Unobserved Preference Heterogeneity in Demand Using Generalized Random Coefficients

Arthur Lewbel and Krishna Pendakur
Boston College and Simon Fraser University

original Sept. 2011, revised April 2013

Abstract

We model unobserved preference heterogeneity in demand systems via random Barten scales in utility functions. These Barten scales appear as random coefficients multiplying prices in demand functions. Consumer demands are nonlinear in prices and may have unknown functional structure. We therefore prove identification of Generalized Random Coefficients models, defined as nonparametric regressions where each regressor is multiplied by an unobserved random coefficient having an unknown distribution. Using Canadian data, we estimate energy demand functions with and without random coefficient Barten scales. We find that not accounting for this unobserved preference heterogeneity substantially biases estimated consumer-surplus costs of an energy tax.

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

For discretely demanded goods, unobserved preference heterogeneity is typically modeled using random coefficients, as in Berry, Levinsohn, and Pakes (BLP 1995). Allowing for substantial unobserved random preference heterogeneity, as BLP does, has proven to be necessary for realistic evaluations of the impacts of price changes on demand. In this paper we propose an analogous way to introduce unobserved preference heterogeneity in continuous demand systems.

The application we consider is energy demand by consumers. Energy is consumed in continuous quantities and displays substantial nonlinearities in income and price effects. Therefore,

energy cannot be appropriately modeled using discrete demand methods like BLP, and instead requires the methodology of continuous demand systems.

We demonstrate the importance of accounting for random coefficient type unobserved preference heterogeneity in energy demand. In particular, we show that failure to do so results in a dramatic underestimate of the variance of impacts of energy price changes across consumers. Accounting for this variation is crucial for correctly assessing the true costs to society of energy policies such as a carbon tax. We show that measures of social welfare that ignore this unobserved preference heterogeneity yield substantially biased estimates of the full costs to society of an energy tax on consumers, by failing to fully account for the tax's distributional impacts.

One of the most commonly used methods for incorporating *observable* sources of preference heterogeneity (such as the impacts of age or family size) in continuous demand systems is via Barten (1964) scales. Barten scales deflate the prices faced by consumers, and so have a structure that is analogous to random coefficients on prices, in that they multiply each price in the demand system. This suggests that a natural way to introduce unobserved preference heterogeneity into continuous demand systems is to allow random variation in the Barten scales via random coefficients on prices.

However, randomly varying Barten scales introduces a substantial econometric difficulty because, unlike discrete demand models such as multinomial logit, realistic continuous demand models are highly nonlinear in prices, due to constraints such as homogeneity and Slutsky symmetry. We therefore require a general type of random coefficients that can be identified and estimated in nonlinear, or even nonparametrically specified, demand functions. We define "generalized random coefficients" to be random coefficients applied to variables in a general nonlinear or nonparametric model, in contrast to ordinary random coefficients that are applied in linear index models.

In this paper we first provide some identification theorems, showing that the distributions of random coefficients can be nonparametrically identified in nonparametric regression models. Our results show identification under relatively weaker assumptions for additive models, and under stronger conditions for general nonparametric regression models.

We then apply these results to identification of random Barten scales in demand systems. This application includes proving a new theorem that nonparametrically characterizes the preferences associated with demand functions having a certain additive structure. This result is relevant because it allows us to exploit the weaker regularity conditions required for nonparametric identification of generalized random Barten coefficients in additive models.

Based on these identification theorems, we estimate energy demand functions for a set of Canadian consumers. To illustrate the importance of allowing for unobserved heterogeneity in Barten scales, we evaluate the (partial equilibrium) impacts of a hypothetical tax on energy goods, like a carbon tax. Among other results, we find that allowing for unobserved preference heterogeneity has a large impact on the estimated distribution of the relative costs (consumer surplus impacts) of the tax. For example, we find that this distribution across consumers has a standard deviation that is six times larger in our model than it is in a model that does not allow for such unobserved preference heterogeneity.

Consider first our proposed generalization of random coefficients models. Suppose an observed variable Y depends on a vector of observed regressors $X = (X_1, ..., X_K)$, and on a vector of unobserved errors $U_0, U_1, ..., U_K$. These errors represent unobserved heterogeneity

2

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

$$Y = G\left(X_1 U_1, ..., X_K U_K\right) + U_0 \tag{1}$$

for some unknown function G. In our empirical application, Y will be a measure of energy demanded by a consumer, G will be a Marshallian demand function, each X_k will be the price of a good k divided by a consumer's total expenditures, and each U_k (other than U_0) will be a Barten scale. All previous empirical implementations of Barten scales have exactly the form of equation (1), but with every U_k other than U_0 specified as deterministic functions of observable characteristics that affect preferences such as age or family size. In contrast, we show that the function G and the distribution functions of each random coefficient U_k can be nonparametrically identified, under low level regularity conditions.

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

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

where the functions $G_1, ..., G_K$ are unknown. For these additive models the identifying assumptions are less restrictive. We then give stronger conditions for identifying the nonadditive components of equation (1). In an econometric appendix 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), and we discuss identification when some X_k components are discretely distributed.

Imposing equation (2) type additivity directly on Marshallian demand functions yields some implausible restrictions on preferences. However, we show that, when $K = 2$, these restrictions can be relaxed by suitably transforming Y. In particular, we prove a theorem showing that when $K = 2$, if Y is defined as a logit transformed budget share, then demands will take the additive form implied by equation (2) if and only if indirect utility has a correspondingly additive form. This theorem also provides closed form expressions for the indirect utility function corresponding to nonparametrically specified demand functions that are additive in this way. These closed form expressions greatly simplify our later consumer surplus and welfare calculations.

We first provide a literature review bearing on the econometric identification of models containing random coefficients and on the modeling of preference heterogeneity in continuous demand systems. We then present our main identification theorems, followed by our theorem characterizing the nonparametric connection between preferences and logit transformed demands. We next provide our empirical implementation of the random Barten scales model, including consumer surplus calculations on the hypothetical impacts of large increase in the price of, or taxes on, energy goods. We then conclude, and in an appendix we provide proofs, and some extensions of our identification theorems.

2 Literature Review

We use generalized random coefficients to 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., Engel (1895), Sydenstricker and King (1921), Rothbarth (1943), Prais and Houthakker (1955), Barten (1964), Pollak and Wales (1981), Jorgenson, Lau, and Stoker (1982), and Ray (1992), and see 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 estimated demand functions and estimated Barten scale distributions to do welfare analyses. In particular, we use a Barten scaled energy demand function to perform consumer surplus calculations for an energy price change (as in Hausman 1981). Our consumer surplus calculations can be interpreted as a variant of 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 linear random coefficients. In contrast, our model follows the prior consumer demand literature by including 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). We also apply our empirical results to estimate Atkinson (1970) type social welfare functions, and thereby analyze the extent to which allowing for unobserved preference heterogeneity affects estimated tradeoffs between mean impacts and inequality impacts of a tax or price change in energy.

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 unobserved 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 related 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.

Nonparametric identification and estimation of ordinary random coefficients models is considered by Beran and Hall (1992), Beran, Feuerverger, 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).

Ordinary random coefficients are the special case of the additive model in 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), and in the particular applications of additivity to consumer demand systems include Gorman (1976) and Blackorby, Primont, and Russell (1978). Particularly relevant for this paper is Matzkin (2003), which in an appendix describes generic, more high level identifying conditions for a large class of additive models with unobserved heterogeneity.

This paper also contributes to the literature on estimation of models with nonseparable er-

rors, 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 particular, Hoderlein, Nesheim, and Simoni (2011) provide high level conditions for identification and estimation of models that, like ours, contain a vector of random parameters. Their results could be applicable to our model if the function G were finitely parameterized instead of nonparametric.

3 Generalized Random Coefficient Model Identification

In this section we first show shows nonparametric identification of additive models given by equation (2), and then, under stronger conditions, we identify the more general model of equation (1). In an econometric appendix we provide additional identification results, including the case where $X = (X_1, ... X_K)$ contains discrete regressors. The connection of this model to Barten scales is provided later.

For any random vectors A and B let $F_{A|B}(a \mid b)$ and $f_{A|B}(a \mid b)$ denote the conditional cumulative distribution function and conditional probability density function, respectively, of A given B. Let e_k be the K vector containing a one in position k and zeros everywhere else. Let $X_{(k)}$ denote the $K - 1$ vector that contains all the elements of X except for X_k.

3.1 Additive Model Identification

ASSUMPTION A1: The conditional distribution $F_{Y|X,Z}(y \mid x, z)$ and the marginal distribution $F_Z(z)$ are identified. $(U_0, U_1, ..., U_K) \perp X \mid Z$ and $(U_1, ..., U_K) \perp U_0 \mid Z$. Either U_0 has a nonvanishing characteristic function or U_0 is identically zero. $supp(U_0) \subseteq supp(Y)$ and $\{0, e_1, ..., e_K\} \subseteq supp(X)$.[1]

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

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 $y_0 \in supp(Y)$ are imposed.

Assumption A1 first assumes identification of $F_{Y|X,Z}(y \mid x, z)$ 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 $F_{Y|X,Z}(y \mid x, z)$ 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 $F_{Y|X,Z}(y \mid x_k e_k, z)$ for $k = 1, ..., K$. However, more information regarding $F_{Y|X,Z}$ is used in Theorem 2 and other extensions.

---

[1]Formally, the condition on U_0 regarding a nonvanishing characteristic function required for the deconvolution step of the proof is only that the set of $t \in \mathbb{R}$ for which $E\left(e^{itU_0}\right) \neq 0$ is dense in $\mathbb{R}$. See, e.g., Meister (2005).

Assumption A1 imposes conditional independence and support requirements on U, X and Z. The role of Z is to permit the error U_0 and random coefficients U_k to be correlated with X, thereby allowing elements of X to be endogenous. See, e.g., the correlated random coefficients model of Heckman and Vytlacil (1998). This allows for Heckman and Robb (1986) control function type endogeneity, with Z being control function residuals as in Blundell and Powell (2003, 2004). In particular, if $X_k = h_k \left(X_{(k)}, Q \right) + Z_k$ for some observed instrument vector Q and some identified function h_k (typically h_k would be $E \left(X_k \mid X_{(k)}, Q \right)$), 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 independent of X and so the regressors X are exogenous. The assumptions also permit Z to be discrete, and place no restriction on the dimension of Z, although control function residuals would generally be continuous and have dimension equal to the number of endogenous elements of X.

Assumption A2 assumes that the regressors and random coefficients are continuously distributed. An alternative to Assumption A2 allowing for discrete distributions is provided in the Appendix. Assumption A2 also calls for a mild relative support assumption on X_k and U_k.

The normalizations in Assumption A3 are free normalizations, 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 $y_0 \in supp (Y)$, 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 $E (U_0) = 0$.

What follows is our first identification theorem, which as noted in the introduction is closely related to results in Matzkin (2003).

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 $k \in \{1, ..., K\}$ such that Assumptions A2 and A3 hold, the function G_k and the distribution function $F_{U_k|Z}$ are nonparametrically identified.

Identification of F_Z was assumed, and Theorem 1 gives identification of $F_{U_0|Z}$ and $F_{U_k|Z}$, and so by combining these 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 $k \in \{1, ..., K\}$, thereby identifying the entire model. However, in the Appendix we 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 continuously distributed regressors and random coefficients.

