ..., d_J$. For a given $p \in \Phi_p$ and $z \in \Phi_z$, assume $E(\eta_j \mid p, d_1, z) \neq 0$ and let $T(p, z)$ be the J by J matrix defined by having $E(\eta_j \mid p, d_k, z) / E(\eta_j \mid p, d_1, z)$ in the row k and column j position. Let Φ_p^* and Φ_z^* be the subsets of elements of p and z in Φ_p and Φ_z having $T(p, z)$ be nonsingular.

The key feature of Assumption A4 is the requirement that our set of distribution factors must take on at least J values, recalling that J is the number of household members. The nonsingularity of $T(p, z)$ required by Assumption A4 will generally hold, failing only when there is some equality coincidence among the expected resource share functions $E(\eta_j \mid p, d_1, z)$. For example, in households with two members, it is straightforward to check that nonsingularity will hold as long the distribution factor affects the mean of η_1 in any way, that is, as long as $E(\eta_j \mid p, d_1, z) \neq E(\eta_j \mid p, d_2, z)$

THEOREM 2: Let Assumptions A1, A2, A3 and A4 hold. Then $F_\eta(\eta_1, ..., \eta_J \mid p, d, z)$ is identified for all $d \in \Phi_d$, $p \in \Phi_p^*$ and $z \in \Phi_z^*$.

The way Theorem 2 works is by exploiting the fact that resource shares must sum to one within a household. This places J equality constraints on the set of functions $E(\eta_j \mid p, d, z)$, one for each of the J values that the distribution factors can take on. By Theorem 2, the $E(\eta_j \mid p, d, z)$ are identified up to J unknown functions, and the J equality constraints allow us to recover these J unknown functions, $c_1(p, z), ..., c_J(p, z)$. Given these $c_j(p, z)$, by Theorem 2 the entire joint distribution of the resource shares is identified.

Note that Theorems 1 and 2 do not require any price variation, and so can all be applied to Engel curve type data where all observations are drawn from a single price regime.

An immediate extension of Theorem 2 is that Assumptions A3 and A4 could also have been used to identify the levels of deterministic resource shares, e.g., in traditional nonstochastic collective household models, having resource shares independent of total expenditures, and having some distribution factors suffices to identify the level of resource shares and thereby overcome the classic nonidentification problem. So, e.g., the SAP and SAT preference restrictions employed by DLP could have been replaced with Assumption A4.

The models in Theorems 1 and 2 assume that all the random variation in expenditures on private assignable goods is due to variation in resources shares, and none to variation in preferences. This may be unrelastic, so in an Appendix we consider extending the model to allow for additional unobserved random variation due to preference heterogeneity. In this model, a vector of random utility parameters is included in the model. To permit identification where both random resource shares and random utility parameters may have unknown distributions, we restrict attention to a semiparametric family of demand functions, namely, functions where household budget shares X_j/x are polynomials in $\ln x$. Many of the most popular demand system models for empirical work are in this family, including Deaton and Muellbauer's (1980) Almost Ideal Demand System, and Banks, Blundell, and Lewbel's (1997) Quadratic Almost Ideal Demand System. Our empirical application will be a model in this class.

3 Empirical Application: The Effects of Microcredit in Malawi

The data for our application come from the second and the third waves of the Malawi Integrated Household Survey (IHS) conducted by the National Statistics Office of Malawi. DLP used the second wave to estimate the children's resource shares and intra-household inequality and reported salient facts about the second wave data there. Since the writing of that paper, a third wave of data has become available which is largely similar in structure to the data collected during the second wave. In this section we will provide a brief overview of the second wave data which draws substantially from DLP and then provide a fuller description of the third wave.

Before describing the IHS surveys, we first provide an overview of Malawi. It is a former British protectorate in southern Africa which achieved independence in 1964. Today it is one of the poorest countries on earth, with an average per capita income level of less than one US dollar per day. The population of Malawi is roughly 16 million as of 2009 with a population density of approximately 120 persons per sq. km. It is one of the most densely populated countries in Africa. Half of Malawians live in the Southern region, 40% in the Central region and 10% in the Northern region, with more than 90% of the population living in rural areas. In 2005, Malawi received almost $600 million in foreign aid, equivalent to roughly 50 percent of government spending. According to Mixmarket, the online microcredit website, in 2013 there was $46 million in outstanding microcredit loans to almost 450,000 borrowers.

The second wave data come from the second Malawi Integrated Household Survey, conducted in 2004-2005, made available for purchase to us by the National Statistics Office of Malawi (NSO). The Survey was designed by the National Statistics Office of the Government of Malawi with assistance from the International Food Policy Research Institute and the World Bank in order to better understand poverty at the household level in Malawi. The survey includes roughly 11,000 households, drawn randomly from a stratified sample of roughly 500 strata in 28 districts.[3] The second wave is not freely available but one can apply to the NSO to purchase it. The third wave data come from the third Malawi IHS conducted in 2010-2011 by the National Statistics Office which is publicly available. The third wave includes roughly 12,200 households drawn randomly from a stratified sample of 768 strata in 31 districts.

Enumerators were sent to individual households to collect the data. Enumerators were monitored by Field Supervisors in order to ensure that the random samples were followed and also to ensure data quality. For the second wave there were 47 field enumerators, 15 team leaders, 12 data entry clerks and zone supervisors provided by the National Statistics Office. For the third wave, there were 75 field enumerators, 16 team leaders and 22 who worked in data capturing. These survey workers were divided into teams. In the second wave, zone supervisors were responsible for driving teams to enumeration areas where team leaders

[3]For computational reasons, we do not use the complex sampling information associated with stratification in our estimation.

on motorcycles would supervisor the field enumerators who were on foot. In the third wave, each team included a team leader, 4 enumerators, one data entry clerk and a driver and were thus more mobile across enumeration areas. The data entry clerks in the field had laptops in the third wave which was a significant difference between the survey methodologies for waves 2 and 3. The fieldwork began in the 2nd week of March 2004 and ended in March 2005 for the 2nd wave. The third wave of the survey began in the 3rd week of March in 2010 and ended in March 2011. In an effort to minimize data entry errors, data were checked at the EA-level so that teams could revisit households if necessary before moving to a new district. The response rate for the third survey was approximately 99 per cent although 6 per cent of households in the final data were replacement households because the originally targeted household was unavailable or refused. This non-response is similar to that observed in the second wave when approximately 5 per cent of the original sample was unavailable or refused. In both surveys roughly 0.5 per cent of households refused to complete the survey so endogenous selection of reporting households is unlikely to be a practical concern.

In the Surveys, households are asked questions from a number of modules relating to health, education, employment, fertility and consumption. Households are asked to recall their food consumption (one week recall) and their non-food expenditure broken into four recall categories (one week, one month, three months and one year). Consumption amounts also include the value of home produced goods and services imputed at the value of those services consumed in the market.

The consumption data include (in the three month recall questionnaire) household expenditures on clothing and shoes for the household head, spouse(s), boys and girls. These are our assignable goods which we construct for each household from the detailed module data. For almost all the empirical work, we use a single private assignable good for each person equal to the sum of clothing and footwear expenditures for that person. As distribution and demographic factors, we use information from the remaining modules to construct measures of education, age, marital status, etc.

Table 1 reports some summary data for our sample across our survey years. We use consumer price data from the National Statistics Office of Malawi to deflate nominal expenditures between survey years. NSO data report that nominal prices increased 70 per cent for our sample.

