ESTIMATING HOUSEHOLD WELFARE FROM DISAGGREGATE EXPENDITURES

ETHAN LIGON

Abstract. Existing models of life-cycle demand often (perhaps implicitly) assume that Engel curves are linear. This allows one to specify utilities as a function of nothing more than real expenditures, but is sharply at odds with strong empirical evidence against linearity, including Engel's Law. Here we show how one can use data on disaggregate expenditures to estimate demand systems that may feature flexible and highly non-linear Engel curves; this same estimation procedure yields an index of household welfare closely related to the marginal utility of expenditures within a period. We illustrate the use of these methods using data from Uganda to estimate an incomplete demand system, and to estimate household welfare in different periods.

Regarded as an index of the marginal utility of expenditures, our measure plays a central role in models involving dynamics and risk. We use our estimates to look for evidence of either borrowing or savings constraints in Uganda, and find no such evidence; separately, we find strong evidence of heterogeneity in relative risk aversion.

1. Introduction

Measures of household-level consumption expenditures are central to a wide variety of important research questions in many fields of economics, and particularly in models involving risk or life-cycle behavior. The household surveys which collect these data almost invariably record disaggregate expenditures;

. *Date*: July 6, 2017.

. This paper has been in the works for several years, and I've benefited from discussions with economists too numerous to mention. However, I must particularly thank researchers at the World Bank for their encouragement and help with data. I'd also like to particularly acknowledge Orazio Attanasio, Richard Blundell and their colleagues at the University College of London, where I learned a great deal about demand whilst on sabbatical.

that is, expenditures on particular kinds of goods or services. However, empirical work employing these data to measure changes in household welfare typically focuses on the sum of these disaggregate expenditures, divided by a price index, or *total* real household consumption expenditures (Browning, Crossley, and Winter 2014).[1] Information on the *composition* of the household's consumption portfolio is generally discarded.

There are two questions one might raise about this procedure: First, when will it be *valid*, in the sense that the resulting measure of real total consumption is what actually matters to the household? And second, when will it be *efficient*, in the sense that there's no further useful information regarding welfare in the *composition* of the consumption portfolio?

That there's a single answer to these two questions has long been known, if not always widely acknowledged: both validity and efficiency hinge on within-period household utility being both homothetic and having an additively separable representation. This is more evidence, if any were needed, that homothetic preference structures have desirable theoretical properties. It's rather a pity that the evidence against homotheticity (e.g., Engel 1857; Houthakker 1957) is so extremely overwhelming.

In this paper we abandon homotheticity and devise methods to exploit information on the composition of households' consumption portfolios.[2] Effectively, we estimate the Lagrange multiplier λ associated with the budget constraint of the consumer's problem, using only cross-sectional information on demand for different kinds of non-durable consumption.

As a welfare measure, this Lagrange multiplier has many desirable properties because of its relationship to the consumer's marginal utility of expenditures. But the Lagrange multiplier in the consumer's problem depends not only on the budget constraint, but also on the cardinal properties of the consumer's utility function. Given a particular cardinalization, the Lagrange multiplier

1. To take an arbitrary sample, of the 69 empirical papers published in the *American Economic Review* in 2013, one third used some sort of data on consumer expenditures. Of these 23, 13 relied exclusively on a measure of real total household consumption expenditures.

2. In this we follow a strategy related to that pursued by Aguiar and Bils (2015) and Young (2012), who both estimate Engel curves for different goods, but in the Marshallian framework adopted by those papers the estimated Engel curves aren't consistent with utility maximization unless the Engel curves are trivial and the utility function homothetic.

can be interpreted as the household's marginal utility of expenditures within a period. But purely cross-sectional data on expenditures sheds little light on risk attitudes or intertemporal choice, since any monotonic transformation of the utility function would yield the same demands, but might yield different marginal utilities of expenditure (Deaton and Muellbauer 1980, pp. 140–41). Thus, when we estimate λ given a particular cardinal utility function, one should think of our estimates as being an index of the consumer's 'true' marginal utility of expenditures, or IMUE.

This paper proceeds by first sketching a simple model of household demand behavior, and using this model to derive a set of λ-constant or "Frischian" demands (e.g., Heckman and MaCurdy 1980; Browning, Deaton, and Irish 1985), using a strategy related to that taken by Attfield and Browning (1985). The class of Frischian demands we estimate do not generally have an explicit Marshallian representation, but when separable can be regarded as an instance of the non-homothetic implicitly-additive Marshallian demands studied by Hanoch (1975), and recently exploited empirically by Comin, Lashkari, and Mestieri (2015).

We next show how the Frischian demand system we derive can be estimated using one or more rounds of cross-sectional data on disaggregate household expenditures, in a specification involving logarithms of those disaggregate expenitures. This estimator delivers "Frisch elasticities" and estimates of an index of each household's marginal utility of expenditures, along with estimates of the effects of various observable household characteristics on demand.

Finally, we illustrate our methods using data from four rounds of household expenditure surveys in Uganda. Using nothing more than cross-sectional variation in expenditures, we're able to obtain estimates of both the parameters of the demand system and the households' IMUE. We relate these measures of welfare to traditional expenditure-based measures of headcount poverty, and observe that the IMUE measures avoid serious problems involve price indices that plague expenditure-based approaches. Separately, we use the estimated IMUE to calculate the distribution of households' relative risk aversions, up to unknown location and scale parameters. There is strong evidence of heterogeneity in these relative risk aversions across households.

2. A Frisch Approach

In the rest of this paper we will describe an approach for measuring our index of household marginal utility of expenditures (IMUE) which uses Engel-style facts about the composition of differently-situated consumers' consumption bundles; which has comparatively modest data requirements; which allows us to simply *ignore* expenditures on goods and services which are too difficult or expensive to measure well; and which completely avoids the price index problem by simply avoiding the need to construct price indices. The approach imposes fewer restrictions on the demand system than is usual; avoids the usual sausage factory from which consumption aggregates are extruded; should allow for much less expensive data collection; and directly yields measures of both household marginal utility and functions of shadow prices which can be used in subsequent analysis and model testing.

Our IMUE has a very precise theoretical interpretation: it is the rate at which a particular cardinalization of household utility would increase if the household's expenditures received a small increase in a given period. Provided that the household has a concave momentary utility function (the usual assumption), then IMUE will decrease as resources increase. Given the cardinalization of utility this same quantity goes by other names, but all of them are awkward or imprecise: the "marginal utility of income" (inaccurate, since a change in income will generally affect utility in several different periods); the "Lagrange multiplier on the budget constraint" (mathematically accurate, but devoid of intuition regarding the consequences for the household); the "marginal utility of expenditures" (perhaps the best of a bad lot).[3] These are not only awkward, but also imprecise, since none of these is invariant to a monotonic transformation of the utility function. The construction of any utility-based welfare measure requires some sort of cardinal utility; our index is the marginal utility of expenditures for a cardinalization that happens to have particularly nice properties. This does not mean that we're claiming that households' utility functions necessarily correspond to this particular cardinal utility function; actual momentary utility functions could be any monotonic

3. The problem of naming this quantity has a long history; Irving Fisher was already complaining about it in 1917. Fisher himself offers the coinage "wantab" (Fisher 1927).

transformation of our cardinalization. In this case our measure will be an index of the true marginal utility of expenditures, varying with it one-to-one for any given set of prices.

The question of how what we're calling IMUE is related to consumer demand and welfare was extensively considered by Ragnar Frisch (see esp. Frisch 1959, 1964, 1978). Demand systems which depend on prices and marginal utility were revived by Heckman and MaCurdy in the seventies (J. J. Heckman 1974, 1976; Heckman and MaCurdy 1980; MaCurdy 1981) and somewhat later given the moniker "Frisch demands" by Martin Browning (Browning, Deaton, and Irish 1985). However, previous approaches to estimating Frisch demand systems have imposed much more structure on the underlying consumer Engel curves than is necessary.

In this paper we implement the first step in the "sequential approach" advocated by Blundell (1998) to the estimation and testing of life-cycle models of the household. We take disaggregate data from one or more rounds of a household expenditure survey to estimate a Frischian demand system. Estimating such a system allows us to more or less directly recover estimates of some demand elasticities and households' IMUE in each round, which can then be used as an input to a subsequent (possibly dynamic) analysis.[4]

Flexibly estimating the IMUE sequentially has great potential value in part because the marginal utility of expenditures is a central object in much recent work on risk and dynamics in both low- and high-income countries. A large number of recent papers featuring data from low-income countries assume that a household's marginal utility of expenditures can be modeled as the household's total real household consumption raised to a common negative exponent; examples include Kinnan (2014) and Karaivanov and Townsend (2014). Papers by Chiappori et al. (2014) and Laczó (2015) relax this by allowing different households to have different exponents. But this still involves assuming that utilities are homothetic, and requires the marginal utility of

4. It would also be possible to estimate the demand system and the dynamic model jointly (as in, e.g., Browning, Deaton, and Irish 1985). But since we so often are able to reject the dynamic models we estimate, joint estimation seems likely to result in a mis-specified system; here, we prefer to not impose any dynamic restrictions on the expenditure data so as to allow ourselves to remain comfortably agnostic about what the 'right' dynamic model ought to be.

expenditures to depend on a single parameter which *also* governs the elasticity of intertemporal substitution. Non-parametric approaches such as that of Mazzocco and Saini (2011) are much less restrictive, but this comes at the cost of not allowing for actual measurement of the IMUE.

Some life-cycle studies of consumers in high-income countries also adopt a Frischian approach, following a line of research established in the seventies and eighties (J. Heckman 1974; Heckman and MaCurdy 1980; MaCurdy 1981; Browning, Deaton, and Irish 1985). These include pioneering papers such as Blundell, Fry, and Meghir (1990), Blundell, Browning, and Meghir (1994), and Blundell, Pistaferri, and Saporta-Eksten (2016). But the focus of these papers remains on highly aggregate forms of consumption and leisure; none of these exploit within-period consumer choices among disaggregate consumption goods, as in the present paper.

There is a vast literature on different approaches to estimating demand systems, so it is surprising to discover that none of these approaches seems well-suited to our problem. The first issue is simply that almost all existing approaches are aimed at estimating Marshallian demands, rather than Frischian.[5] Related, demand systems which are nicely behaved (e.g., linear in parameters) in a Marshallian setting are typically ill-behaved in a Frischian. This includes essentially all of the standard demand systems based on a dual approach (e.g., the AID system). Other existing demand systems can be straight-forwardly adapted to estimating Frischian demand systems, such as the Linear Expenditure System (LES), which can be derived from the primal consumer's problem when that consumer has e.g., Cobb-Douglas utility. But such systems are too restrictive, imposing a linearity in demand which is sharply at odds with observed demand behavior.

3. Model of Household Behavior

In this section we give a simple description of the IMUE function λ, which maps prices and expenditures into a welfare function (higher values of the function mean that the household is in greater need), and which also serves

5. Notable exceptions include Browning, Deaton, and Irish (1985), Blundell (1998), and Cooper, McLaren, and Wong (2001).

as the central object for making predictions regarding *future* welfare. Indeed, there exists a particular cardinalization of the utility function such that λ is the household's marginal utility of expenditures; more generally λ can regarded as an index, with which the household's marginal utility of expenditures must vary one to one, for any set of prices.