3.2 General Model Identification

We now show identification of the general model given by equation (1) for some unknown function G. To do so we decompose the function G into an additive component $\sum_{k=1}^{K} G_k$ and a remaining interaction term $\widetilde{G}$ having the property that $\widetilde{G} (X_k e_k) = 0$ for all k. These two

components are then identified separately. Given a function G, define $G_k(X_k) = G(X_k e_k)$, that is, G_k equals G after setting all the elements of X except for X_k equal to zero. Define the function $\widetilde{G}$ by $\widetilde{G}(X) = G(X) - \sum_{k=1}^{K} G_k(X_k)$. Then, by construction,

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

where $\widetilde{G}$ equals zero whenever all but one element of X is zero. It is important to note that equation (3) is not an assumption, but rather is a construction. Given any function G we can construct $\widetilde{G}$ and G_k with the above properties such that equation (3) holds.

ASSUMPTION A4: Assume that $U_0, U_1, ..., U_K, X$ are mutually independent conditional upon Z, and that for some $z \in supp(Z)$, $E\left(U_k^t \mid Z = z\right) \neq 0$ and is bounded for all integers t. Assume the support of X includes a positive measure neighborhood of zero. Assume $\widetilde{G}$ is a real, entire, analytic function.

THEOREM 2: Let $Y = G(X_1 U_1, ..., X_K U_K) + U_0$ and let Assumptions A1, A2, A3, and A4 hold for $k = 1, ..., K$. Then the function G and the distribution functions $F_{U_k|Z}$ for every $k \in \{0, 1, ..., K\}$ are all nonparametrically identified.

3.3   Identification Notes

The connection between the observable function $F_{Y|X,Z}(y \mid x, z)$ and the unknown functions G and f_k to be identified is $F_{Y|X,Z}(y \mid x, z) =$

$$\int_{supp(U_1,...,U_K|Z=z)} f_{0|Z}\left[y - G(x_1 u_1, ..., x_K u_K \mid z)\right] \widetilde{f}\left[f_{1|Z}(u_1 \mid z), ... f_{K|Z}(u_K \mid z) \mid z\right] du_1...du_K \tag{4}$$

where $\widetilde{f}$ is the copula of $U_1, ..., U_K$. Theorem 1 allows $\widetilde{f}$ to be arbitrary and unknown, but does not identify it nonparametrically, while Theorem 2 (in particular, the mutual independence required by Assumption A4) assumes $\widetilde{f}$ is the identity function. This is a strong restriction that is not required by linear random coefficients models, but it is hard to see how both G and $\widetilde{f}$ could be nonparametrically identified.

One could provide generic high level assumptions that suffice for identification, which are essentially completeness type assumptions guaranteeing that the integral equation (4) has a unique solution. We have instead provided low level, readily interpreted assumptions and constructive proofs for our theorems, based primarily on the fact that equation (4) and its derivatives simplify greatly when most of the elements of X equal zero.

Although our identification proofs are constructive, an estimator based on mimicking the steps of our identification proof would likely be very inefficient. This is because data on the entire support of Y and X provides useful identifying information, while our proofs focus on observations where all but one element of X is near zero.[2] Values of X far from zero are all

---

[2]The proof of Theorem 2 involves evaluating the distribution of Y given X where either $X = 0$ or all but one element of X equals zero. This means conditioning on a set of measure zero. The same applies to Theorem 1

informative, but for identification we focus on X elements near zero only because that is where the integral equation (4) has solutions that are tractible to describe. In that sense, our results differ fundamentally from the usual cases of "identification at infinity," where only data near support boundaries are informative. Generically, the model should be identified even if X is bounded away from zero, but in that case the technical conditions ensuring that equation (4) has a unique solution will be more difficult to express and interpret. This generic identification will be important for our empirical application to consumer demand, where real data may not include values of some elements of X near zero.[3] See, e.g., Chiappori and Ekeland (2006) for other examples of generic identification in consumer demand systems.

Theorem 1 did not require the random coefficients U_k to be mutually independent, though only the separate distributions of each U_k were nonparametrically identified. We could generally identify parametric dependence among $U_1,...,U_K$ by parameterizing the copula $\widetilde{f}$, and using equation (4) to identify those parameters, given that all the other functions in this equation are already identified by Theorem 1.

In contrast, for Theorem 2, identification of the nonadditive functions G required mutual independence given by Assumption A4. Theorem 2 also require that each $E\left(U_k^t\right)$ be nonzero and bounded, so Theorem 2 also places the stronger requirements of a nonzero mean and thin tails on each U_k. Theorem 2 places stronger smoothness assumptions on $\widetilde{G}$ than on each G_k. Without decomposing G into $\widetilde{G}$ and G_k terms, a sufficient but stronger than necessary restriction to satisfy the required assumptions on these functions is that G itself 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.

4   Random Barten Scales

Let a "consumer" refer to an individual or household that maximizes a single well behaved utility function. Let Q_j denote the quantity purchased of a good j, and let $S(Q, U)$ denote the direct utility function over the bundle of goods $Q = (Q_1, ..., Q_J)$ of a consumer having a vector of heterogeneity parameters U. Assume S is continuous, non-decreasing, and quasi-concave in Q. Define the reference consumer to be a consumer that has heterogeneity parameters U normalized to equal one, and let $\overline{S}(Q_1, ..., Q_J)$ denote the direct utility function of a reference consumer. Each 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. Write the Marshallian budget share functions that result from maximizing the reference utility function $\overline{S}$ as $W_j^* = \omega_j(P_1/M, ..., P_J/M)$, where $W_j^* = Q_j P_j/M$ is the share of money M that is spent on

---

regarding conditioning on Z at a point if Z is continuous. Note, however, that 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 density and expectation functions. It would be possible to recast the proofs in terms of conditioning on sets $\| X \| \le c$ and taking limits as $c \to 0$.

[3]A separate but related issue is that, in our application, zero will be on the the boundary of the support of X. Formally, extending our identification theorems to the case where zero is on the boundary means that derivatives with respect to X in the proofs would sometimes need to be replaced with one sided derivatives.

good j (called the budget share of good j). Let $V(P_1/M, ..., P_J/M)$ denote the indirect utility function corresponding to $\overline{S}$, obtained by substituting $Q_j = \omega_j(P_1/M, ..., P_J/M) M/P_j$ into $\overline{S}(Q_1, ..., Q_J)$ for $j = 1, ..., J$.

Our empirical application is based on Barten (1964) scales. Barten scales are a longstanding method used to bring preference heterogeneity on the basis of observed variables into continuous demand models. Barten scales are consequently a natural starting point for the incorporation of random utility parameters representing unobserved preference heterogeneity. See, e.g., Lewbel (1997) for a survey of various types of equivalence scales in the consumer demand literature, including Barten scales, and see Jorgenson, Lau, and Stoker (1982) for a prominent empirical application of traditional Barten scales. Deaton and Muellbauer (1980) includes an extensive discussion of parametric identification of Barten Scales.

Barten (1964) proposed the model in which consumers have utility functions of the form $S(Q_1, .., Q_J; \alpha_{h1}, ..., \alpha_{hJ}) = \overline{S}(Q_1/\alpha_{h1}, ..., Q_J/\alpha_{hJ})$, where the Barten scales $\alpha_{h1}, ..., \alpha_{hJ}$ are positive functions of observable household attributes h, such as age or family size, that embody variation in preferences across consumers. 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 α_{hj} 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 provide a measure of the degree to which different households get utility from different goods. This is how we will employ them. Although Barten scales have long been a popular method of modeling preference heterogeneity in empirical work, 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.

We propose random Barten scales, assuming that consumers have utility functions of the form $S(Q_1, .., Q_J; U_1, ..., U_{hJ}) = \overline{S}(Q_1/U_1, ..., Q_J/U_J)$, where $U_1, ..., U_J$ are positive random utility parameters embodying unobserved preference heterogeneity across consumers. More formally, we could write each random Barten scale as $U_j(h)$, since for each good j, the distribution function that U_j is drawn from could depend on observable household attributes h. Barten's original model is then the special case where the distribution of each $U_j(h)$ is degenerate with a mass point at α_{hj}.

Define normalised prices $X_j = P_j/M$ for each good j and rewrite the budget constraint as $\sum_{j=1}^{J} X_j Q_j = 1$. Now $\overline{S}(Q_1, ..., Q_J)$ and $V(X_1, ..., X_J)$ are the direct and indirect utility functions of the reference consumer, and $\omega_j(X_1, ..., X_J)$ is the Marshallian budget share demand function of the reference consumer. It can be immediately verified from the first order conditions for utility maximization that a consumer will have Marshallian demand functions of the form $W_j^* = \omega_j(U_1 X_1, ..., U_J X_J)$ for each good j if and only if the consumer's direct and indirect utility function equal, up to an arbitrary monotonic transformation,

9

$\overline{S}(Q_1/U_1, ..., Q_J/U_J)$ and $V(U_1X_1, ..., U_JX_J)$, respectively. Also, given a specification of reference indirect utility $V(X_1, ..., X_J)$, the corresponding Barten scaled demand functions can be obtained by the logarithmic form of Roy's identity:

$$\omega_j(U_1X_1, ..., U_JX_J) = \frac{\partial V(U_1X_1, ..., U_JX_J)}{\partial \ln X_j} \Big/ \left(\sum_{\ell=1}^{J} \frac{\partial V(U_1X_1, ..., U_JX_J)}{\partial \ln X_\ell} \right) \quad (5)$$

Notice that the functional form of each ω_j only depends on the functional form of $\overline{S}$ or equivalently of $\overline{V}$, so $U_1, ...U_J$ can vary independently of $X_1, ..., X_J$ across consumers. These derivations are exactly those given by Barten (1964) and by later authors who applied Barten scales, e.g., Jorgenson, Lau, and Stoker (1982), except that we put unobserved random variables U_j in place of deterministic functions α_{hj} of observed household characteristics. Random Barten scaled Marshallian demand functions then have precisely the form of our generalized random coefficients given in equation (1).

4.1 Random Barten Scales Identification

The following corollary applies our identification theorems to the Barten scales model.

COROLLARY 2: Assume that consumers have preferences given by an analytic, Barten scaled indirect utility function $V(U_1X_1, ..., U_JX_J)$; no good j exists that is Giffen at some values of X_j and not Giffen at others; the conditional distribution $F_{Q|X}(q \mid x)$ is identified; and Barten scales U_j are each positive with bounded support, satisfy Assumption A2 with Z empty, and are mutually independent of each other and of X. Then the demand functions $\omega_j(U_1X_1, ..., U_JX_J)$ and the distribution functions $F_{U_j}(U_j)$ are identified for $J = 1, ..., J$.

Corollary 2 shows that consumer demand systems with random Barten scales are nonparametrically identified. By standard revealed preference theory, the direct and indirect utility functions are therefore also nonparametrically identified up to an unknown monotonic transformation.

