Supplementary Materials: Online Appendix for Children's Resources in Collective Households: Identification, Estimation and an Application to Child Poverty in Malawi

Geoffrey Dunbar, Arthur Lewbel, and Krishna Pendakur
Simon Fraser University, Boston College and Simon Fraser University

Revised January 2012

Corresponding Author: Arthur Lewbel, Department of Economics, Boston College, 140 Commonwealth Ave., Chestnut Hill, MA, 02467, USA. (617)-552-3678, lewbel@bc.edu, http://www2.bc.edu/~lewbel/

APPENDIX

This online supplementary appendix contains five main sections:

Appendix A.1 states formally our main identification theorems, providing the general conditions for identification under either the Similar Across People (SAP) or Similar Across Types (SAT) conditions.

Appendix A.2 gives proofs of the Theorems in Appendix A.1.

Appendix A.3 provides an example of a class of indirect utility functions that satisfy the assumptions of both of our main identification Theorems, and yields Engel curves having the functional form we assume for our empirical work.

Appendix A.4 provides an example functional form within the general class of models given in appendix A.3. This functional form could be used if one wished to combine our results with other structural analyses, e.g., if one wished to introduce prices into the model.

Appendix A.5 provides the results of extensive statistical tests of the adequacy of our model's assumptions. These are divided into sets of tests focusing on the general BCL methodology, our SAP and SAT restrictions on preferences, invariance of resource shares with respect to total expenditures at low total expenditure levels, and privateness of clothing expenditures.

Appendix A.1: Theorems

Let $h_t^k(p, y)$ denote the Marshallian demand function for good k associated with the utility function $U_t(x_t)$, so an individual t that chooses x_t to maximize $U_t(x_t)$ under the usual linear budget constraint

$p'x_t = y$ would choose $x_t^k = h_t^k(p, y)$ for every purchased good k. Let $h_t(p, y)$ be the vector of demand functions $h_t^k(p, y)$ for all goods k, so $x_t = h_t(p, y)$ and the indirect utility function associated with $U_t(x_t)$ is then defined as the function $V_t(p, y) = U_t(h_t(p, y))$.

For their identification, BCL assumed that for a person of type t, $U_t(x_t)$ was the same as the utility function of a single person of type t living alone, and so $h_t(y, p)$ would be that single person's observed demand functions over goods. We do not make this assumption. Begin with the three person household version of the BCL model, which is

$$\max_{x_f, x_m, x_c, z_s} \widetilde{U}_s \left[U_f\left(x_f\right), U_m\left(x_m\right), U_c\left(x_c\right), p/y\right] \quad \text{such that} \quad z_s = A_s\left[x_f + x_m + sx_c\right] \text{ and } y = z_s^{\prime} p \tag{1}$$

The demand functions for the household s arising from the household's maximization problem, equation (1), can be written as follows. Let A_s^k denote the row vector given by the k'th row of the matrix A_s.

Define $H_s^k(p, y)$ to be the demand function for each good k in a household with s children. Then an immediate extension of BCL (the extension being inclusion of the third utility function U_c) is that the household s demand functions are given by

$$z_s^k = H_s^k(p, y) = A_s^k\left[h_f\left(A_s'p, \eta_{fs}y\right) + h_m\left(A_s'p, \eta_{ms}y\right) + sh_c\left(A_s'p, \eta_{cs}y\right)\right] \tag{2}$$

where η_{ts} denotes the resource share of a person of type t in a household with s children. In general, resource shares η_{ts} will depend on the given prices p and total household expenditures y, however, we will assume that resource shares to do not vary with y, and so for now will denote them $\eta_{ts}(p)$. The resource shares $\eta_{ts}(p)$ may depend on observable household characteristics including distribution factors, which we suppress for now to simplify notation (recall we have also suppressed dependence of all the above functions on attributes such as age that may affect preferences).

Note in equation (2) that each child gets a share $\eta_{cs}(p)$, so the total share devoted to children is $s\eta_{cs}(p)$. By definition, resource shares must sum to one, so for any s

$$\eta_{fs}(p) + \eta_{ms}(p) + s\eta_{cs}(p) = 1 \tag{3}$$

Our first assumption is that the BCL model as described above holds, that is,

ASSUMPTION A1: Equations (1), (2), and (3) hold, with resource shares $\eta_{ts}(p)$ that do not depend upon y.

BCL show generic identification of their model by assuming the demand functions of single men, single women, and married couples (that is, the functions $h_m(r)$, $h_f(r)$, and $H_0(r)$) are observable, and assuming the utility functions $U_f\left(x_f\right)$ and $U_m(x_m)$ apply to both single and married women and men. Their results cannot be immediately extended to children and applied to our application, because unlike men or women we cannot observe demand functions for children living alone. We also do not want to impose the assumption that single and married adults have the same underlying utility functions $U_f\left(x_f\right)$ and $U_m\left(x_m\right)$.

The assumption that resource shares are independent of y is also made by Lewbel and Pendakur (2009). This assumption implies joint restrictions on the preferences of household members and on the household's bargaining or social welfare function $\widetilde{U}_s$ (see, proposition 2 of Browning, Chiappori, and Lewbel 2008). To illustrate the point, we later give an example of a model satisfying all of our assumptions which has resources shares independent of y, in which the household maximizes a Bergson-Samuelson social welfare function. Note that Assumption A1 permits resource shares to vary freely with other observables that are associated with total expenditures y, such as household income or the mother's and father's wages.

Definition: A good k is a private good if, for any household size s, the matrix A_s has a one in positition k,k and has all other elements in row k and column k equal to zero.

This is equivalent to the definition of a private, assignable good in models that possess only purely private and purely public goods. With our general linear consumption technology, this definition means that the sum of the quantities of good k consumed by each household member equals the household's total purchases of good k, so the good is not consumed jointly like a pure public good, or partly shared like the automobile use example.

Definition: A good k is an assignable good if it only appears in one of the utility functions U_f, U_m, or U_c, e.g. a child good is an assignable good that is only appears in U_c, and so is only consumed by children.

ASSUMPTION A2: Assume that the demand functions include a private, assignable child good, denoted as good c, and a private, assignable good for each parent, denoted as goods m and f.

Note that we do not require a separate assignable good for each child, so good c is consumed by all children. Our identification results will only require observing the demand functions for the three private, assignable goods listed in Assumption A2. Examples of child goods could be toys or children's clothes, while examples of adult goods could be alcohol, tobacco, or men's and women's clothing. Private, assignable goods are often used in this literature to obtain identification, or to increase estimation efficiency. See, e.g., Chiappori and Ekeland (2009).

It follows immediately from Assumptions A1 and A2 that, for the private, assignable goods $k = f, m, c$, equation (2) simplifies to

$$z_s^k = H_s^k(p, y) = h_k\left(A_s' p, \eta_{ks}(p)\, y\right) \quad \text{for } k \in \{m, f\} \tag{4}$$

$$\text{and } z_s^c = H_s^c(p, y) = s h_c\left(A_s' p, \eta_{cs}(p)\, y\right) \tag{5}$$

We will now make some assumptions regarding individual's utility functions, that will translate into restrictions on the demand functions for assignable goods. We will show later that these assumptions are at least partly testable.

The first set of assumptions, leading to Theorem 1, will permit identification by imposing an element of similarity across different individual's demand functions for the assignable goods within a household of any given size. A second set of assumptions, leading to Theorem 2, will yield identification by permitting a comparison of the assignable good demand functions of each household member across households of different sizes.

Let $\widetilde{p}$ denote the vector of all prices except $p_m,\ p_f$, and p_c, so $\widetilde{p}$ consists of the prices of all goods except for the three private, assignable goods in Assumption A2. We may correspondingly define a square matrix $\widetilde{A}_s$ such that the set of prices $A_s' p$ is given by $p_m,\ p_f,\ p_c$, and $\widetilde{A}_s' \widetilde{p}$. Let $I(\cdot)$ be the indicator function that equals one when its argument $\cdot$ is true and zero otherwise.

ASSUMPTION A3: For $t \in \{m, f, c\}$ let

$$V_t(p, y) = I\left(y \le y^*(p)\right) \psi_t\left[\upsilon\left(\frac{y}{G_t(p)}\right) + F_t(p), \widetilde{p}\right] + I\left(y > y^*(p)\right) \Psi_t(y, p) \tag{6}$$

for some functions y^*, Ψ_t, ψ_t, υ, F, and G_t where y^* is strictly positive, G_t is nonzero, differentiable, and homogeneous of degree one, υ is differentiable and strictly monotonically increasing, $F_t(p)$ is differentiable, homogeneous of degree zero, and satisfies $\partial F_t(p)/\partial p_t = \varphi(p) \neq 0$ for some function φ. Also, ψ_t and Ψ_t are differentiable and strictly monotonically increasing in their first arguments, and differentiable and homogeneous of degree zero in their remaining (vector valued) arguments.

As we show below, Assumption A3 only restricts people's demand functions for assignable goods at very low total expenditure levels. It places no restriction at all (except for standard regularity conditions) on the demand functions for all other goods, and place no restrictions on the assignable good demand functions anywhere other than at low total expenditure levels.

