

Generalized Random Coefficients With Equivalence Scale Applications

Arthur Lewbel and Krishna Pendakur
Boston College and Simon Fraser University

original Sept. 2011, revised May 2012

Abstract

We propose a generalization of random coefficients models, in which the regression model is an unknown function of a vector of regressors, each of which is multiplied by an unobserved error. We also investigate a more restrictive model which is additive (or additive with interactions) in unknown functions of each regressor multiplied by its error. We show nonparametric identification of these models. In addition to providing a natural generalization of random coefficients, we provide economic motivations for the model based on demand system estimation. In these applications, the random coefficients can be interpreted as random utility parameters that take the form of Engel scales or Barten scales, which in the past were estimated as deterministic preference heterogeneity or household technology parameters. We apply these results to consumer surplus and related welfare calculations.

JEL codes: C14 D12 D13 C21 Keywords: unobserved heterogeneity, nonseparable errors, random utility parameters, random coefficients, equivalence scales, consumer surplus, welfare calculations. The authors wish to thank Jinyong Hahn, Richard Blundell, Stefan Hoderlein, and Rosa Matzkin for helpful discussions and suggestions.

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/>

1 Introduction

Suppose a variable Y depends on a vector of regressors $X = (X_1, \dots, X_K)$, and on a vector of unobserved errors U_0, U_1, \dots, U_K which represent unobserved heterogeneity in

the dependence of Y on X (U_0 could also represent measurement error in Y). We propose a generalized random coefficients model given by

$$Y = G(X_1 U_1, \dots, X_K U_K) + U_0 \quad (1)$$

for some unknown function G . We assume that G is nonparametrically identified, and show nonparametrical identification of the distributions of each random coefficient U_k . Our model permits the random coefficients to be correlated with regressors. Identification in this case uses a control function type assumption, that is, assumes that each U_k is conditionally independent of the corresponding X_k , conditioning on observed covariates (instruments) Z .

We provide examples of economic models that have been widely used in the past to model preference heterogeneity that take the form equation (1) but with U_k specified as deterministic functions of observables. It is therefore an immediate natural generalization of those models to let U_k to embody unobserved heterogeneity.

Our identification results for the general model of equation (1) impose some strong smoothness assumptions, so we will first focus on additive models of the form

$$Y = \sum_{k=1}^K G_k(X_k U_k) + U_0 \quad (2)$$

where the functions G_1, \dots, G_K are unknown. For these additive models the identifying assumptions are less restrictive. We also consider extensions such as adding interaction terms to the model of the form $X_j X_k U_{jk}$ (that is, additional random coefficients on cross terms) or of the form $G_{jk}(X_j U_j X_k U_k)$ (that is, interactions of composite regressors), and we consider identification when some X_k components are discretely distributed.

Ordinary random coefficients are the special case of equation (2) in which each G_k is the identity function. Additive models are a common generalization of linear models; see, Hastie and Tibshirani (1990), Linton (2000), and Wood (2006). Nonparametric identification and estimation of random coefficients models is considered by Beran and Hall (1992), Beran, Feuerwerker, and Hall (1996) and Hoderlein, Klemelae, and Mammen (2010). Recent generalizations include random coefficient linear index models in binary choice, e.g., Ichimura and Thompson (1998), Gautier and Kitamura (2010), and semiparametric extensions of McFadden (1974) and Berry, Levinsohn, and Pakes (1995) type models, e.g., Berry and Haile (2009).

Particularly relevant for this paper is Matzkin (2003), which in an appendix describes generic identifying conditions for additive models with unobserved heterogeneity. Also relevant is Hoderlein, Nesheim, and Simoni (2011), who provide high level conditions for identification and estimation of models that, like ours, contain a vector of random parameters, though their results require the model to be finitely parameterized.

This paper also contributes to the literature on estimation of models with nonseparable errors, in particular where those errors arise from structural heterogeneity parameters such as random utility parameters. Older examples of such models include Heckman and Singer (1984) and Lewbel (2001). More recent work focusing on general identification and estimation results include Chesher (2003), Altonji and Matzkin (2005), Hoderlein, and Mammen (2007), Matzkin (2007a, 2008), and Imbens and Newey (2009).

In our empirical applications, the random coefficients will represent equivalence scales in consumer demand models. There is a long history of using equivalence scales to empirically model observed sources of preference heterogeneity. See, e.g., Lewbel (1997) for a survey. Engel (1895) and Barten (1964) type equivalence scales take the form of multiplying total expenditures or each price in a demand function by a preference heterogeneity parameter, as in equation (1). It is therefore a natural extension of this literature to include unobserved preference heterogeneity in these equivalence scales.

We apply these estimated demand functions to do welfare analyses. In particular, we use a Barten scaled energy demand function to do consumer surplus calculations for an energy price change (as in Hausman 1981). Our welfare application is essentially a variant or application of the ideas in Hoderlein and Vanhems (2010, 2011), who introduce unobserved preference heterogeneity into the Hausman model. The first of these two papers introduced scalar preference heterogeneity into the model nonparametrically, while the latter incorporated heterogeneity in the form of ordinary random coefficients. The difference in our model from Hoderlein and Vanhems (2010, 2011) is that we follow the prior consumer demand literature by including the preference heterogeneity in the form of Barten equivalence scales, differing from the prior demand literature in that our Barten scales include unobserved heterogeneity (a smaller additional difference is the way we also include an additive measurement error).

Other papers that introduce nonseparable unobserved preference heterogeneity in continuous demand systems include Brown and Walker (1989), Lewbel (2001), Beckert (2006) Matzkin (2007b), and Beckert and Blundell (2008). Lewbel and Pendakur (2009) propose a continuous demand system model in which the standard separable errors equal utility parameters summarizing preference heterogeneity, and do welfare calculations showing that accounting for this heterogeneity has a substantial impact on the results. Lewbel and De Nadai (2011) show how preference heterogeneity can be separately identified from measurement errors. A relatively close empirical model to ours is Comon and Calvet (2003), who use repeated cross sections and deconvolution to identify a distribution of unobserved heterogeneity in income effects.

The next two sections provide our main theorems on identification of equation (2) and (1). We then provide two empirical applications of the results. The first is a small analysis of Engel equivalence scales, and the second is a larger study of Barten scales. The latter

application includes a new theorem characterizing the solution to a semiparametric class of Hausman (1981) type consumer surplus models. These applications are followed by some additional extensions of our main results to encompass the case of discrete regressors, and to allow interaction terms with more random coefficients into the additive model. We then conclude.

2 Additive Model Identification

For any random vectors A and B let $F_{A|B}(a | b)$ and $f_{A|B}(a | b)$ denote the conditional cumulative distribution function and conditional probability density function, respectively, of A given B . The identification theorems here assume the distribution function $F_{Y|X,Z}(y | x, z)$ is known. The simplest estimators based on these theorems could be based on iid draws of $\{Y, X, Z\}$.

We first consider the additive model $Y = \sum_{k=1}^K G_k(X_k U_k) + U_0$ where $X = (X_1, \dots, X_K)$ is continuously distributed, and show nonparametric identification. Later we will extend to cases that allow for discrete X 's and interaction terms.

Some notation: Let e_k be the K vector containing a 1 in position k and zeros everywhere else. The $K - 1$ vector that contains all the elements of X except for X_k is denoted $X_{(k)}$.

ASSUMPTION A1: The conditional distribution $F_{Y|X,Z}(y | x, z)$ and the marginal distribution $F_Z(z)$ are identified. $(U_0, U_1, \dots, U_K) \perp X | Z$ and $(U_1, \dots, U_K) \perp U_0 | Z$. Either U_0 has a nonvanishing characteristic function or U_0 is identically zero. $\text{supp}(U_0) \subseteq \text{supp}(Y)$ and $\{0, e_1, \dots, e_K\} \subseteq \text{supp}(X)$.

ASSUMPTION A2: $U_k, X_k | Z$ are continuously distributed, and for every $r \in \text{supp}(X_k U_k)$ there exists an $x_k \in \text{supp}(X_k)$ such that $f_{U_k}(x_k^{-1} r) \neq 0$.

ASSUMPTION A3: G_k is a strictly monotonically increasing function. The location and scale normalizations $G_k(0) = 0$ and $G_k(1) = y_0$ for some known $y_0 \in \text{supp}(Y)$ are imposed.

Assumption A1 first assumes identification of $F_{Y|X,Z}(y | x, z)$ and $F_Z(z)$, which would in general follow from a sample of observations of Y, X, Z with sample size going to infinity. Identification of $F_{Y|X,Z}(y | x, z)$ is actually stronger than necessary for Theorem 1, since only certain features of this distribution are used in the proof. For example, it would suffice to only identify $F_{Y|X,Z}(y | x_k e_k, z)$ for $k = 1, \dots, K$, but it is difficult to

construct situations where these sufficient distributions would be identified without having $F_{Y|X,Z}$ identified.

Assumption A1 also provides conditional independence and support requirements. The role of Z is to permit the error U_0 and random coefficients U_k to be endogenous and hence correlated with X . With this type of endogeneity, Z could be control function residuals as in Blundell and Powell (2003, 2004). In particular, if $X_k = h_k(X_{-k}, Q) + Z_k$ for some observed instrument vector Q and some identified function h_k (typically h_k would be $E(X_k | X_{-k}, Q)$), then the conditional independence assumptions in A1 correspond to standard control function assumptions.

Note that Z can be empty, so all the results given below will hold if there is no Z , in which case U is exogenous and hence independent of X . We do not require that the random coefficients U_k for $k = 1, \dots, K$ be independent of each other, however, the Theorems below prove identification of the marginal distributions of each U_k , not their joint distribution.

Assumption A2 requires continuously distributed regressors and random coefficients. An alternative to Assumption A2 allowing for discrete distributions will be provided later. Assumptions A2 and A3 will be assumed to hold for each $k = 1, \dots, K$.

The normalizations in Assumption A3 are free because first if $G_k(0) \neq 0$ then we can redefine $G_k(r)$ as $G_k(r) - G_k(0)$ and redefine U_0 as $U_0 + G_k(0)$, thereby making $G_k(0) = 0$. Next, given a nonzero $y_0 \in \text{supp}(Y)$, there must exist a nonzero r_0 such that $G_k(r_0) = y_0$. We can then redefine U_k as $r_0 U_k$ and redefine $G_k(r)$ as $G_k(r/r_0)$, thereby making $G_k(1) = y_0$. These particular normalizations are most convenient for proving Theorem 1 below, but in applications others may be more natural, e.g., choosing location to make $E(U_0) = 0$.

THEOREM 1: Let $Y = \sum_{k=1}^K G_k(X_k U_k) + U_0$ and let Assumption A1 hold. Then the distribution function $F_{U_0|Z}$ is nonparametrically identified, and for every $k \in \{1, \dots, K\}$ such that Assumptions A2 and A3 hold, the function G_k and the distribution function $F_{U_k|Z}$ are nonparametrically identified.

Note that given identification of F_Z , $F_{U_0|Z}$, and $F_{U_k|Z}$, the marginal distributions F_{U_0} and F_{U_k} are also identified. In applications we would generally assume that Assumptions A2 and A3 hold for all $k \in \{1, \dots, K\}$, thereby identifying the entire model. However, later in Theorem 3 we will describe conditions for identification when X_k , U_k , or both are discrete, in which case the assumptions of Theorem 1 would be assumed to hold just for indices k corresponding to regressors and random coefficients that are continuously distributed. We also later provide identification for more general models that contain interaction terms.

An immediate corollary of Theorem 1 is the following alternative model. This model could be useful in contexts where Y is always positive, restricting the support of U_0 to be positive.

COROLLARY 1: Let $Y = \prod_{k=1}^K g_k(X_k U_k) + U_0$ with $g_k(X_k U_k) > 0$, and let Assumption A1 hold. Then the distribution function $F_{U_0|Z}$ is nonparametrically identified, and for every $k \in \{1, \dots, K\}$ such that Assumptions A2 and A3 hold with $G_k(X_k U_k) = \ln[g_k(X_k U_k)]$, the function g_k and the distribution function $F_{U_k|Z}$ are nonparametrically identified.

3 General Model Identification

We now show identification of the general model $Y = G(X_1 U_1, \dots, X_K U_K) + U_0$ (equation 1) for some unknown function G . Define $G_k(X_k U_k) = G(0, \dots, 0, X_k U_k, 0, \dots, 0)$, that is, G_k equals G after setting all the elements of X except for X_k equal to zero. Define the function \tilde{G} by $\tilde{G}(X_1 U_1, \dots, X_K U_K) = G(X_1 U_1, \dots, X_K U_K) - \sum_{k=1}^K G_k(X_k U_k)$, so

$$Y = \tilde{G}(X_1 U_1, \dots, X_K U_K) + \sum_{k=1}^K G_k(X_k U_k) + U_0 \quad (3)$$

ASSUMPTION A4: Assume $E(U_k^t) \neq 0$ and is finite for all integers t and U_0, U_1, \dots, U_K, X are mutually independent conditional upon Z . Assume the support of X includes a positive measure neighborhood of zero. Assume \tilde{G} is a real analytic function.

THEOREM 2: Let equation (1) and Assumptions A1, A2, A3, and A4 hold for $k = 1, \dots, K$. Then the function G and the distribution functions $F_{U_k|Z}$ for every $k \in \{0, 1, \dots, K\}$ are all nonparametrically identified.

Theorem 1 did not require the random coefficients U_k to be mutually independent, though only the separate distributions of each U_k were identified. However, to handle nonadditive functions G we require mutual independence. Theorem 2 also places the stronger restrictions of a nonzero mean and thin tails on each U_k .

Theorem 2 places stronger smoothness assumptions on \tilde{G} than on each G_k . Without decomposing G into \tilde{G} and G_k terms, a sufficient but stronger than necessary restriction to satisfy the required assumptions on these functions is that G be analytic and strictly monotonically increasing in each of its arguments $U_k X_k$ when the other elements of X are set to zero.

The proof of Theorem 2 involves evaluating the distribution of Y given X at $X = 0$, which is conditioning on a set of measure zero. However, issues of nonuniqueness

of the limiting argument (the Borel-Kolmogorov paradox) do not arise here, since the identification proof depends only on transformations of smooth conditional expectation functions.

However, although the identification proofs are constructive, this conditioning argument suggests that estimation based on copying the steps of the identification proof is likely to be inefficient. The distribution of Y given X provides information about the unknown functions at all values of X , which should be employed for efficiency in estimation. However, for obtaining closed form identification arguments, at points other than $X = 0$ this information takes the form of integral equations for g that are difficult to solve.

4 Applications

We have two applications relevant to consumer demand estimation: incorporation heterogeneity via Engel and Barten scaling. These two strategies bring unobserved heterogeneity into demand estimation via scaling total expenditure and prices, respectively. They are two of the most venerable strategies used to bring preference heterogeneity on the basis of observed variables into demand models (see, e.g., Pollack and Wales 1990) and are consequently a natural starting point for the incorporation of unobserved preference heterogeneity.

Let a "consumer" refer to a household or an individual consumer. Let M be money a consumer spends on all goods, and let Q_j denote the quantity purchased of a good j . Let $\bar{S}(Q_1, \dots, Q_J)$ be the direct utility function of a particular consumer, called the reference consumer, over the bundle of goods Q_1, \dots, Q_J . Assume without loss of generality that the reference consumer has an unobserved preference heterogeneity parameter equal to one. Other consumers will have different values of for unobserved preference heterogeneity parameter(s). Let S refer to the direct utility function for other consumers. Assume $S(\bar{S})$ is continuous, non-decreasing, and quasi-concave.

The consumer chooses quantities to maximize utility subject to the standard linear budget constraint $\sum_{j=1}^J P_j Q_j = M$ where P_j is the price of good j and M is the total amount of money the consumer spends on this bundle of goods. Define normalised prices $X_j = P_j/M$ for each good j and rewrite the budget constraint as $\sum_{j=1}^J X_j Q_j = 1$. Write the Marshallian budget share functions that result from maximizing utility as $W_j^* = \omega_j(X_1, \dots, X_J)$, where $W_j^* = Q_j P_j / M = Q_j X_j$ is the share of money M that is spent on good j (called the budget share of good j). Let $V(X_1, \dots, X_J)$ denote the indirect utility function corresponding to S , obtained by substituting $Q_j = \omega_j(X_1, \dots, X_J) / X_j$ into $S(Q_1, \dots, Q_J)$ for $j = 1, \dots, J$.

4.1 Unobserved Heterogeneity in Engel Scales

We consider a 2-good system for food and non-food budget shares. Let W^* denote the fraction of M that the household spends on food, so W^* is the food budget share. Let the function $W^* = g(M)$ denote an individual's budget share Engel curve for food.