3.1. **The household's one-period consumer problem.** To fix concepts, suppose that in a particular period t a household with some vector of characteristics z_t faces a vector of prices for goods p_t and has budgeted a quantity of the numeraire good x_t to spend on contemporaneous consumption, from which it derives utility via an increasing, concave, continuously differentiable cardinala utility function U. Within that period, the household uses this budget to purchase non-durable consumption goods and services $c \in X \subseteq \mathbb{R}^n$, solving the classic consumer's problem

$$V(p_t, x_t; z_t) = \max_{\{c_i\}_{i=1}^n} U(c_1, \ldots, c_n; z_t) \tag{1}$$

subject to a budget constraint

$$\sum_{i=1}^{n} p_{it} c_i \leq x_t. \tag{2}$$

The solution to this problem is characterized by a set of n first order conditions which take the form

$$u_i(c_1, \ldots, c_n; z_t) = \lambda_t^* p_{it} \tag{3}$$

(where u_i denotes the ith partial derivative of the momentary utility function U), along with the budget constraint (2), with which the Lagrange multiplier λ_t^* is associated.

So long as U is well-behaved the solution to this problem delivers a set of demand functions, the Marshallian indirect utility function V, and a Frischian measure of the marginal value of additional expenditures in period t, $\lambda_t^* = \lambda^*(p_t, x_t; z_t)$.

It is this last object which is of central interest for our purposes. By the envelope theorem, the quantity $\lambda_t^* = \partial V/\partial x_t$; it's thus positive but decreasing in x_t, so that marginal utility decreases as the total value of per-period expenditures increase.

3.2. **The household's intertemporal problem.** Since we are ultimately interested in the welfare of households in a stochastic, dynamic environment, we relate the solution of the static one-period consumer's problem above to a multi-period stochastic problem; at the same time we introduce a simple form of (linear) production (this could be easily generalized).

We assume the household has time-separable von Neumann-Morgenstern preferences, and that it weights future utility using a discount factor β_t (allowed to vary across periods). The resulting additive separability across dates and states means that can treat the household's global problem using a two-stage budgeting approach (Gorman 1959). As above, within a period t, a household is assumed to allocate funds for total expenditures in that period obtaining a total momentary utility described by the Marshallian indirect utility function $V(p_t, x_t; z_t)$, where p_t are time t prices, x_t are time t expenditures, and z_t are time t characteristics of the household. Note that the indirect utility function V inherits the cardinality of the utility function U; this is the household's "true" indirect utility function.

The household brings a portfolio of assets with total value $R_t b_t$ into the period, and realizes a stochastic income y_t. Given these, the household decides on investments b_{t+1} for the next period, leaving x_t for consumption expenditures during period t. More precisely, the household solves

$$\max_{\{b_{t+1+j}\}_{j=1}^{T-t}} \mathrm{E}_t \sum_{j=0}^{T-t} \beta_j V(p_{t+j}, x_{t+j}; z_{t+j})$$

subject to the intertemporal budget constraints

$$x_{t+j} = R_{t+j} b_{t+j} + y_{t+j} - b_{t+1+j}$$

and taking the initial b_t as given.

The solution to the household's problem of allocating expenditures across time will satisfy the Euler equation

$$\frac{\partial V}{\partial x}(p_t, x_t; z_t) = \frac{\beta_{t+j}}{\beta_t}\mathrm{E}_t R_{t+j}\frac{\partial V}{\partial x}(p_{t+j}, x_{t+j}; z_{t+j}).$$

But by definition, these partial derivatives of the indirect utility function are equal to the functions λ^* evaluated at the appropriate prices and expenditures, so that we have

$$\lambda^*(p_t, x_t; z_t) = \frac{\beta_{t+j}}{\beta_t}\mathrm{E}_t R_{t+j}\lambda^*(p_{t+j}, x_{t+j}; z_{t+j}). \tag{4}$$

This expression tells us, in effect, that the household's marginal utility or marginal utility of expenditures λ_t^* satisfies a sort of martingale restriction, so that the current value of λ_t^* play a central role in predicting *future* values λ_{t+j}^*.

If we know the Frisch demand functions for a consumer with utility function U and observe prices and quantities demanded for some of these goods, then we can invert the demand relationship to obtain the consumer's λ_t^*.

3.3. **Differentiable Demand.** We now turn our attention to the practical problem of specifying a Frischian demand relation that can be estimated using the kinds of data we have available on disaggregated expenditures. Attfield and Browning (1985) take a so-called "differentiable demand" approach to a related problem; their method yields Frischian (aggregate) demands without requiring separability. These demands will, in general, depend on all prices, yet one need only estimate demand equations for a select set of goods.

Our analysis here follows that of Attfield and Browning (1985) in outline, but where they arrive at a Rotterdam-like demand system in quantities, we obtain something importantly different in expenditures. This overcomes an important shortcoming of Attfield and Browning's demand system, which is that it is integrable only in the homothetic case.

It's easiest here to work with the consumer's profit function (Gorman 1976; Browning, Deaton, and Irish 1985),

$$\pi(p, r, z) = \max_c rU(c; z) - pc,$$

where r has the interpretation of being the "price" of utility. Let subscripts to the π function denote partial derivatives. Some immediate properties of importance: the price r is equal to the quantity $1/\lambda^*$ from our earlier analysis; the profit function is linearly homogeneous in p and r; by the envelope theorem $\pi_i(p,r,z) = -c_i$ for all $i = 1,\dots,n$ and for any z; and (since we want to work with expenditures) $-p_i\pi_i = x_i$.

Using this last fact and taking the total derivative yields

$$dx_i = -\pi_i dp_i - p_i \sum_{j=1}^{n} \pi_{ij} dp_j - p_i \pi_{ir} dr - p_i \sum_{l=1}^{\ell} \pi_{iz_l} dz_l.$$

Since $d\log x = dx/x$ for $x > 0$, this can be written as

$$x_i d\log x_i = -\pi_i p_i d\log p_i - p_i \sum_{j=1}^{n} \pi_{ij} p_j d\log p_j - p_i \pi_{ir} r d\log - p_i \sum_{l=1}^{\ell} \pi_{iz_l} z_l.$$

Recalling that $-\pi_i p_i = x_i$

$$d\log x_i = d\log p_i + \sum_{j=1}^{n} \frac{\pi_{ij}}{\pi_i} p_j d\log p_j + \frac{\pi_{ir}}{\pi_i} r d\log r + \sum_{l=1}^{\ell} \frac{\pi_{iz_l}}{\pi_i} z_l d\log z_l. \tag{5}$$

Now, let $\theta_{ij} = -\frac{\pi_{ij}}{\pi_i} p_j$ denote the (cross-) price elasticities of demand holding r constant (Frisch 1959, called these "want elasticities"); let δ_{il} denote the elasticity of demand for good i with respect to changes in the characteristic z_l; and let $\beta_i = \frac{\pi_{ir}}{\pi_i} r$ denote the elasticity of demand with respect to r. Using the fact that $1/r = \lambda^*$ we can rewrite this as

$$d\log x_i = d\log p_i - \sum_{j=1}^{n} \theta_{ij} d\log p_j - \sum_{l=1}^{\ell} \delta_{il} d\log z_l - \beta_i d\log \lambda^*. \tag{6}$$

Using the linear homogeneity of the profit function, it follows that $\beta_i = \sum_{j=1}^{n} \theta_{ij}$.

Equation (6) gives us an exact description of how expenditures will change in response to infinitesimal changes in prices for a consumer with the utility function U and characteristics z.

Now we make an assumption which is important for reasons both involving principle and practice: that the elasticities $\Theta = [\theta_{ij}]$ (and so $\beta = [\beta_i]$) and $\delta = [\delta_{il}]$ are *constant*, and not functions of prices (p, r) or characteristics z. With this assumption, the matrix of parameters δ summarizes the effects of the consumer's characteristics z_t on demand; *conditional* on these characteristics, the term involving λ^* indicates the rate at which changes in welfare influence changes in expenditures on particular goods. Because the β_i are simply equal to the row sums of the matrix of elasticities $\Theta = (\theta_{ij})$, in this case the Θ matrix summarizes all the pertinent information for understanding changes in demand (conditional on changes in z); we call Θ the matrix of "Frisch elasticities," and refer to the result as the Constant Frisch Elasticity (CFE) demand system.

Setting aside the possibility of error when the matrix of parameters Θ is constant, we can integrate (6) to obtain an exact expression for the *level* of demand and expenditures. In particular, let α be an n-vector of constant parameters, which arise as constants of integration from (6). Then the Frischian demand for good i is given by

$$c_i = \alpha_i \exp(\delta_i^\top \log z) \left[\lambda^{\beta_i} \prod_{j=1}^{n} p_j^{\theta_{ij}} \right]^{-1}. \tag{7}$$

One way of thinking about the richness of this demand system is to consider its *rank* (Lewbel 1991). The marginal utility λ can be regarded as a function of total expenditures x and prices p. Then the budget constraint can be written in the form

$$\sum_{i=1}^{n} a_i(p)\lambda^{-\beta_i} = x,$$

with the function $\lambda(p,x)$ the solution to this equation. Using the same notation, expenditures for good i are $x_i(p,x) = a_i(p)\lambda(p,x)^{-\beta_i}$. Expressed in matrix form, the right hand side of this equation takes the form $\mathbf{a}(p)\mathbf{g}(p,x;z)$, with $\mathbf{g}(p,x;z)$ a diagonal matrix with rank equal to the number of distinct values of β_i. Thus, the rank of a demand system with n goods may be as great as n. This compares with rank 1 for any homothetic system, or rank 2 for any PIGL system, such as the well-known AID system (Lewbel 1987). Further, one can show that the Engel curves of this demand system are a flexible functional form (Diewert 1971), with symmetry and homogeneity which can be tested or imposed by way of linear restrictions on the matrix Θ. Further, when these restrictions are satisfied the demand system is globally regular, and implies a simple parametric form for the direct utility function.

4. Estimation with (Possibly Repeated) Cross-Sectional Data

Suppose we have data on disaggregate expenditures for T cross-sections of households facing the same prices. We want to use these data to estimate the parameters of (7). However, those equations describe only the demand system for a single household. Adapting it, let $j = 1, \dots, N$ index different households, and assume that household characteristics for the jth household at time t include both observable characteristics z_t^j and time-varying unobservable characteristics ϵ_{it}^j. Then we can write our structural estimating equation as

$$\log x_{it}^j = \log \alpha_i + \left(\log p_{it} - \sum_{k=1}^{n} \theta_{ik} \log p_{kt} \right) + \beta_i \delta_i^\top \log z_t^j - \beta_i \log \lambda_t^j + \beta_i \epsilon_{it}^j. \tag{8}$$

We assume that prices are unknown to the econometrician, but that all households face the same prices.[6] Expressed in a reduced form, we write

$$y_{it}^j = a_{it} + d_i^\top (\log z_t^j - \overline{\log z}_t) + b_i w_t^j + e_{it}^j, \tag{9}$$

where

$$\begin{aligned}
y_{it}^j &= \log x_{it}^j \\
a_{it} &= \log \alpha_i + \left[\log p_{it} - \sum_{k=1}^{n} \theta_{ij} \log p_{kt} \right] - \beta_i \overline{\log \lambda}_t + \beta_i \bar{\epsilon}_{it} \\
d_i &= \delta_i \\
e_{it}^j &= \beta_i (\epsilon_{it}^j - \bar{\epsilon}_{it}) \\
b_i w_t^j &= -\beta_i (\log \lambda_t^j - \overline{\log \lambda}_t).
\end{aligned}$$

We obtain the reduced form parameters (a_{it}, d_i) simply by using least squares to estimate (9), treating the a_{it} as a set of good-time effects.