In Assumption A3, $y^*(p)$ is this low but positive threshold level of total expenditures. Households having total expenditures $y > y^*(p)$ have demand functions given by an arbitrary, unconstrained indirect utility function $\Psi_t(y, p)$. Assumption A3 only requires that $\Psi_t(y, p)$ have the standard homogeneity and differentiability properties of any regular indirect utility function. Assumption A3 therefore permits individuals to have any regular preferences at all over bundles of goods that cost more than some minimal level $y^*(p)$, and therefore the demand functions for all goods can have any smooth parametric or nonparametric functional form at total expenditure levels $y > y^*(p)$.

The key restriction in Assumption A3 is that the functions υ and φ do not vary across people. The function $\upsilon(y/g_t(p)) + F_t(p)$ with $\partial F_t(p)/\partial p_t = \varphi(p)$, if it were the entire indirect utility function, would, induce shape invariance on the Engel curves of the private, assignable goods. See Pendakur (1999), Blundell, Duncan, and Pendakur (1998), Blundell, Chen, and Kristensen (2007), and Lewbel (2010). However, the demand functions that arise from equation (6) are only constrained to satisfy same invariance shape at low expenditure levels, because this restriction is only imposed for $y \leq y^*(p)$. The result of this restriction will be that the Engel curves for assignable goods can have any shape, but they will all need to have the same shape at low total expenditure levels.

Also, even at low expenditure levels, shape invariance is only imposed on the demand functions of the private, assignable goods. The role of the function ψ_t and the lack of restriction on cross derivatives $\partial F_t(p)/\partial p_k$ for all $k \neq t$ is to remove constraints on the shapes of Engel curves of goods other than the private, assignable ones.

The restriction that $\partial F_k(p)/\partial p_k$ be the same for k equal to m, f, and c limits either how $F(p)$ can depend on the prices of these goods, or on how the prices of these goods can covary. It follows from assignability that the indirect utility function for each person t will depend on p_t but not on the other two elements of the set $\{p_m, p_f, p_c\}$. Therefore, given assignability, it holds without loss of generality that $F_t(p) = \widetilde{F}_t(p_t, \widetilde{p})$ for some function $\widetilde{F}_t$ (a similar restriction must also hold for the function G_t). If the prices of the assignable goods are perfectly correlated over time, meaning they are Hicks aggregable, then $p_m = p_f = p_c$ (after appropriately rescaling units quantities are measured in if necessary) and it will follow automatically that $\partial F_k(p)/\partial p_k = \varphi(p)$ for the assignable goods k for any $F_k(p) = \widetilde{F}_k(p_k, \widetilde{p})$ function. Alternatively, if we have the functional form $F_t(p) = p_t \widetilde{\varphi}(\widetilde{p})$, then regardless of how the relative prices of the assignable goods vary, the constraint that $\partial F_k(p)/\partial p_k = \varphi(p)$ for k equal to m, f, and c will hold with $\varphi(p) = \widetilde{\varphi}(\widetilde{p})$

The role of the function ψ_t is to impose this low expenditure shape invariance only on the assignable goods, so the shapes of the Engel curves of all other goods are not restricted to be shape invariant anywhere. In short, although Assumption A4 looks complicated, it basically just says the budget share Engel curves of the household member's assignable goods all have same shape (differing only by translations) at low total expenditure levels, and are otherwise unrestricted.

To show this formally, apply Roy's identity to equation (6). The result is that, for person t and any good k, when $y > y^*(p)$, the demand function will be given by applying Roy's identity to $\Psi_t(y, p)$ giving $h_t(y, p) = -\left[\partial \Psi_t(y, p)/\partial p_k\right] / \left[\partial \Psi_t(y, p)/\partial y\right]$. However, when $y \leq y^*(p)$,applying Roy's identity to equation (6) gives

$$
\begin{aligned}
h_t(y, p) \quad = \quad & \frac{\psi_t'\left[\upsilon\left(\frac{y}{G_t(p)}\right) + F_t(p), \widetilde{p}\right]\left[\upsilon'\left(\frac{y}{G_t(p)}\right)\frac{y}{G_t(p)^2}\frac{\partial G_t(p)}{\partial p_k} - \frac{\partial F_t(p)}{\partial p_k}\right]}{\psi_t'\left[\upsilon\left(\frac{y}{G_t(p)}\right) + F_t(p), \widetilde{p}\right]\upsilon'\left(\frac{y}{G_t(p)}\right)\frac{1}{G_t(p)}} \\
& - \frac{\partial \psi_t\left[\upsilon\left(\frac{y}{G_t(p)}\right) + F_t(p), \widetilde{p}\right]/\partial p_k}{\psi_t'\left[\upsilon\left(\frac{y}{G_t(p)}\right) + F_t(p), \widetilde{p}\right]\upsilon'\left(\frac{y}{G_t(p)}\right)\frac{1}{G_t(p)}} \quad \text{for } y \leq y^*(p)
\end{aligned}
$$

Where ψ_t' and υ' denote the derivatives of ψ_t and υ with respect to their first elements.

For the assignable goods $k \in \{m, f, c\}$, the derivative $\partial \psi_t/\partial p_k$ is zero and $\partial F_k(p)/\partial p_k = \varphi(p)$, which makes the above demand function simplify to

$$h_k(y,p) = \frac{y}{G_k(p)} \frac{\partial G_k(p)}{\partial p_k} - \frac{\varphi(p) \frac{G_k(p)}{y}}{v'\left(\frac{y}{G_k(p)}\right)} y \ \text{ for } y \leq y^*(p) \tag{7}$$

which we can write more simply as

$$h_k(y,p) = \delta_k(p)\, y + g\left(\frac{y}{G_k(p)}, p\right) y \ \text{ for } y \leq y^*(p) \tag{8}$$

for functions δ_k and g. Substituting this into equation (4) gives household demand functions for the assignable goods

$$z_s^k = H_s^k(p,y) = \delta_k\left(A_s' p\right) \eta_{ks}(p)\, y + g\left(\frac{\eta_{ks}(p)\, y}{G_k\left(A_s' p\right)}, A_s' p\right) \eta_{ks}(p)\, y \ \text{ when } y \leq y^*(p),\, k \in \{m,\, f\}$$

and, for children

$$z_s^c = H_s^c(p,y) = \delta_c\left(A_s' p\right) s\eta_{cs}(p)\, y + g\left(\frac{\eta_{cs}(p)\, y}{G_c\left(A_s' p\right)}, A_s' p\right) s\eta_{cs}(p)\, y \ \text{ when } y \leq y^*(p).$$

Now consider Engel curves. For the given price regime p we can write the above equation more concisely as

$$\begin{aligned} z_s^k &= H_s^k(y) = \delta_{ks}\eta_{ks} y + g_s\left(\frac{\eta_{ks} y}{G_{ks}}\right)\eta_{ks} y \ \text{ for } y \leq y^*(p),\, k \in \{m,\, f\} \\ \text{and } z_s^c &= H_s^c(y) = \delta_{cs} s\eta_{cs} y + g_s\left(\frac{\eta_{cs} y}{G_{cs}}\right) s\eta_{cs} y \ \text{ for } y \leq y^*(p). \end{aligned}$$

ASSUMPTION A4: The function $g_s(y)$ is twice differentiable. Let $g_s'(y)$ and $g_s''(y)$ denote the first and second derivatives of $g_s(y)$ Either $\lim_{y\to 0} y^{\zeta} g_s''(y)/g_s'(y)$ is finite and nonzero for some constant $\zeta \neq 1$ or $g_s(y)$ is a polynomial in $\ln y$

Polynomials in $\ln y$ can require $\zeta = 1$ to have $\lim_{y\to 0} y^{\zeta} g_s''(y)/g_s'(y)$ be finite and nonzero, which is why Assumption A4 requires a separate statement to identify the polynomial case. The main implication of Assumption A4 is that identification requires some nonlinearity in the demand function, otherwise $g_s''(y)$ would be zero.

For the formal proof it is easiest to have that nonlinearity be present in the neighborhood of zero as in Assumption A4, but in practice nonlinearity over other ranges of y values would generally suffice. Empirically, all points along the engel curves (or at least those below y^*) will generally contribute to the precision of estimation, not just data around zero.

A sufficient, but stronger than necessary, condition for the twice differentiability of g_s in Assumption A4 is that υ be three times differentiable.

THEOREM 1: Let Assumptions A1, A2, A3, and A4 hold. Assume the household's Engel curves of private, assignable goods $H_s^k(y)$ for $k \in \{m, f, c\}$, $y \leq y^*(p)$ are identified. Then resource shares η_{ks} for all household members $k \in \{m, f, c\}$ are identified.

Notes:

1. Theorem 1 says that just from estimates of the household's Engel curves (that is, demand functions in a single price regime) for assignable goods at low expenditure levels, we can identify the fraction of total household resources for all goods that are spent on each household member. Even though resource shares η_{ks} are the fractions of all the household's resources devoted to each household member, we only need to observe their expenditures on three assignable goods (one for each household member type) to identify these resource shares.

2. Many sharing rule identification results in the literature require the existence of "distribution factors," that is, observed variables that affect the allocation of resources within a household but do not affect the preferences and demand functions of individual household members. Theorem 1 does not require the presence of distribution factors. Many identification results also only identify how resource shares change in response to changes in distribution factors, but do not identify the levels of resource shares. Theorem 1 identifies the levels of resource shares, which are important for many policy related calculations such as poverty lines.