A serious limitation in applying our identification theorems to Barten scales is the assumption that the scales U_j for different goods j are distributed independently of each other. We note first that our results permit Barten scales to have correlation with each other through observable characteristics, e.g., larger households could have scales that are all smaller than those of smaller households, because household size is observable. It is only the remaining random variation in scales after conditioning on observable characteristics that would need to be independent across goods. Second, even with independently distributed scales, we are still proposing a significant advance over the previous literature, in which Barten scales were deterministic and therefore had no unobserved random variation at all. Finally, it might be possible to relax this condition to allow for known dependence across the random coefficients (e.g., a parametric copula), but we were not able to show complete nonparametric identification with nonparametric demand functions with unknown dependence across the random coefficients.

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, very little empirical evidence has been found for the existence of Giffen goods, and particularly not for the types of goods we consider in our application.[4]

Regarding other 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 bounded.[5] The issue regarding some $X_j = 0$ points used for identification being on the boundary of the support of X was discussed in the previous section.

Applying Theorem 2 separately to each demand function would require that each demand function be monotonic in all prices, which does not hold in general. 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 standard monotonicity of own price effects, leaving the signs of cross price effects unconstrained, to identify the Barten scale distributions. This is done in Corollary 2 by using the demand function of each good j to just identify the distribution of the Barten scale U_j. 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, assuming 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 applied here because there are actually only $J - 1$ distinct demand functions $\omega_1,...,\omega_{J-1}$, with the remaining demand function ω_J determined by the adding up constraint that $\sum_{j=1}^{J} \omega_j = 1$.

To simplify our empirical analysis, we 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 goods 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 rationalizable 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) provides conditions on the distribution of prices (stochastic hicksian aggregation) under which Marshallian demand functions have the same properties with nonseparable utility as with separable utility.

With $J = 2$ goods, our model is $W_1^* = \omega_1 (U_1 X_1, U_2 X_2)$ and $W_2^* = 1 - W_1^*$, and we can

---

[4]The only example 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.

[5]Boundedness rules out lexicographic preferences in which consumers would not prefer an unlimited quantity of the good with the unbounded Barten scale over an infinitesimal amount any other good. This is extemely unlikely to hold when goods are defined as broad categories like food, energy, clothing, etc.

rewrite Roy's identity as

$$\lambda\left(W_1^*\right) = \ln\left(\frac{\partial V\left(U_1 X_1, U_2 X_2\right)}{\partial \ln X_1}\right) - \ln\left(\frac{\partial V\left(U_1 X_1, U_2 X_2\right)}{\partial \ln X_2}\right) \tag{6}$$

where $\lambda\left(W_1^*\right)$ is the logit transformation $\lambda\left(W_1^*\right) = \ln\left[W_1^* / \left(1 - W_1^*\right)\right]$. This 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 (6) 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.

4.2 Additive Model Random Barten Scales

The regularity conditions for identifying random coefficients in the nonparametrically additive model, given by Theorem 1, are milder than for the fully nonparametric model of Theorem 2, so we 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 $\omega_1\left(X_1, X_2\right)$ results in extreme restrictions on behavior. See, e.g., Blackorby, Primont, and Russell (1978). So we instead impose additivity on the logit transformation of $\omega_1\left(X_1, X_2\right)$ (later this will be relaxed to allow for interaction terms), thereby assuming demands have the additive form

$$\lambda\left(W_1\right) = \lambda\left[\omega_1\left(U_1 X_1, U_2 X_2\right)\right] + U_0 = g_1\left(U_1 X_1\right) + g_2\left(U_2 X_2\right) + U_0 \tag{7}$$

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

$$W_1^* = \omega_1\left(U_1 X_1, U_2 X_2\right) = \left(1 + e^{-g_1\left(U_1 X_1\right) - g_2\left(U_2 X_2\right)}\right)^{-1} \tag{8}$$

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, $\lambda\left(W_1\right)$ 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 (8) and hence (7) exists, and is given by Theorem 3.

THEOREM 3: The demand function ω_1 satisfies $\lambda\left[\omega_1\left(U_1 X_1, U_2 X_2\right)\right] = g_1\left(U_1 X_1\right) + g_2\left(U_2 X_2\right)$ for some functions g_1 and g_2 if and only if ω_1 is derived from an indirect utility function of the form

$$V\left(U_1 X_1, U_2 X_2\right) = H\left[h_1\left(U_1 X_1\right) + h_2\left(U_2 X_2\right), U_1, U_2\right].$$

for some montonic in its first element function H and some differentiable functions h_1 and h_2. The functions g_1, g_2, h_1, and h_2 are related by

$$h_1 (U_1 X_1) + h_2 (U_2 X_2) = \int_{-\infty}^{\ln X_1} e^{g_1 (U_1 x_1)} d \ln x_1 + \int_{-\infty}^{\ln X_2} e^{-g_2 (U_2 x_2)} d \ln x_2 \qquad (9)$$

and

$$g_1 (U_1 X_1) + g_2 (U_2 X_2) = \ln \left(\frac{\partial h_1 (U_1 X_1)}{\partial \ln X_1} \right) - \ln \left(\frac{\partial h_2 (U_2 X_2)}{\partial \ln X_2} \right) \qquad (10)$$

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 for individual consumer's demand functions, and is present only because utility functions are ordinal and therefore unchanged by monotonic transformations.[6] We can therefore just write the indirect utility function in Theorem 3 as

$$V (U_1 X_1, U_2 X_2) = h_1 (U_1 X_1) + h_2 (U_2 X_2) . \qquad (11)$$

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 (8) 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.

Using Theorem 3, and in particular equation (10), we can nonparametrically estimate g_1 and g_2 by nonparametrically specifying h_1 and h_2 in terms of sieve basis functions (imposing the shape restrictions possesed by indirect utility functions if desired). For efficiency, it is desirable to choose a basis for sieve expansions having the property that low order terms are equivalent to good parametric models. We therefore consider a polynomial in logs sieve basis

$$\ln h_k (U_k X_k) = \sum_{s=0}^{S} \beta_{ks} (\ln (U_k X_k))^s \qquad (12)$$

with constants β_{ks}, for $k = 1, 2$, letting $S \to \infty$ as $n \to \infty$. Logarithmic specifications like these are common in demand models, e.g., with $S = 1$ equations (11) and (12) 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. Having $S = 3$ without Barten scales corresponds to an additively separable version of the third-order Translog budget share function as in Nicol (1984).

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 $V (U_1 X_1, U_2 X_2)$

---

[6]Later we will reintroduce the function H to construct a money metric representation of utility for use in social welfare calculations.

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 (7) to this model gives the demand function

$$
\begin{aligned}
\lambda\left(W_{1}\right) & =\lambda\left[\omega_{S 1}\left(U_{1} X_{1}, U_{2} X_{2}, \beta\right)\right]+U_{0} \\
& =\beta_{10}+\left(\sum_{s=1}^{S}\left[\ln\left(U_{1} X_{1}\right)\right]^{s} \beta_{1 s}-\left[\ln\left(U_{2} X_{2}\right)\right]^{s} \beta_{2 s}\right)+\ln\left(\frac{\sum_{s=1}^{S}\left(\ln\left(U_{1} X_{1}\right)\right)^{s-1} s \beta_{1 s}}{\sum_{s=1}^{S}\left(\ln\left(U_{2} X_{2}\right)\right)^{s-1} s \beta_{2 s}}\right)+U_{0}.
\end{aligned}
\tag{13}
$$

where $\omega_{S 1}\left(U_{1} X_{1}, U_{2} X_{2}, \beta\right)$ denotes the sieve representation of $\omega_{1}\left(U_{1} X_{1}, U_{2} X_{2}\right)$ with S terms in the parameters β. Here, $\lambda\left(W_{1}\right)$ is additive as in (7) since the logged ratio may be written as a difference of logs.

We nonparametrically model the density functions of U_{0} and U_{k} for $k = 1, 2$, by using Hermite polynomial seive densities[7]

$$
f_{0 J}\left(U_{0}, \delta, \sigma_{0}\right)=\frac{1}{\sigma_{0}} p_{J}\left(\frac{U_{0}}{\sigma_{0}}, \delta\right) \quad \text { and } \quad f_{k J}\left(U_{k}, \gamma_{k}, \sigma_{k}\right)=\frac{1}{U_{k} \sigma_{k}} p_{J}\left(\frac{\ln U_{k}}{\sigma_{k}}, \gamma_{k}\right) \tag{14}
$$

For a given consumer with observed values x_{1} and x_{2}, the conditional density function of W_{1} is then

$$
\begin{aligned}
& f_{W_{1} \mid X_{1}, X_{2}}\left(w_{1} \mid x_{1}, x_{2} ; \beta, \sigma, \delta, \gamma\right) \\
= & \int_{0}^{\infty} \int_{0}^{\infty} f_{0 J}\left[\ln\left(\frac{w_{1}}{1-w_{1}}\right)-\omega_{S 1}\left(u_{1} x_{1}, u_{2} x_{2}, \beta\right), \delta, \sigma_{0}\right] f_{1 J}\left(u_{1}, \gamma_{1}, \sigma_{1}\right) f_{2 J}\left(u_{2}, \gamma_{2}, \sigma_{2}\right) d u_{1} d u_{2}
\end{aligned}
$$

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

$$
\sum_{i=1}^{n} \ln f_{W_{1} \mid X_{1}, X_{2}}\left(w_{1 i} \mid x_{1 i}, x_{2 i} ; \beta, \sigma, \delta, \gamma\right) \tag{15}
$$

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.[8]

---

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

[8]Depending on the supports and tail thickness of the model errors and regressors, it is possible in problems like ours for identification to 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 framework similar to ours, though in their model only the error distributions are nonparametric. The assumptions required for standard sieve maximum likelihood inference may rule out at least some forms of weak identification, though our use of sieves could even then be interpreted as a choice of regularization for structural function estimation.

In the version of the model described here, the Barten scale distributions do not depend on observable characteristics of consumers. It would be straightforward to replace each $U_j X_j$ term with $U_j X_j \alpha_j (Z, \theta)$, where $\alpha_j (Z, \theta)$ is a traditional deterministic Barten scale parametrically specified as a function of observed household characteristics Z and a parameter vector θ. More generally, each density function f_{kJ} could be specified as an arbitrary conditional demand function, conditioning on Z. To simplify our empirical analysis, we instead focus on a single group of consumers that have identical values of some important characteristics Z.

4.3   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.

Table 1: Summary Statistics: Canadian Energy Shares				
9413 observations	mean	std dev	min	max
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

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 both 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 1 gives summary statistics for budget shares, expenditures, prices, and normalised prices.

We estimate equation (15) in Stata, plugging in equations (13) and (14) with $S = 3$ and $J = 2$. Estimated coefficients are given in Table 2 below. Standard errors are provided with the

caveat that they treat the sieve basis functions as finite model parameterizations.[9]

Table 2: Estimated Parameters

h_1 and h_2	est	se	U_0, U_1, and U_2	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 1 and 2 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 1 and 2 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 substantively to observed variation in demand.