We select non-urban, non-polygamous, households with 1-4 children under 14 years of age and a male household head. We exclude households with either a husband or a wife over 64 years of age and households with other household members present (such as grandparents or aunts or uncles). We also trim our sample of households who have total expenditure in the lowest 1.25 percentile or the top 1.25 percentile. We also drop households with any missing data for any of our distribution factors. Our final sample of households is 5,829, with 2,774 from the second wave and 3,055 from the third wave.

Overall, there does not seem too much difference in reported values across the surveys with the exception of the share of food expenditure. The share of household expenditure on food is lower in the third wave than in the second although total expenditure does not appear to have risen much in real terms. Analysis of the consumption data suggests that some of the food expenditure appears to shifted to durables in the third wave. We are careful to control for survey year effects in our empirical work which follows. Finally, there does not appear to be much difference in the age composition or family size distribution of our sample across survey years.

Table 1: Descriptive Statistics: IHS2

		Number of Children				all
		1	2	3	4	
Number of Observations		851	816	652	455	2774
clothing plus	men	1.41	1.29	1.19	1.0	1.26
footwear	women	2.04	1.89	1.59	1.47	1.80
(in per cent)	children	1.04	1.56	1.57	1.80	1.44
food (in per cent)		60.5	60.0	60.1	60.0	60.2
log-total-expenditure		-0.22	-0.16	-0.09	0.014	-0.14
(median adjusted)						
age	men	30.1	32.5	34.9	37.9	33.2
	women	24.8	26.6	29.0	31.4	27.4
Descriptive Statistics: IHS3						
Number of Observations		828	908	791	528	3055
clothing plus	men	0.88	0.79	0.55	0.50	0.70
footwear	women	1.33	1.09	0.86	0.82	1.12
(in per cent)	children	0.95	1.23	1.28	1.21	1.30
food (in per cent)		35.7	35.8	34.2	32.7	34.8
log-total-expenditure		0.09	0.17	0.23	0.25	0.18
(median adjusted)						
age	men	29.5	31.8	33.9	36.4	32.5
	women	24.8	26.6	28.6	30.5	27.3

In Table 5 we present summary statistics for our distribution factors for our full sample. Almost half of our sample resides in the southern region of Malawi, nearly 40 per cent in the north and the remaining 15 per cent in the central region. Most of the households are in villages several kilometres from main roads and daily markets. There does not appear to be any gender bias in the distribution of children in our sample as 51 per cent of children are girls. Husbands have more years of schooling than wives and are also older. Slightly over 80 per cent of our sample is christian, 12 per cent muslim and the remaining households are typically animist. Approximately 7 per cent of both husbands and wives have a long-term chronic illness. Finally, roughly 13 per cent of our sample has a loan.

3.1 Microcredit in Malawi

There are many credit providers in Malawi, including microcredit institutions, farm credit, formal banks and moneylenders. The data we use includes the name of each lender associated with a particular loan to a household. The household questionnaire also asks each borrower to classify the lender as one of 11 (12) types in the 2004 (2011) survey wave. Although this classification system is useful to us, careful examination of the detailed records suggested several classification inconsistencies as the same lender was variously identified as a neighbour, moneylender or a relative. In addition, the same organization, for example the Malawi Rural Finance Corporation, was classified as both an agricultural credit provider and as a microcredit provider by different borrowers. There are two obvious reasons for such classification errors. First, the borrowers may not know enough about lenders to determine whether a particular lender is a NGO or a government-sponsored rural bank. Second, even if they do have specific knowledge about borrowers, it does not appear that the survey technicians provided a definition of the categories of lenders. Since this classification into type by the borrower is likely to be endogenously determined in such cases, we re-categorize the lender types to avoid such inconsistencies.

By visual inspection of the lender names, we categorize lenders as either microcredit providers, agricultural credit providers or other credit provider. The key difference between microcredit providers and agricultural credit providers is that the loan funds from the latter are intended to provide inputs into farm production. The main agricultural credit provider is the Malawi Rural Finance Corporation (MRFC) although there are several less common agricultural lenders: Agricultural Productivity Investment Programme (APIP), the National Association of Smallholder Farmers (NASFAM), Agora, Farmer's World, Development Aid from People to People (DAPP) and the Food and Agricultural Organization (FAO). To highlight our classification method, we note that DAPP is a development agency that is not limited to agricultural finance. Thus to categorize a loan from DAPP as being part of its Farmer's Club programme, we use the additional survey responses with regards to the stated purpose and timing of the loan to identify it as agricultural credit.

There are numerous microcredit agencies operating in Malawi during the period of our study and thus we do not list each one. The main microcredit providers in Malawi during the first wave of our study was Finca Malawi which was founded in Blantyre, Malawi, in 1994 and subsequently grew to serve all 28 districts in the country. Finca closed operations in 2009. Between the survey waves, the Malawian government started a microcredit provider, the Malawi Rural Development Fund (MARDEF), in 2005 and this is a common lender in the 2011 survey wave. Other common microcredit providers are: Finance Trust for the Self-Employed (FITSE), Concern Universal Microfinance Operations (CUMO), Pride, World Vision and then a number of NGOs with religious affliations.

The remaining credit providers are classified as other. These mainly include moneylenders, relatives, neighbors, large financial institutions (these are relatively uncommon in our sample) and also informal clubs which appear to be similar to rotating savings and credit organizations (ROSCAs) such as those studies by Anderson and Baland (2002) and Besley, Coates and Loury (1993). We also include a small number of credit providers which we could not otherwise classify, either because coding inconsistencies meant we could not differentiate between two or more types or because we could find no information about the organization listed in any public record.

Table 2 presents summary statistics for household loans by our categories. Roughly 14 per cent of our sample of 5829 households received some type of loan. The most common loan type was agricultural credit which accounted for 9 percentage points. We report the average loan size and median loan size in real (2005) Kwacha (the national currency of Malawi) for each category. We use publicly available CPI inflation data from the National Statistics Office of Malawi to convert nominal to real. Other loans are on average larger than either microcredit or agricultural credit, in part because of a few large loans from formal banks. However, microcredit loans have the highest median loan size. Agricultural credit is the smallest in both average and median terms. In terms of the gender of the loan recipient, women receive roughly 15 per cent of loans. While loans to women are on average smaller in average and median terms, this is not true for microcredit loans. In terms of microcredit, women receive average and median loans that are higher than those for men. Finally, median real income in Kwacha for our sample is approximately 80,000 so, in general, these loans are a small fraction of household income.