3. Theorem 1 assumes that all children in a family are treated equally, and so get equal resource shares. The theorem can be immediately extended to allow and identify, e.g., different shares for older versus younger children, or for boys versus girls, as long as expenditures on a separate assignable good can be observed for each type of child.

4. Theorem 1 applies to households with any number of children, including zero, and so could be used in place of the theorems in Browning, Chiappori, and Lewbel (2008) or Lewbel and Pendakur (2009) for identifying resource shares.

5. The assumptions in Theorem 1 imply that the household Engel curve functions for the assignable goods, $H_s^k(y)$, are shape invariant at low levels of total expenditures y. This can be empirically tested using, e.g., Pendakur (1999).

6. Shape invariance is is often assumed to hold for all goods and all total expenditures, not just assignable goods at low expenditures levels as we require (see, e.g., Blundell, Duncan, and Pendakur (1998), and Blundell, Chen, and Kristensen (2007)). If the assignable good Engel curves do satisfy the required shape invariance at all total expenditure levels, then everything above having to do with the cut off expenditure level $y^*(p)$ can be ignored. This will also help estimation precision, since in this case demand functions at all levels of y, not just those below some $y^*(p)$, will help identify the resource shares.

Now we consider alternative identifying assumptions, based on comparing demand functions across households of different sizes, instead of across individuals within a household. We maintain Assumptions A1 and A2, but in place of Assumption A3 now assume the following:

ASSUMPTION B3: Define $\overline{p}$ to be the vector of prices of all goods that are private other than p_f, p_m, and p_c. Assume $\overline{p}$ is not empty, and for $t \in \{m, f, c\}$ assume

$$V_t(p, y) = I\left(y \leq y^*(p)\right) \psi_t \left[u_t\left(\frac{y}{G_t(\widetilde{p})}, \frac{\overline{p}}{p_t}\right), \widetilde{p}\right] + I\left(y > y^*(p)\right) \Psi_t(y, p) \tag{9}$$

for some functions y^*, u_t, ψ_t, F_t, and G_t where y^* is strictly positive, G_t is nonzero, differentiable, and homogeneous of degree one, F_t can be vector valued, is differentiable, and is homogeneous of degree zero, and ψ_t and u_t are differentiable and strictly monotonically increasing in their first arguments, and are differentiable and homogeneous of degree zero in their remaining (vector valued) arguments.

The goods in the price vector $\overline{p}$ are assumed to be private, and so have no economies of scale in household consumption, but they need not be assignable, so for example $\overline{p}$ might include food products that are consumed by all household members. Being private means that the elements of $A_s' p$ corresponding to $\overline{p}$ will just equal $\overline{p}$, so the term $\overline{p}/p_t$ will not change when p is replaced by $A_s' p$.

The difference between Assumption A3 and B3 is that the indirect utility function in B3 has the term $u_t\left[y/G_t(\widetilde{p}), \overline{p}/p_t\right]$ in place of $\upsilon\left(y/G_t(p)\right) + F_t(p)$. So A3 requires some similarity across individual's preferences, in that the function υ is the same for all types of individuals t. In contrast, with B3 the u_t expression describing preferences can freely differ across types of individuals, so B3 allows men, women, and children to have completely different demand functions for their own private goods. However, B3 places more limits on how prices can appear inside u_t versus inside υ and F_t, which will translate into strong restrictions on cross price effects in the demand functions of the private goods.

Other than replacing $\upsilon + F_t$ with u_t, Assumptions A3 and B3 are the same. In particular, the role of the function ψ_t in both cases is to allow the demand functions for all goods other than the private assignable goods to take on any shape, and the role of y^* and Ψ_t is to impose restrictions on preference only for low total expenditure households, leaving the demand functions at higher levels of y completely unconstrained.

To obtain demand functions corresponding to the indirect utility function in Assumption B3, apply Roy's identity to equation (9). As before, for person t and any good k, when $y > y^*(p)$, the demand function will be given by applying Roy's identity to $\Psi_t(y, p)$ giving $h_t(y, p) = -\left[\partial \Psi_t(y, p)/\partial p_k\right] / \left[\partial \Psi_t(y, p)/\partial y\right]$. However, when $y \leq y^*(p)$,applying Roy's identity to equation (9) gives

$$h_t(y,p) = \frac{\psi_t'\left[u_t\left(\frac{y}{G_t(\widetilde{p})},\frac{\overline{p}}{p_t}\right),\widetilde{p}\right]\left[u_t'\left(\frac{y}{G_t(\widetilde{p})},\frac{\overline{p}}{p_t}\right)\frac{y}{G_t(\widetilde{p})^2}\frac{\partial G_t(\widetilde{p})}{\partial p_k} - \frac{\partial u_t\left(\frac{y}{G_t(\widetilde{p})},\frac{\overline{p}}{p_t}\right)'}{\partial(\overline{p}/p_t)}\frac{\partial(\overline{p}/p_t)}{\partial p_k}\right]}{\psi_t'\left[u_t\left(\frac{y}{G_t(\widetilde{p})},\frac{\overline{p}}{p_t}\right),\widetilde{p}\right]u_t'\left(\frac{y}{G_t(\widetilde{p})},\frac{\overline{p}}{p_t}\right)\frac{1}{G_t(\widetilde{p})}} - \frac{\psi_{tk}\left[u_t\left(\frac{y}{G_t(\widetilde{p})},\frac{\overline{p}}{p_t}\right),\widetilde{p}\right]}{\psi_t'\left[u_t\left(\frac{y}{G_t(\widetilde{p})},\frac{\overline{p}}{p_t}\right),\widetilde{p}\right]u_t'\left(\frac{y}{G_t(\widetilde{p})},\frac{\overline{p}}{p_t}\right)\frac{1}{G_t(\widetilde{p})}}$$

Where ψ_t' and u_t' denote the derivatives of ψ_t and u_t' with respect to their first elements, ψ_{tk} denotes the partial derivative of ψ_t with respect to price p_k, and in a small abuse of notation $\partial u_t/\partial\left(\overline{p}/p_t\right)$ is the gradient vector of u_t with respect to the vector $\overline{p}/p_t$.

For the assignable goods $k \in \{m, f, c\}$ these simplify to

$$h_k(y,p) = \frac{\partial u_k\left(\frac{y}{G_k(\widetilde{p})},\frac{\overline{p}}{p_k}\right)'}{\partial\left(\overline{p}/p_k\right)}\frac{\overline{p}}{p_k^2}\frac{G_k\left(\widetilde{p}\right)}{u_k'\left(\frac{y}{G_k(\widetilde{p})},\frac{\overline{p}}{p_k}\right)} \text{ for } y \le y^*\left(p\right) \tag{10}$$

which we can write simply as

$$h_k(y,p) = \widetilde{f}_k\left(\frac{y}{G_k\left(\widetilde{p}\right)}, p_k, \overline{p}\right)y \text{ for } y \le y^*\left(p\right)$$

for functions $\widetilde{f}_k$. Recalling that p_k and $\overline{p}$ do not change when p is replaced with $A_s'p$, substituting this $h_k(y,p)$ expression into equation (4) gives household demand functions for the assignable goods

$$z_s^k = H_s^k\left(p,y\right) = \widetilde{f}_k\left(\frac{\eta_{ks}\left(p\right)y}{G_k\left(\widetilde{A}_s'\widetilde{p}\right)}, p_k, \overline{p}\right)\eta_{ks}\left(p\right)y \text{ when } y \le y^*\left(p\right), k \in \{m, f\}$$

and the same expression multiplied by s for $k = c$.

Now consider Engel curves. For the given price regime p we can write the above equation more concisely as

$$\begin{aligned} z_s^k &= H_s^k\left(y\right) = f_k\left(\frac{\eta_{ks}y}{G_{ks}}\right)\eta_{ks}y \text{ for } y \le y^*\left(p\right), k \in \{m, f\} \\ \text{and } z_s^c &= H_s^c\left(y\right) = f_c\left(\frac{\eta_{cs}y}{G_{cs}}\right)s\eta_{cs}y \text{ for } y \le y^*\left(p\right). \end{aligned}$$

Define the matrix Ω by

$$\Omega = \begin{pmatrix} \frac{\eta_{m1}}{\eta_{m3}} & 0 & -1 & 0 & 0 & 0 \\ 0 & \frac{\eta_{m1}}{\eta_{m2}} & -1 & 0 & 0 & 0 \\ 0 & \frac{\eta_{m1}}{\eta_{m2}} - \frac{\eta_{c1}}{\eta_{c2}} & 0 & 0 & \frac{\eta_{f1}}{\eta_{f2}} - \frac{\eta_{c1}}{\eta_{c2}} & 0 \\ 0 & 0 & 0 & \frac{\eta_{f1}}{\eta_{f3}} & 0 & -1 \\ 0 & 0 & 0 & 0 & \frac{\eta_{f1}}{\eta_{f2}} & -1 \\ \frac{\eta_{m1}}{\eta_{m3}} - \frac{\eta_{c1}}{\eta_{c3}} & 0 & 0 & \frac{\eta_{f1}}{\eta_{f3}} - \frac{\eta_{c1}}{\eta_{c3}} & 0 & 0 \end{pmatrix}.$$

ASSUMPTION B4: The matrix Ω is finite and nonsingular. $f_k(0) \neq 0$ for $k \in \{m, f, c\}$

Finiteness of Ω only requires that in households with two or three members, no member has a zero resource share. Violating Assumption B4 by having Ω singular would require a perfect coincidence relating the values of resource shares across households of different sizes. One of the few interpretable ways this could happen is if parents in households with two children each have the exact same resources shares as parents in households with three children. These statements, and the matrix Ω, have for simplicity been written using households consisting of s equal to 1, 2, and 3 children (with $s = 1$ shares as numerators), but in fact nonsingularity is only required to hold for any one set of three different household sizes.