Based on empirical regularities noted by Engel (1895), one method of modeling how Engel curves vary across households is through Engel equivalence scales. See, e.g., Lewbel and Pendakur (2007) for a survey of various types of equivalence scales in the consumer demand literature, including Engel scales. The traditional Engel scale for a household of size h is a scalar constant α_h that multiplies total expenditures in each Engel curve, that is, $W^* = g(\alpha_h M)$.

The Engel scale α_h is a direct measure of the economies of scale of household consumption. Normalizing $\alpha_1 = 1$, if α_2 equals 1.5 then assuming that the same indifference curve for the two households corresponds to the same utility level for them, it can be shown that the two person household can attain the same indifference curve and hence the same utility level as a single person with only 50% more income. Traditional Engel scales assume that all households of the same size have the same Engel scale. A more reasonable assumption is that households vary in their ability to share goods, and even one person household vary in their ability to derive utility from a given level of consumption, and hence Engel scales should vary in unobserved ways across households. This is especially true since, as noted by Hildenbrand (1994), Browning and Carro (2007), and Lewbel (2007, 2008) among others, there is empirically a great deal of unexplained heterogeneity in consumption across households, in fact, the unexplained variation in budget shares across households is often larger than the variation explained by observed covariates.

In terms of utility functions with observed heterogeneity embodied in α_h and unobserved heterogeneity embodied in U_1 , Engel scaling is satisfied if and only if $S(Q_1, \dots, Q_J; \alpha_h, U_1) = \bar{S}(Q_1 U_1 \alpha_h, \dots, Q_J U_1 \alpha_h)$. This model for utility implies that budget share functions take the form $W^* = g(U_1 \alpha_h M)$, where U_1 varies randomly across households. The equivalence scale is then $U_1 \alpha_h$, so α_h is the systematic, observable variation due to variation in household size, and U_1 is unobserved heterogeneity in the Engel scales.

Let W denote a household's observed food budget share, which may be observed with error. Let $\lambda(W) = \ln[W/(1 - W)]$, which denotes the logit transformation of the household's budget share. For convenience we will assume that $\lambda(W) = \lambda(W^*) + U_0$, where U_0 is measurement (or specification) error in the observed W . The advantage of this formulation is that W and W^* have supports on $[0, 1]$, while $\lambda(W)$ and $\lambda(W^*)$ have supports on the whole real line, so U_0 can have support on the whole real line. In contrast, if we had specified the measurement error to be additive in W^* , then inequality constraints on W and W^* would impose support constraints on the measurement error that would make

it difficult or impossible for the measurement error to be independent of the regressor M or the Engel scale U_1 .

Define $Y = \lambda(W)$ and $G(M) = \lambda(g(M))$. Then we have the logit transformed budget share Engel curve model

$$Y = G(U_1 \alpha_h M) + U_0.$$

This model is in the class discussed in Equation (2). To show identification of this model, replace M with X_1 , let $K = 1$, and apply Theorem 1 to households of size $h = 1$ (using $\alpha_h = 1$) to show identification of the function G and distributions of U_1 and U_0 . Given these functions, identification of the constants α_h for other household sizes h is immediate using data from those other households. Given an estimate of the function G , constants α_h , and the distribution of U_1 , the estimated budget share will then be $W^* = \lambda^{-1}(G(U_1 M)) = 1 / (1 + e^{-G(U_1 M)})$.

The unobservables U_1 and U_0 are assumed to be independent, because the former is a structural random utility parameter and the latter is measurement error. Since Engel's original observations, virtually all empirical studies have found food budget share Engel curves to be monotonic in M , so it is safe to assume as required by Theorem 1 that the food budget share function g and hence the logit transformed budget share G , is strictly monotonic in M .

We estimate the model $Y = G(U_1 \alpha_h M) + U_0$ nonparametrically using sieve maximum likelihood. For modeling densities using sieves, define the density function $p_J(v, \theta)$ given by squared Hermite polynomials

$$p_J(v, \theta) = \phi(v) \left(\Lambda_J(\theta) + \sum_{j=1}^J \theta_j H_j(v) \right)^2$$

where $\phi(v)$ is the standard normal density function and $H_j(v)$ are the Hermite polynomials

$$H_j(v) = \frac{(-1)^j}{\phi(v)} \frac{d^j \phi(v)}{dv^j},$$

so, e.g., $H_1(v) = v$, $H_2(v) = v^2 - 1$, and $H_3(v) = v^3 - 3v$. To make the density integrate to one set

$$\Lambda_J(\theta) = \left[1 - \sum_{j=1}^J (j) \theta_j^2 \right]^{1/2}$$

Following Gallant and Nychka (1987), we model univariate densities using the sieve or semi-nonparametric expansion $p_J(v, \theta)$, letting $J \rightarrow \infty$ as $n \rightarrow \infty$ for sample size n . Noting that U_0 has support $(-\infty, \infty)$ and U_1 has support $(0, \infty)$, we model the densities

of the standardized variables U_0/σ_0 and $\ln U_1/\sigma_1$ using the expansion $p_J(v, \theta)$. This makes the sieve basis density functions of U_0 and U_1 be f_{0J} and f_{1J} where

$$f_{0J}(U_0, \delta, \sigma_0) = \frac{1}{\sigma_0} p_J\left(\frac{U_0}{\sigma_0}, \delta\right) \quad \text{and} \quad f_{1J}(U_1, \gamma, \sigma_1) = \frac{1}{U_1 \sigma_1} p_J\left(\frac{\ln U_1}{\sigma_1}, \gamma\right)$$

respectively, for parameter vectors δ and γ (only J terms of which are estimated).

We model G using polynomial basis functions

$$G_L(m, \beta) = \beta_0 + \beta_1 m + \dots + \beta_L m^L$$

letting $L \rightarrow \infty$ as $n \rightarrow \infty$. Note that in terms of identifying normalizations, we can if desired interpret β_0 as adding to the mean of U_0 rather than as the constant term in G to enforce $G(0) = 0$.

For a given household size h , the conditional density function of Y is then

$$f_{Y|M}(y | m; \alpha_h, \beta, \gamma, \delta) = \int_0^\infty f_0[y - G(u_1 \alpha_h m, \beta), \delta, \sigma_0] f_1(u_1, \gamma, \sigma_1) du_1$$

Let $y_i = \lambda(w_i)$, and let m_i be the household expenditure of household i . Assuming iid observations y_i, m_i, h_i of consuming households i , estimation then proceeds by substituting in the sieve functional forms G_L, f_{0J} , and f_{1J} for the unknown functions G, f_0 , and f_1 , and searching over parameter vectors $\alpha, \beta, \gamma, \delta$ and σ to maximize the log likelihood function

$$\sum_{i=1}^n \ln f_{Y|M}(y_i | m_i; \alpha_{h_i}, \beta, \gamma, \delta, \sigma)$$

We do not list here the formal assumptions for consistency and asymptotic inference of sieve maximum likelihood estimation in this application, because the generic conditions for validity of these estimators in an independently, identically distributed data setting are well established. See, e.g., Chen (2007) and references therein. However, depending on the supports and tail thickness of the model errors and regressors, our identification might be weak, in the sense that recovering the structural functions of the model could entail ill-posed inverse problems. See, e.g., Hoderlein, Nesheim, and Simoni (2011), who document these issues in a more general framework than ours. We do not formalize these conditions here, because the novelty of our paper is only in the identification and economic applications. However, we note in passing that our use of sieves can be interpreted as a choice of regularization for structural function estimation.

4.2 Empirical Engel Curve Results

We bring this model to the data with Indian household expenditure microdata from the 2003 National Sample Survey of India (Expenditures Module). Our data file consists of

3173 observations of households comprised of one or two adults aged 19 to 64. Expenditures are expressed in proportion to average household expenditures of all households (not just those comprised of one or two adults). We drop households with expenditures below the 5th and above the 95th percentile of the expenditure distribution. We follow Engel's original example in estimating budget shares for food. Table 1 gives summary statistics for food shares w_i , total expenditures m_i and an indicator d_i that the household has 2 members.

| 3173 observations | mean | std dev | min | max |
|----------------------------|------|---------|------|------|
| food share, w_i | 0.55 | 0.14 | 0.02 | 0.92 |
| total expenditures, m_i | 0.44 | 0.24 | 0.13 | 1.18 |
| two-adult household, d_i | 0.54 | 0.50 | 0 | 1 |

A standard parametric functional form for food Engel curves going back to Working (1943) is W linear in the log of M . Allowing for traditional constant Engel scales with households having one or two members gives the functional form $W = \beta_0 + \ln(\alpha_h M) \beta_1 + \varepsilon = \beta_0 + (\ln M + d \ln \alpha_2) \beta_1 + \varepsilon$ where α_2 is the traditional Engel scale for a two person household, and the Engel scale for one person household is $\alpha_1 = 1$. So α_2 could in this model be recovered from a linear regression of W on $\ln M$ and on the two person household dummy d . This regression yields coefficients of -0.14 on $\ln M$ and 0.08 on d , giving an equivalence scale of $\alpha_2 = 0.57$. The Engel scale interpretation of this parameter is that a two person household needs to spend $1/0.57 = 1.75$ times as much money as an individual living alone to attain the same level of utility as the individual.

While roughly plausible, this model has two serious drawbacks: it imposes a parametric structure on W as a function of M and it requires that the equivalence scale be constant across all households of each size. Our model relaxes these restrictions.

We estimate the sieve maximum likelihood described in the previous section, with $L = J = 3$. This makes the likelihood function to be maximized be

$$\sum_{i=1}^n \ln \int_0^{\infty} f_0 \left(y_i - \beta_0 - u_1 \alpha_{h_i} m_i \beta_1 - (u_1 \alpha_{h_i} m_i)^2 \beta_2 - (u_1 \alpha_{h_i} m_i)^3 \beta_3, \delta, \sigma_0 \right) f_1(u_1, \gamma, \sigma_1) du_1$$

where $\alpha_{h_i} = d_i \alpha_2 + (1 - d_i)$,

$$\begin{aligned} f_{1J}(u_1, \gamma) &= \frac{1}{u_1 \sigma_1} \phi \left(\frac{\ln u_1}{\sigma_1} \right) \left(1 - \gamma_1^2 - 2\gamma_2^2 - 6\gamma_3^2 \right)^{1/2} + \left(\frac{\ln u_1}{\sigma_1} \right) \gamma_1 \\ &\quad + \left(\left(\frac{\ln u_1}{\sigma_1} \right)^2 - 1 \right) \gamma_2 + \left(\left(\frac{\ln u_1}{\sigma_1} \right)^3 - 3 \left(\frac{\ln u_1}{\sigma_1} \right) \right) \gamma_3^2 \end{aligned}$$

$$f_{0J}(u_0, \delta, \sigma_0) = \frac{1}{\sigma_0} \phi\left(\frac{u_0}{\sigma_0}\right) \left(1 - \delta_1^2 - 2\delta_2^2 - 6\delta_3^2\right)^{1/2} + \left(\frac{u_0}{\sigma_0}\right) \delta_1 \\ + \left(\left(\frac{u_0}{\sigma_0}\right)^2 - 1\right) \delta_2 + \left(\left(\frac{u_0}{\sigma_0}\right)^3 - 3\left(\frac{u_0}{\sigma_0}\right)\right) \delta_3^2$$

and $\phi(\cdot)$ is the standard normal density function. Evaluating this likelihood requires numerical integration, but just a single one dimension integral is involved, which is not numerically onerous. We implemented this model in Stata. Estimated coefficients are given below. Standard errors are provided with the caveat that they treat the sieve basis functions as finite model parameterizations.

| Parameters of G | est | <i>sd</i> | Parameters of U_0, U_1 distributions | est | <i>sd</i> |
|-------------------|-------|-----------|--|-------|-----------|
| β_0 | 1.21 | 0.99 | σ_0 | 0.52 | 0.04 |
| β_1 | -7.21 | 0.94 | δ_1 | -0.12 | 1.30 |
| β_2 | 9.21 | 2.44 | δ_2 | -0.21 | 0.10 |
| β_3 | -4.06 | 1.67 | δ_3 | 0.01 | 0.21 |
| α_2 | 0.57 | 0.02 | σ_1 | 0.44 | 0.04 |
| | | | γ_1 | -0.55 | 0.03 |
| | | | γ_2 | 0.29 | 0.05 |
| | | | γ_3 | -0.06 | 0.02 |

The equivalence scale is now $U_1 \alpha_h$. The deterministic part of these scales is $\alpha_1 = 1$ (a normalization) and $\alpha_2 = .57$, which surprisingly is the same numerical value we obtained in the simple parametric model above. Thus, for a given value of U_1 , the Engel equivalence scale is 0.57.

We find considerable variation in U_1 and hence in the equivalence scales across households of each size. Figures 1 and 2 show the estimated distributions of U_0 and $\ln U_1$. The estimated distribution of U_0 is roughly symmetric, and somewhat heavy-tailed. The estimated distribution of $\ln U_1$ is skewed to the left. The implied distribution of U_1 is less skewed, with a mean of 0.54, median of 0.47, and its 5th and 95th percentiles are 0.25 and 1.02, respectively.

The estimated distribution f_{1J} of U_1 has a standard deviation of 0.29 (which is smaller than σ_1 , because the higher order terms in the hermite expansion reduce the variance in this case). This can be compared to the standard deviation of $\alpha_{h_i} = d_i \alpha_2 + (1 - d_i) \alpha_1$, which is 0.21, showing that variation in the traditional equivalence scales α_h due just to variation in household size is somewhat smaller than the variation due to unobserved heterogeneity U_1 across households of each given size.

Figures 3 and 4 show the estimated budget share functions, W^* , for single- and two-adult households, respectively, where $W^* = \lambda^{-1} (G (U_1 a_h M)) = 1 / (1 + e^{-G(U_1 a_h M)})$ and G is a third-order polynomial. These are shown at various quantiles of the U_1 distribution, illustrating how the Engel curves shift as the Engel scale varies.

5 Random Barten Scales

Our second, larger application of generalized random coefficients will use Barten scales.

Now suppose that non-reference consumers have utility functions of the form $S(Q_1, \dots, Q_J; U_1, \dots, U_J) = \bar{S}(Q_1/U_1, \dots, Q_J/U_J)$, where U_1, \dots, U_J are positive parameters that vary across consumers, embodying variation in preferences across consumers. Let the reference consumer have $U_1 = U_2 = \dots = U_J = 1$. In the consumer demand literature the parameters U_1, \dots, U_J were introduced by Barten (1964), and are known as Barten scales. See, e.g., Lewbel (1997) for a survey of various types of equivalence scales in this literature, including Barten scales. Barten scales are a generalization of Engel scales, specifically, the Barten scale model becomes equivalent to the Engel scale model when $U_1 = U_2 = \dots = U_J$.

It can be immediately verified from the first order conditions for utility maximization that a consumer will have Marshallian demand functions the form $W_j^* = \omega_j (U_1 X_1, \dots, U_J X_J)$ for each good j if and only if the consumer's direct and indirect utility function equal, up to an arbitrary monotonic transformation, $\bar{S}(Q_1/U_1, \dots, Q_J/U_J)$ and $\bar{V}(U_1 X_1, \dots, U_J X_J)$, respectively, where \bar{V} is the indirect utility function of the reference consumer. Also, given a specification of indirect utility $V(X_1, \dots, X_J)$, the corresponding Barten scaled demand functions can be obtained by the logarithmic form of Roys identity:

$$\omega_j (U_1 X_1, \dots, U_J X_J) = \frac{\partial V (U_1 X_1, \dots, U_J X_J)}{\partial \ln X_j} / \left(\sum_{\ell=1}^J \frac{\partial V (U_1 X_1, \dots, U_J X_J)}{\partial \ln X_\ell} \right) \quad (4)$$

Notice that the functional form of each ω_j only depends on the functional form of \bar{S} or equivalently of \bar{V} , so U_1, \dots, U_J can vary independently of X_1, \dots, X_J across consumers. The Barten scaled Marshallian demand functions have precisely the form of our generalized random coefficients given in equation (1).

For households with multiple members, Barten scales can be interpreted as representing the degree to which each good is shared or jointly consumed. The smaller the Barten scale U_j is, the greater the economies of scale to consumption of good j within the household. This is then reflected in the demand functions, where smaller Barten scales have the same effect on demands as lower prices. For example, if a couple with one car rides together some of the time, then in terms of total distance each travels by car, sharing has

the same effect as making gasoline cheaper. The more they drive together instead of alone, the lower is the effective cost of gasoline, and the smaller is the couple's Barten scale for gasoline.

More generally, Barten scales can provide a measure of the degree to which different households get utility from different goods. Barten scales are a popular method of modeling preference heterogeneity in empirical work. However, up until now, Barten scales have always been modeled as deterministic functions of observable characteristics of consumers. Here we consider using Barten scales to embody unobserved heterogeneity of preferences across consumers.