Table 2: Descriptive Statistics: Malawian Credit (in Kwacha)

Type of Credit	Number of Cases	Mean Loan	Median Loan
All loans	792	4453	1925
Agricultural	529	2195	1176
Microcredit	112	6300	4795
Other	151	7117	3465
Female Loan	112	3325	1385
Female Microcredit	29	7773	6000
Female Farm	62	857	588

3.2 Estimation

Since we do not use variation across prices to identify resource shares, the model (1) may be estimated with Engel curve data. As is common in consumer demand econometrics, we will consider a model of budget share equations. Let $W_j(x, z, d) = X_j(x, \overline{p}, z, d)/x$ be the observed household budget share function at a fixed price vector $\overline{p}$, which without loss of generality we normalize to equal one. Let $w_j(x, z) = h_j(x, \overline{p}, z)$ similarly give the unobserved personal budget share function at that same fixed price vector $\overline{p}$. Under our identifying restriction that resource shares are independent of household expenditure, we may write resource shares suppressing both the expenditure and price arguments as $\eta_j(x, p, z.d) = \eta_c(z.d)$, and re-express (1) in

Engel curve form as

$$\begin{aligned} W_c(x,z,d) &= k\eta_c(z.d) \cdot w_c\left(\eta_c(z.d)x, z\right), \\ W_m(x,z,d) &= \eta_c(z.d) \cdot w_m\left(\eta_m(z,d)x, z\right), \\ W_f(x,z,d) &= \eta_c(z.d) \cdot w_f\left(\eta_f(z,d)x, z\right). \end{aligned} \tag{2}$$

Given Theorem 2, the distribution of the resource shares could be estimated nonparametrically, and it would also be possible to estimate the demand functions incorporating preference heterogeneity semiparametrically using Theorem 3 in the Appendix. However, we have a large number of relevent preference shifters z and distribution factors d relative to our sample size, so we pursue a parametric estimation strategy.

Based on Theorems 2 and 3 in the appendix, we allow for both random variation (i.e., unobserved distribution factors) in resource shares η_j, and unobserved preference heterogeneity in the demand functions in w_j. Our demand functions have Engel curves linear in $\ln x$, which based on Theorem 3 also allows for additive preference heterogeneity, that is,

$$w_j = \alpha_j(z) + \beta_j(z)\ln x + \varepsilon_j.$$

The ε_j terms are unobserved preference heterogeneity parameters. Engel curves having budget shares linear in $\ln x$ are known as price independent generalized logarithmic demands (see Muellbauer 1977), and are consistent with many commonly used demand systems, including Deaton and Muellbauer's (1980) Almost Ideal demand system, and Jorgenson, Lau, and Stoker's (1982) Translog demand system. We specify the functions α_j and β_j to be linear indices in the complete set of preference shifters z.

Putting these together yields the model

$$\begin{aligned} W_c(x,z,d) &= \left(\alpha_c(z) + \beta_c(z)\ln\left(\eta_c x\right) + \varepsilon_c\right) k\eta_c, \\ W_m(x,z,d) &= \left(\alpha_m(z) + \beta_m(z)\ln\left(\eta_m x\right) + \varepsilon_m\right) \eta_m, \\ W_f(x,z,d) &= \left(\alpha_f(z) + \beta_f(z)\ln\left(\eta_m x\right) + \varepsilon_f\right) \eta_f. \end{aligned} \tag{3}$$

Resource shares must lie between zero and one, and sum to one, so we specify them as logit transforms of linear indices, that is, we let

$$\begin{aligned} \eta_m &= \exp\left(\gamma_m(z,d) + \nu_m\right) / (1 + \exp\left(\gamma_m(z,d) + \nu_m\right) + \exp\left(\gamma_f(z,d) + \nu_f\right)), \\ \eta_f &= \exp\left(\gamma_f(z,d) + \nu_f\right) / (1 + \exp\left(\gamma_m(z,d) + \nu_m\right) + \exp\left(\gamma_f(z,d) + \nu_f\right)), \\ \eta_c &= 1/k(1 + \exp\left(\gamma_m(z,d) + \nu_m\right) + \exp\left(\gamma_f(z,d) + \nu_f\right)), \end{aligned} \tag{4}$$

where γ_j are linear indices in the complete set of preference shifters z and distribution factors d and the ν_j are random, unobserved distribution factors. The model that we estimate is the set of equations (3) with resource shares η_m, η_f, and η_c given by the set of equations (4).

The resulting model that we estimate has three equations but five unobserved error terms, namely, ν_f, ν_m, ε_c, ε_f, and ε_m. It follows from Theorem 3 that the distributions of these errors are identified (assuming the unobserved distribution factors ν_j are distributed independently of the preference heterogeneity parameters ε_j resource share error). This identification is possible because the two types of errors interact differently with x in how they affect each budget share W_j. Specifically, ν_j (through η_j) multiplies x while the additive ε_j does not.

To estimate model (3) conveniently and efficiently given its nonseparable errors, we parameterize the error distributions and use maximum likelihood. We assume ε_c, ε_f, and ε_m are jointly normally distributed, and ν_f, and ν_m are also jointly normal, so

$$\begin{matrix} \varepsilon_c \\ \varepsilon_m \\ \varepsilon_f \\ \nu_m \\ \nu_f \end{matrix} \,\tilde{}N \left(0, \begin{matrix} \sigma^2_{\varepsilon c} & \rho_{cm}\sigma_{\varepsilon c}\sigma_{\varepsilon m} & \rho_{cf}\sigma_{\varepsilon c}\sigma_{\varepsilon f} & 0 & 0 \\ & \sigma^2_{\varepsilon m} & \rho_{mf}\sigma_{\varepsilon m}\sigma_{\varepsilon f} & 0 & 0 \\ & & \sigma^2_{\varepsilon f} & 0 & 0 \\ & & & \sigma^2_{\nu m} & \rho_{\nu}\sigma_{\nu m}\sigma_{\nu f} \\ & & & & \sigma^2_{\nu f} \end{matrix} \right).$$

Given values for the unobserved distribution factors ν_m and ν_f, we can easily solve the demand system for the preference heterogeneity terms ε_c, ε_f, and ε_m. So, to compute the likelihood for each observation, we can write the likelihood for the ε_j terms, and numerically integrate over values of ν_m and ν_f. We estimate this using ml in Stata, and code is available on request.

Because the recall period is for reporting consumption is months, and clothing is purchased relatively infrequently, we correct for infrequency of purchase as follows. First, we estimate a probit model for each person's indicator of a nonzero clothing expenditure using $x, z, d, x \cdot z$, and $x \cdot d$ as regressors. Then, for each person with a nonzero purchase, we multiply the expenditure X_j by the predicted probit probability of purchase. This scaled expenditure reflects an assumption that the consumption flow from clothing is proportional to the frequency of purchase given purchase frequencies that are constant for each person and conditionally random with respect to the conditioning variables in the probit. The maximum likelihood model is then estimated using all observations with nonzero purchases. Given the above scaling adjustments, for households that have have some zero and some nonzero clothing purchases, the nonzero observations are appropriately kept in the model. For comparison we also estimated the model without infrequency adjustments, dropping all households that had any zeros. The resulting estimates were quite similar as one would expect, though with larger standard errors due to the loss of observations.

3.3 Results

Table 3 presents our estimates of the resource shares and the variances and covariances of the joint distribution of the unobserved preference and distribution factors. Focussing first on the resources shares of fathers, we see similar patterns broadly similar to those observed by DLP with two exceptions. We find fathers' resources shares are lower than those reported in DLP and also do not observe fathers' resource shares to fall after the fourth child (note however that the sample size for four or more children is rather small). However, like DLP, we find that mothers and children appear to share a fixed proportion of household resources and that as the number of children increases the mother's share tends to fall. As well, our estimation approach appears to yield reasonably tight standard errors.

Turning next to the parameters of the joint distribution we see that the observed distribution factors between husbands and wives are fairly highly correlated. The correlation between the predicted index functions $\hat{\gamma}_m(z,d)$ and $\hat{\gamma}_f(z,d)$ is 0.82. Although the focus of this paper is not on marital sorting, this relatively high correlation could be evidence of assortative matching in marriage. The estimated correlation between the unobserved distribution factors for fathers and mothers is also quite high, 0.597. The similarity between the correlations of observed distribution factors across fathers and mothers and the unobserved distribution factors across fathers and mothers suggests that the modeling of unobserved distribution factors is reasonable.