The condition in Assumption B4 that $f_k(0) \neq 0$ will hold if the Engel curves for the private, assignable goods, written in budget share form, are continuous and bounded away from zero. This means that the budget shares will not be in a neighborhood of zero for very small total expenditure levels, and by continuity will not hit zero as y gets arbitrarily small. As with Theorem 1 and Assumption A4, the demand functions at all $y \leq y^*(p)$ help in identifying the model, but the technical conditions are easiest to prove in the neighborhood of zero.

THEOREM 2: Let Assumptions A1, A2, B3, and B4 hold for all household sizes s in some set S that has at least three elements. Assume the household's Engel curves of private, assignable goods $H_s^k(y)$ for $k \in \{m, f, c\}$, $y \leq y^*(p)$, $s \in S$ are identified. Then resource shares η_{ks} for all household members $k \in \{m, f, c\}$ and all $s \in S$ are identified.

Notes 1, 2, 3, and 4 listed after Theorem 1 also apply to Theorem 2.

It is possible to have models that satisfy the restrictions of both Theorems 1 and 2, by restricting the function $G_t(p)$ in Assumption A3 to only depend on $\widetilde{p}$ and restricting $F_t(p)$ in A3 to only depend on p_t and $\overline{p}$. Such models will be able to exploit comparisons of individuals both within and across households to strengthen the identification. For examples of models that satisfy such restrictions, see sections A.3 and A.4 of this appendix.

Appendix A.2: Proofs

Proof of Theorem 1: We have already in the above text derived the household Engel curve functions for the assignable goods at low expenditure levels, that is, for $y \le y^*$, $H_s^k(y) = \delta_{ks}\eta_{ks}y + g_s\left(\frac{\eta_{ks}y}{G_{ks}}\right)\eta_{ks}y$ for $k \in \{m, f\}$, and the same equation multiplied by s for $k = c$. Define $\widetilde{h}_s^k(y) = \partial\left[H_s^k(y)/y\right]/\partial y$ and define $\lambda_s = \lim_{y\to 0}\left[y^{\zeta}g_s''(y)/g_s'(y)\right]^{\frac{1}{1-\zeta}}$, where by assumption $\zeta \neq 1$ (the alternative log polynomial case is considered below). Since the functions $H_s^k(y)$ are identified, we can identify $\kappa_{ks}(y)$ for $y \le y^*$, defined by

$$\begin{aligned}
\kappa_{ks}(y) &= \left(y^{\zeta}\frac{\partial\widetilde{h}_s^k(y)/\partial y}{\widetilde{h}_s^k(y)}\right)^{\frac{1}{1-\zeta}} \\
&= \left(\left(\frac{\eta_{ks}}{G_{ks}}\right)^{-\zeta}\left(\frac{\eta_{ks}y}{G_{ks}}\right)^{\zeta}\left[g_s''\left(\frac{\eta_{ks}y}{G_{ks}}\right)\frac{\eta_{ks}^3}{G_{ks}^2}\right]/\left[g_s'\left(\frac{\eta_{ks}y}{G_{ks}}\right)\frac{\eta_{ks}^2}{G_{ks}}\right]\right)^{\frac{1}{1-\zeta}} \\
&= \frac{\eta_{ks}}{G_{ks}}\left[\left(\frac{\eta_{ks}y}{G_{ks}}\right)^{\zeta}g_s''\left(\frac{\eta_{ks}y}{G_{ks}}\right)/g_s'\left(\frac{\eta_{ks}y}{G_{ks}}\right)\right]^{\frac{1}{1-\zeta}} = \frac{\eta_{ks}}{G_{ks}}\left(y_{ks}^{\zeta}\frac{g_s''(y_{ks})}{g_s'(y_{ks})}\right)^{\frac{1}{1-\zeta}}
\end{aligned}$$

and, in particular,

$$\kappa_{ks}(0) = \frac{\eta_{ks}}{G_{ks}}\lambda_s$$

so for any $y \le y^*$ we can identify $\rho_{ks}(y)$ defined by

$$\rho_{ks}(y) = \frac{\widetilde{h}_s^k(y/\kappa_{ks}(0))}{\kappa_{ks}(0)} = g_s'\left(\frac{y}{\lambda_s}\right)\frac{\eta_{ks}}{\lambda_s}$$

and by equation (3), we can then identify the resource shares η_{ks} for each household member k by $\eta_{ks} = \rho_{ks}/\left(\rho_{ms} + \rho_{fs} + s\rho_{cs}\right)$.

Now consider the case where g_s is a polynomial of some degree λ in logarithms, so

$$g_s\left(\frac{\eta_{ks}y}{G_{ks}}\right) = \sum_{\ell=0}^{\lambda}\left(\ln\left(\frac{\eta_{ks}}{G_{ks}}\right) + \ln(y)\right)^{\ell}c_{s\ell}$$

for some constants $c_{s\ell}$, and therefore for any $y \le y^*$ we can identify $\widetilde{\rho}_{ks}$ defined by

$$\widetilde{\rho}_{ks} = \frac{\partial^{\lambda}\left[h_s^k(y)/y\right]}{\partial(\ln y)^{\lambda}} = c_{s\lambda}\eta_{ks}$$

which identifies resource shares by $\eta_{ks} = \widetilde{\rho}_{ks}/\left(\widetilde{\rho}_{ms} + \widetilde{\rho}_{fs} + s\widetilde{\rho}_{cs}\right)$.

Proof of Theorem 2: In the text we derived the household Engel curve functions for the assignable goods at low expenditure levels, which are, for $y \le y^*$, $H_s^k(y) = f_k\left(\frac{\eta_{ks}y}{G_{ks}}\right)\eta_{ks}y$ for $k \in \{m, f\}$, and the same equation multiplied by s for $k = c$. Let s and 1 be two elements of S. Since the functions $H_s^k(y)$

and $H_1^k(y)$ are identified, we can identify ς_{ks} defined by $\varsigma_{ks} = \lim_{y\to 0} H_1^k(y)/H_s^k(y)$, and

$$\varsigma_{ks} = \frac{f_k(0)\,\eta_{k1} y}{f_k(0)\,\eta_{ks} y} = \frac{\eta_{k1}}{\eta_{ks}} \text{ for } k \in \{m, f\}, \text{ and } \varsigma_{cs} = \frac{f_k(0)\,\eta_{c1} y}{f_k(0)\, s\eta_{cs} y} = \frac{\eta_{c1}}{s\eta_{cs}}$$

so

$$\begin{aligned} \varsigma_{ms}\eta_{ms} + \varsigma_{fs}\eta_{fs} + \varsigma_{cs} s\eta_{cs} &= \eta_{m1} + \eta_{f1} + \eta_{c1} = 1 \\ \varsigma_{ms}\eta_{ms} + \varsigma_{fs}\eta_{fs} + \varsigma_{cs}\left(1 - \eta_{ms} - \eta_{fs}\right) &= 1 \\ \left(\varsigma_{fs} - \varsigma_{cs}\right)\eta_{fs} + \left(\varsigma_{ms} - \varsigma_{cs}\right)\eta_{ms} &= 1 - \varsigma_{cs} \end{aligned}$$

These equations for $k \in \{m, f\}$ and for $s \in \{2, 3\}$ give the matrix equation

$$\begin{pmatrix} \varsigma_{m3} & 0 & -1 & 0 & 0 & 0 \\ 0 & \varsigma_{m2} & -1 & 0 & 0 & 0 \\ 0 & \varsigma_{m2} - \varsigma_{c2} & 0 & 0 & \varsigma_{f2} - \varsigma_{c2} & 0 \\ 0 & 0 & 0 & \varsigma_{f3} & 0 & -1 \\ 0 & 0 & 0 & 0 & \varsigma_{f2} & -1 \\ \varsigma_{m3} - \varsigma_{c3} & 0 & 0 & \varsigma_{f3} - \varsigma_{c3} & 0 & 0 \end{pmatrix} \begin{pmatrix} \eta_{m3} \\ \eta_{m2} \\ \eta_{m1} \\ \eta_{f3} \\ \eta_{f2} \\ \eta_{f1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 - \varsigma_{c2} \\ 0 \\ 0 \\ 1 - \varsigma_{c3} \end{pmatrix}$$

The six by six matrix in this equation equals $\mathbf{\Omega}$ in the text using $\varsigma_{ks} = \eta_{k1}/\eta_{ks}$. Since $\mathbf{\Omega}$ is nonsingular, the above equation can be solved for η_{ms} and η_{fs} for $s \in \{1, 2, 3\}$, meaning that these resource shares are identified because they can be written entirely in terms of the identified parameters ς_{ks}. Children's resource shares are then identified for these household types by $\eta_{cs} = \left(1 - \eta_{ms} - \eta_{fs}\right)/s$, and resource shares for households of other types s are identified by $\eta_{ks} = \eta_{k1}/\varsigma_{ks}$ for any s.