COROLLARY 2: Assume consumers have preferences given by an analytic, Barten scaled indirect utility function $V(U_1 X_1, \dots, U_J X_J)$. Assume that no good j exists that is Giffen at some values of X_j and not Giffen at others. Assume the conditional distribution $F_{Q|X}(q | x, z)$ is identified and $\{0, e_1, \dots, e_K\} \subseteq \text{supp}(X)$. Assume that Barten scales U_j are each positive with bounded support, satisfy Assumption A2, and are mutually independent of each other and of X , conditioning upon Z . Then the demand functions $\omega_j(U_1 X_1, \dots, U_J X_J)$ and the distribution functions $F_{U_j|Z}$ of U_j are identified for $J = 1, \dots, J$.

Corollary 2 shows that consumer demand systems with random Barten scales are non-parametrically identified. By standard revealed preference theory, the direct and indirect utility functions are therefore also nonparametrically identified up to an unknown monotonic transformation. Regarding assumptions in Corollary 2, virtually all empirically implemented consumer demand systems assume functional forms for indirect utility V that are analytic. Barten scales must be positive to preserve the standard property that utility is increasing in quantities. Similarly, the economic rationale for Barten scales suggest that they would be bounded¹

Corollary 2 refers to Giffen goods. A Giffen good is a good that has a positive own price elasticity in its Marshallian quantity demand function, and hence an upward sloping demand curve. Corollary 2 rules out goods that have sometimes positive and sometimes negative own price elasticities, and hence rules out goods that are sometimes but not always Giffen. While possible in theory, almost no empirical evidence has been found for the existence of goods that are ever Giffen.²

¹Boundedness rules out lexicographic preferences in which consumers would prefer an infinitesimal amount any other good over having an unlimited quantity of the good with the unbounded Barten scale. This is extremely unlikely to hold when goods are defined as broad categories like food, energy, clothing, etc.

²The only exception we know of is Jensen and Miller (2008), who show that some grains may have been Giffen goods for extremely poor households in rural China. Note that, despite the name, it is not the good itself that is Giffen, but rather the demand function for the good, meaning that the existence of a Giffen good is a statement only about preferences.

Applying Theorem 2 separately to each demand function would require that each demand function be monotonic in all prices. However, monotonicity in all prices is not necessary for identification here because we have multiple demand functions, and each contains the same Barten scales. We can therefore use just the monotonicity of own price effects, leaving the signs of cross price effects unconstrained, to identify the Barten scale distributions by using the demand function of each good j to just identify the distribution of the Barten scale U_j . Still, although monotonicity just applies to one scale per equation, the fact that we have multiple equations each containing the same scales means that the system of equations provides overidentifying information, relative to a single equation model.

Matzkin, (2007a), (2007b), (2008) discusses identification of systems of equations where the number of equations equals the number of random parameters, and so it is possible to invert the reduced form of the system to express the random parameters as functions of observables. Although our model has J Barten scales U_j and J demand equations, Matzkin's identification method for systems of equations cannot be directly applied here because there are actually only $J - 1$ distinct demand functions $\omega_1, \dots, \omega_{J-1}$, with the remaining demand function ω_J given by the adding up constraint that $\sum_{j=1}^J \omega_j = 1$.

To simplify our empirical analysis, we will let ω_1 be the budget share of a single good of interest, and let ω_2 denote the share of all other goods, so $\omega_2 = 1 - \omega_1$, corresponding to the general Barten scaled model with $J = 2$, and hence only requiring estimation of a single equation. Allowing for more goods would provide overidentifying information. This decomposition of consumption into two good is often done in empirical work when one wishes to focus on the welfare effects of price changes on a particular good, as we will do empirically. See, e.g., Hausman (1981), Hausman and Newey (1995), Blundell, Horowitz, and Parey (2010), and Hoderlein and Vanhems (2010,2011). This construction is formally rationalized by assuming utility is separable into good 1 and a subutility function of all other goods (see, e.g., Blackorby, Primont, and Russell (1978). Alternatively Lewbel (1996) shows that even if utility functions are not separable, conditional mean Marshallian demand functions will have all the same properties as separable demands if prices obey a stochastic hicksian aggregation condition. We therefore have the model $W_1^* = \omega_1 (U_1 X_1, U_2 X_2)$ and $W_2^* = 1 - W_1^*$.

With $J = 2$ goods, we can rewrite Roys identity to give the logit transformation of budget shares, as

$$\lambda (W_1^*) = \ln \left(\frac{\partial V (U_1 X_1, U_2 X_2)}{\partial \ln X_1} \right) - \ln \left(\frac{\partial V (U_1 X_1, U_2 X_2)}{\partial \ln X_2} \right) \quad (5)$$

where $\lambda (W_1^*)$ again is the logit transformation $\lambda (W_1^*) = \ln [W_1^* / (1 - W_1^*)]$. This addi-

tive model falls into the class covered by equation (2). With $J = 2$ and the adding up constraint $\omega_1 + \omega_2 = 1$, the single demand equation (5) embodies all the information in the demand system, and so we don't need to consider multiple equations as in the proof of Corollary 2.

5.1 Additive Model Random Barten Scales

The regularity conditions for identifying random coefficients in the additive model, given by Theorem 1, are milder than for the full model of Theorem 2, so we will first consider identification and estimation of random Barten scales in an additive model, and later consider more general nonadditive specifications.

Due to the constraints of Slutsky symmetry, additivity in Marshallian demand functions $\omega_1(X_1, X_2)$ results in extreme restrictions on behavior (see, e.g., Blackorby, Primont, and Russell 1978). So we will instead for now impose additivity on the logit transformation of demand functions (later this will be relaxed to allow for interaction terms), thereby assuming demands have the additive form

$$\lambda(W_1) = \lambda[\omega_1(U_1X_1, U_2X_2)] + U_0 = g_1(U_1X_1) + g_2(U_2X_2) + U_0 \quad (6)$$

Here the functions g_1 and g_2 are nonparametric and U_0 is interpreted as measurement error in the observed budget share W_1 relative to the true budget share W_1^* . This implies that the underlying demand function is given by

$$W_1^* = \omega_1(U_1X_1, U_2X_2) = \left(1 + e^{-g_1(U_1X_1) - g_2(U_2X_2)}\right)^{-1} \quad (7)$$

Use of the logit transformation here, and assumed additivity in logit transformed budget shares, has as far as we know not been considered before in the estimation of continuous demand functions. However, this logit transformed model has a number of advantages. First, $\lambda(W_1)$ has support on the whole real line, so the measurement error U_0 has unrestricted support, instead of a support that necessarily depends on covariates. Second, with this transform no constraints need to be placed on the range of values the nonparametric functions g_1 and g_2 take on. Third, unlike all other semiparametric or nonparametric applications of the Hausman (1981) consumer surplus type methodology (such as those cited above), a closed form expression for the indirect utility function that gives rise Marshallian demands (7) and hence (6) exists, and is given by Theorem 3.

THEOREM 3: The demand function ω_1 satisfies $\lambda[\omega_1(U_1X_1, U_2X_2)] = g_1(U_1X_1) + g_2(U_2X_2)$ for some functions g_1 and g_2 if and only if ω_1 is derived from an indirect utility function of the form

$$V(U_1X_1, U_2X_2) = H[h_1(U_1X_1) + h_2(U_2X_2), U_1, U_2].$$

for some functions h_1, h_2 , and H . The functions g_1, g_2, h_1 , and h_2 are related by

$$h_1(U_1 X_1) + h_2(U_2 X_2) = \int_{-\infty}^{\ln X_1} e^{g_1(U_1 X_1)} d \ln X_1 + \int_{-\infty}^{\ln X_2} e^{-g_2(U_2 X_2)} d \ln X_2 \quad (8)$$

and

$$g_1(U_1 X_1) + g_2(U_2 X_2) = \ln \left(\frac{\partial h_1(U_1 X_1)}{\partial \ln X_1} \right) - \ln \left(\frac{\partial h_2(U_2 X_2)}{\partial \ln X_2} \right) \quad (9)$$

Also, the functions $h_1(U_1 P_1/M)$ and $h_2(U_2 P_2/M)$ are each nonincreasing, and their sum is strictly increasing in M and quasiconvex in P_1, P_2 , and M .

The function H has no observable implications, and is present only because utility functions are ordinal and therefore unchanged by monotonic transformations. So in practice we can just write the indirect utility function in Theorem 3 as

$$V(U_1 X_1, U_2 X_2) = h_1(U_1 X_1) + h_2(U_2 X_2). \quad (10)$$

Preferences $V(X_1, X_2)$ are defined to be indirectly additively separable (see, e.g., Blackorby, Primont, and Russell 1978) if, up to an arbitrary monotonic transformation, $V(X_1, X_2) = h_1(X_1) + h_2(X_2)$ for some functions h_1, h_2 . So an equivalent way to state the first part of Theorem 3 is that ω_1 satisfies equation (7) if and only if preferences are given by a Barten scaled indirectly additively separable utility function. The second part of Theorem 3 then provides closed form expressions for the indirect utility function given the nonparametric (additive in the logit transformation) demand function and vice versa.

The fact that we have a closed form expression for indirect utility V means that the shape restrictions required for utility maximization are satisfied as long as V has the standard properties of an indirect utility function (monotonically increasing in X_1 and X_2 , homogeneity, and quasiconcavity). For example, since g_1 and g_2 are nonparametric we could nonparametrically specify h_1 and h_2 by sieve basis functions that preserve the shape restrictions implied by indirect utility functions, and then use equation (9) to get the demand functions g_1 and g_2 . Sufficient conditions for satisfying these properties are that h_1 and h_2 each have a positive first and second derivative.

To illustrate, consider a polynomial in logs sieve basis

$$\ln h_k(U_k X_k) = \sum_{s=0}^S \beta_{ks} (\ln(U_k X_k))^s \quad (11)$$

with constants β_{ks} , for $k = 1, 2$, letting $S \rightarrow \infty$ as $n \rightarrow \infty$. Logarithmic specifications like these are common in demand models, e.g., with $S = 1$ equations (10) and (11) correspond to Barten scaled Cobb Douglas preferences, and with $S = 2$ this gives a separable

version of the Translog indirect utility function of Jorgenson, Lau, and Stoker (1982), though in their model the Barten scales have the traditional form of being functions only of observable characteristics.

In this model we impose the free normalization $\beta_{20} = 0$. This is imposed without loss of generality, because if $\beta_{20} \neq 0$ then we can multiply the indirect utility function $V(U_1X_1, U_2X_2)$ by $e^{-\beta_{20}}$ (which is a monotonic transformation of V) and redefine β_{10} as $\beta_{10} - \beta_{20}$ to get an observationally equivalent representation of indirect utility that has $\beta_{20} = 0$. Applying Theorem 3 and equation (6) to this model gives the demand function

$$\begin{aligned} \lambda(W_1) &= \omega_{S1}(U_1X_1, U_2X_2, \beta) + U_0 \\ &= \beta_{10} + \left(\sum_{s=1}^S [\ln(U_1X_1)]^s \beta_{1s} - [\ln(U_2X_2)]^s \beta_{2s} \right) + \ln \left(\frac{\sum_{s=1}^S (\ln(U_1X_1))^{s-1} s \beta_{1s}}{\sum_{s=1}^S (\ln(U_2X_2))^{s-1} s \beta_{2s}} \right) + U_0. \end{aligned} \quad (12)$$

where $\omega_{S1}(U_1X_1, U_2X_2, \beta)$ denotes the sieve representation of $\omega_1(U_1X_1, U_2X_2)$ with S terms in the parameters β . Here, $\lambda(W_1)$ is additive as in (6) since the logged ratio may be written as a difference of logs.

As in the Engel curve application, we model the density functions of U_0 and U_k for $k = 1, 2$, by using Hermite polynomial seive densities³

$$f_{0J}(U_0, \delta, \sigma_0) = \frac{1}{\sigma_0} p_J \left(\frac{U_0}{\sigma_0}, \delta \right) \quad \text{and} \quad f_{kJ}(U_k, \gamma_k, \sigma_k) = \frac{1}{U_k \sigma_k} p_J \left(\frac{\ln U_k}{\sigma_k}, \gamma_k \right)$$

For a given consumer with observed values x_1 and x_2 , the conditional density function of W_1 is then

$$\begin{aligned} &f_{W_1|X_1, X_2}(w_1 | x_1, x_2; \beta, \sigma, \delta, \gamma) \\ &= \int_0^\infty \int_0^\infty f_0 \left[\ln \left(\frac{w_1}{1-w_1} \right) - \omega_{S1}(u_1x_1, u_2x_2, \beta), \delta, \sigma_0 \right] f_1(u_1, \gamma_1, \sigma_1) f_2(u_2, \gamma_2, \sigma_2) du_1 du_2 \end{aligned}$$

Assuming independently, identically distributed observations w_{1i}, x_{1i}, x_{2i} of consuming households i , estimation then proceeds by searching over parameter vectors β, σ, δ , and γ to maximize the sieve log likelihood function

$$\sum_{i=1}^n \ln f_{W_1|X_1, X_2}(w_{1i} | x_{1i}, x_{2i}; \beta, \sigma, \delta, \gamma)$$

³We impose the usual assumption that the additive model error U_0 is mean zero. In our applications we did not find it useful empirically to include more than $J = 3$ terms. Some algebra reveals that the expected value of U_0 for a 3rd order expansion is $\sigma_0 \left(2 \left((1 - \delta_1^2 - 2\delta_2^2 - 6\delta_3^2)^{1/2} - \delta_2 \right) \delta_1 + 6\delta_1\delta_2 + 12\delta_2\delta_3 \right)$, so we translate U_0 in the density f_{0J} by this function to generate a mean zero distribution for U_0 .

5.2 Empirical Additive Model Random Barten Scales

We estimate the model of the previous subsection using Canadian household expenditure microdata from the 1997 to 2008 Surveys of Household Spending. We consider households comprised of one adult (as of 31 Dec) aged 25-45 residing in provinces other than Prince Edward Island (due to data masking). We consider the share of total nondurable expenditures commanded by energy goods, and drop observations whose expenditures on energy goods are zero, and those whose total nondurable expenditures are in the top or bottom percentile of the total nondurable expenditure distribution. This leaves 9413 observations for estimation.

Total nondurable expenditures are comprised of the sum of household spending on food, clothing, health care, alcohol and tobacco, public transportation, private transportation operation, and personal care, plus the energy goods fuel oil, electricity, natural gas and gasoline. Total nondurable expenditures are scaled to equal one at its mean value, which is a free normalization of units.

Prices vary by province (9 included) and year (12 years) yielding 108 distinct price vectors for the underlying commodities comprising nondurable consumption. These underlying commodity prices are normalised to equal one in Ontario in 2002. To maximize price variation, following Lewbel (1989) and Hoderlein and Mihaleva (2008), we construct P_1 as the Stone price index using within group household specific budget shares of energy goods, and P_2 is constructed similarly for non-energy goods. These price indices all have a value of one in Ontario in 2002. Finally, the regressors X_1 and X_2 are defined as the prices for energy and non-energy divided by total nondurable expenditure for the households.

Table 3 gives summary statistics for budget shares, expenditures, prices and normalised prices.

| | mean | std dev | min | max |
|-----------------------------------|------|---------|------|-------|
| 9413 observations | | | | |
| energy share, W | 0.14 | 0.09 | 0.00 | 0.73 |
| total nondurable expenditure, M | 1.00 | 0.50 | 0.1 | 2.90 |
| price of energy goods, P_1 | 1.00 | 0.23 | 0.43 | 2.28 |
| price of nonenergy goods, P_2 | 0.96 | 0.08 | 0.76 | 1.35 |
| energy normalised price, X_1 | 1.31 | 0.92 | 0.19 | 10.27 |
| nonenergy normalised price, X_2 | 1.30 | 0.94 | 0.29 | 9.41 |

We estimate equation (12) with $S = 3$ and $J = 2$. Here, we use a 2nd order hermite expansion around the normal distribution for U_0 , $\ln U_1$ and $\ln U_2$. Higher order terms in these expansions were jointly insignificantly different from zero.