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

4.4   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 $\omega_1 (U_1 X_1, U_2 X_2)$, and identify the function ω_1 and distribution of the associated Barten scales U_1 and U_2. However, in doing so we would lose the benefits we gained from Theorem 3 of having closed form expressions for the corresponding indirect utility function $V (U_1 X_1, U_2 X_2)$, 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

---

[9]Given that our our data ended up implying low order sieve polynomials, it is reasonable to interpret our estimates parametrically.

consider adding second and third order interaction terms to the model of Theorem 3, giving an indirect utility function of the form

$$V(X_1, X_2) = h_1(X_1) + h_2(X_2) + X_1 X_2 \alpha_0 + X_1^2 X_2 \alpha_1 + X_1 X_2^2 \alpha_2 \tag{16}$$

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, indeed, we could interpret these interactions as the first terms in a sieve expansion for an arbitrary indirect utility function. Barten scaling this indirect utility function, substituting the result into equation (6), and adding the error term U_0 as before gives the demand model

$$\lambda(W_1) = \ln\left[g_1(U_1 X_1) + M_1(U_1 X_1, U_2 X_2, \alpha)\right] - \ln\left[g_2(U_2 X_2) + M_2(U_1 X_1, U_2 X_2, \alpha)\right] + U_0 \tag{17}$$

where $g_k(U_k X_k) = U_k X_k \partial h_k'(U_k X_k) / \partial(U_k X_k)$ for $k = 1, 2$ and

$$M_1(U_1 X_1, U_2 X_2, \alpha) = U_1 X_1 U_2 X_2 \alpha_0 + 2U_1^2 X_1^2 U_2 X_2 \alpha_1 + U_1 X_1 U_2^2 X_2^2 \alpha_2, \tag{18}$$

$$M_2(U_1 X_1, U_2 X_2, \alpha) = U_1 X_1 U_2 X_2 \alpha_0 + U_1^2 X_1^2 U_2 X_2 \alpha_1 + 2U_1 X_1 U_2^2 X_2^2 \alpha_2. \tag{19}$$

Identification of this demand model follows directly from Theorem 3.[10]

For estimation of the model, we let the functions h_k in equation (16) be represented by the same polynomial in logs sieve basis functions as before. Barten scaling this indirect utility function gives, by equation (17), the demand function

$$
\begin{aligned}
\lambda(W_1) &= \omega_{S1}(U_1 X_1, U_2 X_2, \beta) + U_0 \tag{20}\\
&= \ln\left[\left(e^{\beta_{10} + \sum_{s=1}^{S}(\ln(U_1 X_1))^s \beta_{1s}}\right)\left(\sum_{s=1}^{S}(\ln(U_1 X_1))^{s-1} s\beta_{1s}\right) + M_1(U_1 X_1, U_2 X_2, \alpha)\right]\\
&\quad - \ln\left[\left(e^{\sum_{s=1}^{S}(\ln(U_2 X_2))^s \beta_{2s}}\right)\left(\sum_{s=1}^{S}(\ln(U_2 X_2))^{s-1} s\beta_{2s}\right) + M_2(U_1 X_1, U_2 X_2, \alpha)\right] + U_0
\end{aligned}
$$

The demand function given by equation (20) is the same as (13), except for the addition of the functions M_1 and M_2 given by equations (18) and (19), which embody the additional desired price interaction terms. We estimate equation (20) using the same sieve maximum likelihood method as before.

4.5   Empirical Barten Scales with Interaction Terms

Table 3 presents estimated parameters for the demand equation (20), 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

---

[10]It's possible to directly prove identification of the demand model of equations (17), (18), and (19) under weaker conditions than those of 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. A proof appears in earlier working paper version of this paper.

make the argument of the log function in $\lambda(W_1)$ negative. We therefore restrict α_0, α_1, and α_2 to be non-negative.

As Table 3 shows, two of the interaction terms are statistically significant (and all three are jointly significant). For comparison, we also estimated the model, denoted "without heterogeneity," imposing the constraint that $U_1 = U_2 = 1$ and 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.

Figures 3 and 4 show 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. Finding that U_2 has a larger variance that U_1 means that consumers have more heterogeneity in their preferences for non-energy goods than for energy goods, which is not surprising, given the extent to which energy goods are necessities.

Recall $X_j = P_j/M$ where M is total expenditures. Figure 5 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 the combinations of each quartile of U_1 with each 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 variation in the random Barten scales U_1 and U_2 causes substantial shifts in the Engel curves. For comparison, Figure 5 also displays, with a thick grey line, the Engel curve from a model without heterogeneity which imposes $U_1 = U_2 = 1$.

Table 3: Interaction Terms in Utility: Estimated Parameters

h_1 and h_2	est	se	U_0, U_1, and U_2	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			

The shape of the Engel curve without unobserved preference heterogeneity in Figure 5 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 neccessity 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 6 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, as a thick gray line, the simulated shares from the model without heterogeneity.

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.

4.6   Consumer Surplus Effects of a Carbon Tax

We now apply our model to evaluate the partial equilibrium effects of a large change in the price of energy, as might result from a carbon tax.[11] Using equation (16), even with non-parametric 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 (16) to obtain the cost of living impact of a price change. We would otherwise need to numerically solve a differential equation as in Hausman and Newey (1995), but such a solution would need to be calculated for every value on the continuum of points that U_1 and U_2 can take on.

For an individual facing initial prices $\overline{P}_1$, $\overline{P}_2$, having total expenditures M and preferences indexed by Barten scales U_1, U_2, the cost-of-living impact of moving to new prices P_1, P_2 is $\pi\left(U_1, U_2, M, P_1, P_2, \overline{P}_1, \overline{P}_2\right)$, defined as the solution to

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

Here π is the proportionate change in costs M needed to compensate for the price change, that is, the amount by which M would need to be scaled up to bring an individual facing prices P_1, P_2 (and having preferences given by U_1, U_2) back to the same indifference curve they were on when facing prices $\overline{P}_1$, $\overline{P}_2$. To show price effects clearly, we consider a large price change: doubling the price of energy. So we solve for the π function given the initial price vector $\overline{P}_1 = \overline{P}_2 = 1$ and the new price vector $P_1 = 2$, $P_2 = 1$. Figure 7 shows the resulting 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 6, the thick gray line gives estimates of π based on the model no preference heterogeneity.

Table 4 gives summary statistics of distributions presented in Figure 7. The "without heterogeneity" estimates in Table 4 are statistics for the marginal distribution of π, obtained from

---

[11]Our model is not a general equilibrium model, so we are only estimating the consumer's responses to a change in energy prices. Moreover, these should only be interpreted as short run responses, since in the longer run consumers could change their energy elasticities and demand by, e.g., buying more energy efficient cars and appliances. Also, we just consider a change in the overall price of energy, and so do not consider impacts of possible changes in the composition of energy goods.

sieve maximum likelihood (SML) estimates of a model without unobserved preference heterogeneity (the same estimates that generate the thick gray line in Figure 7). The variation in these without-heterogeneity statistics comes only from variation in M.

The second set of estimates in Table 4 are based on the SML estimates given in Table 3, that is, our main model that includes the random Barten scales. The "ignore heterogeneity" estimates in Table 4 describe the marginal distribution of π, conditioning on U_1 and U_2 equaling their estimated medians. These estimates differ from "without heterogeneity" in the estimated parameter values of the demand functions, but still have variation that only comes from variation in M. If our main model is correctly specified, then "without heterogeneity" demand function estimates will be inconsistent, while the "ignore heterogeneity" statistics will show the impacts of ignoring heterogeneity, evaluated at consistently estimated demand parameters.

Table 4: Summary Statistics for Cost of Living Change and Social Welfare					
	Mean	Std Dev	Lower Qtl	Median	Upper Qtl
log-cost of living impacts for individuals					
Overall–without heterogeneity	0.111	0.013	0.105	0.113	0.120
std err	*0.020*	*0.006*	*0.018*	*0.018*	*0.021*
Overall-ignore heterogeneity	0.128	0.016	0.122	0.131	0.135
std err	*0.016*	*0.003*	*0.015*	*0.015*	*0.017*
Overall–with heterogeneity	0.128	0.072	0.073	0.120	0.172
std err	*0.013*	*0.004*	*0.011*	*0.014*	*0.016*
At lower Qtl of M ($\ln M = -0.38$)	0.140	0.079	0.070	0.137	0.204
At median of M ($\ln M = -0.03$)	0.130	0.071	0.070	0.123	0.181
At upper Qtl of M ($\ln M = 0.26$)	0.113	0.057	0.068	0.110	0.151

The remaining set of estimates in Table 4, labeled "with-heterogeneity," are based on our main model (SML estimates from Table 3), showing features of the marginal distribution of π accounting for the estimated variation in U_1 and U_2, as well as variation in M. The last three rows of Table 4 give the conditional distribution of π in the with-heterogeneity model, conditioning on the quartiles of M. The variation in these last three rows comes only from variation in U_1 and U_2.

It has long been known that first order approximations to the cost of living effects of marginal price changes can be evaluated without estimating demand functions and associated demand elasticities (see, e.g., Stern 1987). These theoretical results have been used to argue that demand function estimation is not required for marginal policy analyses. In our data, the average value of the budget share for energy is 0.14, so if there were no substitution effects in response to a price change, doubling the price of energy would increase the cost of living by $\pi = 0.14$. This would be the first order approximation based estimate of π. However, the estimated average cost-of-living impacts given in Table 4 are much less 0.14, showing substantial relative price substitution effects. This supports findings in, e.g., Banks, Blundell, and Lewbel (1996) that, contrary to the first order approximation theory, it is empirically 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. Moreover, one goal of an

20

energy tax would be to reduce energy consumption, so it's important to account for the impact on welfare of this reduction.

In place of the first order approximation of .14, Table 4 shows that our main model gives an estimated average cost of living impact of 0.128, which is over nine percent smaller. This is true whether we ignore heterogeneity in our main model or not, but if we misspecify the demand functions by estimating them without heterogeneity, we get a value of 0.111, thereby underestimating the average impact by over ten percent, a substantial amount. This is because, in a nonlinear model like ours, estimation without-heterogeneity leads to bias in the estimated parameters. Thus, even if we were not interested in the distribution of effects of a price change, just getting the average impact correct requires accounting for unobserved preference heterogeneity when estimating demand function parameters.

Far larger than these differences in the mean of π is the effect of preference heterogeneity on the *variation* of impacts in cost of living π across consumers. This can be seen in Figure 7, where consumers near the top of the contour plot have their costs of living go up by over 25%, while those near the bottom have impacts near zero. This variation can be seen in the standard deviations and interquartile ranges of estimates of π in Table 4. Particularly notable is that, when allowing for heterogeneity, the standard deviation increases 450%, from .016 when ignoring heterogeneity to .072 when allowing for preference heterogeneity. The increase relative to the without-heterogeneity model is even larger, an increase of almost sixfold.

The effects of preference heterogeneity on cost of living are even larger than the impacts of varying M. The cost of living impact of the tax is, in percentage terms, bigger for poor households than for rich ones, as can be seen in the first column and last three rows of Table 4. The cost of living impact drops from 0.14 for consumers at the bottom quartile level of total expenditures M to 0.113 for those at the top quartile, giving an interquartile range of 0.027. In contrast, looking across the "median of M" row of Table 4 shows that the interquartile range in impacts just from variation in preferences (that is, holding M fixed) is 0.111 (0.181 minus 0.070). This is more than four times larger than the interquartile range of 0.027 stemming from variation in M. Overall, variation in preferences has a dramatically larger impact than variation in income on who is hurt the most or least by our energy tax experiment.