Appendix A.3: An Example Model

In this example, we assume that at low total expenditure levels, individual's Engel curves for the assignable private goods m, f, and c, are linear in $\ln(y)$. This requires that the subutility function $\upsilon\left(Y/G_t(p)\right) + F_t(p)$ in equation (6) be in Muellbauer's (1976) Price Independent Generalized Logarithmic (PIGLOG) functional form. This form is usually written as $\ln\left(Y/G_t(p)\right)/\widetilde{F}_t(p)$ for consumer t, for arbitrary (up to regularity) price functions G_t and $\widetilde{F}_t$. However, by ordinality of individual's utility functions, the same demand functions will be obtained using the monotonic transformation $\ln(\ln\left(Y/G_t(p)\right)) + F_t(p)$, where $F_t(p) = -\ln \widetilde{F}_t(p)$. We therefore suppose that the Assumptions of Theorem 1 hold, with the function υ in equation (6) given by

$$\upsilon\left(\frac{y}{G_t(p)}\right) = \ln\left[\ln\left(\frac{y}{G_t(p)}\right)\right] \tag{11}$$

Then by equations (7) and (8), we can define a function $\widetilde{\delta}_k(p)$ such that

$$h_k(y,p) = \frac{y}{G_k(p)}\frac{\partial G_k(p)}{\partial p_k} - \varphi(p)\frac{G_k(p)}{y}\left[\frac{y\ln y}{G_t(p)} - \frac{y\ln G_t(p)}{G_t(p)}\right]y \tag{12}$$
$$= \widetilde{\delta}_k(p)\,y - \varphi(p)\ln y \quad \text{for } y \le y^*(p).$$

This then yields private assignable good Engel curves having the functional form

$$\frac{z_s^k}{y} = \widetilde{\delta}_{ks}\eta_{ks} + \varphi_s\eta_{ks}\ln y \text{ for } y \le y^*, k \in \{m, f\} \tag{13}$$
$$\text{and } \frac{z_s^c}{y} = \widetilde{\delta}_{cs}s\eta_{cs} + s\varphi_s\eta_{cs}\ln y \text{ for } y \le y^*(p).$$

with unknown constants $\widetilde{\delta}_{ks}$, φ_s, and η_{ks} for $k \in \{m, f, c\}$. It follows from Theorem 1 that η_{ks} are identified from these Engel curves, but in this case that is easily directly verified. One could simply project (i.e., regress) the observed private assignable good household budget shares z_s^k/y on a constant and on $\ln y$, just using household's having s children and low values of y, to identify the $\ln y$ coefficients $\rho_m = \varphi_s\eta_{ms}$, $\rho_f = \varphi_s\eta_{fs}$, and $\rho_c = \varphi_s\eta_{cs}$ (this last is the coefficient of $s\ln y$ for children) and then use $\eta_{ks} = \rho_{ks}/\left(\rho_{ms} + \rho_{fs} + s\rho_{cs}\right)$ for $k \in \{m, f, c\}$ to identify each η_{ks}.

In this example if $\varphi(p)$ only depends on the prices of private goods $\overline{p}$, then Assumption B3 will also be satisfied. In this case the assignable good Engel curves will be given by equation (13) with $\varphi_s = \varphi$, the same constant for all household sizes s. In this case, identification can be obtained by either Theorem 1 or Theorem 2, specifically, we can compare the coefficient of $\ln y$ both across individuals within a household and across households of different sizes to identify and hence estimate the resource shares η_{ts}.

Appendix A.4: A Fully Specified Example Model

The information and derivation in the previous section is all that is required to apply our estimator empirically. However, to clarify how our assumptions work and interact, we will now provide an example of functional forms for the entire household model that incorporate the above piglog private goods, and in particular verify that resource shares can be independent of y. If desired, this example model could be used for deeper structural analyses, such as estimation that includes price variation.

First assume each household member t has utility given by Muellbauer's piglog model so, the function υ is given by equation (11), and let $\ln F_t(p) = \ln p_t - a'\ln\widetilde{p}$ for some constant vector a with elements a_k that sum to one. This is a simple example of a function that is homogeneous as required and is a special case of $F_t(p) = p_t\widetilde{\varphi}(\widetilde{p})$ as described in the text after Assumption A3. As noted there, if all the private assignable goods have the same price, then we could instead take F_t to be any suitably regular price function, instead of requiring $F_t(p) = p_t\widetilde{\varphi}(\widetilde{p})$.

For simplicity let $y^*(p)$ be larger than any household's actual y, so the functional forms of $y^*(p)$ and of $\Psi_t(y,p)$ are irrelevant and drop out of the model. This assumption makes private assignable good Engel curves be piglog, hence linear in $\ln y$, at all total expenditure levels, not just at low levels as the theorem requires. Also for simplicity let the function $\psi_t(\upsilon + F_t, \widetilde{p}) = \exp(\upsilon + F_t)$, which by not depending upon $\widetilde{p}$ makes individual Engel curves for all goods be the same as those of the private assignable goods, and exponentiating provides a convenient cardinalization for pareto weighting utility within the household. Finally, in a small abuse of notation let $G_t(p) = G_t(p_t, \widetilde{p})$, which makes explicit the assumption that the goods p_t are assignable, so e.g. the price p_m of the good that is assignable to the father does not appear in a child's utility function, and hence does not appear in $G_c(p_c, \widetilde{p})$.

The combination of all these assumptions means that the indirect utility functions for each household member t are given by

$$\ln V_t(p, y) = \ln\left[\ln\left(\frac{y}{G_t(p_t, \widetilde{p})}\right)\right] + p_t e^{-a' \ln \widetilde{p}} \tag{14}$$

Let the function $\widetilde{U}_s$, which describes how the household weighs together the utility functions of its members, be a general Bergson-Samuelson social welfare function

$$\widetilde{U}_s\left(U_f, U_m, U_c, p/y\right) = \omega_f(p)\left[U_f + \rho_f(p)\right] + \omega_m(p)\left[U_m + \rho_m(p)\right] + \left[U_c + \rho_c(p)\right]\omega_c(p) \tag{15}$$

Note that the positive Pareto weight functions $\omega_t(p)$ and the utility transfer or externality functions $\rho_f(p)$ must be homogenous of degree zero by our Assumptions, so e.g. $\omega_t(p) = \omega_t(p/y)$, but otherwise these functions are unrestricted.

Assume the matrix A_s, which defines the extent to which goods are consumed jointly rather than privately, is diagonal, and let A_{sk} denote the k'th element along the diagonal. In the terminology of Browning, Chiappori, and Lewbel (2008), this is a Barten type consumption technology, so each A_{sk} gives the degree of publicness vs privateness of the good k in a household with s children.

Substituting this structure for A_s and equation (15) into equation (1) gives a household with s children the maximization problem

$$\max_{x_f, x_m, x_c, z_s} \omega(p) + \omega_f(p) U_f\left(x_f\right) + \omega_m(p) U_m(x_m) + \omega_c(p) U_c(x_c)$$

$$\text{such that } z_s^k = A_{sk}\left[x_{fk} + x_{mk} + s x_{ck}\right] \text{ for each good } k\text{, and } y = z_s' p$$

where $\omega(p) = \omega_f(p)\rho_f(p) + \omega_m(p)\rho_m(p) + \rho_c(p)\omega_c(p)$. This maximization can be decomposed into two steps as follows. Define resource shares η_{ts} for $t = m, f, c$ by $\eta_{ts} = x_t' A_s p/y = \sum_k A_{sk} p_k x_{tk}/y$, evaluated at the optimized level of expenditures x_t. In a lower step, conditional upon knowing η_{ts}, each household member can choose their optimal bundle x_t by maximizing $U_t(x_t)$ subject to the constraint $\sum_k A_{sk} p_k x_{tk} = \eta_{ts} y$. This is identical to standard utility maximization facing a linear budget constraint with prices $A_{sk} p_k$ and total expenditure level $\eta_{ts} y$. The resulting optimized utility level is then given by the individual's indirect utility function V_t evaluated at these shadow (Lindahl) prices, that is, $V_t\left(A_s' p, \eta_{ts} y\right)$.