4.1. Identification of the Parameters of Interest.

Identification of the Parameters of Interest. What other parameters and unobserved quantities can we identify? We first consider the Frisch elasticities β_i and the associated IMUEs (the $\log \lambda$s). Variation in expenditures is enough to identify these up to a scale factor ϕ, and to identify the IMUEs up to a set of location parameters $\overline{\log \lambda}_t$. Observing variation over time in a

6. This can be easily extended; for example, in our application below we allow for different prices in different regions.

price index allows us to identify these location parameters. We discuss these in turn.

4.1.1. *Frisch elasticities* β_i *and* $\log\lambda$. The residuals from (9) are equal to $b_i w_t^j + e_{it}^j$. The first term of this sum is what we're interested in. Arrange the residuals as an $n \times NT$ matrix $\mathbf{Y}$. The first term in the equation captures the role that variation in marginal utility λ plays in explaining variation in expenditures. Because it's equal to the outer product of two vectors, this first term is at most of rank one. The second term captures the further role that other unobservables (e.g., unobservable household characteristics, measurement error) play in changes in demand; if there are m such unobservable factors, then this second term is of at most rank $\bar{m} = \min(m, n-1)$.

We proceed by considering the singular value decomposition (SVD)[7] of $\mathbf{Y} = \mathbf{U}\Sigma\mathbf{V}^\top$, where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices, and where Σ is a diagonal matrix of the singular values of $\mathbf{Y}$, ordered from the largest to the smallest. Then the rank one matrix that depends on λ is $\mathbf{b}\mathbf{w}^\top = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^\top$, while the second matrix (of at most rank $\bar{m}$) is $\mathbf{d}\mathbf{Z}^\top = \sum_{k=2}^{\bar{m}} \sigma_k \mathbf{u}_k \mathbf{v}_k^\top$, where σ_k denotes the kth singular value of $\mathbf{Y}$, and where the subscripts on $\mathbf{u}$ and $\mathbf{v}$ indicate the column of the corresponding matrices $\mathbf{U}$ and $\mathbf{V}$. The sum of these matrices is equal to $\mathbf{Y}$, and the truncated sum of the first $k \leq \bar{m}$ matrices is the optimal k rank approximation to $\mathbf{Y}$, in the sense that by the Eckart-Young theorem this is the solution to the problem of minimizing the Frobenius distance between $\mathbf{Y}$ and the approximation. Accordingly, this is also the least-squares solution (Golub and Reinsch 1970).

The singular value decomposition thus identifies the structural parameters β_i and changes in log marginal utility up to an unknown scalar ϕ, so that we obtain estimates of $\phi\beta_i$ and of $(\log\lambda_t^j - \overline{\log\lambda}_t)/\phi$.

4.1.2. *Identifying* $\overline{\log\lambda}_t$. Suppose between two periods we see an increase in average expenditures across all households. Our expenditure data are in nominal terms, and so far we have assumed that we don't observe prices: is the

7. Data on some expenditures is absent for some households, so our SVD must somehow contend with missing data. The algorithm we've developed for doing this is described in the appendix.

increase because the population has become wealthier (a decrease in $\overline{\log \lambda_t}$), or because the general price level has increased?

To address this question we need to describe what is meant by the 'general price level.' Our non-homothetic demand system poses a problem here; if it's a rank $m \leq n$ system, then in fact we would need m *different* price indices to construct an indirect utility or expenditure function (Lewbel 1991).

However, we don't need to construct a Marshallian indirect utility function to do welfare analysis in this Frischian setting. We can just choose any of the m price indices; any one will do. Different choices will simply imply different units of measurement for λ.

The simplest choice is simply to adopt a particular good as *numéraire*, and this simple approach is the one we take in our application below. If we choose, say, rice as numéraire, then of course this affects the units in which other quantities are measured: we would then speak of milk expenditures in terms of some amount of rice.[8] The units of $\log \lambda$ will also be determined in part by our choice of *numéraire*: ignoring household characteristics, the first order condition from the consumer's problem for a *numéraire* good implies that $\log c_1 = -\beta_1 \log \lambda$, and the aggregate value $\overline{\log \lambda_t} = -\overline{\log c_{1t}}/\beta_1$.

5. Data

To illustrate some of the methods and issues discussed above, we use data from four rounds of surveys conducted in Uganda (in 2005–06, 2009–10, 2010-11, and 2011-12).[9] We first give a descriptive account of some of the data on household characteristics and expenditures from these surveys.

5.1. **Summary Statistics.** Table 1 gives some information on household characteristics. In each of four rounds, there are about 3000 households; of these, between 70–80% are rural. There is a panel aspect to these data. There

8. It's pleasing that rice actually *has* been often used as a numéraire good, often defined in terms of a quantity of rice necessary to feed an adult for a particular period of time. Examples include the *masu* (rice for one day) or *koku* (rice for one year) in feudal Japan (Beasley 1972) which were used as standard measures of value, or the *kolaga* in parts of South India (Srinivasan 1979).

9. These datasets are available at `http://go.worldbank.org/MO5MSKCQS0`, with documentation available at `http://go.worldbank.org/S233P3YC30`.

Table 1. Characteristics of households in Uganda. Figures in parentheses are standard deviations.

Year	N	Boys	Girls	Men	Women	Rural
2005	3115	1.48	1.48	1.12	1.24	0.72
		(1.45)	(1.44)	(0.89)	(0.86)	(0.45)
2009	2927	1.70	1.67	1.21	1.33	0.74
		(1.55)	(1.50)	(0.97)	(0.89)	(0.44)
2010	2639	1.77	1.78	1.26	1.40	0.78
		(1.57)	(1.56)	(1.01)	(0.95)	(0.41)
2011	2795	1.70	1.72	1.22	1.37	0.80
		(1.53)	(1.53)	(0.97)	(0.86)	(0.40)

are a total of 3727 distinct households observed across the four rounds; of these 2151 are observed in every round.

The average household size consists of 5.8 people; the average rural household is larger, at 5.9, while the average urban household consists of 5.5 people.[10]

5.2. **Expenditure Data.** Excluding durables, taxes, fees & transfers, there are 110 categories of expenditure in the data, of which 72 are different food items or categories, and 38 are other nondurables or services. Food codes and items are fairly consistently recorded across rounds, but not perfectly so; further, some are clearly sensibly treated as substitutes (e.g., different size bunches of matoke). Other food items are treated separately in some rounds (e.g., "Watermelon" in 2010 and 2011) but assigned to an aggregate (e.g., "Other Fruits") in other rounds, necessitating the use of the coarser aggregate to achieve consistency across rounds. Appendix Table B.1 gives a precise accounting of all codes and aggregation.

The aggregation in Table B.1 results in total of 49 different items. Most of these are straightforward types of food, such as peas, mangoes, ground nuts, maize, or sugar. Food consumed in restaurants is a category of its own, however, and may be thought of an aggregate bundle of food and services. Alcoholic beverages account for two additional categories, "beer" and "other

10. For our purposes a person is a household member if they've lived in the household for at least one month of the previous twelve. People identified as 'guests' who satisfy these criteria must also have spent the night prior to the interview.

alcoholic drinks." And then finally there are two non-food categories included, "cigarettes" and "other tobacco." Altogether there are five categories which are explicitly undifferentiated aggregates: "other fruits", "other vegetables", "other alchoholic drinks", "other drinks", and "other tobacco." Other categories may be implicitly aggregated: for example, "ground nut" includes nuts shelled, unshelled, and made into paste. Finally, even after aggregation some of these categories contain very few positive observations in at least some years; dropping these yields a total of 41 categories.[11]

Table 2 paints a picture of aggregate expenditure shares across these categories, listing mean and aggregate expenditure shares for all foods, ordered by the size of their aggregate expenditure share in 2005. A glance reveals that shares of aggregates is fairly stable across the period 2005–2011, with only a handful of goods changing their aggregate shares by as much as one percentage point (the only exceptions are cassava, sugar, and "other foods."). It should be noted, however, that stability of shares over time is not a prediction of theory, as it would be in a homothetic demand system—changes in incomes or relative prices can be expected to cause changes in shares.

Table 2: Aggregate food expenditure shares and household mean shares in 2005 and 2011.

Goods	Aggregate 2005	Aggregate 2011	Mean 2005	Mean 2011
Matoke	0.113	0.114	0.097	0.099
Maize	0.079	0.084	0.092	0.091
Cassava	0.077	0.092	0.090	0.105
Restaurant	0.074	0.073	0.061	0.070
Sweet potato	0.070	0.063	0.078	0.075
Beans	0.066	0.074	0.080	0.090
Beef	0.058	0.058	0.044	0.046
Sugar	0.055	0.044	0.051	0.042

Continued on next page

11. Excluded goods include beer, infant formula, butter & margarine, ghee, ground nuts, pork, other juice, other drinks, and other meat.

Continued from previous page

Goods	2005	2011	2005	2011
Fresh milk	0.041	0.040	0.033	0.034
Rice	0.027	0.027	0.022	0.021
Fresh fish	0.024	0.022	0.022	0.020
Ground nut	0.022	0.024	0.023	0.024
Cooking oil	0.021	0.022	0.022	0.022
Dried fish	0.020	0.022	0.022	0.021
Chicken	0.019	0.026	0.014	0.019
Tomatoes	0.019	0.017	0.020	0.017
Other Alcohol	0.018	0.016	0.021	0.020
Soda	0.016	0.012	0.012	0.008
Bread	0.015	0.015	0.010	0.010
Other foods	0.015	0.003	0.021	0.003
Millet	0.015	0.014	0.016	0.015
Other Fruit	0.012	0.019	0.010	0.017
Goat meat	0.012	0.015	0.010	0.010
Irish potato	0.012	0.012	0.010	0.011
Other Veg.	0.011	0.016	0.018	0.020
Sorghum	0.009	0.012	0.017	0.018
Cigarettes	0.009	0.004	0.009	0.004
Dodo	0.008	0.007	0.010	0.010
Onions	0.008	0.010	0.009	0.011
Passion fruit	0.007	0.003	0.005	0.002
Mangos	0.006	0.004	0.007	0.004
Sweet Banana	0.006	0.005	0.004	0.003
Salt	0.006	0.005	0.008	0.007
Eggs	0.006	0.005	0.004	0.004
Cabbages	0.005	0.006	0.005	0.007
Tea	0.005	0.003	0.005	0.004
Peas	0.005	0.005	0.006	0.006
Sim sim	0.004	0.004	0.007	0.006

Continued on next page

Continued from previous page

Goods	2005	2011	2005	2011
Oranges	0.002	0.002	0.002	0.002
Other Tobacco	0.002	0.001	0.004	0.002
Coffee	0.001	0.001	0.001	0.000

What one *can* say about these data on shares is that they do not seem consistent with a model in which consumers have homothetic utility. Such a model would predict equal aggregate and mean expenditure shares, and for many goods there is a larger difference between these than there is in shares across six years.