4.7 Social Welfare Implications of a Carbon Tax

The above analyses summarized the distribution of welfare implications of our energy tax experiments across consumers. To evaluate the implications for aggregate welfare, we consider the impacts of our experiment on a range of possible social welfare functions. To proceed, we require interpersonally comparable and cardinal measures of individual utility. To make utility functions interpersonally comparable, we follow the standard procedure in this literature of constructing money metric cardinalizations of utility. A money metric utility cardinalization $\widetilde{V}$ of a given indirect utility function V is the monotonic transformation $H(V)$ having the property that, at base prices $\overline{P}_1, \overline{P}_2$, the function $\widetilde{V} = H(V)$ equals total expenditures. Since in our application preferences also vary by Barten scales, we also need to choose a base level of Barten scales $\overline{U}_1, \overline{U}_2$ at which $H(V)$ equals total expenditures. Formally, $\widetilde{V}$ is defined by

$$\widetilde{V}\left(U_1 P_1/M, U_2 P_2/M\right) = H\left[V\left(U_1 P_1/M, U_2 P_2/M\right), \overline{U}_1, \overline{U}_2, \overline{P}_1, \overline{P}_2\right].$$

where V is given by equation (16) and H is defined to satisfy

$$\widetilde{V}\left(\overline{U}_1\overline{P}_1/M, \overline{U}_2\overline{P}_2/M\right) = M$$

for all values of M. Strict monotonicity of H as a function of V, and of V as a function of M, ensures that such a function H exists and that the resulting function $\widetilde{V}$ is unique.

We let base prices be $\overline{P}_1 = \overline{P}_2 = 1$ and we take $\overline{U}_1$ and $\overline{U}_2$ to equal the medians of the estimated distributions of U_1 and U_2, respectively. We could have allowed H to depend on U_1 and U_2 directly (as in the statement of Theorem 3), in addition to depending on $V, \overline{U}_1, \overline{U}_2, \overline{P}_1$, and $\overline{P}_2$, but this turns out to be unnecessary for constructing the money metric utility $\widetilde{V}$.[12] As a result $\widetilde{V}$, like our original expression of the utility function V, depends on Barten scales only through their interactions with prices.[13]

Let $\widetilde{V}_i$ denote the money metric utility of a consumer i. The range of social welfare functions over money-metric utilities that we consider are in the Atkinson (1970) Mean-of-Order-r class, defined by

$$S_r(\widetilde{V}_1...\widetilde{V}_N) = \left(\frac{1}{N}\sum (\widetilde{V}_i)^r\right)^{1/r} \text{ for } r \neq 0, \text{ and } S_r(\widetilde{V}_1...\widetilde{V}_N) = \exp\left(\frac{1}{N}\sum \ln \widetilde{V}_i\right) \text{ for } r = 0.$$

We use $r = -1, 0, 1$ corresponding to the harmonic, geometric and arithmetic mean of individual money metric utility. The social welfare function S_1 is inequality neutral, while S_0 and S_{-1} are inequality averse.

Table 5 provides estimates of the value of these social welfare functions defined over utilities when facing initial prices $\overline{P}_1 = 1, \overline{P}_2 = 1$ and new prices $P_1 = 2, P_2 = 1$, holding the population distribution of M_i U_{1i} and U_{2i} fixed. Simulated parametric standard errors are provided in *italics*.[14] Table 5 includes estimates of the social welfare functions S_r at both initial and new prices, and also reports the social welfare loss associated with the energy price increase, defined as the difference in the estimated social welfare function S_r between the two price regimes. In addition, we report the loss difference, defined as the difference in estimated losses based on S_r versus the estimated losses based on S_1. The estimated loss is a social welfare measure of the cost of the hypothetical energy tax, while the loss difference measures the impact of accounting for inequality aversion on these estimated costs.

As one would expect, the estimated welfare losses based on the S_1 social welfare function (the arithmetic mean of money metrics) confirm the patterns associated with the mean estimates

---

[12]This convenient feature does not always hold, e.g., the social welfare calculations in Jorgenson, Lau, and Stoker (1982) use representations of Barten scaled utility functions that do not have this property.

[13]We do one final adjustment to money-metric utilities, which is to scale them by a factor which makes the average money metric at initial prices the same for the models with and without unobserved preference heterogeneity. This makes the inequality-neutral social welfare function S_1 have the same value for both models at base prices, which makes comparisons across the specifications easier to interpret.

[14]Note that for the model without heterogeneity, the estimated level of social welfare at base prices does not have sampling variability induced by the estimation of the parameter vector. This is because, by construction, heterogeneity is only induced by variation in M_i, and in particular will for S_1 just equal the mean of M_i at base prices, which is 1.035. To facilitate comparisons across models with and without heterogeneity, our normalization of individual utilities likewise normalizes the S_1 estimates at base prices with heterogeneity to equal 1.035, the average of M_i.

of π in Table 4. In particular, the estimated loss based on S_1 increases when we account for unobserved heterogeneity in the model. As reported in the first column of Table 5, the estimated welfare loss based on S_1 increases from 10.3% without heterogeneity to 14.9% with heterogeneity.

Table 5: Social Welfare Values

		arithmetic		geometric		harmonic	
		S_1	*std err*	S_0	*std err*	S_{-1}	*std err*
without heterogeneity	initial prices	1.035	–	0.907	–	0.763	–
	new prices	0.928	*0.035*	0.809	*0.024*	0.675	*0.017*
	loss	0.103	*0.033*	0.108	*0.026*	0.115	*0.023*
	loss difference			-0.005	*0.009*	-0.012	*0.019*
with heterogeneity	initial prices	1.035	–	0.698	*0.005*	0.430	*0.007*
	new prices	0.880	*0.013*	0.606	*0.005*	0.387	*0.002*
	loss	0.149	*0.013*	0.132	*0.013*	0.101	*0.012*
	loss difference			0.017	*0.003*	0.048	*0.005*

However, a different story emerges when we include inequality aversion in the social welfare calculations. Without accounting for unobserved preference heterogeneity, we see that inequality in total expenditures M_i does not affect estimated welfare losses very much. Welfare loss is 10.3% under S_1, 10.8% under S_0, and 11.5% under S_{-1}, and the differences between these measures are statistically insignificant. When we fail to account for unobserved heterogeneity, it appears greater inequality aversion is associated with no change, or possibly a small increase, in the estimated social welfare cost of the energy tax.

In contrast, once we take unobserved preference heterogeneity into account, estimated welfare losses *decrease* as the degree of inequality aversion increases, from 14.9% under S_1 to 13.2% under S_0 to 10.1% under S_{-1}. The loss differences between these numbers are strongly statistically significant.

Essentially, the impacts of inequality due to variation in preferences mitigates the impacts of inequality due to variation in income. The reason is that, without allowing for unobserved preference variation, every poor consumer appears to be hit harder by an energy tax than every rich consumer. Accounting for preference variation shows that some wealthy people who consume a lot of energy or are more energy price inelastic will be hurt more, in relative terms, than some poorer consumers who use relatively less energy or who are more energy price elastic.

These patterns can be seen in Figure 8. In the model without unobserved preference heterogeneity, the induced consumer surplus loss increases somewhat with total expenditures M, but does not have much variation overall. However, in the model with unobserved heterogeneity, some rich households have very large utility (money-metric) impacts, which drives down S_1 substantially. Since risk averse social welfare functions downweight the utilities of the rich, these individuals in the upper right of the distribution shown in Figure 8 don't influence social welfare as much, and so the social welfare loss based on inequality averse measures like S_0 and S_{-1} is reduced. The main point here is that failing to account for preference heterogeneity completely misses the potential mitigating effect that variation in preferences from causes other than income has on inequality averse social welfare calculations.

23

Overall, our energy tax experiment yields two major conclusions. First, accounting for unobserved preference heterogeneity has a big impact on how much variation we find in the cost-of-living impacts of price changes. In our example, the standard deviation of cost-of-living impacts due to a 100% energy tax is 5 times as large in the model with unobserved preference heterogeneity as in the model without unobserved preference heterogeneity. Second, we find that accounting for unobserved preference heterogeneity reverses how the estimated social welfare loss of an energy tax varies with inequality aversion.

5   Conclusions and Extensions

We have shown nonparametric identification of a generalized random coefficients model, and provided an empirical application 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 application to Barten scales, allowing for general forms of unobserved heterogeneity is shown to be important for empirically evaluating the welfare effects of potential policy interventions such as a carbon tax. For example, we find that failure to account for preference heterogeneity would result in underestimating the total cost of an energy tax (measured as the effect on total consumer surplus) by over ten percent, and in underestimating the variation in impacts (measured as the standard deviation in cost of living impacts across consumers) by over eighty percent. We also find that unobserved preference heterogeneity strongly interacts with inequality aversion in social welfare calculations, reversing conclusions that would have been made in models that failed to account for preference heterogeneity.

Our application focused on consumers with single utility functions. A possible extension would be to consider collective household models, e.g., Barten scales have been applied to household models in, e.g., Browning, Chiappori, and Lewbel (2010). It would also be useful to place our carbon tax analyses into a general equilibrium setting.

Useful areas for further work on the theory of generalized random coefficients would be extensions to nonparametrically 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.

6   Appendix A: Econometric Extensions

Here we present some extensions of our identification theorems that are not directly relevant to Barten scales, but would be useful in other applications. One 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. This could have been applied to quantity demands as an alternative to our logit transformed budget shares.

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 $k \in \{1, ..., K\}$ 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.

Another straightforward extension of our model is to combine Theorems 1 or 2 with a special regressor as in Lewbel (2000) to identify discrete choice models like $D = I[V + G(X_1 U_1, ..., X_K U_K) + U_0 \geq 0]$, where V is an exogenous regressor with large support and D is an observed binary dependent variable. In this case, even though Y is not observed, if $V \perp U \mid X$ Then $1 - E(D \mid V = v, X = x, Z = z) = F_{Y|X,Z}(-v \mid x, z)$ identifies $F_{Y|X,Z}$, and this paper's theorems can then by applied to identify G and each U_k distribution. This structure is related to that of Berry and Haile (2009), and could be useful in applications that involve discrete choices with both substantial unobserved heterogeneity and substantial nonlinearity in the latent index.

We now consider some lengthier extensions to Theorem 1. 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 $\widetilde{Y}_k$, the function δ_k, and the set Ψ_k as follows. $\widetilde{Y}_k = G_k(X_k U_k)$. Let $\delta_k(x_k, z) = 1$ if and only if there exists a $\widetilde{y}_k \in supp(\widetilde{Y}_k \mid X = x_k e_k, Z = z)\}$ such that $F_{\widetilde{Y}_k|X,Z}(y_0 \mid e_k, z) = F_{\widetilde{Y}_k|X,Z}(\widetilde{y}_k \mid e_k x_k, z)$ if $x_k > 0$ or $F_{\widetilde{Y}_k|X,Z}(y_0 \mid e_k, z) = 1 - F_{\widetilde{Y}_k|X,Z}(\widetilde{y}_k \mid e_k x_k, z)$ if $x_k < 0$. Let $\Psi_k = \{x_k : \text{for some } z \in supp(Z), \delta_k(x_k, z) = 1\}$.