Substituting these maximum attainable utility levels for each individual into the household's maximization problem then reduces the household's problem to determining optimal resource share levels by

$$\max_{\eta_{ms},\eta_{fs},\eta_{cs}} \omega(p) + \omega_f(p) V_f\left(A_s' p, \eta_{fs} y\right) + \omega_m(p) V_m\left(A_s' p, \eta_{ms} y\right) + \omega_c(p) V_c\left(A_s' p, \eta_{cs} y\right) \tag{16}$$

$$\text{such that } \eta_{ms} + \eta_{fs} + s\eta_{cs} = 1$$

Given our chosen functional form for utility, substituting equation (14), into equation (16) gives

$$\max_{\eta_{ms},\eta_{fs},\eta_{cs}} \omega(p) + \widetilde{\omega}_{fs}(p) \ln\left(\frac{\eta_{fs} y}{G_f\left(A_s' p\right)}\right) + \widetilde{\omega}_{ms}(p) \ln\left(\frac{\eta_{ms} y}{G_m\left(A_s' p\right)}\right)$$

$$+\widetilde{\omega}_{cs}(p) \ln\left(\frac{\eta_{cs} y}{G_c\left(A_s' p\right)}\right) \text{ such that } \eta_{ms} + \eta_{fs} + s\eta_{cs} = 1$$

where $\widetilde{\omega}_{ts}(p) = \omega_t(p) \exp\left(A_{st} p_t e^{-a'\left(\ln \widetilde{p} + \ln \widetilde{A}_s\right)}\right)$. Using a Lagrange multiplier for the constraint that resource shares sum to one, the first order conditions for this maximum are

$$\frac{\widetilde{\omega}_{fs}(p)}{\eta_{fs}} = \frac{\widetilde{\omega}_{ms}(p)}{\eta_{ms}} = \frac{\widetilde{\omega}_{cs}(p)}{s\eta_{cs}}$$

which has the solution

$$\begin{aligned} \eta_{ks}(p) &= \frac{\widetilde{\omega}_{ks}(p)}{\widetilde{\omega}_{fs}(p) + \widetilde{\omega}_{ms}(p) + \widetilde{\omega}_{cs}(p)} \text{ for } k \in \{m, f\} \\ \eta_{cs}(p) &= \frac{\widetilde{\omega}_{cs}(p)/s}{\widetilde{\omega}_{fs}(p) + \widetilde{\omega}_{ms}(p) + \widetilde{\omega}_{cs}(p)} \end{aligned}$$

These explicit formulas for the resource shares in this example do not depend on y, as required by Assumption A1.

Given these resource shares, the household's demand functions can now be obtained by having each household member choose their optimal bundle x_t by maximizing $U_t(x_t)$ subject to the constraint $\sum_k A_{sk} p_k x_{tk} = \eta_{ts} y$, which by standard utility duality theory is equivalent to applying Roys identity to the member's indirect utility function evaluated at prices $A_s' p$ and total expenditure level $\eta_{ts} y$, that is, $V_t\left(A_s' p, \eta_{ts} y\right)$, where the function $V_t(p, y)$ is given by equation (14).

Applying Roy's identity to equation (14) gives individual's demand functions

$$h_t^k(y, p) = \frac{y}{G_t(p_t, \widetilde{p})} \frac{\partial G_t(p_t, \widetilde{p})}{\partial p_k} - \frac{\partial\left(p_t e^{-a' \ln \widetilde{p}}\right)}{\partial p_k} \left[\ln y - \ln G_t(p_t, \widetilde{p})\right] y \tag{17}$$

for each good k and any individual t. Recalling that the sharing technology matrix A_s is diagonal, the household's quantity demand functions satisfy

$$z_s^k = A_{sk}\left[h_f^k\left(A_s' p, \eta_{fs}(p) y\right) + h_m^k\left(A_s' p, \eta_{ms}(p) y\right) + s h_c^k\left(A_s' p, \eta_{cs}(p) y\right)\right] \tag{18}$$

The demand functions of a household having s children, for each good k, are therefore obtained by substituting equation (17), and the above derived expression for $\eta_{ts}(p)$, for $t = f, m, c$, into equation (18).

Equation (17) can be written more simply as

$$h_t^k(y, p) = \widetilde{\delta}_{kt}(p)\, y - \varphi_t^k(p)\, y \ln y$$

which, when substituted into equation (18) gives household demand equations of the form

$$\begin{aligned}
\frac{z_s^k}{y} = {} & \left(\widetilde{\delta}_{kf}\left(A_s' p\right) + \widetilde{\delta}_{km}\left(A_s' p\right) + s\widetilde{\delta}_{kc}\left(A_s' p\right)\right) A_{sk} \\
& - \left(\varphi_f^k\left(A_s' p\right) \ln \eta_{fs}(p) + \varphi_m^k\left(A_s' p\right) \ln \eta_{ms}(p) + s\varphi_m^k\left(A_s' p\right) \ln \eta_{cs}(p)\right) A_{sk} \\
& - \left(\varphi_f^k\left(A_s' p\right) + \varphi_m^k\left(A_s' p\right) + s\varphi_m^k\left(A_s' p\right)\right) A_{sk} \ln y
\end{aligned}$$

For the private, assignable goods, this expression simplifies to the demand functions given earlier. Evaluating this equation in a single price regime shows that, in this model, the resulting Engel curves for all goods have the piglog form

$$\frac{z_s^k}{y} = \delta_{ks} + \varphi_s^k \eta_{ks} \ln y.$$

Appendix A.5: Empirical Tests of Model Assumptions

Theorems 1 and 2 show identification of the resource shares of individual household members from household-level Engel curve data on private assignable goods. Identification rests on four crucial assumptions: (1) that the collective household model satisfies the BCL assumptions regarding joint consumption and pareto efficiency; (2) that one of two restrictions on preferences, SAP or SAT, is true; (3) that the resource shares of individual household members do not vary with expenditure (at least at low levels of expenditure); and (3) that we observe demands for private assignable goods for each person. Violation of any of these four assumptions could imply that our estimates fail to carry the meaning implied by our structural model. Economic theory is more or less silent on the validity of these assumptions—they are all restrictive, but none violate economic principles. So we focus on empirical evaluation. Each of the four assumptions is partly testable, and we consider their validity in this section. For the first two assumptions (BCL and SAP/SAT), we use auxilliary data and invoke nonstructural tests of our structural assumptions. For the second two assumptions (invariance and private assignability), we assess the assumptions in the context of the structural model.

Appendix A.5.1. Testing BCL

We first consider tests of the BCL structure of household demand functions (which arise from BCL's assumptions regarding joint consumption and pareto efficiency). BCL implement their model for households without children, using the behaviour of single adult men and women to provide information about the preferences of men and women in married couples. In our empirical application we do not impose the BCL assumptions regarding comparability of preferences of single versus married adults. For our first set of tests we consider imposing this additional assumption regarding singles to our other assumptions, which provides a number of overidentifying implications for our model.

Using information only on the demands of single men and women and those of married childless households, BCL derive the model

$$\begin{aligned} W_t^{married}(y,p) &= \eta_t(y,p)\, w_t\left(\eta_t(y,p)\, y, A'p\right), \\ W_t^{alone}(y,p) &= w_t(y,p), \end{aligned}$$

for $t = m, f$. Here, A does not have a subscript because there is only one consumption technology, that of a married couple household. Given PIGLOG preferences for both men and women, we have $w_t(y,p) = d_t(p) + \beta_t(p)\ln y$. Assuming in addition that $\eta_t(y,p)$ does not depend on y results in the following model of Engel curves:

$$\begin{aligned} W_t^{married}(y,p) &= \eta_t\left(\delta_t + \beta_t \ln \eta_t\right) + \eta_t \beta_t \ln y, \\ W_t^{alone}(y,p) &= d_t + b_t \ln y, \end{aligned}$$

for $t = m, f$ and where $\eta_t = \eta_t(y,p)$, $\delta_t = d_t\left(A'p\right)$, $\beta_t = \beta_t(A'p)$ and $b_t = \beta_t(p)$. Given SAP, $\beta_m = \beta_f$ and $b_m = b_f$; given SAT $\beta_m = b_m$ and $\beta_f = b_f$.

BCL is a model of household demands, which are connected via the structural model to singles' demands. Without SAP or SAT, the Engel curves above are linear in $\ln y$, and any observed slopes of individual or household budget shares with respect to $\ln y$ could be rationalised with suitable choices of b_t and β_t. However, given either SAP or SAT, we can test BCL in this context because we can directly observe the preferences of individuals. Given SAP, household Engel curves are constrained only by the restriction that the slopes of men's and women's private assignable have the same sign (because η_t cannot be negative). Given SAT, household Engel curves are constrained differently: the slopes of household demands must proportional to those of singles' demands, with factors of proportionality that sum to 1.

Consider first the restriction that the slopes of household demands for men's and women's have the same sign. Using a sample of 484 married childless households from the same Malawian database as above, we run a linear SUR regression of the men's and women's clothing budget share on the log of total expenditure, interacting the intercept and slope in each equation with all demographic variables except those relating to children. This regression results in predicted slopes with respect to log-expenditure for men's and women's clothing for all 484 households. These slopes may differ across men's and women's clothing shares, and across values of the demographic variables. The slopes have a mean of 0.004 and 0.003 for men's and women's clothing, respectively, which satisfies the 'same sign' restriction. However, these slopes vary substantially across households (due to variation in demographic variables), so that for both men's clothing shares and women's clothing shares, we observe both positive and negative values of this slope. The slopes of men's and women's clothing shares are highly correlated, with a correlation coefficient of 0.86, so they tend to either both be positive or both be negative: 30% have both negative and 56% have both positive. However, 14% of observations of married couples have a positive slope for one clothing share and a negative slope for the other, though in many of these cases these slopes are not statistically significantly different from zero, so one would have difficulty rejecting the hypothesis that many of these households actually do have the same signs.