This general point is graphically borne out in Figure 1. For this figure we construct a statistic ρ_{it} which is the logarithm of aggregate shares minus the logarithm of mean shares, or, for good i at time t,

$$\rho_{it} = \log\left(\frac{\sum_{j=1}^N x_{it}^j}{\sum_{j=1}^N \sum_{k=1}^n x_{kt}^j}\right) - \log\left(\sum_{j=1}^N \frac{x_{it}^j}{\sum_{k=1}^n x_{kt}^j}\right).$$

We then produce a scatterplot of this statistic, ordered by the size of the statistic in 2005. Thus, each good (labelled on the left axis) has associated with it a statistic for each of four years, each with (overlapping) confidence intervals.

With homothetic preferences, this statistic must always be equal to zero, but we can reject this equality for most of the 41 goods in the figure. Instead, a positive value of the statistic identifies goods which play an outsized role in the consumption portfolios of wealthier (i.e., higher expenditure) households, and include passion fruit, bread, chicken, soda, and sweet bananas, among others. Conversely, when the statistic is negative we identify goods that are particularly important in the portfolios of households with lower food expenditure. Here we see “other tobacco”, sorghum, “other vegetables”, sim sim (sesame) and salt.

The figure also argues against the usual *quasi*-homothetic specification of preferences, interpreted as though some positive “subsistence” level of the good is necessary for survival. Subsistence requirements of this case could account

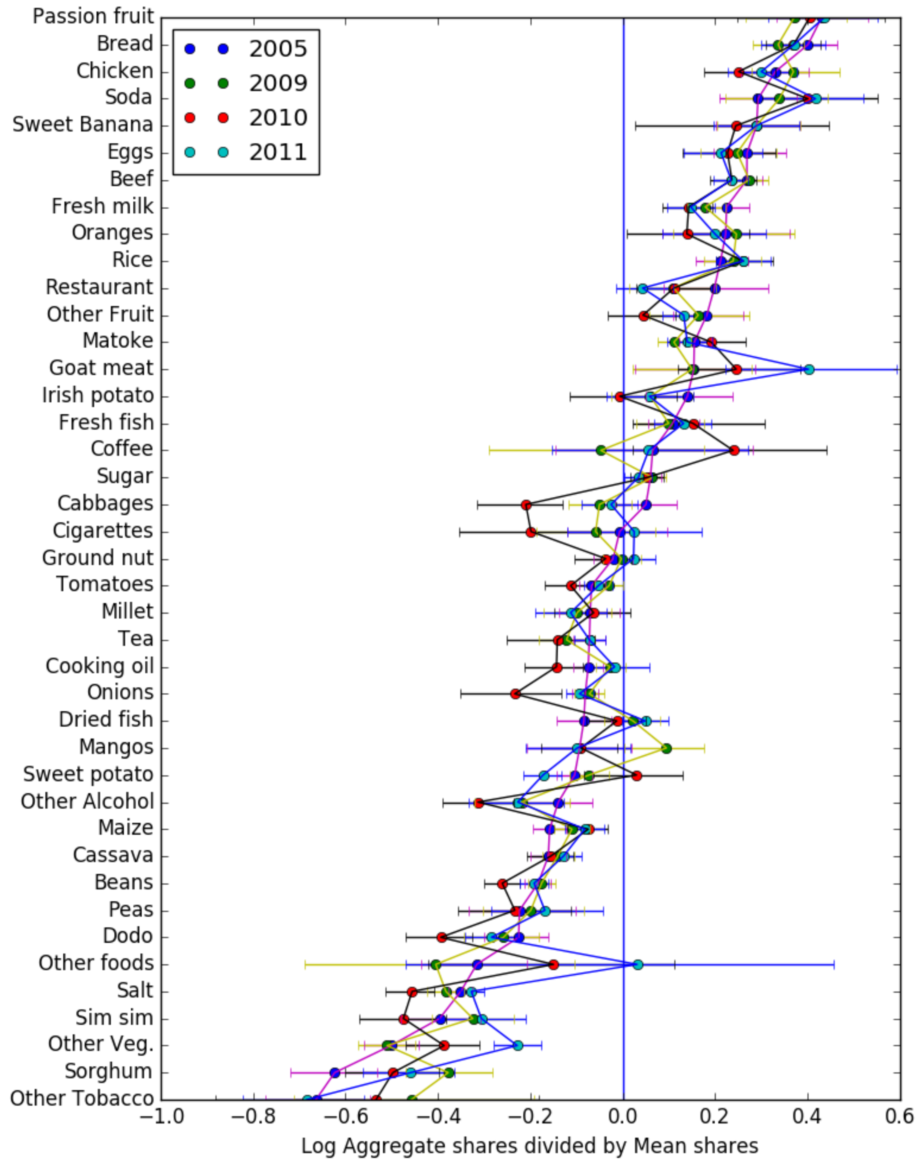

FIGURE 1. Log of mean shares minus log of aggregate shares for different years (ordered by ranking in 2005), with 95% confidence intervals.

for the goods for which negative statistics are observed, such as salt, cassava, or maize, but as total expenditures increase, quasi-homothetic utility implies that budget shares should converge to a fixed constant. This implies that in Figure 1 the plotted statistics should converge to zero as one moves from bottom to top. This predicted pattern is not at all evident.

6. Results

6.1. **Estimates of Demand Elasticities.** We now turn to estimates of some of the parameters of the the demand system (8), estimated using the four rounds of data from Uganda discussed above. Table 3 presents results from our baseline specification. In this specification we obtain results for a system of 41 minimally aggregated consumption goods, assuming that all households face the same relative prices. We take as observable characteristics the number of men, women, boys and girls in each household, as well as the logarithm of total household size. We also include a dummy indicating a rural or urban location for the household. One interpretation of this dummy is that allows for differences in the general price level and average $\log \lambda$ between rural and urban areas, but it also can accommodate differences in the marginal utility functions across rural and urban areas (i.e., heterogeneity in the α_i parameters).

Where recorded values of consumption expenditure are equal to zero, we regard these as missing and dropped from the analysis. There are two reasons for this treatment of zeros: first, at an entirely practical level, our dependent variable is the logarithm of expenditures, which is undefined at zero. But second, if a household is at a corner when it chooses a particular consumption item, then the first order condition in (3) for that consumption good won't be correct (we'd be missing a multiplier related to non-negativity). By simply dropping observations for goods where consumption is zero we are effectively dropping observations where expenditures do not correctly reveal the index $\log \lambda$.

Table 3: Estimates of expenditure system assuming a single market. Controls include the numbers of boys, girls, men, and women in household, along with the log of household size. Figures in parentheses are estimated standard errors.

	$\phi\beta_i$	$\log \alpha_i$	Rural	Boys	Girls	Men	Women	log Hsize
Passion fruit	0.65***	−1.18***	−0.37***	−0.01	−0.01	0.03	0.09***	0.23**
	(0.05)	(0.03)	(0.06)	(0.03)	(0.03)	(0.03)	(0.03)	(0.10)
Other Fruit	0.61***	−1.43***	−0.11***	0.03*	0.01	0.10***	0.05**	0.28***
	(0.04)	(0.03)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.07)
Oranges	0.59***	−1.75***	−0.18***	0.09***	0.00	0.11***	0.16***	0.02
	(0.06)	(0.05)	(0.06)	(0.03)	(0.03)	(0.04)	(0.04)	(0.13)
Bread	0.59***	−0.94***	−0.45***	0.01	0.04*	0.11***	0.09***	0.17**
	(0.04)	(0.02)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.08)
Coffee	0.56***	−2.79***	−0.22***	0.06*	0.02	0.16***	0.16***	−0.14
	(0.08)	(0.06)	(0.08)	(0.03)	(0.03)	(0.04)	(0.04)	(0.13)
Other Tobacco	0.56***	−1.90***	−0.16	0.04	−0.01	0.05	0.09*	−0.11
	(0.13)	(0.04)	(0.12)	(0.03)	(0.03)	(0.04)	(0.05)	(0.12)
Other foods	0.56***	−0.48***	−0.34***	0.00	0.08	0.02	0.13*	−0.14
	(0.13)	(0.05)	(0.12)	(0.07)	(0.06)	(0.07)	(0.07)	(0.23)
Sweet Banana	0.53***	−1.57***	−0.28***	−0.00	0.02	0.12***	0.06**	0.28***
	(0.05)	(0.03)	(0.05)	(0.02)	(0.03)	(0.03)	(0.03)	(0.10)
Soda	0.52***	−0.23***	−0.32***	−0.00	0.03	0.08***	0.06**	0.03
	(0.05)	(0.03)	(0.04)	(0.02)	(0.02)	(0.03)	(0.03)	(0.08)
Fresh milk	0.51***	−0.50***	−0.31***	0.00	0.00	0.15***	0.05***	0.25***
	(0.03)	(0.02)	(0.03)	(0.01)	(0.01)	(0.02)	(0.02)	(0.06)
Cigarettes	0.48***	−0.40***	−0.34***	0.03	0.06*	0.20***	0.15***	−0.20
	(0.09)	(0.05)	(0.08)	(0.04)	(0.04)	(0.05)	(0.05)	(0.14)
Matoke	0.47***	−0.14***	−0.04	0.01	0.02*	0.08***	0.08***	0.45***
	(0.03)	(0.02)	(0.03)	(0.01)	(0.01)	(0.01)	(0.02)	(0.05)

Continued on next page

Continued from previous page

	$\phi\beta_i$	$\log\alpha_i$	Rural	Boys	Girls	Men	Women	log Hsize
Rice	0.45***	−0.71***	−0.18***	0.04***	0.03**	0.08***	0.05***	0.33***
	(0.03)	(0.02)	(0.03)	(0.01)	(0.01)	(0.02)	(0.02)	(0.06)
Eggs	0.43***	−1.20***	−0.29***	−0.01	−0.02	0.09***	0.04	0.22***
	(0.03)	(0.03)	(0.04)	(0.02)	(0.02)	(0.02)	(0.03)	(0.08)
Cooking oil	0.43***	−1.25***	−0.44***	−0.01	0.01	0.09***	0.03**	0.21***
	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.04)
Mangos	0.43***	−1.58***	0.10*	−0.00	−0.03	0.07**	−0.04	0.52***
	(0.06)	(0.05)	(0.06)	(0.03)	(0.03)	(0.03)	(0.04)	(0.13)
Sugar	0.43***	−0.61***	−0.38***	0.02**	0.03***	0.09***	0.08***	0.27***
	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.04)
Tomatoes	0.42***	−1.38***	−0.42***	−0.00	0.02**	0.08***	0.06***	0.13***
	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.04)
Other Alcohol	0.40***	−0.20***	−0.34***	0.03	0.00	0.15***	0.03	0.02
	(0.06)	(0.04)	(0.07)	(0.02)	(0.02)	(0.03)	(0.03)	(0.10)
Ground nut	0.40***	−1.19***	−0.01	0.05***	0.05***	0.10***	0.12***	0.06
	(0.03)	(0.02)	(0.03)	(0.01)	(0.01)	(0.01)	(0.02)	(0.05)
Goat meat	0.39***	−0.10**	−0.18***	0.02	0.02	0.09**	0.07*	0.28**
	(0.05)	(0.04)	(0.06)	(0.03)	(0.03)	(0.03)	(0.04)	(0.12)
Other Veg.	0.39***	−1.76***	−0.06*	0.04**	0.02	0.06***	0.06***	0.17***
	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.07)
Onions	0.39***	−2.19***	−0.40***	−0.01	−0.00	0.08***	0.07***	0.14***
	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)
Restaurant	0.39***	0.92***	−0.62***	0.01	−0.01	0.13***	0.06**	−0.04
	(0.05)	(0.04)	(0.05)	(0.02)	(0.02)	(0.03)	(0.03)	(0.08)
Irish potato	0.38***	−0.71***	0.20***	0.09***	0.05**	0.08***	0.05*	0.13
	(0.05)	(0.04)	(0.05)	(0.02)	(0.03)	(0.03)	(0.03)	(0.09)
Fresh fish	0.38***	−0.31***	−0.11***	0.05**	0.02	0.12***	0.06***	0.12
	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.07)
Beef	0.38***	0.16***	−0.18***	0.03***	0.02	0.11***	0.07***	0.19***