Table 4 presents our estimates of the marginal effects of the credit distribution factors. We report estimation results for two specifications of our model: one with no unobserved distribution factors, $\nu_j = 0$ for all j, and; one with unobserved distribution factors. Recall that we have three credit origination sources: microcredit, agricultural credit and other credit; and also the loan size and whether mothers are the loan recipients.

Loans and credit can have two separate effects on household spending. One effect is to potentially raise total household expenditures x, and the second is change the allocation of x among household members. In the following summary of results, it is important to bear in mind that we are only estimating the latter effect. So, e.g., it is possible for a household member to become better off when accessing credit increases x, even when having credit also causes his or her own share of x to decline.

Of the three types of credit we analyze, consider microcredit first. We find that microcredit lowers the resource share of children by roughly 2.8 percentage points. This estimate is, however, only marginally significant, with an asymptotic p-value of 0.069 and with statistically insignificant corresponding offsets in mothers and fathers shares. In contrast, if we estimate the model without allowing for random variation in resource shares (forcing all v_j to equal zero), then we find mother's shares increasing at the expense of children's shares (with men's shares staying constant). This shows that allowing for unobserved distribution factors doesn't just allow for heterogeneity, rather, it results in qualitatively different mean responses.

Some intuition for this effect on mean estimates comes from seeing that not accounting for unobserved distribution factors is analogous to the effect of not accounting for measurement errors in regressors. For example if $\gamma_j(z,d)$ were observed instead of estimated, then omitting unobserved distribution factors v_j would be equivalent to estimating the model with mismeasured resource shares, given as logit functions of

Table 3: Estimated Resource Shares

			mean	sd	min	max
	No Unobserved Distribution Factors					
			mean	sd	min	max
values of η_j	all households	men	0.222	0.040	0.100	0.402
$\nu_j = 0$		women	0.364	0.050	0.181	0.598
		children	0.414	0.070	0.195	0.656
	one child	men	0.244	0.032	0.143	0.402
		women	0.403	0.052	0.254	0.598
		children	0.353	0.060	0.195	0.520
	two children	men	0.231	0.034	0.120	0.385
		women	0.353	0.039	0.203	0.506
		children	0.416	0.050	0.263	0.617
	three children	men	0.196	0.039	0.100	0.342
		women	0.345	0.040	0.181	0.522
		children	0.460	0.064	0.282	0.645
	four children	men	0.209	0.036	0.114	0.379
		women	0.343	0.035	0.218	0.466
		children	0.448	0.049	0.303	0.656
	With Unobserved Distribution Factors					
values of η_j	all households	men	0.234	0.048	0.066	0.413
$\nu_j = 0$		women	0.380	0.056	0.243	0.604
		children	0.386	0.081	0.164	0.633
	one child	men	0.258	0.043	0.095	0.413
		women	0.431	0.050	0.303	0.604
		children	0.311	0.060	0.164	0.501
	two children	men	0.231	0.042	0.084	0.398
		women	0.383	0.043	0.269	0.525
		children	0.386	0.059	0.222	0.571
	three children	men	0.209	0.046	0.066	0.383
		women	0.343	0.036	0.243	0.517
		children	0.448	0.064	0.267	0.633
	four children	men	0.235	0.049	0.088	0.406
		women	0.344	0.034	0.251	0.467
		children	0.421	0.062	0.231	0.615
index values	γ_m		-0.499	0.403	-2.181	0.823
	γ_f		-0.002	0.349	-0.916	1.054
variances	$\sigma_{\nu m}$		0.955	*0.024*		
	$\sigma_{\nu f}$	ν_f	0.807	*0.016*		
correlations	(γ_m, γ_f)		0.821			
	ρ_ν		0.597	*0.017*		

Italicized values are asymptotic standard errors.

$\gamma_j(z, d)$, in place of true resource shares given as logit functions of $\gamma_j(z, d) + v_j$.

Turning next to agricultural credit, we find that agricultural credit shifts resources from children to fathers in our preferred specification and that both effects are significant at the 5 per cent level. That men

Table 4: Estimated Marginal Effects, Couple with 1 Child, z=0

		no unobserved d		with unobserved d	
		estimate	*std err*	estimate	*std err*
microcredit	male	-0.004	*0.015*	0.021	*0.021*
	female	-0.032^{**}	*0.016*	0.007	*0.020*
	children	0.036^{**}	*0.016*	-0.028^{*}	*0.015*
farm	male	0.012^{*}	*0.007*	0.016^{*}	*0.009*
	female	-0.016^{**}	*0.007*	0.004	*0.010*
	children	0.004	*0.007*	-0.020^{**}	*0.008*
other	male	-0.014	*0.013*	-0.013	*0.016*
	female	0.012	*0.014*	0.008	*0.017*
	children	0.001	*0.013*	0.005	*0.016*
woman	male	-0.020	*0.015*	-0.102^{***}	*0.013*
	female	0.032^{*}	*0.016*	0.065^{***}	*0.019*
	children	-0.012	*0.015*	0.037^{**}	*0.017*
ln loan size	male	0.009^{*}	*0.005*	0.015^{**}	*0.007*
	female	-0.022^{***}	*0.006*	-0.009	*0.007*
	children	0.012^{**}	*0.006*	-0.006	*0.006*

gain is consistent across both specifications. This does not necessarily indicate discrimination within the household if fathers are the main source of labor on the family farm. Agricultural loans could increase farm inputs thus necessitating greater labor supply by fathers and the shift in resource shares towards fathers may be in compensation for their greater caloric expenditures (noting that a very percentage of x goes toward food). This interpretation would seem somewhat implausible, however, as women constitute almost 70 per cent of the small-scale agricultural labour force (African Development Fund, 2005) and at best there seems to be no appreciable gain in their resource shares.

Women's resources shares are significantly increased when they are the loan recipient. Our estimates suggest that if the mother receives the loan then her resource share and those of her children increase at the expense of fathers in our preferred specification. Nor are these increases small. We estimate that if the mother receives the loan her resource share rise 6.5 percentage points and that of children rises 3.5 percentage points. The resource share of fathers thus falls 10 percentage points. These are rather large changes in the resource allocations within families. We note that adjusting for the frequency of purchase appears particularly important in uncovering these effects, possibly because receiving a loan empowers mothers in consumption decisions. If we did not adjust for zero purchases as we do then if a mother decided to purchase clothes only for herself and her children but for fathers then we would not include these observations (assuming no clothing expenditures otherwise) and we would understate the benefits of a mother receiving a loan (see Table 6). If we do not include unobserved distribution factors then we still conclude that women's resources shares increase but, it would seem, at the expense of both fathers and children.

Finally, we find some evidence that the size of the loan matters for resource shares within the household.

The resource shares of men rise as the loan size increases although our estimates are not sufficiently precise to determine who in the household loses resources in our preferred specification. Nevertheless, this result is interesting because of the distribution of average loan sizes. As Table 1 indicates, the average loan size for microcredit and other credit is at least double that of agricultural credit (where we also saw a positive effect for men's resource shares). Thus, this result does not appear to mis-attribute agricultural loan effects as loan size effects and so it does appear that men are able to appropriate greater resource shares from larger loans. Our model is silent as the the possible mechanism through which such redistribution could occur but we note that loan pipelining, where women cede control of loans to men, was documented by Goetze and Gupta (1996). Our results hint at this possibility for larger loans.