ASSUMPTION A2': The function $F_{U_k|Z}(u_k \mid z)$ is invertible in u_k for all $u_k \in supp(U_k \mid Z = z)$. 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 $\widetilde{y}_k$ such that the distribution $F_{\widetilde{Y}_k|X,Z}$ evaluated at $\widetilde{y}_k$ 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. This will generally require that X_k and U_k have comparable supports. 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 $\widetilde{y}_k$ such that $F_{\widetilde{Y}_k|X}(y_0 \mid e_k) = F_{\widetilde{Y}_k|X_k}(\widetilde{y}_k \mid 0)$. 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 $\widetilde{y}_k$),

Assumption A2' will be satisfied, because only three points are required to identify a quadratic function. 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 $k \in \{1, ..., K\}$ 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

$$Y = \left(\sum_{k=1}^{K} G_k (X_k U_k)\right) + \left(\sum_{j=1}^{K-1} \sum_{k=j+1}^{K} X_j X_k U_{jk}\right) + U_0 \qquad (21)$$

Equation (21) 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 $F_{Y|X,Z} (y \mid x, z)$ and the marginal distribution $F_Z (z)$ are identified. $(U_0, U_1, ..., U_K, U_{12}, U_{13}, ...U_{K-1,K}) \perp X \mid Z$,
$(U_1, ..., U_K, U_{12}, U_{13}, ...U_{K-1,K}) \perp U_0 \mid Z$, the variables $U_0, U_1, ..., U_K$ 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 nonvanishing characteristic function or U_0 is identically zero. $supp (U_0) \subseteq supp (Y)$ and $\{0, e_{11}, e_{12}, ..., e_{KK}\} \subseteq supp (X)$. For all $j \neq k$, $G_k (U_k) + G_j (U_j) + U_0$ has a nonvanishing characteristic function.

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

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 $U_0, U_1, ..., U_K$ be mutually independent as in Assumption A4, 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 (21) and Assumptions B1, B2, and (for $k = 1, ..., K$) A3 hold. Then the distribution function $F_{U_0|Z}$, the regression and distribution functions G_k and $F_{U_k|Z}$ for every $k \in \{0, 1, ..., K\}$, and the distribution functions $F_{U_{jk}|Z}$ for every $j = 1, ..., K - 1$ and $k = j + 1, ..., K$, 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   Appendix B: Proofs

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

LEMMA 1: Let $\widetilde{Y}_k = G_k(X_k U_k)$ where G_k is a strictly monotonically increasing function. Assume $U_k \perp X \mid Z$. 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 $\widetilde{Y}_k$, and for every r such that $G_k(r)$ is on the support of $\widetilde{Y}_k$ there exist an $x_k \neq 0$ on the support of X_k such that $f_{U_k}\left(x_k^{-1}r\right) \neq 0$. 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 $\widetilde{Y}_k$ are imposed. Let $r = H_k(\widetilde{y}_k)$ be inverse of the function G_k where $\widetilde{y}_k = G_k(r)$. Define $X_{(k)}$ to be the vector of all the elements of X except for X_k. Define the function $S_k(\widetilde{y}_k, \widetilde{x})$ by

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

Then

$$H_k(\widetilde{y}_k) = sign\left(sign(x_k) \frac{\partial S_k\left(\widetilde{y}_k, x_k^{-1}\right)}{\partial x_k^{-1}}\right) \exp\left(\int_{y_0}^{\widetilde{y}} \frac{x_k \partial S_k\left(\widetilde{y}_k, x_k^{-1}\right)/\partial \widetilde{y}}{\partial S_k\left(\widetilde{y}_k, x_k^{-1}\right)/\partial x_k^{-1}} d\widetilde{y}_k\right) \quad (22)$$

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 $S_k(\widetilde{y}_k, \widetilde{x}) = F_{\widetilde{Y}_k|X_k, X_{(k)}}\left(\widetilde{y}_k \mid \widetilde{x}^{-1}, 0\right)$. The main implication of Lemma 1 is that if the distribution $F_{\widetilde{Y}_k|X, Z}$ is identified, then the function H_k is identified by construction.

PROOF of Lemma 1: For any $\widetilde{y}_k = G_k(x_k U_k)$ and any $x_k > 0$ we have

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

27

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

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

Together these equations say

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

So

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

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

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

and

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

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

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

so

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

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

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

So the right side of equation (22) equals $sign\,(H_k\,(\widetilde{y}_k))\,|H_k\,(\widetilde{y}_k)\,| = H_k\,(\widetilde{y}_k)$ as claimed.

LEMMA 2: If Assumption A1 holds and the normalization $G_k\,(0) = 0$ for all k holds, then $F_{U_0|Z}$ and the distribution function $F_{\widetilde{Y}|X,Z}\left(\widetilde{Y}\mid x, z\right)$ are identified, where $\widetilde{Y} = \sum_{k=1}^{K} G_k\,(X_k U_k)$.

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 $\widetilde{Y} = Y - U_0$. Then since $Y = \widetilde{Y} + U_0$ 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 $\widetilde{Y}\mid X, Z$, where $\widetilde{Y} = \sum_{k=1}^{K} G_k\,(X_k U_k)$.

PROOF of Theorem 1: When $X_{(k)} = 0$ (equivalently, when $X = e_k x_k$ for some x_k) we get $\widetilde{Y} = G_k\,(X_k U_k) + \sum_{j\neq k} G_j\,(0) = G_k\,(X_k U_k)$. Define $\widetilde{Y}_k = G_k\,(X_k U_k)$. It follows that $F_{\widetilde{Y}_k|X_k,X_{(k)},Z}\,(\widetilde{y}_k\mid x_k, 0, z) = F_{\widetilde{Y}|X,Z}\,(\widetilde{y}_k\mid x_k e_k, z)$, so the distribution function on the left of this identity is identified, given by Lemma 2 that $F_{\widetilde{Y}|X,Z}$ is identified. Let $r = H_k\,(\widetilde{y}_k)$ denote the inverse of the function G_k where $\widetilde{y}_k = G_k\,(r)$. It follows by construction from Lemma 1 that $H_k\,(\widetilde{y}_k)$ is identified for every value of $\widetilde{y}_k$ on the support of $\widetilde{Y}_k$ satisfying the property that, for some x_k on the support of X_k, $f_{U_k}\left[x_k^{-1}H\,(\widetilde{y}_k)\right] \neq 0$. This identification of $H_k\,(\widetilde{y}_k)$ in turn means that the function $G_k\,(r)$ is identified for every r such that $G_k\,(r)$ is on the support of $\widetilde{Y}_k$ 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 $F_{\widetilde{Y}|X,Z}$ and of $H_k\,(\widetilde{y}_k)$, the distribution function $F_{U_k|Z}$ is identified by $F_{U_k|Z}\left[H\,(\widetilde{y})/x_k\mid z\right] = F_{\widetilde{Y}|X_k,X_{(k)},Z}\,(\widetilde{y}\mid x_k, 0, z)$ for $x_k > 0$ and $F_{U_k|Z}\left[H\,(\widetilde{y})/x_k\mid z\right] = 1 - F_{\widetilde{Y}|X_k,X_{(k)},Z}\,(\widetilde{y}\mid x_k, 0, z)$ 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 $F_{\widetilde{\widetilde{Y}}|X,Z}\left(\widetilde{\widetilde{Y}}\mid x, z\right)$ are identified, where $\widetilde{\widetilde{Y}} = \prod_{k=1}^{K} g_k\,(X_k U_k)$. It therefore follows that $F_{\widetilde{Y}|X,Z}\left(\widetilde{Y}\mid x, z\right)$ is identified where $\widetilde{Y} = \ln\left(\widetilde{\widetilde{Y}}\right) = \sum_{k=1}^{K} \ln\left[g_k\,(X_k U_k)\right] = \sum_{k=1}^{K} G_k\,(X_k U_k)$, and the remainder of the identification therefore follows applying the proof of Theorem 1.

PROOF of Theorem 2: By construction, the function $\widetilde{G}\,(X_1 U_1, ..., X_K U_K)$ 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 conditional distributions of each U_k given Z, and identification of each function G_k only using $X = 0$ and $X = X_k e_k$, so these functions are also identifed for equation (3). What remains is to identify the function $\widetilde{G}$. Define $R\,(X, Z) = E\left[\widetilde{G}\,(X_1 U_1, ..., X_K U_K)\mid X, Z\right]$. The function $R\,(X, Z)$ is identified for all X because $R\,(X, Z) = E\left[Y - \sum_{k=1}^{K} G_k\,(X_k U_k) - U_0\mid X, Z\right]$,

which depends only on distributions and functions that have already been identified. For non-negative integers $t_1, ...t_K$ define $R_{t_1,...t_K}$ by

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

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

$$R_{t_1,...t_K}(x, z) = E\left(U_1^{t_1} \times ... \times U_K^{t_K} \widetilde{G}_{t_1,...t_K}(x_1 U_1, ..., x_K U_K) \mid X = x, Z = z\right)$$

and $E\left(U_1^{t_1} \times ... \times U_K^{t_K} \mid X = x, Z = z\right) = E\left(U_1^{t_1} \mid Z = z\right) \times ... \times E\left(U_K^{t_K} \mid Z = z\right)$ has already been identified from Theorem 1. Therefore $\widetilde{G}_{t_1,...t_K}(0) = R_{t_1,...t_K}(0) / E\left(U_1^{t_1} \times ... \times U_K^{t_K} \mid Z = z\right)$ is identified for all sets of nonnegative integers $t_1, ...t_K$. Now $\widetilde{G}$ is entire and analytic, so it equals its Maclaurin series expansion

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

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

PROOF of Corollary 2: For a given $j \in \{1, ..., J\}$ 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 $-\omega_j(U_1 X_1, ..., U_J X_J) / U_j X_j$ which makes the function and G_j in Theorems 1 and 2 be $-\omega_j(0, ..., 0, U_j X_j, ..., 0) / U_j X_j$ (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 $j \in \{1, ..., J\}$ 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 $Q_j X_j = \omega_j(U_1 X_1, ..., U_J X_J)$ for $j \in \{1, ..., J\}$ 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 (5), given any indirect utility function V, the corresponding demand function ω_1 is given by

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

30

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

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

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

$$V\left(X_1, X_2\right) = \int_{-\infty}^{\ln X_1} e^{g_1(x_1)} d\ln x_1 + \int_{-\infty}^{\ln x_2} e^{-g_2(X_2)} d\ln x_2. \tag{24}$$

Substituting equation (24) into equation (23) gives

$$\lambda\left[\omega_1\left(X_1, X_2\right)\right] = g_1\left(X_1\right) + g_2\left(X_2\right) \tag{25}$$

which shows that, up to monotonic transformation, equation (24) is the indirect utility function that generates the demand equation (25). Since equation (24) is additive, this shows that the indirect utility function that generates the demand equation (25) is additive.

To go the other direction, given any differentiable functions $h_1\left(X_1\right)$ and $h_2\left(X_2\right)$, if $V\left(X_1, X_2\right) = h_1\left(X_1\right) + h_2\left(X_2\right)$ equation (23) equals

$$\lambda\left[\omega_1\left(X_1, X_2\right)\right] = \ln\left(\frac{\partial h_1\left(X_1\right)}{\partial \ln X_1}\right) - \ln\left(\frac{\partial h_2\left(X_2\right)}{\partial \ln X_2}\right) \tag{26}$$

which is in the form of equation (25), showing that any additive indirect utility function generates a demand equation in the form of (25).