Continued on next page

Continued from previous page

	$\phi\beta_i$	$\log\alpha_i$	Rural	Boys	Girls	Men	Women	log Hsize
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.05)
Peas	0.38***	−1.11***	−0.20***	−0.00	−0.02	0.03	0.03	0.35**
	(0.06)	(0.04)	(0.06)	(0.03)	(0.03)	(0.03)	(0.04)	(0.14)
Dried fish	0.34***	−0.83***	−0.17***	0.03*	0.02	0.12***	0.05**	0.18**
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.07)
Dodo	0.34***	−1.68***	−0.11***	0.06***	0.04***	0.05***	0.07***	0.08
	(0.03)	(0.02)	(0.03)	(0.01)	(0.01)	(0.02)	(0.02)	(0.06)
Maize	0.33***	−0.78***	0.08***	0.07***	0.08***	0.09***	0.04**	0.28***
	(0.03)	(0.02)	(0.03)	(0.01)	(0.01)	(0.01)	(0.02)	(0.05)
Cabbages	0.33***	−1.44***	−0.14***	0.00	0.01	0.04***	0.03**	0.22***
	(0.02)	(0.02)	(0.03)	(0.01)	(0.01)	(0.02)	(0.02)	(0.06)
Chicken	0.30***	0.77***	−0.23***	−0.01	−0.02	0.07***	0.04	0.20**
	(0.04)	(0.03)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.09)
Millet	0.30***	−1.21***	0.14***	0.03	0.00	0.05*	0.05*	0.40***
	(0.05)	(0.03)	(0.05)	(0.02)	(0.02)	(0.02)	(0.03)	(0.09)
Tea	0.28***	−2.71***	−0.19***	0.02**	0.03***	0.13***	0.09***	0.09**
	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.04)
Beans	0.28***	−0.81***	−0.01	0.03***	0.04***	0.08***	0.04***	0.32***
	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.04)
Sim sim	0.27***	−1.22***	0.02	0.01	−0.01	0.09***	0.06**	0.13
	(0.04)	(0.04)	(0.06)	(0.02)	(0.02)	(0.03)	(0.03)	(0.11)
Sweet potato	0.26***	−0.83***	0.31***	0.07***	0.06***	0.08***	0.05***	0.35***
	(0.03)	(0.02)	(0.03)	(0.01)	(0.01)	(0.01)	(0.02)	(0.05)
Sorghum	0.24***	−1.33***	0.03	−0.03	−0.07***	−0.07**	−0.09**	0.77***
	(0.05)	(0.04)	(0.07)	(0.03)	(0.03)	(0.03)	(0.04)	(0.12)
Salt	0.14***	−3.02***	0.04***	0.03***	0.02***	0.05***	0.03***	0.24***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)
Cassava	0.14***	−0.68***	0.27***	0.05***	0.02	0.05***	−0.00	0.43***
	(0.02)	(0.02)	(0.03)	(0.01)	(0.01)	(0.01)	(0.02)	(0.05)

In its first column Table 3 presents estimates of the Frisch elasticities, ordered in descending order, and identified only up to an unknown parameter ϕ.[12] However, since ϕ is constant, ratios of these estimated parameters can be interpreted as ratios of elasticities. Thus, the most elastic good (with respect to λ) is "passion fruit" followed by oranges and "other fruit", with elasticities roughly twice that of millet, or four times that of salt. Estimated demand parameters are only reported for those goods where enough data is non-missing to reliably estimate the covariance matrix of the elasticities. All estimated elasticities (including those for unreported goods) are positive; thus, there is no evidence that any of these goods is inferior, with demand increasing as $\log \lambda$ decreases. Standard errors for these elasticities are obtained via a block bootstrap.[13]

The second column of Table 3 gives estimates of $\log \alpha_i$, where α_i is a multiplicative preference parameter. With homothetic utility, i.e., $\beta_i = \beta$, α_i would be equal to n times the expenditure share of good i, and elasticities would be constant across all goods. In our non-homothetic case expenditure shares depend on the parameters α_i, elasticities β_i, and prices. The parameters $\log \alpha_i$ vary positively with expenditure share, and are set equal to mean log expenditures in our first round of data, 2005. Goods with positive and significant values of $\log \alpha_i$ are beef, chicken, and food in restaurants, while the good with the smallest significant value is salt. Estimated standard errors for these parameters are simply equal to the standard deviation of residuals in 2005 divided by the square root of the number of observed positive expenditures in that year.

The third column of the table reports estimates of the effect of being a rural rather than an urban household. Associated standard errors are clustered by round, as are the standard errors associated with other household characteristics (Arellano 1987). The effect of being 'rural' is negative and significant for all but a few goods, consistent with the fact that total food expenditures

12. Here ϕ is determined by a normalization that makes the standard deviation of the estimated $\log \lambda/\phi$ in the first round equal to one.

13. We've also computed standard errors by calculating the inter-quartile range of the bootstrapped estimates, and scaling these up under the hypothesis of normality to provide an estimate of standard errors which is more robust to outliers; both estimators deliver very similar results.

are roughly 12% less than in urban areas. A handful of exceptions stand out: mangos, ground nuts, sorghum, millet, maize, Irish and sweet potatoes, sim sim, cassava, and salt expenditures are all (though not all significantly) greater in rural areas, other things equal.

The next several columns report indicate how expenditures vary with household size and composition. Here we've included the log of household size, but also a count of the number of boys, girls (both under the age of 18), women, and men in the household. This allows for variation in expenditures to respond to household composition, but in a way which also allows for varying returns to scale. Interpretation is slightly complicated, but let m_k denote the number of boys, girls, women and men in a particular household, with k indexing these groups; thus, m_{boys} is the number of boys, while total household size is $m = \sum_k m_k$. Then the elasticity associated with a small increase in the number of boys in the household is given by $\delta_{\text{boys}} m_{\text{boys}} + \delta_{\text{Hsize}} m_{\text{boys}}/m$.

Table 4: Estimated household expenditure elasticities associated with adding additional household members, evaluated at the mean. The final column reports the percentage of observations which are missing or zero. Standard errors reported in parentheses.

	Boys	Girls	Men	Women
Passion fruit	0.04	0.05	0.10***	0.17***
	(0.03)	(0.03)	(0.03)	(0.03)
Other Fruit	0.12***	0.09***	0.19***	0.13***
	(0.02)	(0.02)	(0.02)	(0.02)
Oranges	0.15***	0.01	0.14***	0.21***
	(0.04)	(0.04)	(0.04)	(0.04)
Bread	0.06***	0.10***	0.18***	0.16***
	(0.02)	(0.02)	(0.02)	(0.02)
Coffee	0.06	−0.01	0.16***	0.18***
	(0.04)	(0.04)	(0.05)	(0.04)

Continued on next page

Continued from previous page

	Boys	Girls	Men	Women
Other Tobacco	0.04	−0.05	0.04	0.09**
	(0.04)	(0.04)	(0.04)	(0.04)
Other foods	−0.03	0.10	−0.01	0.13**
	(0.08)	(0.07)	(0.08)	(0.07)
Sweet Banana	0.06**	0.10***	0.21***	0.16***
	(0.03)	(0.03)	(0.03)	(0.03)
Soda	0.00	0.07**	0.10***	0.09***
	(0.03)	(0.03)	(0.03)	(0.03)
Fresh milk	0.07***	0.07***	0.24***	0.14***
	(0.02)	(0.02)	(0.02)	(0.02)
Cigarettes	−0.00	0.06	0.19***	0.15***
	(0.05)	(0.04)	(0.05)	(0.05)
Matoke	0.13***	0.15***	0.21***	0.22***
	(0.02)	(0.02)	(0.02)	(0.02)
Rice	0.14***	0.14***	0.17***	0.15***
	(0.02)	(0.02)	(0.02)	(0.02)
Eggs	0.04	0.02	0.16***	0.12***
	(0.03)	(0.03)	(0.03)	(0.03)
Cooking oil	0.03***	0.07***	0.16***	0.10***
	(0.01)	(0.01)	(0.01)	(0.01)
Mangos	0.12***	0.08**	0.22***	0.08**
	(0.04)	(0.03)	(0.04)	(0.03)
Sugar	0.10***	0.12***	0.17***	0.18***
	(0.01)	(0.01)	(0.01)	(0.01)
Tomatoes	0.02**	0.06***	0.13***	0.11***
	(0.01)	(0.01)	(0.01)	(0.01)
Other Alcohol	0.05	0.01	0.19***	0.05
	(0.03)	(0.03)	(0.03)	(0.03)
Ground nut	0.10***	0.09***	0.13***	0.17***

Continued on next page

Continued from previous page

	Boys	Girls	Men	Women
	(0.01)	(0.02)	(0.02)	(0.02)
Goat meat	0.10**	0.10***	0.17***	0.16***
	(0.04)	(0.03)	(0.04)	(0.04)
Other Veg.	0.10***	0.08***	0.11***	0.13***
	(0.02)	(0.02)	(0.02)	(0.02)
Onions	0.01	0.03***	0.13***	0.13***
	(0.01)	(0.01)	(0.01)	(0.01)
Restaurant	0.01	−0.02	0.15***	0.07**
	(0.03)	(0.03)	(0.03)	(0.03)
Irish potato	0.18***	0.12***	0.12***	0.11***
	(0.03)	(0.03)	(0.03)	(0.03)
Fresh fish	0.10***	0.07***	0.18***	0.11***
	(0.02)	(0.02)	(0.02)	(0.02)
Beef	0.10***	0.08***	0.18***	0.14***
	(0.01)	(0.01)	(0.01)	(0.01)
Peas	0.08**	0.05	0.12***	0.13***
	(0.04)	(0.04)	(0.04)	(0.04)
Dried fish	0.09***	0.08***	0.19***	0.11***
	(0.02)	(0.02)	(0.02)	(0.02)
Dodo	0.12***	0.08***	0.08***	0.12***
	(0.02)	(0.02)	(0.02)	(0.02)
Maize	0.18***	0.20***	0.17***	0.12***
	(0.01)	(0.01)	(0.02)	(0.02)
Cabbages	0.06***	0.08***	0.11***	0.10***
	(0.02)	(0.02)	(0.02)	(0.02)
Chicken	0.03	0.02	0.14***	0.10***
	(0.02)	(0.02)	(0.03)	(0.02)
Millet	0.15***	0.10***	0.15***	0.17***
	(0.03)	(0.03)	(0.03)	(0.03)

Continued on next page

Continued from previous page

	Boys	Girls	Men	Women
Tea	0.05***	0.07***	0.17***	0.15***
	(0.01)	(0.01)	(0.01)	(0.01)
Beans	0.13***	0.14***	0.17***	0.13***
	(0.01)	(0.01)	(0.01)	(0.01)
Sim sim	0.05*	0.02	0.14***	0.12***
	(0.03)	(0.03)	(0.03)	(0.03)
Sweet potato	0.21***	0.18***	0.18***	0.15***
	(0.02)	(0.01)	(0.02)	(0.02)
Sorghum	0.14***	0.07**	0.10***	0.08**
	(0.03)	(0.03)	(0.04)	(0.04)
Salt	0.11***	0.10***	0.12***	0.10***
	(0.01)	(0.01)	(0.01)	(0.01)
Cassava	0.19***	0.14***	0.16***	0.11***
	(0.01)	(0.01)	(0.02)	(0.02)

Table 4 reports these estimated elasticities, evaluated at the "average" household composition (e.g., for a household composed of the mean number of boys, 1.66, having total size equal to the average of 5.84). For most goods the addition of an adult has a larger effect on household expenditures than does the addition of a child: if we take a simple average of elasticities across goods we obtain 0.09 for boys, 0.08 for girls, 0.15 for men, and 0.13 for women. We can further identify particular "adult goods" where the difference in elasticities between adults and children are greatest, such as coffee, cigarettes, fresh milk and food consumed in restaurants. But adult-child differences are smaller for staples such as millet, rice, and beans, and are even reversed for starchy staples such as maize, cassava, and both sweet and Irish potatoes. There are also a handful of goods which seem to be differentially preferred by females: goods for which point estimates of elasticities are greater for women than for men, and for girls than for boys, are passion fruit, matoke, and sugar.