4 Conclusion

We provide theorems that identify the distribution across households of the shares of household resources devoted to each member of a household. We identify the levels of these resource shares, and how they vary with demographic variables and distribution factors. The random variation in resource shares we identify can be interpreted as the effects of unobserved distribution factors such as might commonly be encountered in empirical applications. We apply our results to estimate the effects of microcredit on household's allocation of resources in Malawi for two waves (2004 and 2010) of the Integrated Household Survey. We find that agricultural credit diverts household resources to men and also that large loans tend to divert resources to men. As well, we find that loans to woman reduce men's resource shares, as expected, and that microcredit loans may divert resources away from children.

References

[1] Ashraf, N., Karlan, D. and W. Yin, 2010. "Female Empowerment: Impact of a Commitment Savings Product in the Phillipines," *World Development*, 38(3), 333-344.

[2] Baird, S.J., E. Chirwa, J. de Hoop and B. ´Ozler, 2013. "Girl Power: Cash Transfers and Adolescent Welfare. Evidence from a Cluster-Randomized Experiment in Malawi," NBER Working Paper 19479.

[3] Baland, J.-M. and J.A. Robinson, 2000. "Is Child Labor Inefficient?," *Journal of Political Economy*, 108(4), 663-679.

[4] Banerjee, A., 2013. "Microcredit Under the Microscope: What Have We Learned in the Past Two Decades,and What Do We Need to Know?," *Annual Review of Economics*, 5, 487-519.

[5] Banerjee, A., E. Duflo, R. Glennester and C. Kinnan, 2013. "The miracle of micro finance? Evidence from a randomized evaluation," MIT Working Paper.

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

[7] Bargain, Olivier and Olivier Donni, 2009. "The Measurement of Child Costs: A Rothbarth-Type Method Consistent with Scale Economies," *Institute for the Study of Labor (IZA) Discussion Paper*, No. 4654.

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

[9] Bourguignon, F. and P.-A. Chiappori, 1992. "Collective models of household behavior : An introduction," *European Economic Review*, 36(2-3), 355-364.

[10] Bourguignon, F., M. Browning, and P.-A. Chiappori, 2009. "Efficient Intra-household Allocations and Distribution Factors: Implications and Identification," *Review of Economic Studies*, 76: 503-528.

[11] Browning, M., F. Bourguignon and P.-A. Chiappori and V. Lechene, 1994. "Income and Outcomes: A Structural Model of Intrahousehold Allocation," *Journal of Political Economy*, 102(6), pages 1067-96.

[12] Browning, M., P.-A. Chiappori and A. Lewbel, 2013. "Estimating Consumption Economies of Scale, Adult Equivalence Scales, and Household Bargaining Power," *Review of Economic Studies*, forthcoming.

[13] Cherchye, L., B. De Rock and F. Vermeulen, 2012. "Married with Children: A Collective Labor Supply Model with Detailed Time Use and Intrahousehold Expenditure Information," *American Economic Review*, 102(7), 3377-3405.

[14] Cherchye, L., B. De Rock, A. Lewbel and F. Vermeulen, 2013. "Sharing Rule Identification for General Collective Consumption Models," mimeo.

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

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

[17] Chiappori, P.-A., B. Fortin and G. Lacroix, 2002. "Marriage Market, Divorce Legislation, and Household Labor Supply," *Journal of Political Economy*, 110(1), 37-72.

[18] Chiappori, P.-A. and J. H. Kim, 2013. "Identifying Heterogeneous Sharing Rules," mimeo

[19] Couprie, H., E. Peluso and A. Trannoy, 2010. "Is power more evenly balanced in poor households?," *Journal of Public Economics*, 94(7-8), pages 493-507.

[20] Crepon, B., F. Devoto, E. Duflo and W. Pariente, 2011. "Impact of Microcreidt inn rural areas of Morocco: evidence from a randomized evaluation," MIT Working Paper.

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

[22] Donni, Olivier, 2009. "A Simple Approach to Investigate Intrahousehold Allocation of Private and Public Goods," *The Review of Economics and Statistics*, 91: 617-628.

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

[24] Goetz, M.-A. and R. S. Gupta, 1996. "Who takes the credit? Gender, power, and control over loan use in rural credit programs in Banglasdesh," *World Development*, 24(1), 45-63.

[25] Hazarika, G. and B. Guha-Khasnobis, 2008. "Household Access to Microcredit and Children's Food Security in Rural Malawi: A Gender Perspective," IZA Working Paper 3793.

[26] Kaboski, J. and R. Townsend. 2005. "Policies and Impact: An Analysis of Village Microfinance Institutions, *Journal of the European Economic Association*, 3, 1-50.

[27] Kaboski, J. and R. Townsend. 2011. "A Structural Evaluation of a Large-Scale Quasi-Experimental Microfinance Initiative," *Econometrica*, 79(5), 1357-1406.

[28] Lewbel, A. and K. Pendakur, 2008. "Estimation of Collective Household Models With Engel Curves," *Journal of Econometrics*, 147: 350-358.

[29] Lewbel, A. and K. Pendakur, 2013. "Unobserved Preference Heterogeneity in Demand Using Generalized Random Coefficients," Boston College Working Paper.

[30] Ligon, E., 2002. Dynamic bargaining in households (with an application to Bangladesh). Giannini Foundation Working Paper.

[31] Lise, J. and S. Seitz. 2004. "Consumption Inequality and Intra-Household Allocations," Queen's University Department of Economics Working Paper 1019.

[32] Matzkin, R. and J. Pérez-Estrada, 2011. Heterogeneous Collective Demand: Exploring Differential Generosity within Elderly Couples, mimeo.

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

[34] Morduch, J. 1998. "Does Microfinance Really Help the Poor? New Evidence from Flagship Programs in Bangladesh," Working Paper 198, Princeton University, Woodrow Wilson School of Public and International Affairs.

[35] Morduch, J. 1999. "The Microfinance Promise," *Journal of Economic Literature*, XXXVII, 1569-1614.

[36] Morduch, J. and D. Roodman. 2011. "The Impact of Microcredit on the Poor in Bangladesh: Revisiting the Evidence," NYU Wagner Research Paper 2011-04.

[37] Ngo, T. M.-P. and Z. Wahhaj, 2012. "Microfinance and Gender Empowerment," *Journal of Development Economics*, 99, 1-12.

[38] Pitt, M. and S. Khandker. 1998, "The Impact of Group-Based Credit Programs on Poor Households in Bangladesh: Does the Gender of Participants Matter? *Journal of Political Economy*, 106, 958-996.

[39] Vermeulen, F., 2002. "Collective Household Models: Principles and Main Results," *Journal of Economic Surveys*, 16(4), 533-64.

5 Appendix A: Additional Error Terms

Here we consider including unobserved preference heterogeneity into our demand models, in addition to the unobserved heterogeneity across households in how they allocate resource shares. To obtain identification while allowing for additional preference heterogeneity, we restrict attention to models where the demand functions have budget shares that are polynomials in $\ln x$.