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 (10) and (9). 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 $h_1\left(U_1 P_1 / M\right) + h_2\left(U_2 P_2 / M\right)$ must possess the standard properties of all indirect utility functions, i.e., homogeneity and quaisconvexity 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 $\widetilde{Y} = G_k\left(X_k U_k\right) + \sum_{j \neq k} G_k\left(0\right) = G_k\left(X_k U_k\right)$. Define $\widetilde{Y}_k = G_k\left(X_k U_k\right)$. It follows that $F_{\widetilde{Y}_k | X_k, X_{(k)}, Z}\left(\widetilde{y}_k \mid x_k, 0, z\right) = F_{\widetilde{Y} | X, Z}\left(\widetilde{y}_k \mid x_k e_k, z\right)$, so $F_{\widetilde{Y}_k | X_k, X_{(k)}, Z}\left(\widetilde{y}_k \mid x_k, 0, z\right)$ is identified, given by Lemma 1 that $F_{\widetilde{Y} | X, Z}$ is identified. Let $r = H_k\left(\widetilde{y}_k\right)$ be inverse of the function G_k where $\widetilde{y}_k = G_k\left(r\right)$. Now consider any particular positive $x_k \in \Psi_k$. For that x_k we have $F_{\widetilde{Y}_k | X, Z}\left(y_0 \mid e_k, z\right) = F_{\widetilde{Y}_k | X, Z}\left(\widetilde{y}_k \mid e_k x_k, z\right)$ and since the function $F_{\widetilde{Y}_k | X, Z}$ is identified, the particular value $\widetilde{y}_k$ that satisfies this equation is identified. Then

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

31

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

$$
\begin{aligned}
1 - \Pr\left(G_k\left(x_k U_k\right) \le \widetilde{y}_k \mid X = x_k e_k, Z = z\right) &= \Pr\left(G_k\left(U_k\right) \le y_0 \mid X = x_k e_k, Z = z\right) \\
1 - \Pr\left(U_k \ge H_k\left(\widetilde{y}_k\right)/x_k \mid X = x_k e_k, Z = z\right) &= \Pr\left(U_k \le H_k\left(y_0\right) \mid Z = z\right) \\
F_{U_k|Z}\left[H_k\left(\widetilde{y}_k\right)/x_k, z\right] &= F_{U_k|Z}\left[H_k\left(y_0\right), z\right]
\end{aligned}
$$

By invertibility of $F_{U_1|Z}$ these equations show that for any $x_k \in \Psi_K$ we get $H_k\left(\widetilde{y}_k\right)/x_k = H_k\left(y_0\right)$ where the $\widetilde{y}_k$ corresponding to the given x_k is known. Now $G_k\left(1\right) = y_0$ means that $H_k\left(y_0\right) = 1$, so $H_k\left(\widetilde{y}_k\right) = x_k$, and therefore $\widetilde{y}_k = G_k\left(x_k\right)$, 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 $F_{\widetilde{Y}|X,Z}$ and of $H_k\left(\widetilde{y}\right)$, the distribution function $F_{U_k|Z}$ is identified by $F_{U_k|Z}\left[H\left(\widetilde{y}_k\right)/x_k \mid z\right] = F_{\widetilde{Y}_k|X,Z}\left(\widetilde{y}_k \mid e_k x_k, z\right)$ for $x_k > 0$ and $F_{U_k|Z}\left[H\left(\widetilde{y}_k\right)/x_k \mid z\right] = 1 - F_{\widetilde{Y}_k|X,Z}\left(\widetilde{y}_k \mid e_k x_k, z\right)$ 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 $V_{jk} = U_0 + G_k\left(U_k\right) + G_j\left(U_j\right)$ At this stage the distribution of $V_{jk} \mid Z$ is identified (because each component is identified), so $F_{U_{jk}|Z}$ can be identified by a deconvolution of $Y \mid Z$ with $V_{jk} \mid Z$.

8 References

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

Atkinson, A. B., (1970), "On the Measurement of Inequality," Journal of Economic Theory, 2, 244-263.

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 Identi cation 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 Coeffient 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. Parey, (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.

Chiappori, P.A. and Ekeland, I., (2006) "The micro economics of group behavior: General characterization," Journal of Economic Theory, 130, 1-26.

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

Crossley, T. F. and K. (2010), "The Common-Scaling Social Cost of Living Index", Journal of Business and Economic Statistics, 28, 523-538.

Deaton, A. and J. Muellbauer (1980), Economics and Consumer Behavior, Cambridge University Press, Cambridge.

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.

Gorman, W. M., (1976) "Tricks with Utility Functions," in Essays in Economic Analysis, edited by M. Artis and R. Nobay, Cambridge University Press, Cambridge.

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 R. Robb, (1985) "Alternative Methods for Estimating The Impact of Interventions," In James J. Heckman and Burton Singer (Eds.), Longitudinal Analysis of Labor Market Data, Cambridge: Cambridge University Press.

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

Heckman, J. J. and E. Vytlacil, (1998), "Instrumental Variables Methods for the Correlated Random Coefficient Model: Estimating the Average Rate of Return to Schooling When the Return is Correlated with Schooling," Journal of Human Resources, 33, 974-987.

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 coeffcients 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. (2000), "Semiparametric Qualitative Response Model Estimation With Unknown Heteroscedasticity or Instrumental Variables," Journal of Econometrics, 97, 145-177.

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.

Meister, A. (2005), "Non-estimability in spite of identifiability in density deconvolution,". Mathematical Methods of Statistcs, 14, 479-487.

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: Additive Utility

Distribution of lnU1


U0, lnU1, lnU2 are 2nd order hermite expansions from normal
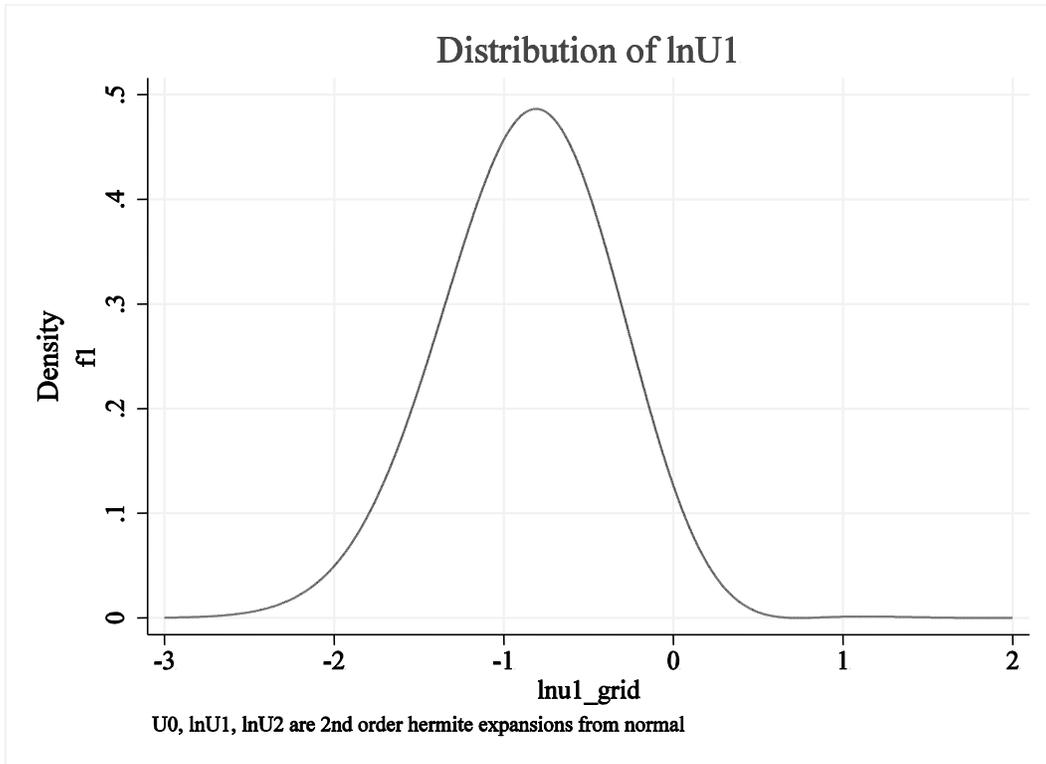
Figure 2: Additive Utility

Distribution of lnU2

U0, lnU1, lnU2 are 2nd order hermite expansions from normal
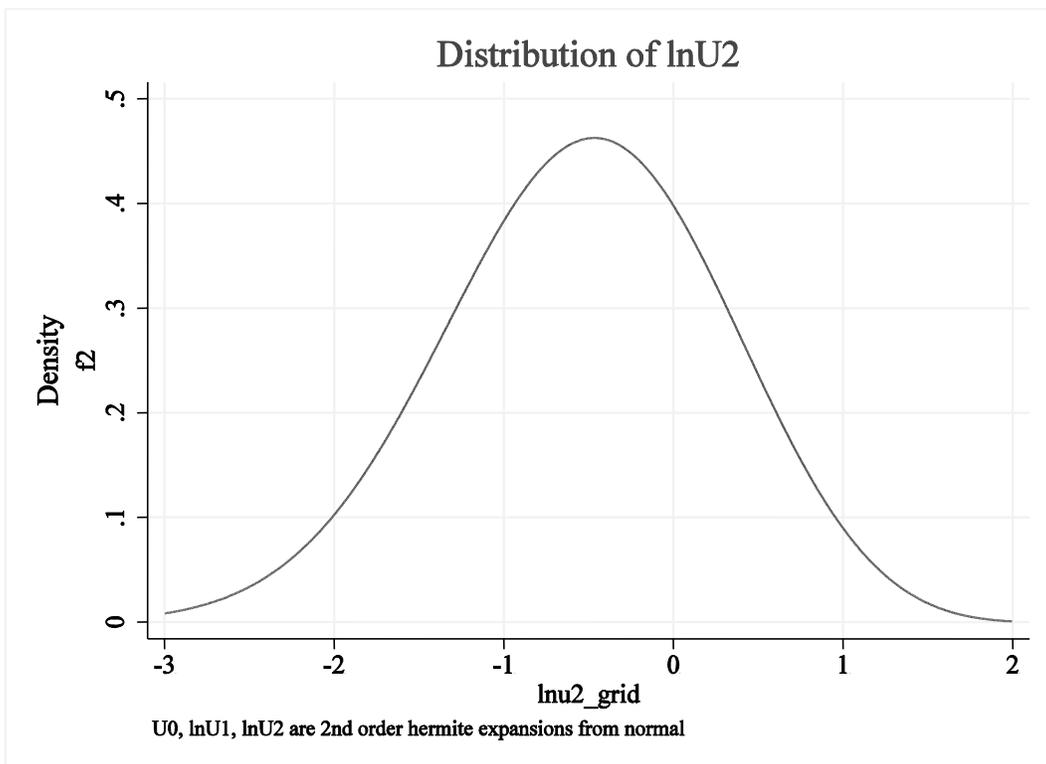
Figure 3: Additive Utility With Interaction Terms