The fit of the estimated demand equations varies only moderately, with the R^2 for bread 0.49, while other goods such as passion fruit, sugar, rice, beef, tomatoes, matoke and oranges also have R^2 statistics exceeding 0.40. At the other end of the scale, sim sim (sesame), sorghum, and cassava all have an R^2 statistic of 0.17.

6.2. **Estimates of** $\log \lambda$**.** Figure 2 presents histograms of the estimated $\log \lambda$ for each round of data. The scale of these is identified by the normalization that the distribution in 2005 should have a standard deviation of one. The location of each distribution depends on changes in average consumption of the numéraire good (rice); under the assumption that this good is separable in the utility function it follows that demand (conditional on characteristics) can vary only with λ.

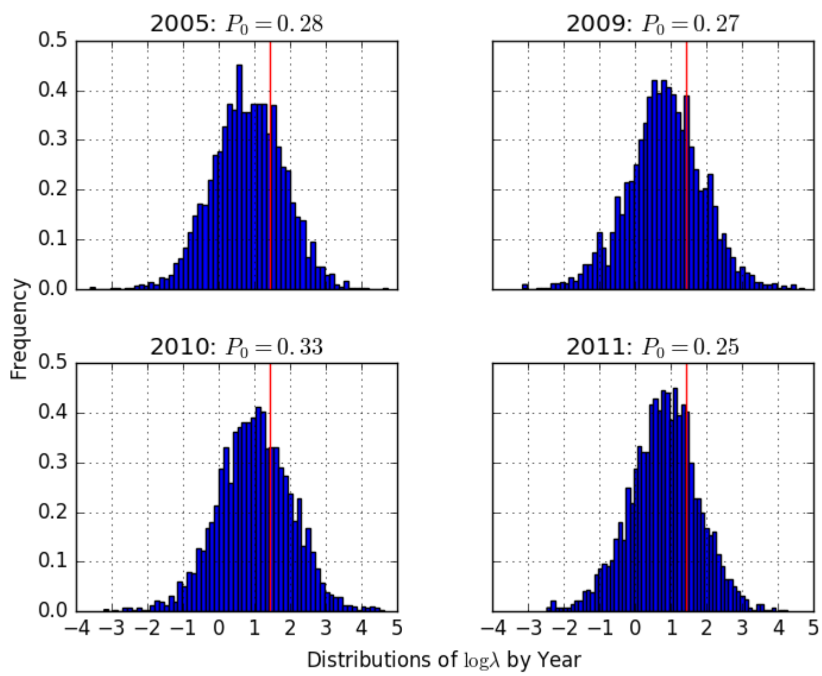

FIGURE 2. Distribution of $\log \lambda$ by Year.

So what can we say about changes in welfare in Uganda over this period? First, the location of these distributions is almost equal in 2005, and 2011, at roughly 0.81. The years in between followed on the food price crisis of

2008, and include what Brunori, Palmisano, and Peragine (2015) call Uganda's "great recession" year of 2010, when average $\log\lambda$ increased 23% to 1.00. The variance of the distribution tended to increase then fall slightly over time, with the standard deviation falling from 1.0 to 0.97 at the end of the period.

6.2.1. *Relation of* $\log\lambda$ *to poverty measures.* Though this is a topic to be explored in greater depth elsewhere, it's instructive to compare how these kinds of welfare comparisons match up with conventional measures. One easy comparison involves the construction of headcount poverty measures (the P_0 measure of Foster, Greer, and Thorbecke 1984). Uganda seems to have chosen a poverty line which is meant to equate to a level of per capita expenditures equal to $1.25 in PPP-adjusted terms, a level and method of adjustment recommended by Ravallion, Chen, and Sangraula (2009). The World Bank's online PovCalNet uses the same underlying datasets for calculating welfare statistics as in this paper, and recommends a PPP-adjustment of 946.89. Using this adjustment and the the $1.25 poverty line, the World Bank's figure for headcount poverty in 2005 is 28%.[14] We use this proportion to pin down a poverty line expressed not in terms of total real expenditures, but instead in terms of $\log\lambda$—in the 2005 distribution of these statistics pictured in Figure 2, this is 1.44 (the vertical line in the 2005 panel of the Figure).

Now, in the usual expenditure-based approach to poverty measurement one would need to update the poverty line as the price level changed; the World Bank does this by using a country-specific consumer price index (CPI). These are Laspeyres indices, which assume that consumers have the same budget shares regardless of wealth (i.e., that utilities are homothetic). We've already seen evidence that this isn't true in Uganda, so the procedure of recomputing the poverty line by using the CPI is suspect. However, the 'poverty line' value of $\log\lambda$ doesn't depend on prices—changes in nominal prices in the Frischian approach are automatically accounted for by the good-time effects a_{it} in (9), so once chosen, the $\log\lambda$ poverty line will never change.

The results of fixing the $\log\lambda$ poverty line at 1.44 (to match the 2005 headcount poverty reported by PovCalNet) can be seen in Figure 2; from the initial headcount poverty of 28% in 2005 it decreases slightly to 27% in 2009, increases

14. See `http://iresearch.worldbank.org/PovcalNet`; downloaded July 2017.

to 33% in the "great recession" year of 2010, and then falls to 25% in the recovery of 2011–12. This all seems sensible. However, these changes are very sharply at odds with the PovCalNet calculations, based on measures of total expenditures. The World Bank reports that the headcount poverty rate of 28% in 2005 falls dramatically to 17% in 2009, and falls further to 13% in 2011 (PovCalNet fails to report 2010 recession year values, for reasons unknown). The fall of 4 percentage points between 2009 and 2011 isn't shockingly different from the fall of 2 percentage points we calculate using our method, but the magnitude of the change from 2005 to 2009 is quite different across the two methods. Which method (if either) is correct?

As noted above, the food price crisis lead to sharply increased prices for staple foods world-wide. Benson, Mugarura, and Wanda (2008) confirms that these increases are also observed in Uganda, consistent with the data we have on prices paid by households in 2005 and 2009. These large nominal price increases are considerably larger than the change in the CPI used in the PovCalNet calculation, so we can confidently conclude that relative prices for most kinds of food increased during this period. Further, the large increase in relative food prices corresponds to sizeable decreases in the *quantities* of food consumed in our Ugandan data—on average across food items we observe a 3% decrease in quantities consumed between 2005 and 2009.

Engel's Law makes it very difficult to reconcile the decline in the quantity of food consumed over 2005–09 with a claim that poverty fell dramatically over the same period, and so we are skeptical of the PovCalNet figures. The most likely cause of problems is that the CPI is inappropriate for households at (or below) the poverty line, and places too little weight on food prices.[15]

6.3. **Validation: Estimated Aggregate Shares versus Mean Shares.** In Figure 1 we used data on observed expeditures to produce a plot of a statistic equal to the logarithm of mean shares minus the logarithm of aggregate shares, ordered by the size of the statistic in 2005, and observed that the pattern of observed in that figure could not be generated by any demand system featuring

15. We're not alone in our skepticism; Duponchel, McKay, and Ssewanyana (2014) also conduct an expenditure-based poverty analysis in Uganda using the same datasets, but allowing for the CPI to vary with regional prices, and obtain results closer to ours than to the PovCalNet figures.

homothetic preferences, and also did not give evidence of being generated by a quasi-homothetic preference structure.

The question naturally arises: is the non-quasi-homothetic demand system we've estimated here capable of delivering the pattern of expenditure shares pictured in Figure 1?

This is a challenging test, because although we've used the observed data to estimate the demand system, our estimation procedure is designed to fit conditional expectations of log expenditures to the data, while the shares statistics we've constructed is built using logs of means and sums of expenditures. Jensen's inequality alone tells us that our ability to match the share statistics will depend not only on the estimated equation, but also on the distribution of residuals.

Let $h_i(p, \lambda, z) = \mathrm{E}(\log x_i | p, \lambda, z)$. Assume that the residuals e^j_{it} in the estimating equation (9) are independent and identically normally distributed for each good, with mean zero and variance σ^2_i. Then a simple estimator of $\mathrm{E}(x^j_{it}|p_t, \lambda^j_t, z^j_t)$ is $\exp(h_i(\hat{p}_t, \hat{\lambda}^j_t, z^j_t) + \hat{\sigma}^2_i/2)$, where $\hat{\sigma}^2_i$ is the maximum likelihood estimate of the variance of the residuals for good i, and where $\hat{p}_t$ and $\hat{\lambda}^j_t$ are estimates of price indices and $\log \lambda$ as described in Section 4.1.

We next simply substitute our estimates $\hat{x}^j_{it}$ into the expression defining the statistics ρ_i, and plot the values of these statistics predicted by our model of demand and estimates of prices and $\log \lambda$. The result is picture in the left hand panel of Figure 3.

The left panel of Figure 3 reproduces Figure 1, except with predicted rather than actual shares. The general pattern evidencing non-quasi-homotheticity is readily apparent. But beyond this, the ρ_i statistics calculated using our predicted expenditures have a Spearman correlation coefficient of 0.97 with statistics calculated using the observed data. The right-hand panel of Figure 3 provides a scatter plot of observed vs. predicted values of the statistic, along with a 45 degree line. The scatterplot confirms the success of our demand system at reproducing even patterns in the data that our estimator wasn't designed to fit.