It can be shown (see, e.g., Gorman (1981) and Lewbel (1991, 1997) that almost all utility derived demand systems that have budget shares which are polynomials in $\ln x$ have indirect utility functions of the form $e^{\beta(p)} H \left[\ln x - \gamma(p)\right] + \delta(p)$ for some functions $\beta(p)$, $\gamma(p)$, $\delta(p)$ and H, where H is a polynomial. Examples include Deaton and Muellbauer's (1980) Almost Ideal Demand System, Jorgenson, Lau, and Stoker's (1983) Translog demand system, and Banks, Blundell, and Lewbel's (1997) Quadratic Almost Ideal Demand System. Rank two polynomial demand functions have $\delta(p) = 0$, while rank three polynomial demands have all three functions $\beta(p)$, $\gamma(p)$, and $\delta(p)$ nonzero. See Lewbel (1991) and references therein for more details.

Applying Roy's identity to this class of indirect utility functions gives Marshallian demand functions in budget share form $W_j = \omega_j \left[\ln x - \gamma(p), \beta(p), \delta(p)\right]$ for the operator ω_j defined by

$$\omega_j \left[\ln x - \gamma(p), \beta(p), \delta(p)\right] = \frac{\partial \beta(p)}{\partial \ln p_j} H \left[\ln x - \gamma(p)\right] + \frac{\partial \gamma(p)}{\partial \ln p_j} \frac{\partial H \left[\ln x - \gamma(p)\right]}{\partial \ln x} + \frac{\partial \delta(p)}{\partial \ln p_j} H \left[\ln x - \gamma(p)\right]^2$$

Suppose each individual j in a household has preferences described by the indirect utility function $e^{\beta_j(p,z)} H \left[\ln x - \gamma_j(p, z, \varepsilon_j), z\right] + \delta_j(p, z)$ for some functions $\beta_j(p, z)$, $\gamma_j(p, z, \varepsilon_j)$, $\delta_j(p, z)$ and H, where

H is a polynomial in its first argument and ε_j is a vector of random utility parameters representing unobserved preference heterogeneity. This ε_j is in addition to allowing for observable preference heterogeneity by permitting all the functions in the model to vary with observable characteristics z. It then follows immediately that the corresponding demand function for the private assignable good j in the household will be

$$W_j = \frac{X_j}{x} = \omega_j \left[\ln(\eta_j x) - \gamma_j \left(A(z)p, z\right) + \varepsilon_j, \beta_j \left(A(z)p, z\right), \delta_j \left(A(z)p, z\right)\right] \eta_j$$

Much more generally, we assume that the demand functions for the private, assignable goods in the household have the form

$$W_j = \frac{X_j}{x} = b_j(p,z) \ln(\eta_j x)^K \eta_j + \sum_{k=0}^{K-1} \left(a_{jk}(p,z,\varepsilon_j) \ln(\eta_j x)^k\right) \eta_j \tag{5}$$

As the above derivations show, equation (5) introduces preference heterogeneity in a way that allows individuals to have arbitrary log polynomial budget shares. The utility functions giving rise to equation (5) could be even more general than the above described utility functions, because we can also have preferences yield demands other than log polynomials for goods other than the private assignable goods.

ASSUMPTION B1: Same as Assumption A1 except that instead of assuming $X_j = h_j(\eta_j x, p, z)$, assume that equation (5) holds for some integer K, some unknown functions α_{jk} and some unknown functions b_j, and some random parameters $(\varepsilon_1, ...\varepsilon_J)$. The terms $(\varepsilon_1, ...\varepsilon_J)$ are independent of x conditional on p, z, d. Assume the functions $a_{jk}(p,z,\varepsilon_j)$ are bounded.

THEOREM 3: Let Assumptions B1, A2, A3, and A4 hold. Then $F_\eta(\eta_1, ..., \eta_J \mid p, d, z)$ is identified for all $d \in \Phi_d$, $p \in \Phi_p^*$ and $z \in \Phi_z^*$.

Theorem 3 has the same conclusion as Theorem 2, and thereby shows that we can identify random resource shares, allowing for additional unobserved preference heterogeneity, at the cost of restricting attention to polynomial demand functions instead of arbitrary unknown demand functions. An analogous result could also have been obtained assuming budget shares are polynomials in x instead of $\ln x$.

While Theorem 3 satisfies the primary goal of identifying resource shares, it may also be useful to identify the demand functions themselves, and the distribution of unobserved preference heterogeneity as well. We now show this can be done in the context of the model we use in our empirical analysis. The demand systems we consider in our empirical application are linear in $\ln x$ and in a scalar $\widetilde{\varepsilon}_j$ for each j. These correspond to indirect utility functions of the form $e^{\beta_j(p,z)}\left(\ln x - \gamma_j(p,z) + \widetilde{\varepsilon}_j\right)$ which gives demand functions for private assignable goods in the household of

$$W_j = \left(\frac{\partial \beta_j\left(A(z)p, z\right)}{\partial \ln p_j}\left(\ln(\eta_j x) - \gamma_j\left(A(z)p, z\right) + \widetilde{\varepsilon}_j\right) + \frac{\partial \gamma_j\left(A(z)p, z\right)}{\partial \ln p_j}\right)\eta_j$$

or, in a form comparable to equation (5),

$$W_j = \left(b_j\left(p,z\right)\ln\left(\eta_j x\right) + a_j\left(p,z\right) + \varepsilon_j\right)\eta_j \tag{6}$$

where $b_j\left(p,z\right) = \partial\beta_j\left(A\left(z\right)p,z\right)/\partial\ln p_j$, $\varepsilon_j = b_j\left(p,z\right)\widetilde{\varepsilon}_j$, and $a_j\left(p,z\right) = \partial\gamma_j\left(A\left(z\right)p,z\right)/\partial\ln p_j - b_j\left(p\right)\gamma_j\left(p,z\right)$.

COROLLARY 1: Let Assumptions B1, A2, A3, and A4 hold, where equation (5) has the form given by equation (6). Assume $(\varepsilon_1,...\varepsilon_J)$ have mean zero and are independent of $x,\eta_1,...\eta_J$, conditioning on p,d,z. Then $F_\eta\left(\eta_1,...,\eta_J \mid p,d,z\right)$ is identified, the functions $b_j\left(p,z\right)$ and $a_j\left(p,z\right)$ for $j=1,...,J$ are identified, and the joint distribution of $(\varepsilon_1,...\varepsilon_J)$ conditioning on p,d,z is identified, for all $d\in\Phi_d$, $p\in\Phi_p^*$ and $z\in\Phi_z^*$.

Corollary 1 shows that everything about the the model of equation (6) is identified, i.e., the unknown demand functions and the joint distributions of unobserved preference heterogeneity and of unobserved resource shares heterogeneity.