With degenerate U_1 and U_2 and $S = 2$, this model corresponds to an additively separable version of the Translog budget share function. With degenerate U_1 and U_2 and $S = 3$, it corresponds to an additively separable version of the third-order Translog budget share function (see, e.g., Nicol 1984). The innovation here is that we relax the restriction that Barten scales U_k are the same for all demographically identical consumer. We implemented this model in Stata. Estimated coefficients are given in Table 4 below. Standard errors are provided with the same caveat as before.

| Parameters of h_1 and h_2 | est | se | Parameters of U_0, U_1, U_2 distributions | est | se |
|-------------------------------|--------|-------|---|--------|-------|
| β_0 | -2.144 | 0.137 | σ_0 | 0.070 | 0.026 |
| β_{11} | 1.046 | 0.033 | σ_1 | 0.636 | 0.031 |
| β_{12} | -0.120 | 0.016 | σ_2 | 0.983 | 0.045 |
| β_{13} | 0.040 | 0.005 | δ_1 | -0.452 | 0.925 |
| β_{21} | 0.554 | 0.048 | γ_{11} | -0.663 | 0.038 |
| β_{22} | 0.087 | 0.010 | γ_{21} | -0.322 | 0.063 |
| β_{23} | 0.024 | 0.004 | δ_2 | -0.348 | 0.482 |
| | | | γ_{12} | 0.154 | 0.038 |
| | | | γ_{22} | -0.056 | 0.029 |

Figures 5 and 6 show the estimated distributions of $\ln U_1$ and $\ln U_2$. We do not show the distribution of U_0 , because it is insignificantly different from a normal (δ_1 and δ_2 are jointly insignificant). These two distributions of unobserved heterogeneity parameters are not far from log normal and hence rather strongly right-skewed, with modes well below zero.

The estimated standard deviations of $\ln U_1$ and $\ln U_2$ in Figures 5 and 6 are 0.52 and 0.84, (these differ from σ_1 and σ_2 because the γ parameters affect the second moments). The standard deviations of $\ln X_1$ and $\ln X_2$ are 0.54, indicating that unobserved preference heterogeneity in the Barten scales contributes variation to energy demand of the roughly the same order of magnitude as that contributed by observed variation in prices and total expenditures across consumers. The standard deviation of the additive error U_0 is 0.26, showing that both additive errors and unobserved preference heterogeneity contribute substantially to observed variation in demand.

We postpone more thorough empirical analyses to later, when we report estimated results from a richer model.

5.3 Interaction Terms in Utility

The additive utility model in Theorem 3, estimated in the previous subsection, restricts price interaction effects. Using identification based on Theorem 2 instead of Theorem 1, we could instead nonparametrically estimate any sufficiently smooth demand function $\omega_1(U_1X_1, U_2X_2)$, and identify the function ω_1 and distribution of the associated Barten scales U_1 and U_2 . However, doing so would lose the benefits we gained from Theorem 3 of having closed form expressions for the corresponding indirect utility function $V(U_1X_1, U_2X_2)$, which is useful for welfare analyses and convenient for imposing constraints associated with utility maximization. We will therefore instead generalize the class of indirect utility functions given by Theorem 3.

Theorem 3 yielded the indirectly additive utility function $V(X_1, X_2) = h_1(X_1) + h_2(X_2)$. To relax the restrictiveness (in terms of cross effects) of additive demand functions, we now consider adding second and third order interaction terms to the model of Theorem 3, giving an indirect utility function of the form

$$V(X_1, X_2) = h_1(X_1) + h_2(X_2) + X_1X_2\alpha_0 + X_1^2X_2\alpha_1 + X_1X_2^2\alpha_2 \quad (13)$$

For unknown functions $h_1(X_1)$ and $h_2(X_2)$ along with unknown constants α_0 , α_1 , and α_2 . Higher order interactions could be similarly identified if necessary. Barten scaling this indirect utility function, substituting the result into equation (5), and adding the error term U_0 as before gives the demand model

$$Y = \ln [g_1(U_1X_1) + M_1(U_1X_1, U_2X_2, \alpha)] - \ln [g_2(U_2X_2) + M_2(U_1X_1, U_2X_2, \alpha)] + U_0 \quad (14)$$

where $Y = \lambda(W_1)$, $g_k(U_kX_k) = U_kX_k \partial h'_k(U_kX_k) / \partial (U_kX_k)$ for $k = 1, 2$ and

$$M_1(U_1X_1, U_2X_2, \alpha) = U_1X_1U_2X_2\alpha_0 + 2U_1^2X_1^2U_2X_2\alpha_1 + U_1X_1U_2^2X_2^2\alpha_2, \quad (15)$$

$$M_2(U_1X_1, U_2X_2, \alpha) = U_1X_1U_2X_2\alpha_0 + U_1^2X_1^2U_2X_2\alpha_1 + 2U_1X_1U_2^2X_2^2\alpha_2. \quad (16)$$

Identification of this demand model follows directly from Theorem 3.⁴

For estimation of the model, we let the functions h_k in equation (13) be represented by the same polynomial in logs sieve basis functions as before. Barten scaling this indirect

⁴It's possible to directly prove identification of the demand model (14) under somewhat weaker conditions than Theorem 3. Specifically, identification follows if Assumptions A1, A2, and A3 hold with $G_k(X_k) = \ln g_k(X_k)$ for $k \in \{1, 2\}$, the functions g_1 and g_2 are differentiable, and either $g'_k(0) E(U_k)$ for $k = 1$ or for $k = 2$ is nonzero and finite. See earlier versions of this paper for a proof.

utility function gives, by equation (14), the demand function

$$\begin{aligned}
\lambda(W_1) &= \omega_{S1}(U_1 X_1, U_2 X_2, \beta) + U_0 & (17) \\
&= \ln \left[\left(e^{\beta_{10} + \sum_{s=1}^S (\ln(U_1 X_1))^s \beta_{1s}} \right) \left(\sum_{s=1}^S (\ln(U_1 X_1))^{s-1} s \beta_{1s} \right) + M_1(U_1 X_1, U_2 X_2, \alpha) \right] \\
&\quad - \ln \left[\left(e^{\sum_{s=1}^S (\ln(U_2 X_2))^s \beta_{2s}} \right) \left(\sum_{s=1}^S (\ln(U_2 X_2))^{s-1} s \beta_{2s} \right) + M_2(U_1 X_1, U_2 X_2, \alpha) \right] + U_0
\end{aligned}$$

The demand function given by equation (17) is the same as (12), except for the addition of the functions M_1 and M_2 given by equations (15) and (16), which embody the additional desired price interaction terms. We estimate equation (17) using the same sieve maximum likelihood method as before.

5.4 Empirical Barten Scales and Consumer Surplus with Interaction Terms

Table 5 presents estimated parameters for the demand equation (17), that is, the Barten Scale model with interaction terms. Again, we use a 2nd order Hermite expansion around the normal for U_0 , $\ln U_1$ and $\ln U_2$, and a 3rd order polynomial in $\ln X_j$ for G_j . In this model, if any of the interaction coefficients α_0 , α_1 , and α_2 are negative, then for large values of either U_1 or U_2 , the utility function will violate monotonicity. In the demand and likelihood functions, this would make the argument of the log function in Y negative. We therefore restrict α_0 , α_1 , and α_2 to be non-negative.

As Table 5 shows, two of the interaction terms are statistically significant (and all three are jointly significant). For comparison, we also estimated the model imposing the constraint that $U_1 = U_2 = 1$, thereby removing unobserved preference heterogeneity. This corresponds to a more traditional demand model in which the only error term is additive, albeit additive in the logit transform of the budget share.

Table 5: Interaction Terms in Utility: Estimated Parameters, Barten Scales

| Parameters of h_1 and h_2 | est | se | Parameters of U_0, U_1, U_2 distributions | est | se |
|-------------------------------|--------|-------|---|--------|-------|
| β_0 | -2.546 | 0.174 | σ_0 | 0.166 | 0.039 |
| β_{11} | 1.084 | 0.056 | σ_1 | 0.540 | 0.046 |
| β_{12} | -0.143 | 0.030 | σ_2 | 0.854 | 0.041 |
| β_{13} | 0.056 | 0.013 | δ_1 | -0.822 | 1.732 |
| β_{21} | 0.947 | 0.064 | γ_{11} | -0.646 | 0.072 |
| β_{22} | 0.276 | 0.032 | γ_{21} | -0.482 | 0.039 |
| β_{23} | 0.066 | 0.008 | δ_2 | 0.002 | 2.517 |
| α_0 | 0.000 | 0.001 | γ_{12} | 0.141 | 0.063 |
| α_1 | 0.006 | 0.002 | γ_{22} | 0.134 | 0.016 |
| α_2 | 0.017 | 0.004 | | | |

In Figures 7 and 8, we give the estimated densities of $\ln U_1$ and $\ln U_2$. Here the standard deviations of $\ln U_1$ and $\ln U_2$ are 0.44 and 0.81, respectively, which is similar to what we observed in the model without interaction terms. So after allowing for interaction terms, unobserved heterogeneity is of similar importance in energy and non-energy preferences.

Recall $X_j = P_j/M$ where M is total expenditures. Figure 9 displays estimated energy budget share functions (Engel curves) evaluated at prices $P_1 = P_2 = 1$, for each quartile of the U_1 and U_2 distribution. Nine Engel curves are displayed, corresponding to each of the combinations of one quartile of U_1 and one quartile of U_2 . Each Engel curve was obtained by simulation, drawing 10,000 observations of total expenditures M from a nonparametric estimate of the distribution of real expenditure (nominal expenditure deflated by the Stone Index) and evaluating the estimated budget share equations for each given U_1 and U_2 quartile at each total expenditure M draw. Here, we see that U_1 and U_2 cause substantial shifts in the Engel curves. For comparison, Figure 9 also displays, with a thick grey line, the Engel curve from a model which imposes $U_1 = U_2 = 1$.

The shape of the Engel curve without unobserved preference heterogeneity (thick grey line) is rather different from those that allow for unobserved preference heterogeneity. For example, at low expenditure levels, allowing for unobserved preference heterogeneity reduces the slope of the energy Engel curve, suggesting that it is not as much of a necessity as would appear in the absence of such heterogeneity.

Because U_1 and U_2 affect budget shares in different ways, it is difficult to see the joint effect of these two unobserved heterogeneity parameters on the distribution of implied behaviour. We address this in our remaining figures. Figure 10 displays a contour plot of the density of estimated energy budget shares evaluated at $P_1 = P_2 = 1$. This is again obtained by simulation, based on 10,000 draws of M as before. This time, for each real expenditure draw we also draw a value of U_1 and U_2 from their estimated distributions,

and evaluate the estimated energy budget share at these drawn values of M , U_1 and U_2 . For comparison, we also display the which have no preference heterogeneity with a thick gray line.

The standard deviation of the marginal distribution of energy budget shares is 0.09 in the model which accounts for both unobserved preference heterogeneity and observed expenditure variation. In contrast, it is only 0.02 in the model which accounts only for observed expenditure variation. Thus, the variation in budget shares due to heterogeneity in preferences is large relative to that due to variation in total expenditures.

Next, we use our model to evaluate the distribution of effects of a large change in the price of energy. Using equation (13), even with nonparametric demand components we have a closed form expression for indirect utility. We can therefore compute consumer surplus effects without approximations of the type proposed by Vartia (1984). Instead, we numerically invert the indirect utility function (13) to obtain the cost of living impact of a price change. We would otherwise need to numerically obtain a differential equation solution as in Hausman and Newey (1995), but such a solution would need to be calculated for every value U_1 , U_2 can take on.

To show price effects clearly, we consider a large price change: doubling the price of energy. We consider an initial price vector $\bar{P}_1 = \bar{P}_2 = 1$ and a new price vector $P_1 = 2, P_2 = 1$. The cost-of-living impact for the change from initial to new prices, $\pi(U_1, U_2, M, P_1, P_2, \bar{P}_1, \bar{P}_2)$, is defined as the solution to

$$V\left(\frac{U_1 \bar{P}_1}{M}, \frac{U_2 \bar{P}_2}{M}\right) = V\left(\frac{U_1 P_1}{\pi M}, \frac{U_2 P_2}{\pi M}\right),$$

which, in our case, is the solution to

$$V\left(\frac{U_1}{M}, \frac{U_2}{M}\right) = V\left(\frac{U_1 2}{\pi M}, \frac{U_2}{\pi M}\right).$$

Here, π is the proportionate change in costs M needed to compensate for the energy price change.

Figure 11 gives the estimated joint distribution (contour plot) of $\ln \pi$ and $\ln M$. This plot is constructed by calculating the surplus for each of 10,000 draws of U_1 , U_2 , and M , and, as in Figure 10, the thick gray line gives estimates based on the model without preference heterogeneity. Figure 12 shows the same information in levels rather than logs.

Table 6 gives a numerical version of the information presented in Figure 11. Here, we present summary statistics on consumer surplus unconditionally for the model with and without unobserved preference heterogeneity, and conditionally at quartiles of the expenditure distribution for the model with unobserved preference heterogeneity.

Table 6: Summary Statistics for log-Cost of Living Change

| | Mean | Std Dev | Lower Qtl | Median | Upper Qtl |
|-------------------------------|-------|---------|-----------|--------|-----------|
| Overall—without heterogeneity | 0.111 | 0.012 | 0.105 | 0.113 | 0.120 |
| Overall—with heterogeneity | 0.128 | 0.072 | 0.073 | 0.120 | 0.172 |
| Lower Qtl ($\ln M = -0.38$) | 0.140 | 0.079 | 0.070 | 0.137 | 0.204 |
| Median ($\ln M = -0.03$) | 0.130 | 0.071 | 0.070 | 0.123 | 0.181 |
| Upper Qtl ($\ln M = 0.26$) | 0.113 | 0.057 | 0.068 | 0.110 | 0.151 |

The estimated energy budget shares shown in Figure 10 have an average of 0.14 for both the model with unobserved preference heterogeneity and the model without. It has long been known that first order approximations to cost of living effects of marginal price changes can be evaluated without estimating demand functions and associated demand elasticities (see, e.g., Stern 1987). However, the estimated average cost-of-living impacts given in Table 6 are much less than the value of 0.14 that would be used for first order approximations, showing that price substitution effects are large. This supports findings in, e.g., Banks, Blundell, and Lewbel (1996) that, contrary to the first order approximation results, it is necessary to estimate demand functions and associated price elasticities to properly evaluate consumer surplus and welfare effects when price changes are large rather than marginal. We find substitution effects exceeding 1 percentage point for our preferred model.

Table 6 also shows that the average cost-of-living impact ascribed to a price change depends quite substantially on whether or not we account for unobserved preference heterogeneity. The model without unobserved preference heterogeneity shows an average cost-of-living impact of 11.1 per cent. In contrast, when we allow for unobserved preference heterogeneity, the estimated average is 12.8 per cent. This difference in the averages is large, accounting for more than a tenth of the overall impact.

Given that our preferred model has variation in both preferences and budgets, it is not surprising that the variance of cost-of-living impacts is higher than in the model with just variation in budgets. However, the magnitude of this difference is surprisingly large. Variation in budgets in the model with no preference heterogeneity induces a standard deviation of cost-of-living impacts of 1.2 percentage points. Variation in budgets and preferences in the model with preference heterogeneity induces a standard deviation in cost-of-living impacts six times as large (7.2 percentage points). Thus, unobserved preference heterogeneity is dramatically more important than variation in budgets in this particular policy experiment.

Another way to see this point is to assess how much variation in cost-of-living impacts there is at a given level of expenditure. The bottom part of Table 6 addresses this. Here, we see that at the median expenditure level, unobserved preference heterogeneity induces a standard deviation of 7.1 percentage points, comparable in magnitude to the unconditional

standard deviation. However, it is interesting to note that for richer consumers, unobserved preference heterogeneity induces less variation in cost-of-living impacts. The standard deviation is 5.7 percentage points at the upper quartile of expenditures in comparison with 7.9 percentage points at the bottom quartile cutoff. This shows that, due to nonseparability of preference heterogeneity, even independently distributed preference heterogeneity may have effects on economic variables that are correlated with observables.

It is clear from Figure 11 that accounting for unobserved heterogeneity makes a big difference in our assessment of the consumer surplus associated with an increase in energy prices. Here, we see that the variance in consumer surplus conditional on total nondurable expenditure, M , is much larger than the variance across values of M , so accounting for unobserved heterogeneity has a bigger impact than accounting for observed heterogeneity in M .

6 Extensions

Theorem 2 provided one extension of our base case additive model of Theorem 1. Here we consider two alternative extensions. The first concerns identification when some regressors are discrete, while the second looks at identification of models containing additional random coefficients on interaction terms.

6.1 Discrete Regressors

Much of the literature on nonseparable errors and unobserved heterogeneity focuses on continuous regressors, but in empirical econometric applications, discrete regressors are common. Here we extend the results of Theorem 1 to allow for discrete regressors. Define the random variable \tilde{Y}_k , the function δ_k , and the set Ψ_k as follows. $\tilde{Y}_k = G_k(X_k U_k)$. Let $\delta_k(x_k, z) = 1$ if and only if there exists a $\tilde{y}_k \in \text{supp}(\tilde{Y}_k | X = x_k e_k, Z = z)$ such that $F_{\tilde{Y}_k | X, Z}(y_0 | e_k, z) = F_{\tilde{Y}_k | X, Z}(\tilde{y}_k | e_k x_k, z)$ if $x_k > 0$ or $F_{\tilde{Y}_k | X, Z}(y_0 | e_k, z) = 1 - F_{\tilde{Y}_k | X, Z}(\tilde{y}_k | e_k x_k, z)$ if $x_k < 0$. Let $\Psi_k = \{x_k : \text{for some } z \in \text{supp}(Z), \delta_k(x_k, z) = 1\}$.