Consider next the restriction given SAT that the slopes of household demands are proportional to individual demands, with factors of proportionality summing to 1. Using a sample of 307 single men, 168 single women, and the same 484 married childless households as above, we run a linear SUR regressions of the men's and women's clothing budget share[1] on the log of total expenditure and all demographic variables except those relating to children and to spousal characteristics. All regressors are interacted with a dummy for married households, so that all coefficients can differ between married couple and single adult households. For men, the ratio of slopes in married versus single households is 0.39; for women, it is 0.55. Although the sum of 0.94 is not exactly 1, it is insignificantly different from 1. Taken together, these results suggest that the model of BCL, either with or without the SAP and SAT restrictions, does not impose violated restrictions on the behaviour of households.

Appendix A.5.2. Testing the SAP and SAT restrictions on preferences.

Here we consider whether the SAT, SAP, or both assumptions are valid. Imposing both SAP and SAT results in overidentifying restrictions, so for our first set of tests we impose SAP and test the additional SAT restrictions on that model, and then vice versa. We first estimate the model (on the combined clothing and footwear private assignable good) under SAP and conduct a Wald test of the hypothesis that the coefficients on the household size dummies inside β are identical for the 4 household types. The sample value of this test statistic is 1.1, and it is distributed as a χ^2_3 which has a 5 per cent critical value of 6 under the null hypothesis that SAP and SAT both hold. Alternatively, we also estimate under SAT and conduct a Wald test of the hypothesis that the β_t are the same for all persons t. Since each of the 3 person-specific β_t functions has 15 parameters, this amounts to testing 30 restrictions. The sample value of this test statistic is 7.4, and it is distributed as a χ^2_{30} with a 5 per cent critical value of 43.8. Thus, the combination of SAP and SAT is not much worse than either SAP or SAT separately, so we favour estimates that combine SAP and SAT.

Now consider testing SAP or SAT separately. Since resource shares are exactly identified given SAP, there are no overidentifying conditions that can be used to directly test SAP by itself with only one assignable good (though, we do obtain testable overidentifying restrictions with two assignable goods, as we show and use below). Unlike SAP, resource shares are overidentified given just SAT when there are more than three household sizes, so our setting with four household sizes allows us to test this overidentifying restriction. Given SAT and four household sizes, there are 12 identifiable slopes of W_{ts} with respect to $\ln y$ (4 household sizes times 3 goods), and they depend on 8 resource share functions (4 household

[1] Eight single men and one single woman had nonzero expenditures for the other sex's clothing. These expenditures were recoded to zero.

sizes times 2 resource share functions, where the third is given by the summation restriction) and 3 latent slopes β_t. Thus, we can add an additional slope parameter to the model, which is "on" for one household size for one person's assignable good, and test the exclusion restriction on this additional parameter. Of course, this additional parameter must be interacted with the 14 demographic parameters as well, yielding a total of 15 parametric restrictions. The sample value of the Wald test statistic for this restriction is 0.4, and it is distributed as a χ^2_{15} with a 5 per cent critical value of 25. Thus, we do not reject SAT against this more general alternative.

Additional overidentifying restrictions can be obtained given additional private assignable goods k. We implement a $k = 2$ private goods model by separating clothing and footwear expenditures, treating each as a separate private assignable good, and so estimate a separate clothing budget share equation and footwear budget share equation for each member of the household, imposing both SAT and SAP to obtain the strongest possible test. In this test we estimate the resource shares η_{ts} for each person using each good k, and then test that the estimated resource shares do not vary by k. The implication of the model that the estimated resource shares η_{ts} recovered from the clothing equations are the same as those obtained from the footwear equations gives a total of 36 restrictions – the 2 resource share functions each have 18 parameters (4 household sizes and 14 demographic variables). The sample value of the likelihood ratio test statistic for this restriction is 15.4, and is distributed as a χ^2_{36} with 5 per cent critical value of 51 under the null hypothesis that our resource shares are unique. In contrast, the sample value of the Wald test statistic for this restriction is 72, so the Wald and likelihood ratio tests disagree regarding rejection at standard significance levels.[2] Footwear as a separate category is a very small component of total expenditures and yielded very erratic estimates, which likely affects the outcome of this test, and is why we have more confidence in our estimates that combine footwear and clothing into a single consumption category.

Next we consider restrictions on SAP and SAT that go beyond our main data set. SAP and SAT are restrictions on the preferences of individuals, so we next test if these restrictions are satisfied by single men and single women living alone. Our main results only require SAP and SAT to hold for couples with children, but we find below that these preference restrictions also appear to hold for single men and women, which strengthens our confidence in the validity of these restrictions. An added advantage of testing with single men and women is that the complications associated with the presence of shared and public goods within a household do not arise with singles.

Consider first a test of the hypothesis that SAP holds across single men and single women. Given PIGLOG preferences for both men and women, we have $w_t(y, p) = d_t(p) + \beta_t(p) \ln y$. SAP implies $\beta_t(p) = \beta(p)$, a restriction on demands at a given price vector. SAT implies $\beta_t(p) = \beta_t$, a restriction on how demands vary across price vectors. To test SAP, we use the same sample of 307 single men and 168 single women we used for testing BCL earlier, and estimate separate regressions for men's and women's clothing budget shares on the log of total expenditure, interacting the intercept and slope in each

[2]We suspect that this big difference between Wald and Likelihood Ratio is due to the size of our model, and suggests that one should be cautious in interpreting our test statistics and confidence bands. Recent work exists on resolving differences between Wald and Likelihood Ratio tests in finite samples and high dimension, but generally applies to specific models. See, e.g., Belloni and Didier (2008), Annals of Statistics, 36, 2377-2408).

equation with the 7 demographic variables that do not relate to children and to spousal characteristics. The sample value of the likelihood ratio test statistic for the hypothesis that the coefficient on $\ln y$ and its 7 demographic interactions are the same for men and for women is 5.6. It is distributed as a χ^2_8 under the Null, with a 5% critical value of 15.5. The sample value of the Wald test statistic for this hypothesis is 9.9. So, the observed behaviour of single men and single women is consistent with SAP.

SAT does not restrict how preferences vary across individuals; rather it restricts how preferences vary across price vectors. In particular, with PIGLOG preferences, SAT implies that the slopes of budget shares for private assignable goods do not vary with prices. We use data on an additional 492 single men and 355 single women from the 1999/2000 wave of the IHS, deflating total expenditure by the change in the World Bank price index for Malawi. (These data are not as good for the analysis of collective households as are the 2004 data, since they lack some demographic covariates and all instruments, but they suffice for the study of single individuals.) We estimate separate regressions for men's and women's clothing budget shares on the log of total expenditure, interacting the intercept and slope in each equation with the year of the survey and with the 7 demographic variables that do not relate to children and to spousal characteristics. If preferences are stable over the 5 years separating the survey waves, then the year dummy in the intercept and slope capture the response to relative price changes. In this case, SAT implies that the year dummy may be excluded from the slope term. For men, the sample value of the z-statistic for this hypothesis is 1.34; for women, it is -1.12. So, the observed behaviour of single men and single women across these two survey years is consistent with SAT.

We conclude on the basis of these tests that both SAT and SAP are reasonable restrictions on our data.

Appendix A.5.3. Testing resource share invariance

The restriction that resource shares are invariant to expenditure has been invoked several times in the literature for reasons of convenience, rather than of economics. We show elsewhere in the appendix that there exist reasonable structural models of household decision-making that imply that resource shares are invariant to expenditure. This tells us only that this sort of invariance is possible, not that it holds in reality. Lise and Seitz (2004) and BCL do not require this restriction for identification, but they both impose it in their empirical work. Lewbel and Pendakur (2008) and Bargain and Donni (2009) invoke the restriction for identification and use it in their empirical work. In this subsection, we consider whether or not it holds empirically in our setting.

To test invariance, we run the same nonlinear SUR as that reported in the rightmost column of Table 2, but with an additional covariate in the η_{ts}, δ_{ts} and β parameters. This regression applies both SAP and SAT. The additional covariate is a dummy variable indicating that the household is in the top half of the total expenditure distribution. Our model would permit this variable to enter the preference parameters δ_{ts} and β, but if it enters the resource share η_{ts} then our identifying restriction is violated. The z-test statistic on its exclusion for the man's resource shares is 2.11; for the woman's resource share, it is -1.87. But, these tests covary, so the sample value of the Wald test statistic for the hypothesis that the high expenditure dummy may be excluded from both the man's and woman's resource share function is 4.6, which is lower than the 5 per cent critical value of 6.0. The p-value of the test statistic is 0.102, so the test may be seen as marginally significant. We take these results as suggesting that invariance of resource shares may be tolerable as a modeling assumption, but that identification results which allow for its relaxation would be welcome. Identification results in BCL suggest that this might be done by introducing price variation, exploiting data over multiple time periods.

Appendix A.5.4. Testing if clothing is a private, assignable good.

The restriction that clothing is private may be violated in at least two ways. First, clothing may have an externality such that some household members derive utility from the clothing worn by other household members. For example, an adult might get utility from their spouse being well-dressed. In this case, the consumption externality renders the BCL model inapplicable, because it implies inefficiency of decentralised decision-making, and the estimates resulting from implementation of our structural model would not correspond to the resource shares of each household member.