6.4. **Measuring Heterogeneity in Risk Attitudes.** When the household-specific λs we've calculated are equal to the marginal utility of expenditures,

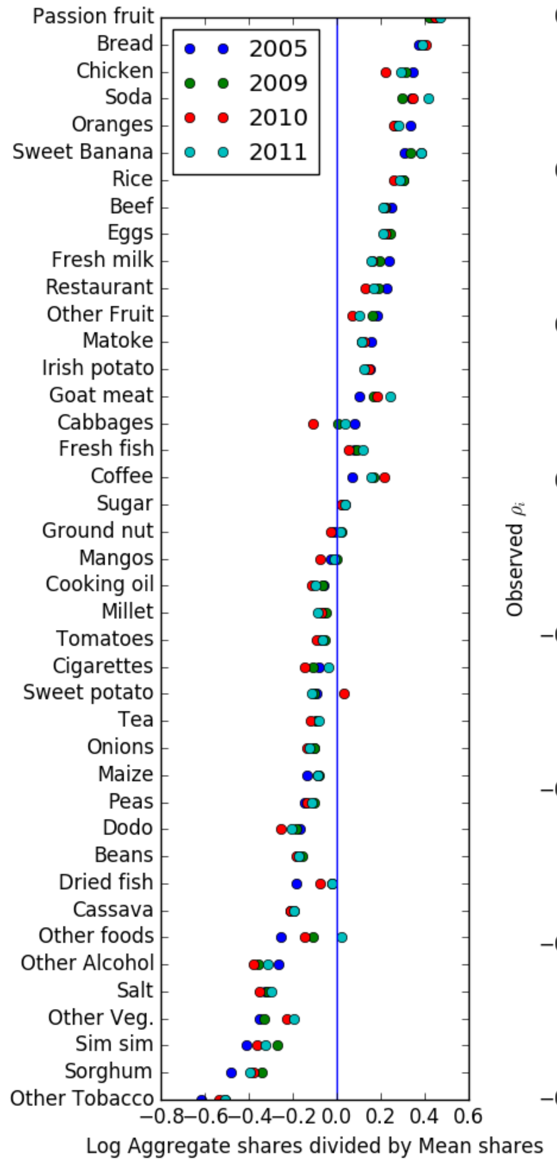

FIGURE 3. Left panel: Predicted log of mean shares minus log of aggregate shares for different years (ordered by ranking in 2005). Right panel: Predicted versus actual, with 45 degree line.

then $\omega = \partial \log \lambda / \partial \log x$ is what Frisch called the household's "money flexibility"; its negative is the household's relative risk aversion; and its negative reciprocal is the intertemporal elasticity of substitution. Further, with data on x we can calculate the quantity ω for each household.

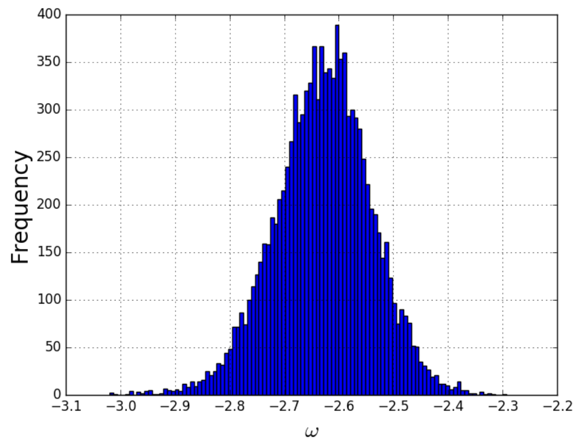

FIGURE 4. Distribution of estimated ω in 2005. Standard deviation of pooled estimates is 0.09; kurtosis is 0.43.

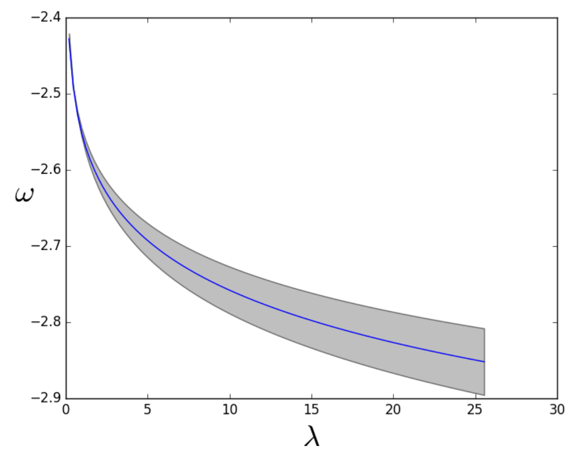

FIGURE 5. Calculated ω versus λ. Varying all estimates of β_i by plus or minus two standard errors yields the shaded area.

Results of calculating households' values of ω in 2005 are shown in the histogram of Figure 4. The mean of this distribution is -2.63, with a standard deviation of 0.09. Households are assumed to have identical β_i parameters and to face identical prices, so differences in ω (and hence in risk aversion) are driven entirely by differences in $\log \lambda$. The estimated values of ω in Figure 4 are well within the range of plausible relative risk aversions, though the range is smaller than one might suppose based on the evidence of Chiappori et al. (2014).

The relationship between ω and $\log \lambda$ is illustrated in Figure 5, where λ is on the horizontal axis. The central plotted line offers calculations of relative risk aversion given the point estimates of β_i elasticities given in Table 3 (taking $\phi = 1$); the shaded region allows each point estimate to vary by plus or minus one standard error, giving some sense of how sensitive our estimated risk aversions are to imprecisely estimated Frisch elasticities. From the figure, one can see that ω is decreasing in λ, and so increasing in total expenditures. Since ω is negative, and has the interpretation of the elasticity of λ with respect to total expenditures, the figure illustrates the point that the utility of wealthier households is less sensitive to variation in total expenditures than is the utility of poorer households; translated into statements about risk aversion, Figure 5 indicates that households have decreasing relative risk aversion with respect to expenditures on food.

Knowing the quantity ω is what we need for estimating within-period demands or indifference curves. However, if we're interested in measuring risk attitudes or more generally the curvature of the momentary utility function this isn't enough—the purely cross-sectional demand behavior we observe simply can't non-parametrically identify the momentary utility function, because any monotonic transformation of utility (say $M(U)$) would generate exactly the same intra-temporal demands.

To see this, let us suppose that the "true" (momentary) indirect utility function is not $V(p, x)$, but a monotonic transformation $V^*(p, x) = M(V(p, x))$. We've estimated $\lambda = \partial V/\partial x$, but if utility is $M(U)$ then the marginal utility of expenditures isn't λ, but rather $\lambda M'(U)$, where M' is the derivative of the monotone transformation. Without knowledge of the transformation M we're limited in what we can say about risk attitudes.

However, with modest additional assumptions it's possible to estimate the empirical *distribution* of households' relative risk aversions, up to two unknown parameters.

Recall that the Arrow-Pratt measure of relative risk aversion for a household with indirect utility $V^*(p,x)$ is given by

$$\mathrm{RRA}(p,x) = -x\frac{\partial^2 V^*/\partial x^2}{\partial V^*/\partial x}.$$

With $V^*(p,x) = M(V(p,x))$, we have $\partial V^*/\partial x = M'(V(p,x))\lambda(p,x)$ (recalling that $\lambda = \partial V/\partial x$). Differentiating again and applying the chain rule allows us to write

$$\mathrm{RRA}(p,x) = -x\frac{M''}{M'}\lambda(p,x) - \frac{\partial \log \lambda(p,x)}{\partial \log x}.$$

The quantities in the second term are the ω elasticities shown in Figure 4, though the value will depend on the unknown factor of proportionality ϕ. The first term involves the first and second derivatives of the unknown function M. A judicious parameterization is

$$M(U) = \frac{U^{1-\sigma} - 1}{1-\sigma};$$

this matches related assumptions used by MaCurdy (1983) or Browning, Deaton, and Irish (1985). With this parameterization of M we have the first term $-x\lambda\frac{M''}{M'} \approx \sigma$ to a first order approximation, so that we have

$$\mathrm{RRA}(p,x) \approx \sigma - \omega(p,x)/\phi.$$

This then allows us to identify households' relative risk aversion up to the unknown constants σ and ϕ. If households all have a common transformation M, then the distribution of ω in Figure 4 will be approximately the same as the distribution of (minus) households' relative risk aversions.

The finding that households have heterogeneous relative risk aversions echoes recent findings for households in Thailand; as here, Chiappori et al. (2014) use observed data on expenditures to estimate the distribution of relative risk aversion up to an unknown parameter. However, their estimates assume homothetic utility and rely on a maintained hypothesis that households are fully insured. We are able to avoid these strong assumptions entirely; the analogous assumptions which allow us to identify the distribution of risk attitudes (up to unknown location and scale parameters) are just the much weaker requirements that elasticities are constant (but may vary across goods) and that the household maximizes utility within the period subject to a budget constraint.

7. Conclusion

In this paper we've outlined some of the key methodological ingredients needed in a recipe to estimate a simple measure of household welfare. This measure is closely related to the household's marginal utility of expenditures; it differs only in that it controls for household characteristics and adopts a particular cardinalization of utility.

The methods described are theoretically coherent, in the sense that they're consistent with a particular utility-derived demand system. Further, our approach lends itself to straightforward statistical inference and hypothesis testing, and is very parsimonious in its data requirements.

Our approach involves estimating an incomplete demand system of a new sort which features a highly flexible relationship between total expenditures and demand. The methods described here involve using one or more cross-sections of data on household expenditures on different nondurable goods and/or services. The limited data requirements suggest that these methods may be useful in constructing programs to inexpensively *measure* and *monitor* households' welfare over both different environments and across time.

In an application of these methods we use four rounds of data from Uganda. We focus on food expenditures in this dataset, estimating a system of 41 demands. We estimate both household $\log \lambda$ and Frischian elasticity parameters from this expenditure system, in addition to other demand parameters. A separate analysis allows us to characterize the distribution of households' relative risk aversions; we find convincing evidence of heterogeneity, though the distribution is not fully identified.

8. References

Aguiar, Mark, and Mark Bils. 2015. Has consumption inequality mirrored income inequality? *American Ecnomic Review* 105, no. 9 (September). `http://pubs.aeaweb.org/doi/pdfplus/10.1257/aer.20120599`.

Arellano, M. 1987. Computing robust standard errors for within-groups estimators. *Oxford Bulletin of Economics and Statistics* 49 (4): 431–34.

Attfield, Clifford LF, and Martin J Browning. 1985. A differential demand system, rational expectations and the life cycle hypothesis. *Econometrica* 53 (1): 31–48.

Beasley, William G. 1972. *The Meiji restoration.* Stanford University Press.

Benson, Todd, Samuel Mugarura, and Kelly Wanda. 2008. Impacts in Uganda of rising global food prices: the role of diversified staples and limited price transmission. *Agricultural Economics* 39 (s1): 513–524.

Blundell, R., M. Browning, and C. Meghir. 1994. Consumer demand and the life-cycle allocation of household expenditures. *Review of Economic Studies* 61:57–80.

Blundell, Richard. 1998. Consumer demand and intertemporal allocations: Engel, Slutsky, and Frisch. *Econometric Society Monographs* 31:147–168.

Blundell, Richard, V Fry, and C Meghir. 1990. Preference restrictions in microeconometric models of life cycle behaviour. Chap. 2 in *Microeconometrics: surveys and applications,* edited by J.P. Florens, M. Ivaldi, J.J. Laffont, and F. Laisney. Oxford: Basil Blackwell.

Blundell, Richard, Luigi Pistaferri, and Itay Saporta-Eksten. 2016. Consumption inequality and family labor supply. *American Economic Review* Forthcoming.

Browning, Martin, Thomas F. Crossley, and Joachim Winter. 2014. The measurement of household consumption expenditures. *Annual Review of Economics* 6:475–501.

Browning, Martin, Angus Deaton, and Margaret Irish. 1985. A profitable approach to labor supply and commodity demands over the life cycle. *Econometrica* 53:503–543.

Brunori, Paolo, Flaviana Palmisano, and Vito Peragine. 2015. *Inequality of opportunity during the great recession in Uganda.* Working Paper 2015/039. WIDER.