ASSUMPTION A2': The function $F_{U_k | Z}(u_k | z)$ is invertible in u_k for all $u_k \in \text{supp}(U_k | Z = z)$. If $G_k(r)$ is known for all $r \in \Psi_k$, then $G_k(r)$ is known for all r on the support of $X_k U_k$.

For a given k , Assumption A2' essentially provides identification for a discrete regressor X_k by taking a value x_k that X_k can take on, and finding a value \tilde{y}_k such that the distribution $F_{\tilde{Y}_k | X, Z}$ evaluated at \tilde{y}_k and $x_k e_k$ matches a known value for the distribution at which G_k is identified by normalization. This is then used to identify the function $G_k(r)$ at the point $r = x_k$. The set Ψ_k is then the set of all such points for which G_k can be

identified by matching. The last part of Assumption A2' then assumes that identifying G_k at all the points in Ψ_k suffices to identify G_k everywhere.

This last condition will hold nonparametrically if Ψ_k contains all the values that $X_k U_k$ can take on. For example, if X_k and U_k are binary (each taking the values zero or one with strictly positive probability) and Z is empty, then it is straightforward to verify that Assumption A2' will hold if there exists a \tilde{y}_k such that $F_{\tilde{Y}_k|X}(y_0 | e_k) = F_{\tilde{Y}_k|X_k}(\tilde{y}_k | 0)$. As this example shows, unlike Assumption A2, Assumption A2' does not require X_k or U_k to be continuously distributed. Assumption A2' could also be used in place Assumption A2 if X_k is continuously distributed and U_k is not.

Assumption A2' can alternatively be satisfied if $G_k(r)$ is parameterized to be identifiable just from the values of $r \in \Psi_k$. So, e.g., if it is known that $G_k(r) = \theta_{k0} + \theta_{k1}r + \theta_{k2}r^2$, then as long as Ψ_k contains at least three elements (associated with three different values of \tilde{y}_k), Assumption A2' will be satisfied, because only three points are required to identify a quadratic. This will suffice for identification even if U_k is continuous and X_k is discrete.

THEOREM 4: Let $Y = \sum_{k=1}^K G_k(X_k U_k) + U_0$ and let Assumption A1 hold. Then the distribution function $F_{U_0|Z}$ is identified, and for every $k \in \{1, \dots, K\}$ such that Assumptions A2' and A3 hold, the function G_k and the distribution function $F_{U_k|Z}$ are nonparametrically identified.

Note that Theorems 1 and 3 can be combined, using Assumption A2' and Theorem 3 to identify G_k and $F_{U_k|Z}$ for indices k in which X_k , U_k , or both are discrete, and using Assumption A2 and Theorem 1 for identification for the remaining continuous regressors and random coefficients.

6.2 Additional random coefficients on interaction terms

Consider models of the form

$$Y = \left(\sum_{k=1}^K G_k(X_k U_k) \right) + \left(\sum_{j=1}^{K-1} \sum_{k=j+1}^K X_j X_k U_{jk} \right) + U_0 \quad (18)$$

Equation (18) relaxes the additivity restriction of Theorem 1 by adding pairs of interacting regressors to interact, and allowing each of these pairs to have their own random coefficients, in addition to the random coefficients in each $G_k(X_k U_k)$ term.

Assumptions A1 and A2 are extended to this model as follows. Let e_{jk} be a K vector that equals one in positions j and k and zero elsewhere. Note that when $j = k$, $e_{jk} = e_k$.

ASSUMPTION B1: The conditional distribution $F_{Y|X,Z}(y | x, z)$ and the marginal distribution $F_Z(z)$ are identified. $(U_0, U_1, \dots, U_K, U_{12}, U_{13}, \dots, U_{K-1,K}) \perp X | Z$,

$(U_1, \dots, U_K, U_{12}, U_{13}, \dots, U_{K-1,K}) \perp U_0 \mid Z$, U_0, U_1, \dots, U_K are mutually independent conditional upon Z , and for all $j < k$, $(U_j, U_k) \perp U_{jk} \mid Z$. Either U_0 has a non-vanishing characteristic function or U_0 is identically zero. $\text{supp}(U_0) \subseteq \text{supp}(Y)$ and $\{0, e_{11}, e_{12}, \dots, e_{KK}\} \subseteq \text{supp}(X)$. For all $j \neq k$, $G_k(U_k) + G_j(U_j) + U_0$ has a nonvanishing characteristic function.

ASSUMPTION B2: For $k = 1, \dots, K$: $U_k, X_k \mid Z$ are continuously distributed, and for every $r \in \text{supp}(X_k U_k)$ there exist an x_k on the support of X_k such that $f_{U_k}(x_k^{-1}r) \neq 0$. For $j = 1, \dots, K - 1$ and $k = j + 1, \dots, K$: $U_{jk}, X_j X_k \mid Z$ are continuously distributed, and for every $r \in \text{supp}(X_j X_k U_{jk})$ there exist an $x_j x_k$ on the support of $X_j X_k$ such that $f_{U_{jk}}(x_j^{-1}x_k^{-1}r) \neq 0$.

These are all direct extensions of Assumptions A1 and A2, and A3 to include the interaction terms. The main additional assumptions we now require are that U_0, U_1, \dots, U_K be mutually independent and that U_{jk} to be independent of (U_j, U_k) conditioning on Z . We then get the following generalization of Theorem 1.

THEOREM 5: Let equation (18) and Assumptions B1, B2, and (for $k = 1, \dots, K$) A3 hold. Then the distribution function $F_{U_0|Z}$, the regression and distribution functions G_k and $F_{U_k|Z}$ for every $k \in \{0, 1, \dots, K\}$, and the distribution functions $F_{U_{jk}|Z}$ for every $j = 1, \dots, K - 1$ and $k = j + 1, \dots, K$, are all nonparametrically identified.

The proof of Theorem 5 immediately extends to identification of triplets like $X_j X_k X_\ell U_{jk\ell}$ added to the model, and similarly for all higher order 'tuples up to the product of all K regressors.

7 Conclusions

We have shown nonparametric identification of a generalized random coefficients model, and provided empirical applications in which the generalized random coefficient structure arises from extending existing commonly used economic models of observed heterogeneity to models of unobserved heterogeneity. In our applications to Engel and Barten scales, allowing for unobserved heterogeneity were shown to be extremely important for empirically evaluating the welfare effects of potential policy interventions such as a carbon tax.

In terms of empirical applications, it would be useful to extend the applications to collective household models, e.g., our Engel scale application resembles a model for unobserved variation in resource share allocations, and Barten scales have applied to collective models in, e.g., Browning, Chiappori, and Lewbel (2010). Useful areas for further work

on the theory of generalized random coefficients would be extensions to identify joint rather than marginal distributions of the random coefficients, and to relax the smoothness assumptions that were imposed for identification of the nonadditive model.

8 Proofs

Before proving Theorem 1, we prove a couple of lemmas.

LEMMA 1: Let $\tilde{Y}_k = G_k(X_k U_k)$ where G_k is a strictly monotonically increasing function. Assume $U_k \perp X | Z$. The marginal distributions of U_k and X_k are continuous. The support of X_k includes zero, the support of U_k is a subset of the support of \tilde{Y}_k , and for every r such that $G_k(r)$ is on the support of \tilde{Y}_k there exist an x_k on the support of X_k such that $f_{U_k}(x_k^{-1}r) \neq 0$. Assume the location and scale normalizations $G_k(0) = 0$ and $G_k(1) = y_0$ for some known y_0 in the support of \tilde{Y}_k are imposed. Let $r = H_k(\tilde{y}_k)$ be inverse of the function G_k where $\tilde{y}_k = G_k(r)$. Define $X_{(k)}$ to be the vector of all the elements of X except for X_k . Define the function $S_k(\tilde{y}_k, \tilde{x})$ by

$$S_k(\tilde{y}_k, \tilde{x}) = E \left[F_{\tilde{Y}_k | X_k, X_{(k)}, Z}(\tilde{y}_k | \tilde{x}^{-1}, 0, Z) \right] = \int_{\text{supp}(Z)} F_{\tilde{Y}_k | X_k, X_{(k)}, Z}(\tilde{y}_k | \tilde{x}^{-1}, 0, z) f_z(z) dz.$$

Then

$$H_k(\tilde{y}_k) = \text{sign} \left(\text{sign}(x_k) \frac{\partial S_k(\tilde{y}_k, x_k^{-1})}{\partial x_k^{-1}} \right) \exp \left(\int_{y_0}^{\tilde{y}_k} \frac{x_k \partial S_k(\tilde{y}_k, x_k^{-1}) / \partial \tilde{y}}{\partial S_k(\tilde{y}_k, x_k^{-1}) / \partial x_k^{-1}} d\tilde{y}_k \right) \quad (19)$$

Note that if Z is discretely distributed, then the integral defining S_k becomes a sum. If Z is empty (so U_k and X are unconditionally independent) then $S_k(\tilde{y}_k, \tilde{x}) = F_{\tilde{Y}_k | X_k, X_{(k)}}(\tilde{y}_k | \tilde{x}^{-1}, 0)$. The main implication of Lemma 1 is that if the distribution $F_{\tilde{Y}_k | X, Z}$ is identified, then the function H_k is identified by construction.

PROOF of Lemma 1: For any \tilde{y}_k and any $x_k > 0$ we have

$$\begin{aligned} F_{\tilde{Y}_k | X_k, X_{(k)}, Z}(\tilde{y}_k | x_k, 0, z) &= \Pr(G_k(x_k U_k) \leq \tilde{y}_k | X_k = x_k, X_{(k)} = 0, Z = z) \\ &= \Pr(U_k \leq x_k^{-1} H_k(\tilde{y}_k) | X_k = x_k, X_{(k)} = 0, Z = z) \\ &= F_{U_k | X_k, X_{(k)}, Z}[x_k^{-1} H_k(\tilde{y}_k) | x_k, 0, z] = F_{U_k | Z}[x_k^{-1} H_k(\tilde{y}_k) | z] \end{aligned}$$

where the last equality uses $U_k \perp X \mid Z$. Similarly for any $x_k < 0$ we have

$$\begin{aligned} F_{\tilde{Y}_k | X_k, X_{(k)}, Z}(\tilde{y}_k \mid x_k, 0, z) &= \Pr(G_k(x_k U_k) \leq \tilde{y} \mid X_k = x_k, X_{(k)} = 0, Z = z) \\ &= \Pr(U_k \geq x_k^{-1} H_k(\tilde{y}) \mid X_k = x_k, X_{(k)} = 0, Z = z) \\ &= 1 - F_{U_k | Z}[x_k^{-1} H_k(\tilde{y}) \mid z] \end{aligned}$$

Together these equations say

$$F_{U_k | Z}[x_k^{-1} H_k(\tilde{y}_k) \mid z] = I(x_k < 0) + \text{sign}(x_k) F_{\tilde{Y}_k | X_k, X_{(k)}, Z}(\tilde{y}_k \mid x_k, 0, z).$$

So

$$\begin{aligned} F_{U_k}[x_k^{-1} H_k(\tilde{y}_k)] &= \int_{\text{supp}(Z)} [I(x_k < 0) + \text{sign}(x_k) F_{\tilde{Y}_k | X_k, X_{(k)}, Z}(\tilde{y}_k \mid x_k, 0, z)] f(z) dz. \\ &= I(x_k < 0) + \text{sign}(x_k) S(\tilde{y}_k, x_k^{-1}) \end{aligned}$$

It follows that for any $x_k \neq 0$,

$$\frac{\partial S(\tilde{y}_k, x_k^{-1})}{\partial x_k^{-1}} = \text{sign}(x_k) f_U[x_k^{-1} H_k(\tilde{y}_k)] H_k(\tilde{y}_k)$$

and

$$\frac{\partial S(\tilde{y}_k, x_k^{-1})}{\partial \tilde{y}_k} = \text{sign}(x_k) f_U[x_k^{-1} H_k(\tilde{y}_k)] x_k^{-1} \frac{\partial H_k(\tilde{y}_k)}{\partial \tilde{y}_k}$$

So for $f_U[x_k^{-1} H_k(\tilde{y}_k)] \neq 0$ it follows that

$$\frac{x_k \partial S(\tilde{y}_k, x_k^{-1}) / \partial \tilde{y}_k}{\partial S(\tilde{y}_k, x_k^{-1}) / \partial x_k^{-1}} = \frac{\partial H_k(\tilde{y}_k) / \partial \tilde{y}_k}{H_k(\tilde{y}_k)} = \frac{\partial \ln |H_k(\tilde{y}_k)|}{\partial \tilde{y}_k}$$

so

$$\begin{aligned} \exp\left(\int_{y_0}^{\tilde{y}_k} \frac{x_k \partial S(\tilde{y}_k, x_k^{-1}) / \partial \tilde{y}_k}{\partial S(\tilde{y}_k, x_k^{-1}) / \partial x_k^{-1}} d\tilde{y}_k\right) &= \exp\left(\int_{y_0}^{\tilde{y}_k} \frac{\partial \ln |H_k(\tilde{y}_k)|}{\partial \tilde{y}_k} d\tilde{y}_k\right) \\ &= \exp(\ln |H_k(\tilde{y}_k)| - \ln |H_k(\tilde{y}_0)|) = |H_k(\tilde{y}_k)| \end{aligned}$$

where $H_k(\tilde{y}_0) = 1$ follows from $G_k(1) = \tilde{y}_0$. Finally

$$\begin{aligned} \text{sign} \left(\text{sign}(x_k) \frac{\partial S(\tilde{y}_k, x_k^{-1})}{\partial x_k^{-1}} \right) &= \text{sign} \left(\text{sign}(x_k) \text{sign}(x_k) f_U \left[x_k^{-1} H_k(\tilde{y}_k) \right] H_k(\tilde{y}_k) \right) \\ &= \text{sign} \left(f_U \left[x_k^{-1} H_k(\tilde{y}_k) \mid z \right] H_k(\tilde{y}_k) \right) = \text{sign}(H_k(\tilde{y}_k)) \end{aligned}$$

So the right side of equation (19) equals $\text{sign}(H_k(\tilde{y}_k)) |H_k(\tilde{y}_k)| = H_k(\tilde{y}_k)$ as claimed.

LEMMA 2: If Assumption A1 holds, then $F_{U_0|Z}$ and the distribution function $F_{\tilde{Y}|X,Z}(\tilde{Y} \mid x, z)$ are identified, where $\tilde{Y} = \sum_{k=1}^K G_k(X_k U_k)$.

PROOF of Lemma 2:

$$F_{Y|X,Z}(y \mid 0, z) = \Pr(G(0) + U_0 \leq y \mid X = 0, Z = z) = F_{U_0|X,Z}(y \mid 0, z) = F_{U_0|Z}(y \mid z)$$

identifies the distribution function $F_{U_0|Z}$ on the support of Y , which contains the support of U_0 . Next define $\tilde{Y} = Y - U_0$. Then since $Y = \tilde{Y} + U_0$ and the distributions of $Y \mid X, Z$ and $U_0 \mid X, Z$ are identified, for each value of $X = x, Z = z$ apply a deconvolution (using the nonvanishing characteristic function of U_0) to identify the distribution of $\tilde{Y} \mid X, Z$, where $\tilde{Y} = \sum_{k=1}^K G_k(X_k U_k)$.

PROOF of Theorem 1: When $X_{(k)} = 0$ (equivalently, when $X = e_k x_k$ for some x_k) we get $\tilde{Y} = G_k(X_k U_k) + \sum_{j \neq k} G_j(0) = G_k(X_k U_k)$. Define $\tilde{Y}_k = G_k(X_k U_k)$. It follows that $F_{\tilde{Y}_k|X_k, X_{(k)}, Z}(\tilde{y}_k \mid x_k, 0, z) = F_{\tilde{Y}|X,Z}(\tilde{y}_k \mid x_k e_k, z)$, so the distribution function on the left of this identity is identified, given by Lemma 1 that $F_{\tilde{Y}|X,Z}$ is identified. Let $r = H_k(\tilde{y})$ denote the inverse of the function G_k where $\tilde{y} = G_k(r)$. It follows by construction from Lemma 1 that $H_k(\tilde{y})$ is identified for every value of \tilde{y}_k on the support of \tilde{Y}_k satisfying the property that, for some x_k on the support of X_k , $f_{U_k} \left[x_k^{-1} H(\tilde{y}) \right] \neq 0$. This identification of $H_k(\tilde{y})$ in turn means that the function $G_k(r)$ is identified for every r such that $G_k(r)$ is on the support of \tilde{Y}_k and there exist an x_k on the support of X_k such that $f_{U_k|Z} \left(x_k^{-1} r \right) \neq 0$. This then implies identification of G_k on its support. Finally, given identification of $F_{\tilde{Y}|X,Z}$ and of $H_k(\tilde{y}_k)$, the distribution function $F_{U_k|Z}$ is identified by $F_{U_k|Z} \left[H(\tilde{y}) / x_k \mid z \right] = F_{\tilde{Y}|X_k, X_{(k)}, Z}(\tilde{y} \mid x_k, 0, z)$ for $x_k > 0$ and $F_{U_k|Z} \left[H(\tilde{y}) / x_k \mid z \right] = 1 - F_{\tilde{Y}|X_k, X_{(k)}, Z}(\tilde{y} \mid x_k, 0, z)$ for $x_k < 0$.