6 Appendix B: Proofs

PROOF OF THEOREM 1: Let $h_j^{-1}\left(x_j,p,z\right)$ denote the inverse of the function h_j with respect to its first argument, which exists by Assumption A1. Let $F_{X_j}\left(x_j \mid p,x,d,z\right)$ be the identified conditional distribution of X_j and let $F_{\eta_j}\left(\eta_j \mid p,d,z\right)$ be the unknown conditional distribution of η_j. Then for any $x_j\in supp\left(X_j \mid p,x,d,z\right)$,

$$\begin{aligned} F_{X_j}\left(x_j \mid p,x,d,z\right) &= \Pr\left(h_j\left(\eta_j x,p,z\right)\leq x_j \mid p,x,d,z\right) = \Pr\left(\eta_j \leq \frac{h_j^{-1}\left(x_j,p,z\right)}{x} \mid p,x,d,z\right) \\ &= F_{\eta_j}\left(\frac{h_j^{-1}\left(x_j,p,z\right)}{x} \mid p,x,d,z\right) = F_{\eta_j}\left(\frac{h_j^{-1}\left(x_j,p,z\right)}{x} \mid p,d,z\right) \end{aligned}$$

where the last equality follows from Assumption A3. By continuity of the distribution of η_j, the distribution function F_{η_j} is differentiable, and

$$\frac{-x\partial F_{X_j}\left(x_j \mid p,x,d,z\right)/\partial x_j}{\partial F_{X_j}\left(x_j \mid p,x,d,z\right)/\partial x} = \frac{-x\partial h_j^{-1}\left(x_j,p,z\right)/\partial x_j}{-xh_j^{-1}\left(x_j,p,z\right)} = \frac{\partial\ln h_j^{-1}\left(x_j,p,z\right)}{\partial x_j}$$

It follows by the continuity assumptions and interval support of x that $supp\left(X_j \mid p,z\right)$ is an interval. Let $\kappa_j\left(p,z\right)$ denote any given nonzero element in this support. Define the identified function r_j by

$$\begin{aligned} r_j\left(x_j,p,z\right) &= \exp\left(\int_{\kappa_j(p,z)}^{x_j} E\left(\frac{-x\partial F_{X_j}\left(x_j \mid p,x\right)/\partial x_j}{\partial F_{X_j}\left(x_j \mid p,x\right)/\partial x} \mid x_j,p,z\right)dx_j\right) = \exp\left(\int_{\kappa_j(p,z)}^{x_j}\frac{\partial\ln h_j^{-1}\left(x_j,p,z\right)}{\partial x_j}dx_j\right) \\ &= h_j^{-1}\left(x_j,p,z\right)c_j\left(p,z\right) \end{aligned}$$

where $c_j\left(p,z\right) = 1/h_j^{-1}\left(\kappa_j\left(p,z\right),p,z\right)$ is an unknown function.

Now $h_j^{-1}(X_j,p,z) = \eta_j x$ so $r_j(X_j,p,z)/x = \eta_j c_j(p,z)$. The joint distribution of $r_1(X_j,p,z)/x$,... $r_J(X_j,p,z)/x$ conditional on p,d,z is identified from identification of $F_X(X_1,...X_j \mid p,x,d,z)$ and of the functions $r_j(x_j,p,z)$. This joint distribution of $r_1(X_j,p,z)/x$,... $r_J(X_j,p,z)/x$ conditional on p,d,z equals the joint distribution of $\eta_1 c_1(p,z),...,\eta_J c_J(p,z)$ conditional on p,d,z, which proves the Theorem.

PROOF OF THEOREM 2: Define the identified function t_j by

$$t_j(p,d,z) = E\left(\frac{r_j(X_j,p,z)}{x} \mid p,d,z\right) = E(\eta_j c_j(p,z) \mid p,d,z) = E(\eta_j \mid p,d,z)\, c_j(p,z)$$

so

$$\frac{t_j(p,d_k,z)}{t_j(p,d_1,z)} E(\eta_j \mid p,d_1,z) = E(\eta_j \mid p,d_k,z)$$

It follows that $T(p,z)$ defined in Assumption A4 equals the J by J matrix that has $t_j(p,d_k,z)/t_j(p,d_1,z)$ in the row k and column j position, and therefore that the matrix $T(p,z)$ is identified. We also have

$$\sum_{j=1}^{J} \frac{t_j(p,d_k,z)}{t_j(p,d_1,z)} E(\eta_j \mid p,d_1,z) = \sum_{j=1}^{J} E(\eta_j \mid p,d_k,z) = E\left(\sum_{j=1}^{J} \eta_j \mid p,d_k,z\right) = 1$$

Let $E(\eta \mid p,d_1,z)$ be the vector of elements $E(\eta_j \mid p,d_1,z)$ for $j=1,...,J$, and let 1_J denote the J vector of ones. Then the above equation for $k=1,...,J$ is equivalent to $T(p,z)\,E(\eta \mid p,d_1,z) = 1_J$, and so, using Assumption A4, $E(\eta \mid p,d_1,z) = T(p,z)^{-1} 1_J$, and therefore $E(\eta_j \mid p,d_1,z)$ is identified for $j = 1,...,J$. We can then identify the functions $c_j(p,z)$ for $j=1,...,J$ by $c_j(p,z) = t_j(p,d_1,z)/E(\eta_j \mid p,d_1,z)$. Identification of $c_j(p,z)$ for $j=1,...,J$ along with Theorem 1 then proves Theorem 2.

PROOF OF THEOREM 3: Rewrite equation (5) as

$$W_j = \frac{X_j}{x} = b_j(p,z)\,\eta_j \ln(x)^K + \eta_j \sum_{k=0}^{K-1}\left(b_j(p,z)\binom{K}{k}\ln(x)^k \ln(\eta_j)^{K-k} + a_{jk}(p,z,\varepsilon_j)\left[\ln(\eta_j)+\ln(x)\right]^k\right) \tag{7}$$

Define the function $t_j(p,d,z)$ by

$$t_j(p,d,z) = E\left(\frac{\partial^K E(W_j \mid x,p,d,z)}{\partial \ln(x)^K} \mid p,d,z\right).$$

This $t_j(p,d,z)$ is by construction identified, and it follows from equation (7) with ε_j are independent of x that $t_j(p,d,z) = b_j(p,z)\,E(\eta_j \mid p,z,d)$. Now apply the proof of Theorem 2, replacing $c_j(p,z)$ there with $b_j(p,z)$ here, thereby showing that the functions $E(\eta_j \mid p,z,d)$ and hence the functions $b_j(p,z)$ are identified for $j=1,...,J$.

Next consider the identified functions

$$t_{k_1,k_2,...k_J}(p,d,z) = E\left(\frac{\partial^K E\left(\left(\frac{W_1}{b_1(p,z)}\right)^{k_1}\left(\frac{W_2}{b_2(p,z)}\right)^{k_2} ... \left(\frac{W_J}{b_J(p,z)}\right)^{k_J} \mid x,p,z,d\right)}{\partial \ln(x)^{(k_1+k_2+...+k_J)K}} \mid p,d,z\right)$$

for any set of integers $k_1, k_2, ...k_J$. It follows from equation (7) that $t_{k_1,k_2,...k_J}(p,d,z) = E\left(\eta_1^{k_1}...\eta_J^{k_J} \mid p,z,d\right)$. This shows that we can identify all the moments of F_η, and therefore this distribution function is identified. the resource shares $\eta_1, ...\eta_J$ are all bounded between zero and one, so all of their moments exist and are finite.

PROOF OF COROLLARY 1: We have that F_η is identified by Theorem 3, and the proof of Theorem 3 shows that the functions $b_j(p,z)$ are also identified. The functions $a_j(p,z)$ are identified by

$$a_j(p,z) = E(W_j \mid x,p,z,d) - b_j(p)\, E(\eta_j \ln(\eta_j) \mid p,z,d) - b_j(p) \ln x E(\eta_j \mid p,z,d)$$

Given identification of $F_\eta(\eta_1,...,\eta_J \mid p,d,z)$ and of the functions $a_j(p,z)$ and $b_j(p,z)$ for $j = 1,...,J$, we can write the demand functions as $W_j/b_j(p,z) = M_j + (\ln x + \widetilde{\varepsilon}_j)\eta_j$, where the joint distribution of M_1,...,M_J,η_1,...η_J given p,d,z is identified and does not depend on x.