Chiappori, Pierre-André, Krislert Samphantharak, Sam Schulhofer-Wohl, and Robert M Townsend. 2014. Heterogeneity and risk sharing in village economies. *Quantitative Economics* 5 (1): 1–27.

Comin, Diego A, Danial Lashkari, and Martí Mestieri. 2015. *Structural change with long-run income and price effects.* Technical report. National Bureau of Economic Research.

Cooper, Russel J., Keith R. McLaren, and Gary K.K. Wong. 2001. On the empirical exploitation of consumers' profit functions in static analyses. *Economics Letters* 72:181–187.

Deaton, Angus, and John Muellbauer. 1980. An almost ideal demand system. *The American Economic Review* 70, no. 3 (June): 312–326.

Development Research Group of the World Bank. Downloaded 2017. *PovcalNet: the on-line tool for poverty measurement.* `http://iresearch.worldbank.org/PovcalNet/`.

Diewert, W. E. 1971. An application of the Shephard duality theorem: a generalized Leontief production function. *Journal of Political Economy* 79:461–507.

Duponchel, Marguerite, Andy McKay, and Sarah Ssewanyana. 2014. The dynamics of poverty in Uganda, 2005/6 to 2011/12: has the progress stalled? Unpublished manuscript, October.

Engel, Ernst. 1857. Die productions-und consumtionsverhaltnisse des königreichs sachsen. *Zeitschrift des Statistischen Biireaus des Koniglich Sachsischen Ministeriums des Innern,* no. 8 9: 1–54.

Fisher, Irving. 1918. Is "utility" the most suitable term for the concept it is used to denote? *American Economic Review* 8 (2): 335–337.

Fisher, Irving. 1927. A statistical method for measuring "marginal utility" and testing the justice of a progressive income tax. In *Economic essays in honor of John Bates Clark,* 157–193. Macmillan Co.

Foster, James, Joel Greer, and Erik Thorbecke. 1984. A class of decomposable poverty measures. *Econometrica* 52 (3): 761–766.

Frisch, Ragnar. 1959. A complete scheme for computing all direct and cross demand elasticities in a model with many sectors. *Econometrica* 27, no. 2 (April): 177–196.

———. 1964. Dynamic utility. *Econometrica:* 418–424.

———. 1978. *New methods of measuring marginal utility.* Porcupine Press.

Golub, G. H., and C. Reinsch. 1970. Singular value decomposition and least squares solutions. *Numerische Mathematik* 14 (5): 403–420.

Gorman, W. M. 1959. Separable utility and aggregation. *Econometrica* 27:469–81.

———. 1976. Tricks with utility functions. In *Essays in economic analysis,* edited by M. Artis and R. Nobay, 211–243. Cambridge: Cambridge University Press.

Hanoch, Giora. 1975. Production and demand models with direct or indirect implicit additivity. *Econometrica* 43 (3): 395–419.

Heckman, James. 1974. Life cycle consumption and labor supply: an explanation of the relationship between income and consumption over the life cycle. *American Economic Review* 64 (1): 188–194.

Heckman, James J. 1974. Shadow prices, market wages, and labor supply. *Econometrica* 42 (4): 679–694.

———. 1976. A life-cycle model of earnings, learning, and consumption. *Journal of political economy:* S9–S44.

Heckman, James J, and Thomas E MaCurdy. 1980. A life cycle model of female labour supply. *The Review of Economic Studies:* 47–74.

Houthakker, Hendrik S. 1957. An international comparison of household expenditure patterns, commemorating the centenary of Engel's law. *Econometrica* 25 (4): 532–551.

Karaivanov, Alexander, and Robert M. Townsend. 2014. Dynamic financial constraints: distinguishing mechanism design from exogenously incomplete regimes. *Econometrica* 82 (3): 887–959.

Kinnan, Cynthia. 2014. Distinguishing barriers to insurance in Thai villages. Unpublished Ms.

Laczó, Sarolta. 2015. Risk sharing with limited commitment and preference heterogeneity: structural estimation and testing. *Journal of the European Economic Association* 13 (2): 265–292.

Lewbel, Arthur. 1987. Characterizing some Gorman Engel curves. *Econometrica* 55:1451–59.

———. 1991. The rank of demand systems: theory and nonparametric estimation. *Econometrica* 59 (3): 711–730.

MaCurdy, Thomas E. 1981. An empirical model of labor supply in a a life-cycle setting. *Journal of Political Economy* 89 (61): 1059–1085.

———. 1983. A simple scheme for estimating an intertemporal model of labor supply and consumption in the presence of taxes and uncertainty. *International Economic Review* 24 (2): 265–289.

Mazzocco, Maurizio, and Shiv Saini. 2011. Testing efficient risk sharing with heterogeneous risk preferences. *American Economic Review* 102 (1): 428–468.

Ravallion, Martin, Shaohua Chen, and Prem Sangraula. 2009. Dollar a day revisited. *The World Bank Economic Review* 23 (2): 163–184.

Srinivasan, Saradha. 1979. *Mensuration in ancient India.* Delhi: Ajanta Publications.

Young, Alwyn. 2012. The African growth miracle. *Journal of Political Economy* 120 (4): 696–739.

Appendix A. A method for computing the singular value decomposition of random matrices with missing elements

Consider a random matrix $\mathbf{X}$ having a continuous distribution with support over some subset of $\mathbb{R}^{n \times m}$ and a second random matrix $\mathbf{M}$ of the same dimension with elements either 1 or 0. Assume without loss of generality that $n \leq m$, and that the matrix $\mathbf{X}$ is "low rank" in the sense that its rank is strictly less than n. We observe only a matrix $\mathbf{A}$ which is the Hadamard product of $\mathbf{X}$ and $\mathbf{M}$. Any zero element of $\mathbf{A}$ is said to be "missing."

We wish to construct a matrix $\hat{\mathbf{X}}$ close to $\mathbf{X}$ in the Frobenius norm. If we assume that the rank of $\mathbf{X}$ is some known number r then we have the compact singular value decomposition of the matrix $\mathbf{X} = \mathbf{U}^* \mathbf{\Sigma}^* \mathbf{V}^{*\top}$, with $\mathbf{U}^*$ $n \times r$, $\mathbf{S}^*$ $r \times r$ and diagonal, and $\mathbf{V}^*$ $m \times r$.

Our strategy involves first using the m columns of $\mathbf{A}$ to estimate the $n \times r$ matrix $\mathbf{U}^* \mathbf{\Sigma}^*$. We construct a matrix

$$\mathbf{P} = n\text{diag}(\mathbf{M}\ell_m)^{-1}\mathbf{A}\mathbf{A}^\top,$$

which can be interpreted as the inner product $\mathbf{A}\mathbf{A}^\top$ scaled to ignore zero elements. Then the square root of the eigenvalues of P is an estimator for the diagonal of $\mathbf{\Sigma}^*$, while the corresponding eigenvectors $\mathbf{U}$ estimate $\mathbf{U}^*$.

With $\mathbf{U}\mathbf{\Sigma}$ in hand we proceed row by row to construct an estimate of $\mathbf{V}^*$: suppose a is a column vector from the matrix $\mathbf{A}$, and that $\mathbf{A}$ has the compact singular value decomposition $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$. The vector a can be partitioned into two parts y and x, while a matrix $\mathbf{U}\mathbf{\Sigma}_x$ can be constructed by selecting just the rows of $\mathbf{U}\mathbf{\Sigma}$ corresponding to the x elements of the vector a. Then the row of $\mathbf{V}^*$ corresponding to a can be estimated by

$$v = (\mathbf{U}\mathbf{S}_x)^+ x,$$

where the + operator here indicates the Penrose-Moore pseudo-inverse. Iterating over all m columns of $\mathbf{A}$ then yields the desired matrix $\hat{\mathbf{X}}$.

Appendix B. Online Appendix: Food items across rounds and aggregation

Table B.1 gives a complete list of codes and labels across rounds (a "0" indicates that a given code wasn't used in the corresponding round).

The following just modifies the output of `item_expenditures` to add a "Preferred label" column (and column headings), and an "Aggregate Label." The "Aggregate Label" need not be unique; expenditures for all items with the same "Aggregate Label" will be summed together, yielding what we call a "minimally-aggregated" set of data of food expenditures. This minimal aggregation confines itself to combining expenditures on different food items which seem to obviously be very close substitutes. Sometimes these differences are just in units: we aggregate "clusters" and "heaps" of Matoke, for example. Othertimes the form of the good is somewhat different: fresh and dried cassava are aggregated, for example.

Table B.1: Labels for various food items in different rounds, with "Preferred" and "Aggregate" labels. Some items do not have a Preferred Label.

Code	Preferred Label	Aggregate Label
100	Matoke (??)	Matoke
101	Matoke (bunch)	Matoke
102	Matoke (cluster)	Matoke
103	Matoke (heap)	Matoke
104	Matoke (other)	Matoke
105	Sweet Potatoes (fresh)	Sweet Potatoes
106	Sweet Potatoes (dry)	Sweet Potatoes
107	Cassava (fresh)	Cassava
108	Cassava (dry/flour)	Cassava
109	Irish Potatoes	Irish Potatoes
110	Rice	Rice
111	Maize (grains)	Maize

Continued on next page

Continued from previous page

Code	Preferred Label	Aggregate Label
112	Maize (cobs)	Maize
113	Maize (flour)	Maize
114	Bread	Bread
115	Millet	Millet
116	Sorghum	Sorghum
117	Beef	Beef
118	Pork	Pork
119	Goat Meat	Goat Meat
120	Other Meat	Other Meat
121	Chicken	Chicken
122	Fresh Fish	Fresh Fish
123	Dry/Smoked fish	Dry/Smoked fish
124	Eggs	Eggs
125	Fresh Milk	Fresh Milk
126	Infant Formula	Infant Formula
127	Cooking oil	Cooking oil
128	Ghee	Ghee
129	Margarine, Butter, etc	Margarine, Butter, etc
130	Passion Fruits	Passion Fruits
131	Sweet Bananas	Sweet Bananas
132	Mangoes	Mangoes
133	Oranges	Oranges
134	Other Fruits	Other Fruits
135	Onions	Onions
136	Tomatoes	Tomatoes
137	Cabbages	Cabbages
138	Dodo	Dodo
139	Other vegetables	Other Vegetables
140	Beans (fresh)	Beans

Continued on next page

Continued from previous page

Code	Preferred Label	Aggregate Label
141	Beans (dry)	Beans
142	Ground nuts (in shell)	Ground nuts
143	Ground nuts (shelled)	Ground nuts
144	Ground nuts (pounded)	Ground nuts
145	Peas	Peas
146	Sim sim	Sim sim
147	Sugar	Sugar
148	Coffee	Coffee
149	Tea	Tea
150	Salt	Salt
151	Soda	Soda
152	Beer	Beer
153	Other Alcoholic drinks	Other Alcoholic drinks
154	Other drinks	Other drinks
155	Cigarettes	Cigarettes
156	Other Tobacco	Other Tobacco
157	Restaurant (food)	Restaurant (food)
158	Restaurant (soda)	Soda
159	Restaurant (beer)	Beer
160	Other juice	Other juice
161	Other foods	Other foods
162	Peas (dry)	Peas
163	Ground nut (paste)	Ground nuts
164		Other Vegetables
165		Other Vegetables
166		Other Fruits
167		Other Vegetables
168		Other Fruits
169		Other Fruits

Continued on next page

Continued from previous page

Code	Preferred Label	Aggregate Label
170		Other Fruits
171		Other Fruits