PROOF of Corollary 1: Applying the proof of Lemma 2 to the model of Corollary 1 shows that $F_{U_0|Z}$ and the distribution function $F_{\tilde{Y}|X,Z}(\tilde{Y} \mid x, z)$ are identified, where

$\tilde{Y} = \prod_{k=1}^K g_k(X_k U_k)$. It therefore follows that $F_{\tilde{Y}|X,Z}(\tilde{Y} | x, z)$ is identified where $\tilde{Y} = \ln(\tilde{Y}) = \sum_{k=1}^K \ln[g_k(X_k U_k)] = \sum_{k=1}^K G_k(X_k U_k)$, and the remainder of the identification therefore follows applying the proof of Theorem 1.

PROOF of Theorem 2: By construction, the function $\tilde{G}(X_1 U_1, \dots, X_K U_K)$ is zero when evaluated at $X = 0$ or at $X = X_k e_k$ for any k , so evaluated at any such value of X , equation (3) is equivalent to equation (2). For equation (2), the proof of Theorem 1 showed identification of the marginal distributions of each U_k and each function G_k only using $X = 0$ and $X = X_k e_k$, so these functions are also identified for equation (3). What remains is to identify the function \tilde{G} . Define $R(X) = E[\tilde{G}(X_1 U_1, \dots, X_K U_K) | X]$. The function $R(X)$ is identified for all X because $R(X) = E\left[Y - \sum_{k=1}^K G_k(X_k U_k) - U_0 | X\right]$, which depends only on the already identified distributions and functions. For nonnegative integers t_1, \dots, t_K define R_{t_1, \dots, t_K} by

$$R_{t_1, \dots, t_K}(x) = \frac{\partial^{t_1 + \dots + t_K} R(x)}{\partial x_1^{t_1} \dots \partial x_K^{t_K}}$$

and similarly for $\tilde{G}_{t_1, \dots, t_K}$. Then

$$R_{t_1, \dots, t_K}(x) = E(U_1^{t_1}, \dots, U_K^{t_K} \tilde{G}_{t_1, \dots, t_K}(x_1 U_1, \dots, x_K U_K))$$

so $\tilde{G}_{t_1, \dots, t_K}(0) = R_{t_1, \dots, t_K}(x) / E(U_1^{t_1}, \dots, U_K^{t_K})$ is identified for all sets of nonnegative integers t_1, \dots, t_K . Now \tilde{G} is analytic so can write the Maclaurin series

$$\tilde{G}_{jk}(r) = \sum_{t_1=0}^{\infty} \dots \sum_{t_K=0}^{\infty} \frac{r_1^{t_1} \dots r_K^{t_K} \tilde{G}_{t_1, \dots, t_K}(0)}{(t_1 + \dots + t_K)}$$

which shows that the function $\tilde{G}(r)$ is identified, since $\tilde{G}_{t_1, \dots, t_K}(0)$ is identified for all sets of nonnegative integers t_1, \dots, t_K .

PROOF of Corollary 2: For a given $j \in \{1, \dots, J\}$ let $Y = -Q_j$ if j is not a Giffen good, otherwise let $Y = Q_j$. Then the function G in Theorem 2 is given by $-\omega_j(U_1 X_1, \dots, U_J X_J) / U_j X_j$ which makes the function and G_j in Theorems 1 and 2 be $-\omega_j(0, \dots, 0, U_j X_j, \dots, 0) / U_j X_j$ (remove the minus signs if the good j was Giffen). Then G_j is strictly monotonically increasing, and we have taken $U_0 = 0$, so by Theorem 1, the distribution function $F_{U_j|Z}$ is identified. Repeating this procedure for each $j \in \{1, \dots, J\}$ identifies all of the $F_{U_j|Z}$ distributions. Given identification of all of the

$F_{U_j|Z}$ distributions, we can now apply the remainder of the proofs of Theorems 1 and 2 to each demand function $Q_j X_j = \omega_j (U_1 X_1, \dots, U_J X_J)$ for $j \in \{1, \dots, J\}$ to identify each function ω_j , observing that by Roys identity (and boundedness of budget shares) each ω_j will be analytic, and having each U_j be positive and bounded makes the remaining assumptions of Theorem 2 hold.

PROOF of Theorem 3: As discussed in the text, a property of Barten scales (which can be readily verified using Roys identity) is that, if $V (X_1, X_2)$ is the indirect utility function corresponding to the demand function $\omega_1 (X_1, X_2)$, then up to an arbitrary monotonic transformation $H (V, U_1, U_2)$ of V , the indirect utility function corresponding to $\omega_1 (U_1 X_1, U_2 X_2)$ is $V (U_1 X_1, U_2 X_2)$, and vice versa. It therefore suffices to prove that the theorem holds with $U_1 = U_2 = 1$.

By equation (4), given any indirect utility function V , the corresponding demand function ω_1 is given by

$$\omega_1 (X_1, X_2) = \frac{\partial V (X_1, X_2) / \partial \ln X_1}{\left[\partial V (X_1, X_2) / \partial \ln X_1 \right] + \left[\partial V (X_1, X_2) / \partial \ln X_2 \right]}$$

Similarly, given any demand function ω_1 , if this equation holds then V equals, up to an arbitrary monotonic transformation, the indirect utility function that corresponds to ω_1 . It follows that

$$\lambda [\omega_1 (X_1, X_2)] = \ln \left(\frac{\partial V (X_1, X_2)}{\partial \ln X_1} \right) - \ln \left(\frac{\partial V (X_1, X_2)}{\partial \ln X_2} \right) \quad (20)$$

Given any functions $g_1 (X_1)$ and $g_2 (X_2)$, define a corresponding function $V (X_1, X_2)$ by

$$V (X_1, X_2) = \int_{-\infty}^{\ln X_1} e^{g_1(X_1)} d \ln X_1 + \int_{-\infty}^{\ln X_2} e^{-g_2(X_2)} d \ln X_2. \quad (21)$$

Substituting equation (21) into equation (20) gives

$$\lambda [\omega_1 (X_1, X_2)] = g_1 (X_1) + g_2 (X_2) \quad (22)$$

which shows that, up to monotonic transformation, equation (21) is the indirect utility function that generates the demand equation (22). Since equation (21) is additive, this shows that the indirect utility function that generates the demand equation (22) is additive.

To go the other direction, given any differentiable functions $h_1 (X_1)$ and $h_2 (X_2)$, if $V (X_1, X_2) = h_1 (X_1) + h_2 (X_2)$ equation (20) equals

$$\lambda [\omega_1 (X_1, X_2)] = \ln \left(\frac{\partial h_1 (X_1)}{\partial \ln X_1} \right) - \ln \left(\frac{\partial h_2 (X_2)}{\partial \ln X_2} \right) \quad (23)$$

which is in the form of equation (22), showing that any additive indirect utility function generates a demand equation in the form of (22).

Together these results prove the first part Theorem 3. Adding back the Barten scales U_1 and U_2 to the functions $g_1, g_2, h_1,$ and h_2 proves equations (9) and (8). The properties of the functions h_1 and h_2 given at the end of Theorem 2 follow from the fact that the indirect utility function $h_1 (U_1 P_1/M) + h_2 (U_2 P_2/M)$ must possess the standard properties of all indirect utility functions, i.e., homogeneity and quasiconvexity in $P_1, P_2,$ and $M,$ nondecreasing in each price, and increasing in $M.$

PROOF of Theorem 4: When $X_{(k)} = 0$ we get $\tilde{Y} = G_k (X_k U_k) + \sum_{j \neq k} G_k (0) = G_k (X_k U_k).$ Define $\tilde{Y}_k = G_k (X_k U_k).$ It follows that $F_{\tilde{Y}_k | X_k, X_{(k)}, Z} (\tilde{y}_k | x_k, 0, z) = F_{\tilde{Y} | X, Z} (\tilde{y}_k | x_k e_k, z),$ so $F_{\tilde{Y}_k | X_k, X_{(k)}, Z} (\tilde{y}_k | x_k, 0, z)$ is identified, given by Lemma 1 that $F_{\tilde{Y} | X, Z}$ is identified. Let $r = H_k (\tilde{y}_k)$ be inverse of the function G_k where $\tilde{y}_k = G_k (r).$ Now consider any particular positive $x_k \in \Psi_k.$ For that x_k we have $F_{\tilde{Y}_k | X, Z} (y_0 | e_k, z) = F_{\tilde{Y}_k | X, Z} (\tilde{y}_k | e_k x_k, z)$ and since the function $F_{\tilde{Y}_k | X, Z}$ is identified, the particular value \tilde{y}_k that satisfies this equation is identified. Then

$$\begin{aligned} \Pr (G_k (x_k U_k) \leq \tilde{y}_k | X = x_k e_k, Z = z) &= \Pr (G_k (U_k) \leq y_0 | X = x_k e_k, Z = z) \\ &= \Pr (G_k (U_k) \leq y_0 | Z = z) \\ F_{U_k | Z} [H_k (\tilde{y}_k) / x_k, z] &= F_{U_k | Z} [H_k (y_0), z] \end{aligned}$$

similarly, if we have a given negative $x_k \in \Psi_k$ then

$$\begin{aligned} 1 - \Pr (G_k (x_k U_k) \leq \tilde{y}_k | X = x_k e_k, Z = z) &= \Pr (G_k (U_k) \leq y_0 | X = x_k e_k, Z = z) \\ 1 - \Pr (U_k \geq H_k (\tilde{y}_k) / x_k | X = x_k e_k, Z = z) &= \Pr (U_k \leq H_k (y_0) | Z = z) \\ F_{U_k | Z} [H_k (\tilde{y}_k) / x_k, z] &= F_{U_k | Z} [H_k (y_0), z] \end{aligned}$$

By invertibility of $F_{U_k | Z}$ these equations show that for any $x_k \in \Psi_K$ we get $H_k (\tilde{y}_k) / x_k = H_k (y_0)$ where the \tilde{y}_k corresponding to the given x_k is known. Now $G_k (1) = y_0$ means that $H_k (y_0) = 1,$ so $H_k (\tilde{y}_k) = x_k,$ and therefore $\tilde{y}_k = G_k (x_k),$ so the value of the function G_k evaluated at this particular x_k is known. This holds for any and hence all $x_k \in \Psi_k,$ so by Assumption A2' this suffices to identify the function G_k everywhere, and hence also identifies the function H_k everywhere.

Given identification of $F_{\tilde{Y} | X, Z}$ and of $H_k (\tilde{y}),$ the distribution function $F_{U_k | Z}$ is identified by $F_{U_k | Z} [H (\tilde{y}_k) / x_k | z] = F_{\tilde{Y}_k | X, Z} (\tilde{y}_k | e_k x_k, z)$ for $x_k > 0$ and $F_{U_k | Z} [H (\tilde{y}_k) / x_k | z] = 1 - F_{\tilde{Y}_k | X, Z} (\tilde{y}_k | e_k x_k, z)$ for $x_k < 0.$

PROOF Theorem 5: First observe that Lemma 2 still holds in this model, identifying $F_{U_0 | Z}$ by taking $X = 0.$ Similarly, all the interaction terms $X_j X_k$ equal zero when $X =$

$e_k x_k$ for any k , so the proof of Theorem 1 goes through to identify each $F_{U_k|Z}$ and G_k function. Next, for each j, k pair evaluate the model at $X = e_{jk}$ to get $Y = V_{jk} + U_{jk}$ where $V_{jk} = U_0 + G_k(U_k) + G_j(U_j)$. At this stage the distribution of $V_{jk} | Z$ is identified (because each component is identified), so $F_{U_{jk}|Z}$ can be identified by a deconvolution of $Y | Z$ with $V_{jk} | Z$.

Altonji, J. G. and Matzkin, R. L. (2005), "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica*, 73, 1053–1102.

Banks, Blundell, and Lewbel (1996), "Tax Reform and Welfare Measurement: Do We Need Demand System Estimation?" *Economic Journal*, 106, 1227-1241.

Barten, A. P. (1964), "Family Composition, Prices, and Expenditure Patterns," in *Econometric Analysis for National Economic Planning: 16th Symposium of the Colston Society*, P. Hart, G. Mills, and J. K. Whitaker, eds., 277-292, London: Butterworth.

Beckert, W. (2006), "Specification and Identification of Stochastic Demand Models," *Econometric Reviews*, 26, 669–683.

Beckert, W. and R. Blundell (2008), "Heterogeneity and the Non-Parametric Analysis of Consumer Choice: Conditions for Invertibility," *Review of Economic Studies*, 75, 1069–1080.

Berry, S. and P. Haile (2009), "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers," NBER Working Paper No. 15276.

Berry, S., J. Levinsohn, and A. Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 60, 889–917.

Beran, Feuerverger, and Hall (1996), "On Nonparametric Estimation of Intercept and Slope Distributions in Random Coefficient Regression," *Annals of Statistics*, 24, 2569-2692.

Beran, R. and Hall, P. (1992), "Estimating Coefficient Distributions in Random Coefficient Regressions," *Annals of Statistics* 20, 1970-1984.

Blackorby, C., D. Primont, and R. Russell, (1978) *Duality, Separability, and Functional Structure: Theory and Economic Applications*. North-Holland.

Blundell, R., and J. Powell (2003) "Endogeneity in Nonparametric and Semiparametric Regression Models" In *Advances in Economics and Econometrics*, eds. M. Dewatripont, L. Hansen, and S. Turnovsky, 294-311, Cambridge: Cambridge University Press.

Blundell, R., and J. Powell (2004) "Endogeneity in Semiparametric Binary Response Models" *The Review of Economic Studies* 71, 581-913.

Blundell, R., Horowitz J., and M. Patey, (2010), "Measuring the Price Responsiveness of Gasoline Demand, Economic Shape Restrictions and Nonparametric Demand Estimation," CeMMAP working paper CWP11/09.

Brown, B.W. and M.B. Walker (1989), "The Random Utility Hypothesis and Inferences in Demand Systems", *Econometrica*, 57, 815-829.

Browning, M. and J. Carro (2007), "Heterogeneity and Microeconometrics Modeling," in R. Blundell, W. Newey and T. Persson, eds., *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*, vol. 3, ch. 3, *Econometric Society Monographs*, 43 Cambridge: Cambridge University Press, 46-74.

Browning, M., P.-A. Chiappori, and A. Lewbel (2010), "Estimating Consumption Economies of Scale, Adult Equivalence Scales, and Household Bargaining power," Boston College Working Paper #588.

Chen, X. (2007), "Large Sample Sieve Estimation of Semi-Nonparametric Models," in J. Heckman and E. Leamer eds. *Handbook of Econometrics*, vol. 6B, Chapter 76, 5549-5632. New York: Elsevier Science.

Chesher, A. (2003), "Identification in Nonseparable Models," *Econometrica*, 71, 1405–1441.

Comon, E. and L. Calvet (2003), "Behavioral Heterogeneity and the Income Effect," *Review of Economics and Statistics*, 85, 653-669.

Engel, E. (1895), "Die Lebenskosten Belgischer Arbeiter-Familien Fruher and jetzt," *International Statistical Institute Bulletin*, vol. 9, pp. 1-74.

Gautier, E. and Y. Kitamura (2010), "Nonparametric estimation in random coefficients binary choice models," Unpublished manuscript.

Hausman, J. A. (1981), "Exact Consumer's Surplus and Deadweight Loss," *The American Economic Review*, 71, 662-676.

Hausman, J. A. and W. K. Newey, (1995), "Nonparametric Estimation of Exact Consumer's Surplus and Deadweight Loss," *Econometrica* 63, 1445-1476.

Heckman, J. J., and Singer, B. (1984) "A method for minimizing the impact of distributional assumptions in econometric models for duration data," *Econometrica*, 52, 271–320.

Hildenbrand, W. (1994) *Market Demand: Theory and Empirical Evidence*, Princeton: Princeton University Press.

Hoderlein, S., Klemelae, J. and Mammen, E. (2010), "Analyzing the Random Coefficient Model Nonparametrically," *Econometric Theory*, 26, 804-837.