Second, clothing might be shared across household members. For example, similarly aged children may share clothing, especially if they are the same gender. In our context, this would imply that the diagonal element of A_s corresponding to children's clothing is less than 1, implying that the private good equivalent of clothing expenditure is greater than the market expenditure on clothing. Identification given SAT uses the restriction that the market price of the private assignable equals its shadow price, which requires that the diagonal element of A_s corresponding to children's clothing be exactly 1 (and all the off-diagonal elements be 0). Thus, if sharing of clothing is important for children, we cannot use SAT identification of children's resource shares. Identification using SAP rests on the assumption that every person in the household faces the same shadow price vector. However, if children can share their clothing a lot, but adults cannot share their clothing, then children face a lower shadow price for clothing than adults. Consequently, if sharing of clothing is important for children, we cannot use SAP identification of children's resource shares.

Thus, if clothing has consumption externalities, or if clothing is shared for some but not all family members, our methods cannot be based on clothing as a private assignable good. This is because a consumption externality from one person's clothing demand to another person's utility violates the assumptions of BCL, and sharing of clothing violates the SAP and SAT conditions required for identification of children's resource shares given BCL.

Consider first the possibility that clothing expenditures cause externalities across household members. If we assume that the major externalities of this form are between husbands and wives, then estimates based on lone parent families should not be polluted in this way. We implement our model on a sample of female-headed lone parent families. Because such families make up less than 10% of the households in our data, the 2004 data do not provide sufficient observations to test our hypothesis. So, we pool the 1999/2000 and 2004 waves of the Integrated Household Survey. We exclude households with more than one person aged 16 or more, and households whose head was aged less than 16 or more than 58. Our sample then consists of 1184 female-headed lone parent households: 390 with 1 child; 362 with 2 children; 293 with 3 children; and 139 with 4 children. The model is the same as that in Table 2, except that we estimate only the resource share of the female, η_{fs}, with the children's share calculated as $1- \eta_{cs}$. The preference parameters β are linear in the following demographic variables: a dummy for the 1999/2000 survey wave; region of residence; the average age of children less 5; the minimum age of children less 5; the proportion of children who are girls; the age of the woman less 22; and the education level of the woman. (The 1999/2000 wave did not collect information on month of collection (dry season dummy), religion or distance to road or daily market.) The preference parameters and resource shares are linear in these demographic variables and a set of household size dummies. Table 3 presents estimated resource shares in lone parent families analogous to those for dual parent families presented in Table 2.

Table 3: Estimates for Female Lone Parent Families

	SAP		SAT		SAP&SAT	
	Estimate	*Std Err*	Estimate	*Std Err*	Estimate	*Std Err*
one child	0.603	0.072	0.544	0.085	0.604	0.069
two children	0.331	0.082	0.207	0.100	0.334	0.078
three children	0.258	0.089	0.087	0.039	0.246	0.086
four children	0.206	0.091	0.154	0.089	0.097	0.041
min. age children	-0.004	0.008	0.003	0.005	-0.001	0.008
avg. age children	-0.007	0.010	-0.006	0.005	-0.007	0.009
prop. girl children	0.022	0.045	0.016	0.032	0.027	0.044
woman age	-0.002	0.002	-0.001	0.001	-0.002	0.002
woman education	-0.011	0.018	0.005	0.014	-0.007	0.019

Here, we see clearly that the children's resource share rises with the number of children, and that it rises more slowly as the number of children increases. For example, given the estimates which impose both SAP and SAT, the children's share is about 0.40 for 1 child, 0.67 for 2 children, 0.75 for three children and 0.90 for 4 children (the standard errors are the same as those for the woman's share). With the other demographic variables, the precision of the estimates is low, so we cannot assess very well whether or not the patterns are the same as for dual parent families.

Given SAT, we may assess whether or not the preferences of women are different depending on whether they are single or dual parents. If SAT is true, and if clothing does not have an associated consumption externality, then women's preferences will be the same whether or not they are single or dual parents. This is not a pure test of privateness, because it tests the joint restriction of SAT and privateness. However, given our other evidence that SAT holds, if women's preferences vary significantly between single and dual parent households, we would take that as suggestive of a consumption externality. To implement this idea, we test whether or not all parameters relating to region of residence, child gender and age, and woman's age and education, are the same in the estimated β_f (the women's latent slope term) in the regressions corresponding to the SAT estimates in Table 2 (dual parents) and Table 3 (single parents). (We do not include the parameters for dry season, distance to road or daily market, or religion because they are not available for single parents.)[3] The sample value of the Wald test statistic for this hypothesis is 5.3. Under the Null, it is distributed as a χ^2_8, with a 5% critical value of 15.5. This test is somewhat

[3]Tests which restrict the dual parent model to exclude these variables reach the same conclusion.

weak, because under SAT alone, the β_t slope parameters are estimated quite imprecisely. Alternatively, we may impose both SAP and SAT to increase the precision of the estimated slope parameters. In this case, males, females and children are all restricted to have the same latent slopes (so that $\beta_t = \beta$), so the test of sameness across single and dual parent households is stronger. The sample value of the Wald test statistic for this hypothesis is 17.3, which exceeds the χ^2_8 5% critical value of 15.5, but not its 1% critical value of 20.1. These results suggest that we may or may not reject the Null hypothesis that women's preferences are the same in single and dual parent households, but the evidence in favour of rejection is not overwhelming.

These estimates and tests compare the behaviour of single mothers, where there is no consumption externality across adults, to the behaviour of married couples with children, where there may be a consumption externality across adults. We see similar patterns in the variation of resource shares across numbers of children for single- and dual-parent households. We find little evidence that mothers' preference parameters are different across these groups. We therefore conclude that the consumption externalities in clothing are not behaviourally important in our Malawian context.

Now consider the second possibility for violation of the private assignability of clothing—sharing of clothing among household members. We assess this possibility by estimating the model for a private assignable good that is a priori less shareable than clothing—footwear. We estimate the model given SAP and SAT, and using the same data as that corresponding to Table 2, with just footwear as a private assignable good, and with just non-footwear clothing as a private assignable good. If non-footwear clothing is substantially polluted by sharing, we would expect the estimated resource shares in the latter model to be quite different from those in the former. The leftmost two panels of Table 4 present estimates of resource shares for men and women, analogous to those presented in Table 2. We suppress reporting of results for children, since these may be computed from the resource shares of adults.

Table 4: Estimates from Malawian Clothing and Footwear Budget Shares, SAP&SAT

		Footwear		Clothing		Both	
		Estimate	*Std Err*	Estimate	*Std Err*	Estimate	*Std Err*
one child	man	0.495	0.083	0.403	0.053	0.420	0.045
	woman	0.290	0.065	0.402	0.049	0.378	0.030
two children	man	0.493	0.089	0.426	0.060	0.448	0.052
	woman	0.222	0.064	0.276	0.052	0.240	0.044
three children	man	0.541	0.093	0.440	0.064	0.470	0.053
	woman	0.145	0.060	0.217	0.056	0.172	0.043
four children	man	0.420	0.116	0.350	0.075	0.375	0.067
	woman	0.171	0.070	0.230	0.066	0.187	0.052
min. age	man	0.013	0.017	0.001	0.010	0.003	0.010
of children	woman	-0.013	0.011	-0.010	0.010	-0.013	0.008
avg. age	man	-0.014	0.017	-0.004	0.010	-0.003	0.010
of children	woman	0.008	0.011	0.017	0.010	0.017	0.007
proportion	man	-0.018	0.047	-0.017	0.035	-0.012	0.028
girl children	woman	0.012	0.038	0.076	0.034	0.071	0.027

It is clear that the estimated resource shares using footwear alone are much noisier than those using clothing alone. The standard errors on resource shares using footwear alone are as much as twice the size of those using clothing alone. However, the broad features of resource shares noted in Table 2 are all visible in these panels of Table 4: men's resource shares are roughly invariant to the number of children and to their characteristics and women's resource shares decline strongly with the number of children. But, it appears that the standard errors in the estimates based on footwear alone are too large to detect the effects of child age and gender proportion (if those effects are similar in size to those reported in Table 2).

Under the model, the resource share function should be the same regardless of which private assignable good we use. The rightmost panel in Table 4 presents estimates corresponding to this model. Here, we use information from both private assignable goods to inform the resource shares, and as a consequence, the standard errors are tighter than in either of the other panels. In this panel, we see all the results from Table 2 again: roughly constant men's shares; women's shares strongly declining in the number of children; and women's shares rising in the average age of children and the proportion of children who are girls.

A formal test that the estimated resource shares estimated from just the clothing shares are the same as those estimated from just the footwear shares has 36 restrictions – the 2 resource share functions (men and women) each have 18 parameters (4 household sizes and 14 demographic variables). The sample value of the likelihood ratio test statistic for this restriction is 28, and is distributed as a χ^2_{36} with 5 per cent critical value of 51 under the null hypothesis that our resource shares are unique. In contrast, the sample value of the Wald test statistic for this restriction is 80. The Bonferroni adjusted p-values for these 36 individual tests suggest that the violations of equality are driven by 2 of the demographic covariates, and not by household size. The sample value of a Wald test statistic for the less restrictive hypothesis that the

4 household size parameters are the same for men and women across the two specifications is 3.2, which is smaller than 15.5, the 5 per cent critical value of the χ^2_8 distribution.

The bottom line from this model assessment exercise is that our four crucial modeling assumptions necessary to achieve the identification of children's resource shares in collective households are for the most part satisfied by the Malawian household expenditure data. In the minority of cases where tests of overidentifying restrictions are rejected, the estimated patterns of behavior implied by the assumptions still generally hold.