Let $\widetilde{\varepsilon}_j = \varepsilon_j/b_j(p,z)$ and for each $j_1 = 1,...,J$ and $j_2 = 1,...,J$, define the identified functions $\vartheta_{j_1,j_2}(x,p,d,z)$ by

$$\begin{aligned}
\vartheta_{j_1,j_2}(x,p,d,z) &= E\left(\frac{W_{j_i}}{b_{j_i}(p,z)}\frac{W_{j_2}}{b_{j_2}(p,z)} \mid x,p,z,d\right) \\
&= E\left([M_{j_1} + (\ln x + \widetilde{\varepsilon}_{j_1})\eta_{j_1}][M_{j_2} + (\ln x + \widetilde{\varepsilon}_{j_2})\eta_{j_2}] \mid x,p,z,d\right) \\
&= E\left(M_{j_1}M_{j_2} + (\ln x + \widetilde{\varepsilon}_{j_1})\eta_{j_1}M_{j_2} + (\ln x + \widetilde{\varepsilon}_{j_2})\eta_{j_2}M_{j_1} + (\ln x + \widetilde{\varepsilon}_{j_1})(\ln x + \widetilde{\varepsilon}_{j_2})\eta_{j_2} \mid x,p,z,d\right) \\
&= E\left(M_{j_1}M_{j_2} + \eta_{j_1}M_{j_2}\ln x + \eta_{j_2}M_{j_1}\ln x + (\ln x)^2\eta_{j_1}\eta_{j_2} \mid x,p,z,d\right) + E(\widetilde{\varepsilon}_{j_1}\widetilde{\varepsilon}_{j_2} \mid p,z,d)\, E(\eta_{j_1}\eta_{j_2} \mid p,z,
\end{aligned}$$

Since resource shares are nonzero, we can solve this expression for $E(\widetilde{\varepsilon}_{j_1}\widetilde{\varepsilon}_{j_2} \mid p,z,d)$ in terms of $\vartheta_{j_1,j_2}(x,p,d,z)$ and functions of x and of the joint distribution of M_1,...,M_J,η_1,...η_J given p,d,z , which are identified. This then identifies all second moments $E(\widetilde{\varepsilon}_{j_1}\widetilde{\varepsilon}_{j_2} \mid p,z,d)$ of the distribution of $\widetilde{\varepsilon}_j$. We may similarly use

$$E\left(\frac{W_{j_i}}{b_{j_i}(p,z)}\frac{W_{j_2}}{b_{j_2}(p,z)}\frac{W_{j_3}}{b_{j_3}(p,z)} \mid x,p,z,d\right)$$

to solve for all of the third moments $E(\widetilde{\varepsilon}_{j_1}\widetilde{\varepsilon}_{j_2}\widetilde{\varepsilon}_{j_3} \mid p,z,d)$ of the distribution of $(\widetilde{\varepsilon}_1, ...\widetilde{\varepsilon}_J)$ in terms of previously identified objects (consisting of the second moments of $(\widetilde{\varepsilon}_1, ...\widetilde{\varepsilon}_J)$, moments of M_1,...,M_J,η_1,...η_J, and functions of x). Continuing in this way we identify all of the moments of $(\widetilde{\varepsilon}_1, ...\widetilde{\varepsilon}_J)$, and hence all of the moments of $(\varepsilon_1, ...\varepsilon_J)$ (since $b_j(p,z)$ is identified), Therefore, since these ε_j variables have bounded support, the conditional distribution of $(\varepsilon_1, ...\varepsilon_J)$ of p,z,d is identified.

7 Appendix C: Tables

appendix table, no zero correction

Table 5: Descriptive Statistics: Distribution Factors

	variable	mean	sd	min	max
z (not k)	2010 wave	0.5241	0.49946	0	1
	south	0.47607	0.49943	0	1
	central	0.1472	0.35433	0	1
	north	0.37674	0.48461	0	1
	avg child age less 5	-0.581	2.9865	-5	9
	min child age less 5	-2.6898	2.75297	-5	9
	proportion girl children	0.50948	0.37291	0	1
	age of husband less 28	4.83788	9.27341	-13	37
	age of wife less 22	5.34414	7.98179	-7	42
	education of husband less 2	-0.6674	1.06546	-2	5
	education of wife less 2	-1.058	0.7886	-2	5
	ln(distance to road)	2.29028	1.41033	0	7.17089
	ln(distance to daily market)	1.61713	1.16571	0	4.39445
	dry season	0.45102	0.49764	0	1
	christian household	0.81095	0.39159	0	1
	muslim household	0.117	0.32145	0	1
	health problem husband	0.07669	0.26611	0	1
	health problem wife	0.07291	0.26001	0	1
distribution	microcredit	0.01921	0.13729	0	1
factors	farm loan	0.09075	0.28728	0	1
	other loan	0.02591	0.15887	0	1
	wife signatory	0.01921	0.13729	0	1
	ln(loan size less median)	-0.004	0.48752	-3.7113	3.82508

Table 6: Estimates: No Frequency Correction: Couple with 1 Child, z=0

Distribution Factors		no unobserved			with unobserved		
		est		*std err*	est		*std err*
microcredit	male	0.022	*	*0.013*	-0.027		*0.023*
	female	-0.016		*0.012*	-0.004		*0.020*
	children	-0.007		*0.011*	0.031	*	*0.017*
farm	male	0.022	*	*0.013*	0.016		*0.012*
	female	-0.016		*0.012*	0.010		*0.009*
	children	-0.007		*0.011*	-0.026	***	*0.008*
other	male	-0.007		*0.012*	-0.001		*0.019*
	female	0.002		*0.011*	-0.015		*0.016*
	children	0.005		*0.010*	0.016		*0.014*
woman	male	0.011		*0.014*	0.015		*0.021*
	female	0.005		*0.013*	-0.001		*0.019*
	children	-0.016		*0.011*	-0.014		*0.014*
ln loan size	male	0.008	*	*0.005*	0.022	***	*0.008*
	female	-0.019	***	*0.004*	-0.006		*0.007*
	children	0.010	**	*0.004*	-0.016	***	*0.006*

Table 7: Resource Share Estimates: No Frequency Correction: Couple with 1 Child, z=0

Table 2: Estimated Model

			mean	sd	min	max
values of η_j	all households	men	0.234	0.048	0.066	0.413
$\nu_j = 0$		women	0.380	0.056	0.243	0.604
		children	0.386	0.081	0.164	0.633
	one child	men	0.258	0.043	0.095	0.413
		women	0.431	0.050	0.303	0.604
		children	0.311	0.060	0.164	0.501
	two children	men	0.231	0.042	0.084	0.398
		women	0.383	0.043	0.269	0.525
		children	0.386	0.059	0.222	0.571
	three children	men	0.209	0.046	0.066	0.383
		women	0.343	0.036	0.243	0.517
		children	0.448	0.064	0.267	0.633
	four children	men	0.235	0.049	0.088	0.406
		women	0.344	0.034	0.251	0.467
		children	0.421	0.062	0.231	0.615
index values	γ_m		-0.499	0.403	-2.181	0.823
	γ_f		-0.002	0.349	-0.916	1.054
variances	$\sigma_{\nu m}$		0.955	*0.024*		
	$\sigma_{\nu f}$		0.807	*0.016*		
correlations	(γ_m, γ_f)		0.821			
	ρ_ν		0.597	*0.017*		