Hoderlein, S. and Mammen, E. (2007), "Identification of Marginal Effects in Nonseparable Models Without Monotonicity," *Econometrica*, 75, 1513–1518.

Hoderlein, S. and S. Mihaleva (2008), "Increasing the price variation in a repeated cross section," *Journal of Econometrics*, 147, 316-25.

Hoderlein, S, L. Nesheim and A. Simoni, (2011), "Semiparametric Estimation of Random Coefficients in Structural Economic Models," Unpublished Manuscript.

Hoderlein, S. and A. Vanhems (2010), "Welfare Analysis using Nonseparable Models," Unpublished Manuscript, Boston College.

Hoderlein, S. and A. Vanhems (2011), "Welfare Analysis using Random Coefficient Models," incomplete manuscript, Boston College.

Ichimura, H. Thompson, T.S. (1998), Maximum Likelihood estimation of a binary choice model with random coefficients of unknown distribution, *J. of Econometrics* 86, 269-295.

Imbens, G. W., and W. K. Newey (2009), "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77, 1481–1512.

Jensen, R. and N. Miller (2008), "Giffen Behavior and Subsistence Consumption," *American Economic Review*, 98, 1553-1577.

Jorgenson, D. W., L. J. Lau, and T. M. Stoker (1982), "The Transcendental Logarithmic Model of Aggregate Consumer Behavior," in: *Advances in Econometrics*, R. Basman and G. Rhodes, Eds., Greenwich: JAI Press.

Lewbel, A., (1989), "Identification and estimation of equivalence scales under weak separability," *Review of Economic Studies* 56, 311-316.

Lewbel, A. (1996), "Aggregation Without Separability: A Generalized Composite Commodity Theorem," *American Economic Review*, 86, 524-543.

Lewbel, A. (1997), "Consumer Demand Systems and Household Equivalence Scales," *Handbook of Applied Econometrics, Volume II: Microeconomics*, M. H. Pesaran and P. Schmidt, eds., Oxford: Blackwell Publishers Ltd.

Lewbel, A. (2001), "Demand Systems with and without Errors," *American Economic Review*, 91, 611-618.

Lewbel, A. (2007), "Modeling Heterogeneity," in R. Blundell, W. Newey and T. Persson, eds., *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*, vol. 3, ch. 5, *Econometric Society Monographs*, 43 Cambridge: Cambridge University Press, 111-119.

Lewbel, A. (2008), "Engel Curves," entry for *The New Palgrave Dictionary of Economics*, 2nd Edition, MacMillan Press.

Lewbel, A. and K. Pendakur (2009), "Tricks With Hicks: The EASI Demand System," *American Economic Review*, 99, 827-863.

Lewbel, A. and M. De Nadai (2011), "Nonparametric Errors in Variables Models with Measurement Errors on both sides of the Equation," unpublished working paper.

Linton, O. (2000), "Efficient estimation of generalized additive nonparametric regression models," *Econometric Theory*, 16, 502-523.

McFadden, D., (1974), "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, ed., *Frontiers in Econometrics*, 105-142, Academic Press: New York.

Matzkin, R. L. (2003), "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 1339–1375.

Matzkin, R. L. (2007a), Nonparametric Identification, in: J. Heckman and E. Leamer,

Eds, Handbook of Econometrics, vol. 6B, chap. 73, 5307-5368, New York: Elsevier Science.

Matzkin, R. L. (2007b), "Heterogeneous Choice," in R. Blundell, W. Newey and T. Persson, eds., Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society, vol. 3, ch. 4, Econometric Society Monographs, 43 Cambridge: Cambridge University Press, 75–110.

Matzkin, R. L. (2008) "Identification in Nonparametric Simultaneous Equations," *Econometrica*, 76, 945–978.

Stern, N. H, (1987), "The Theory of Optimal Commodity and Income Taxation: An Introduction," in D. M. G. Newbery and N. H. Stern, eds., *The Theory of Taxation for Developing Countries*, Oxford: Oxford University Press.

Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*, Chapman & Hall/CRC.

Working, H. (1943), "Statistical Laws of Family Expenditures," *Journal of the American Statistical Association*, 38, 43-56.

Figure 1

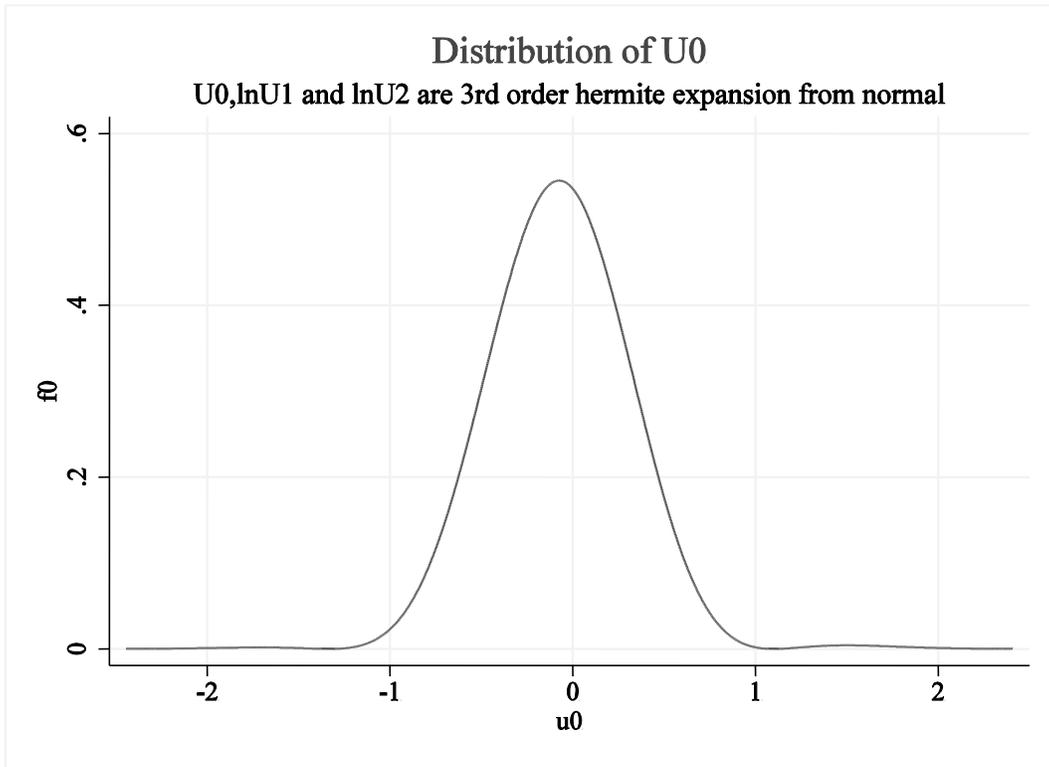


Figure 2

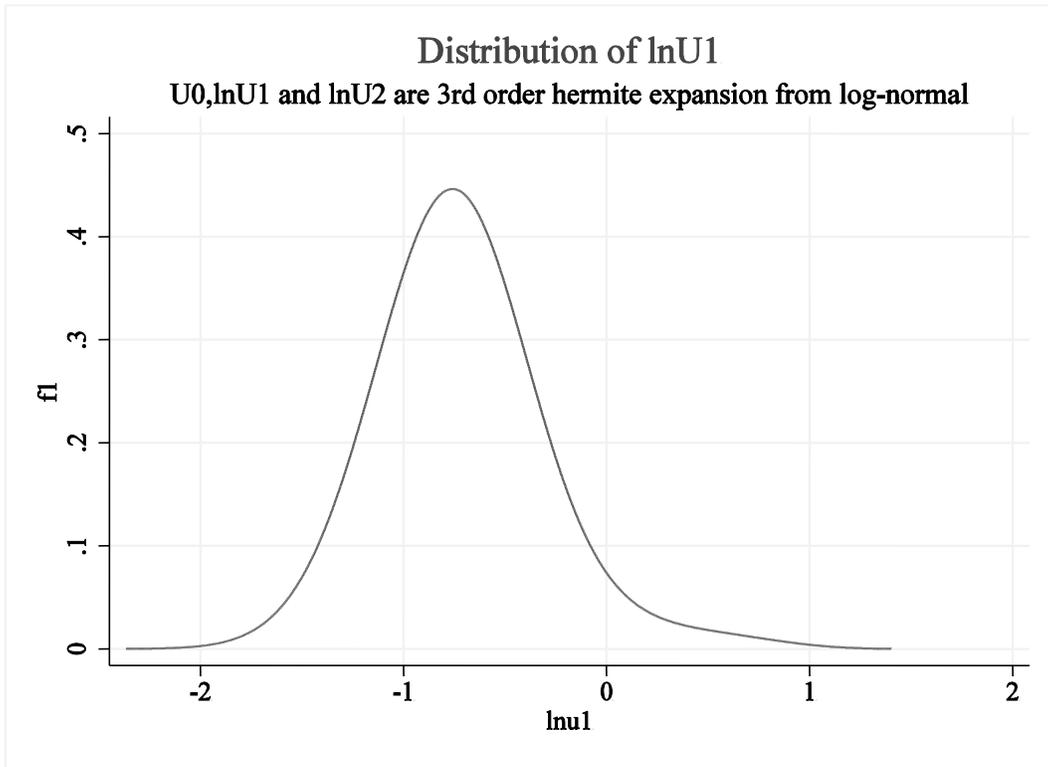
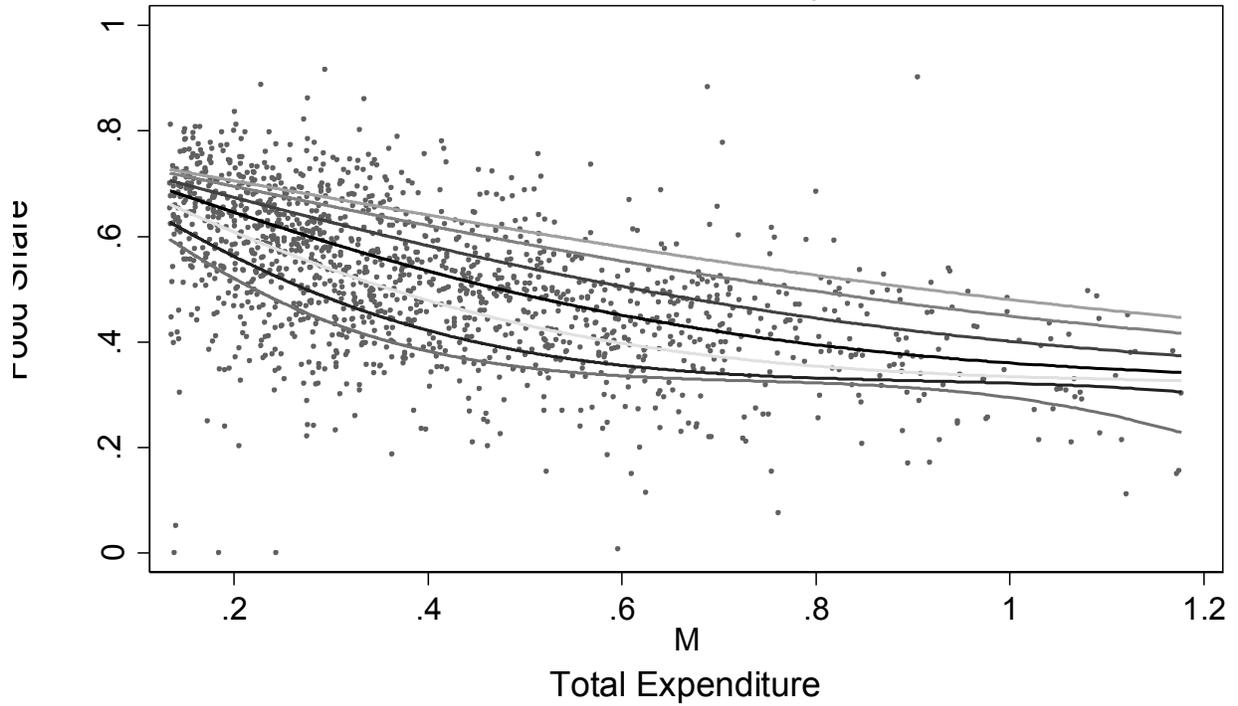


Figure 3

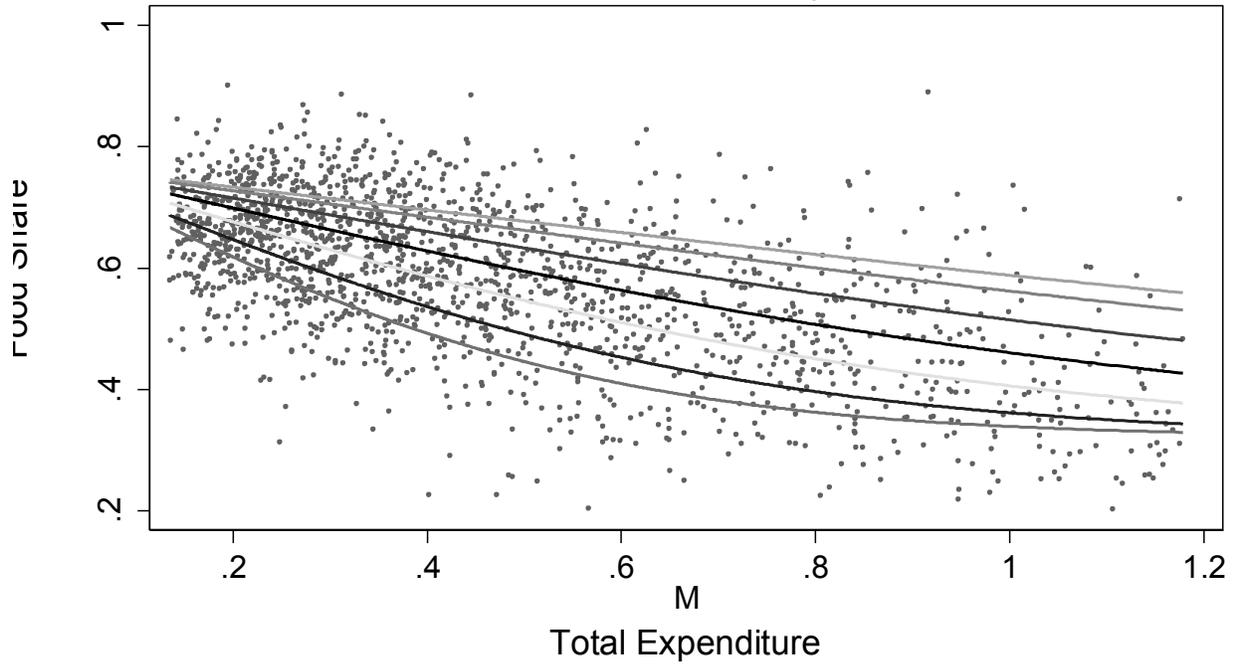
Single-Adult (Reference) Households, India, 2003
5th,10th,25th,50th,75th,90th,95th percentiles of U1



logit(Foodshare)=3rd order poly in M; U0,U1 are 3rd order Hermite expansions

Figure 4

Two-Adult Households, India, 2003
5th,10th,25th,50th,75th,90th,95th percentiles of U1



logit(Foodshare)=3rd order poly in M; U0,U1 are 3rd order Hermite expansions
Eqscale=0.57 (0.02)

Figure 5

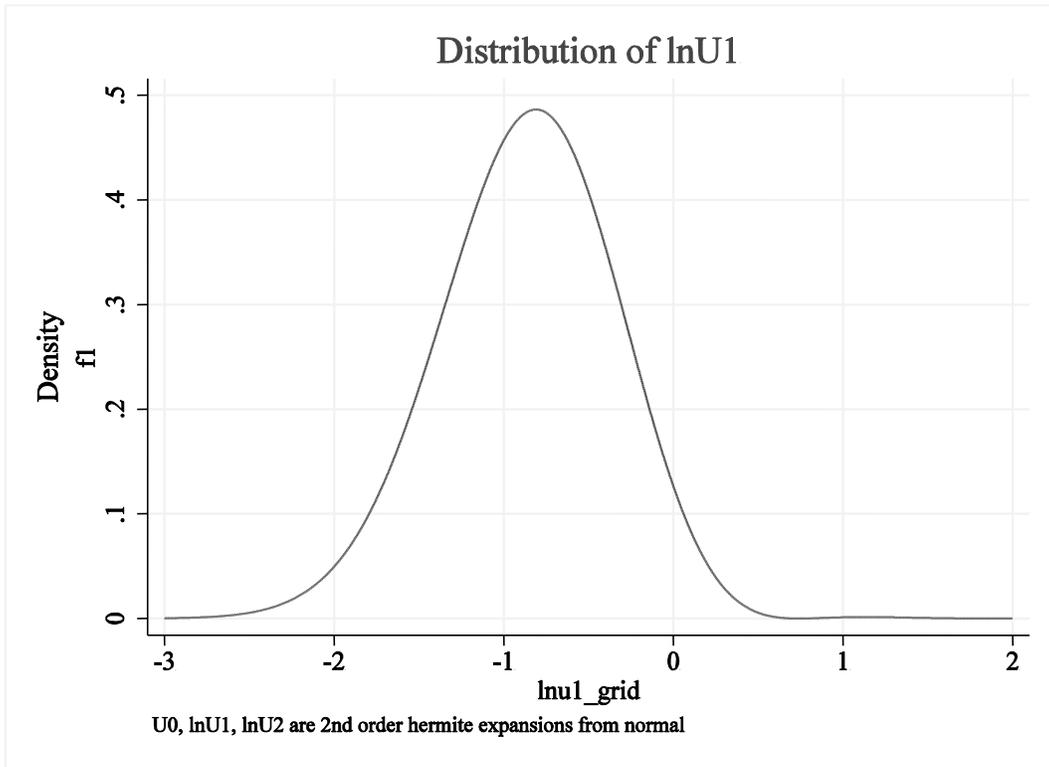


Figure 6

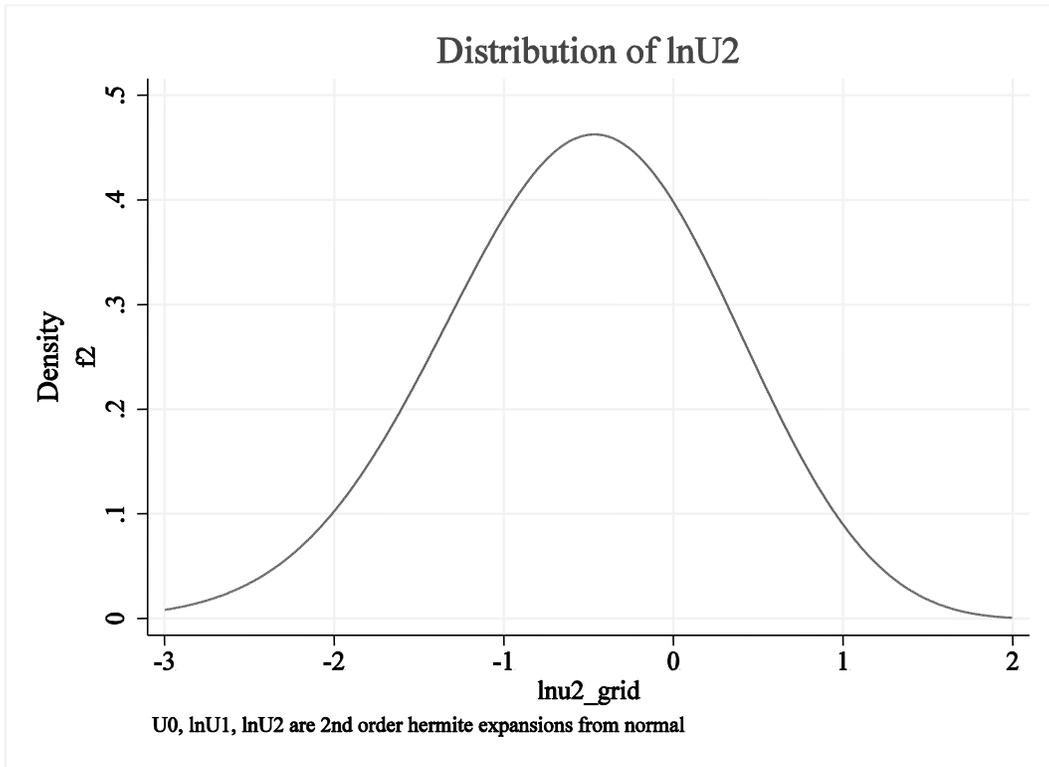


Figure 7

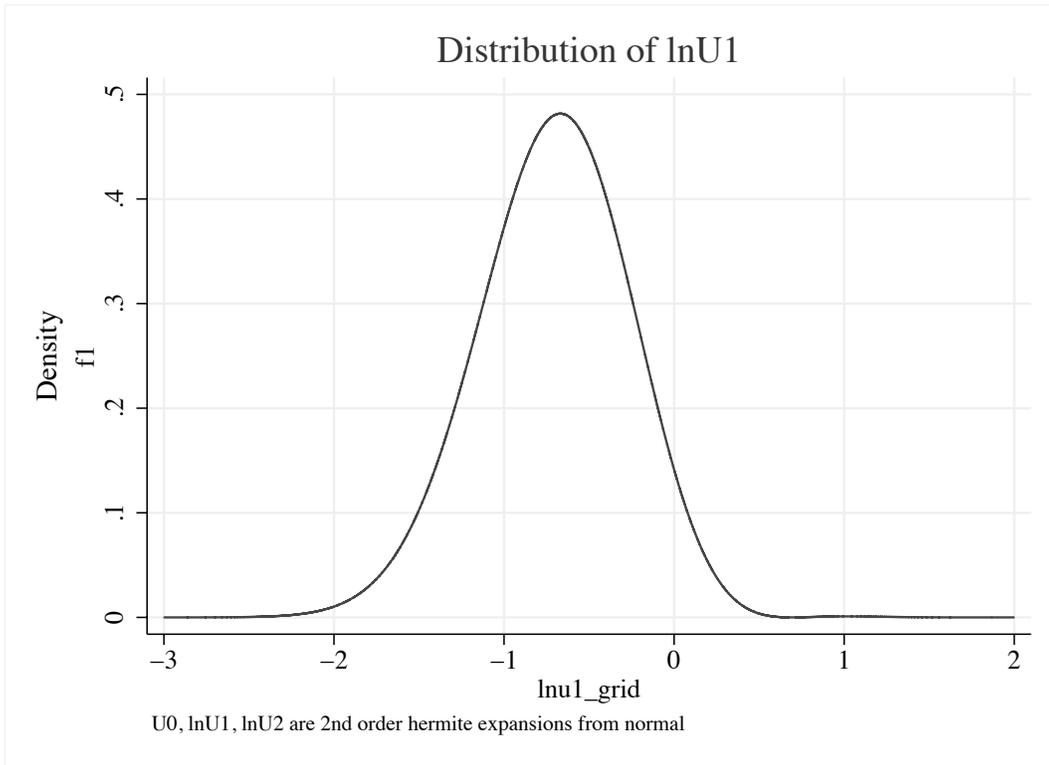


Figure 8

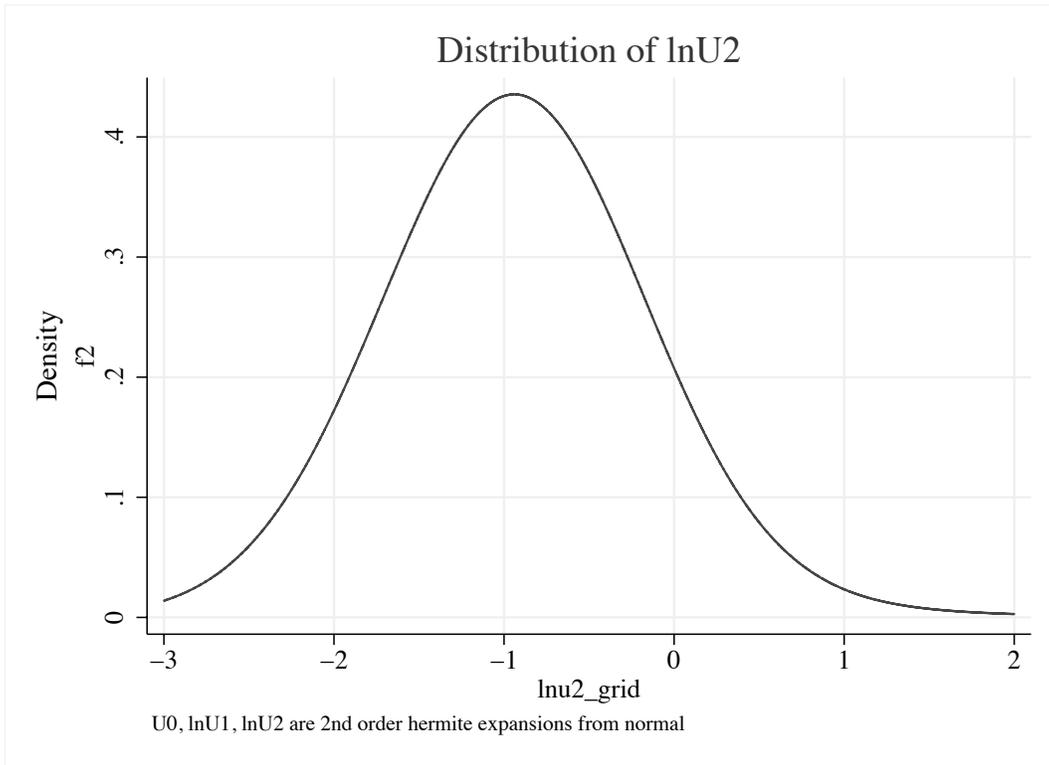


Figure 9

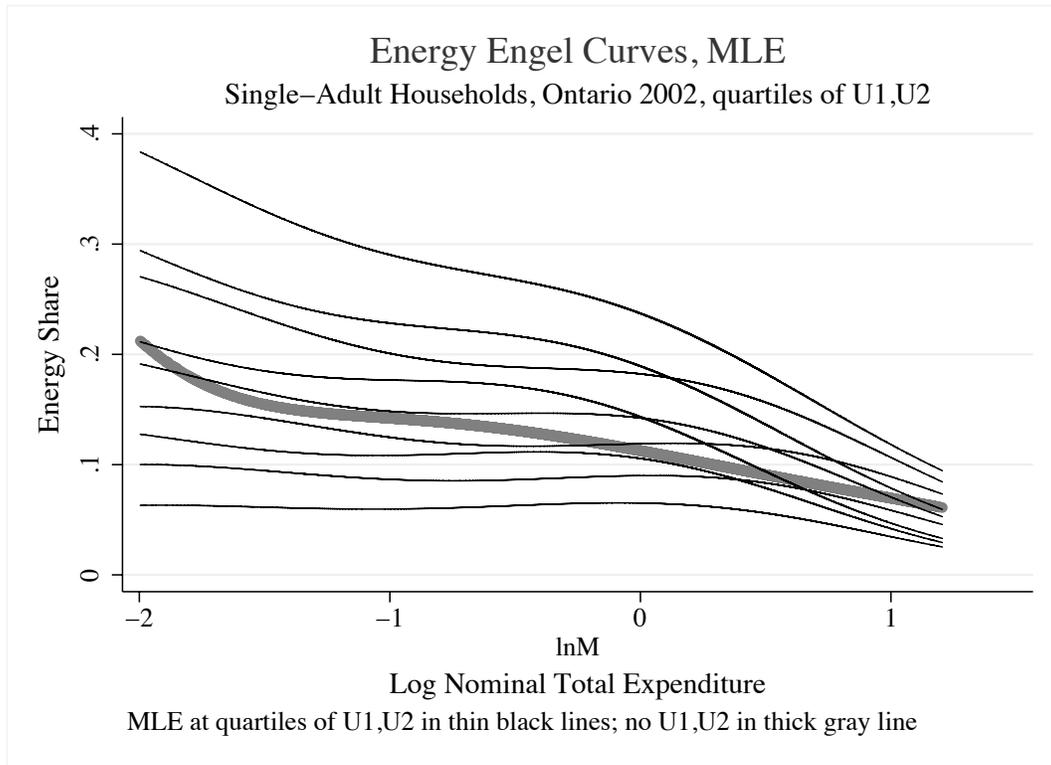


Figure 10

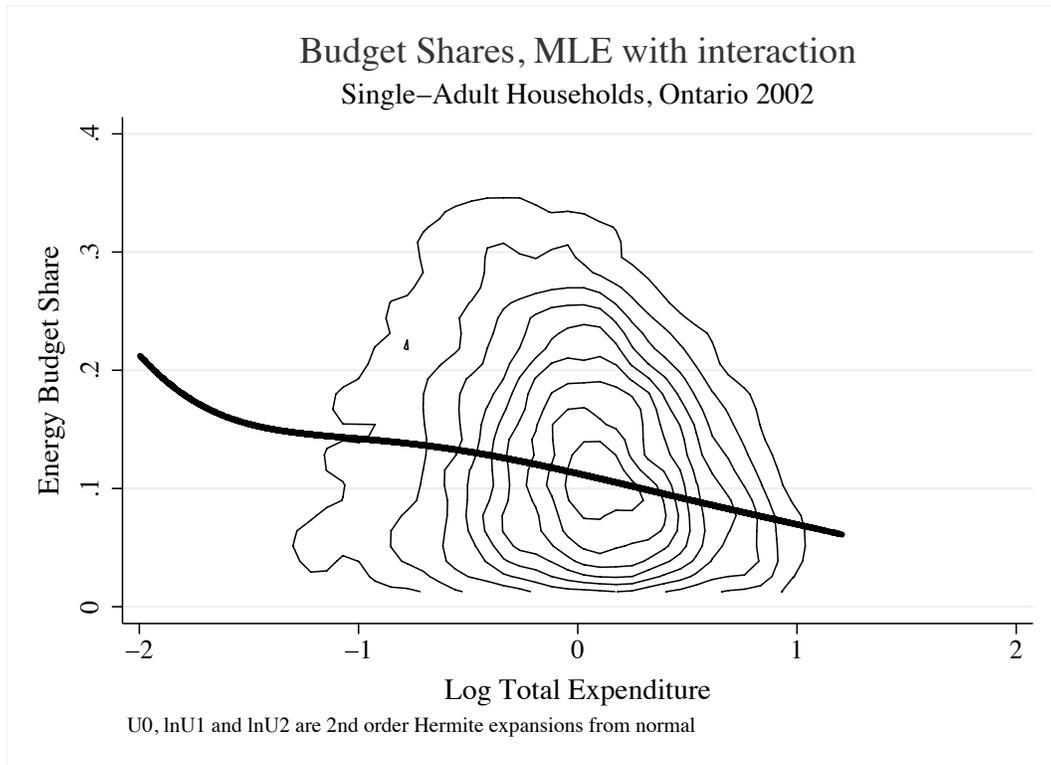


Figure 11

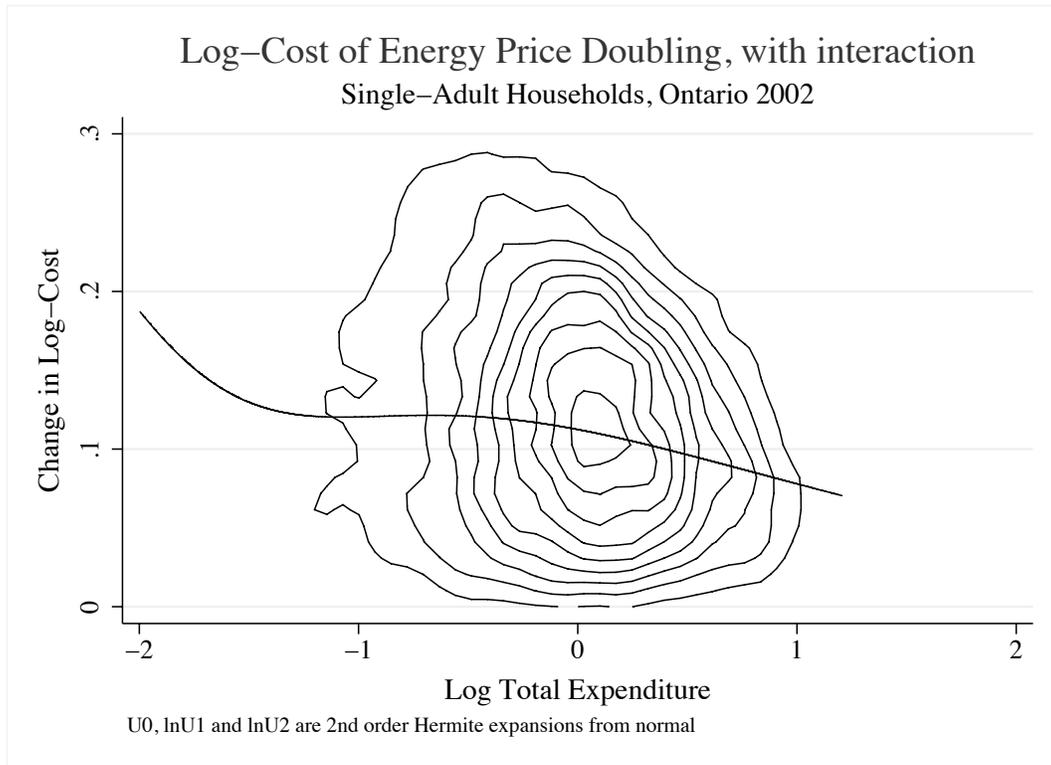


Figure 12