Distribution of lnU1

U0, lnU1, lnU2 are 2nd order hermite expansions from normal
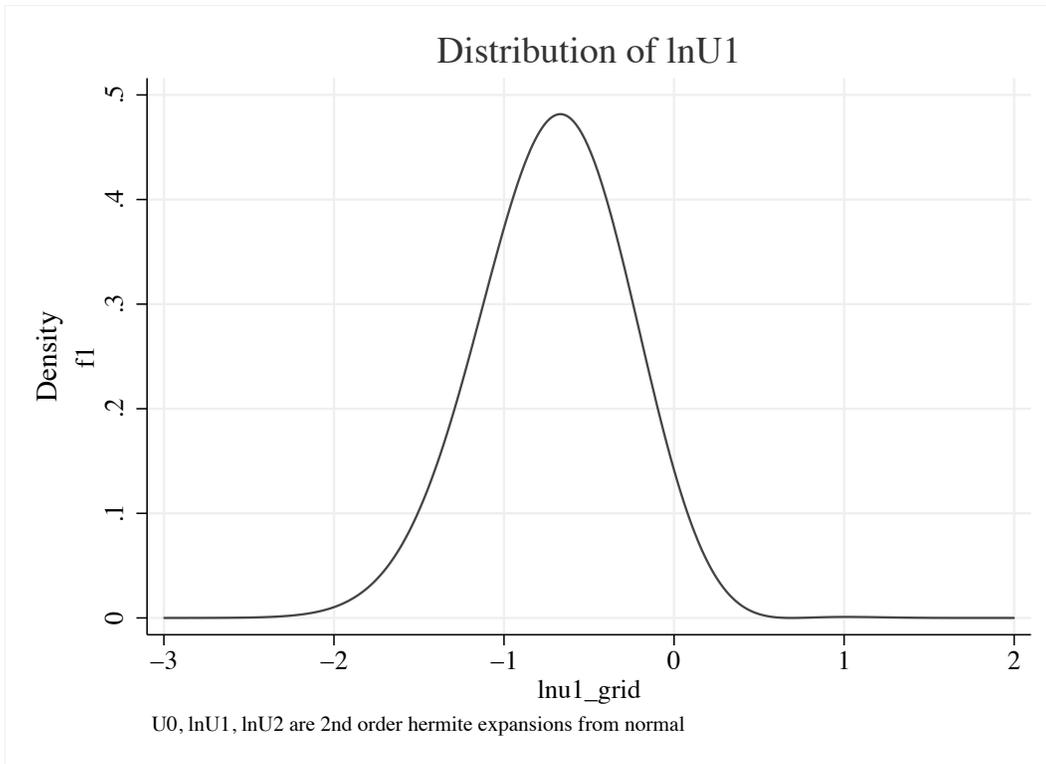
Figure 4: Additive Utility With Interaction Terms

Distribution of lnU2

U0, lnU1, lnU2 are 2nd order hermite expansions from normal
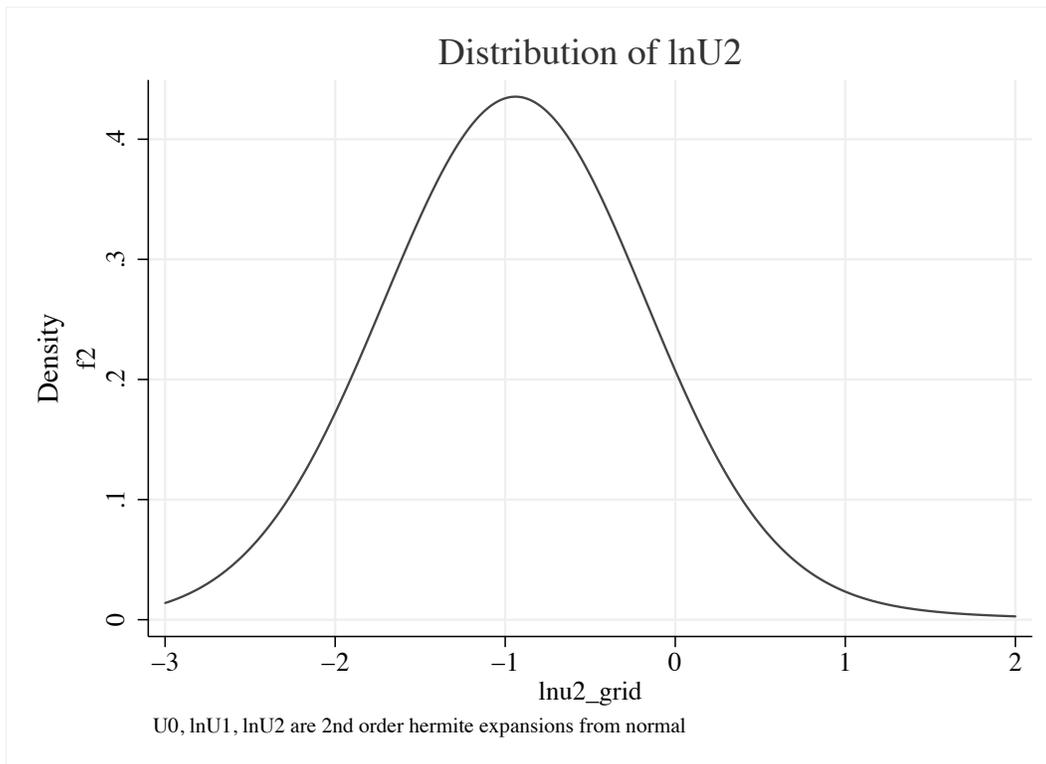
Figure 5: Predicted Engel Curves

SML Estimates with Interaction, Quartiles of U1,U2

Single Adult Households, Ontario 2002

Energy Share

lnM

Nominal Total Expenditure

SML at quartiles of U1,U2 in thin black lines; SML without het. in thick gray line
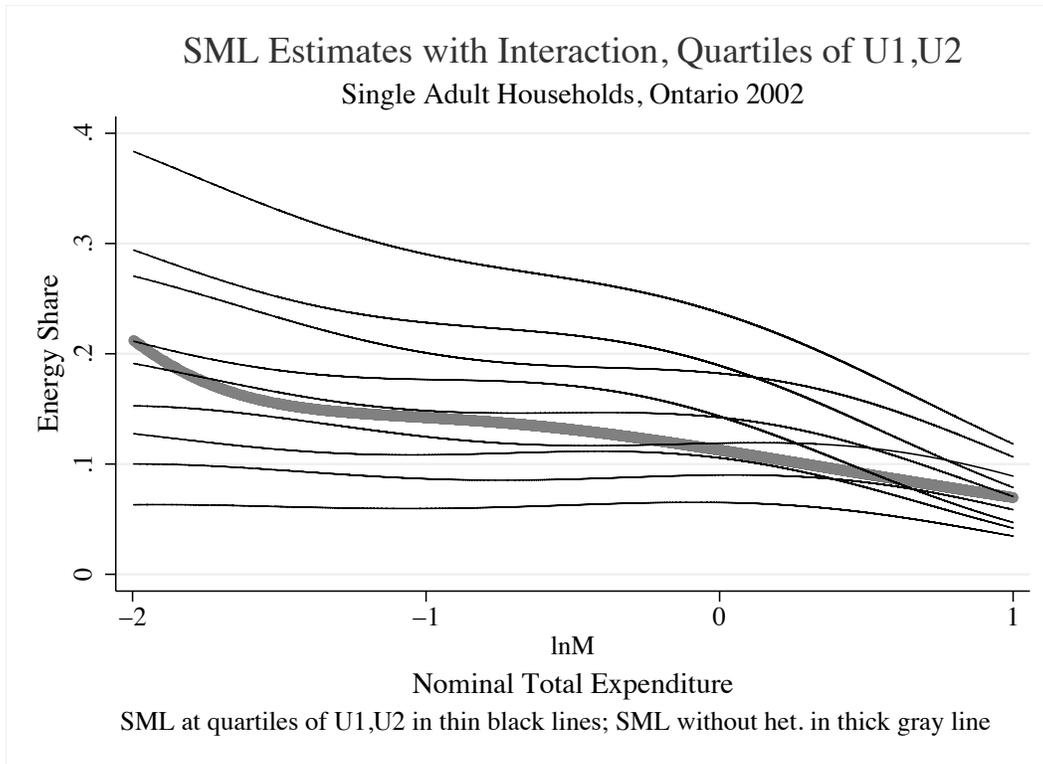
Figure 6: Distribution of Predicted Budget Shares

SML Estimates with Interaction

Single Adult Households, Ontario 2002

Energy Budget Share

Log Total Expenditure

Contours of distribution in thin black lines; SML without het. in thick gray line
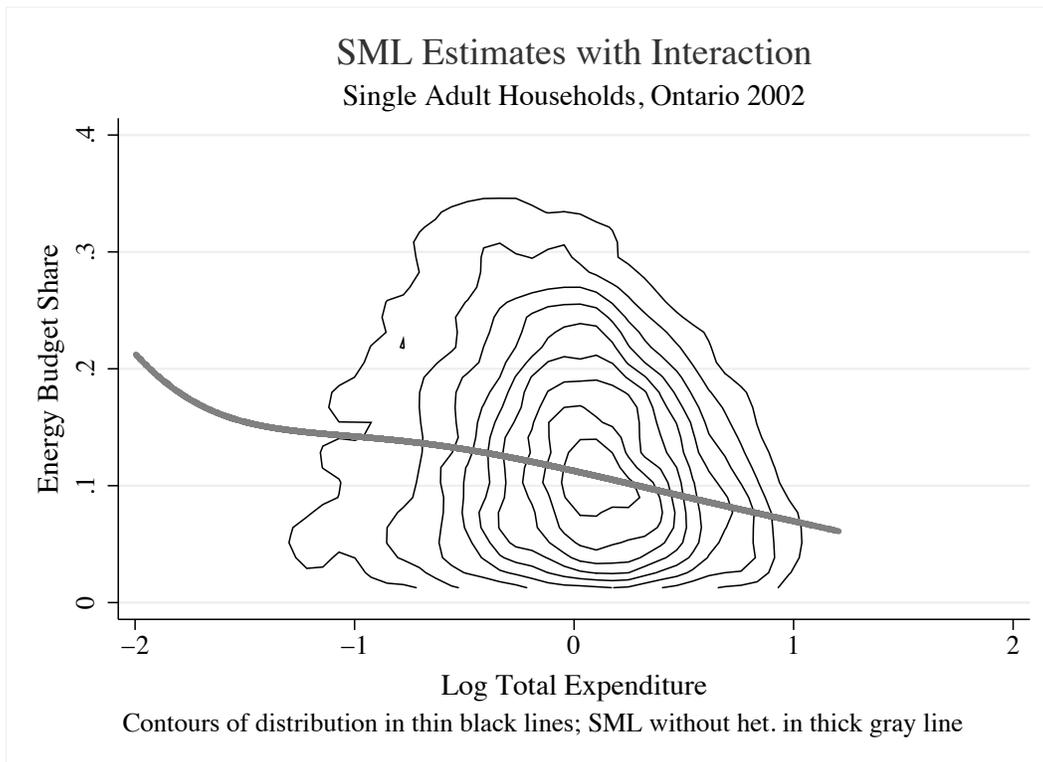
Figure 7: Distribution of Log Cost-of-Living Impact

SML Estimates with Interaction

Single Adult Households, Ontario 2002

Contours of distribution in thin black lines; SML without het. in thick gray line
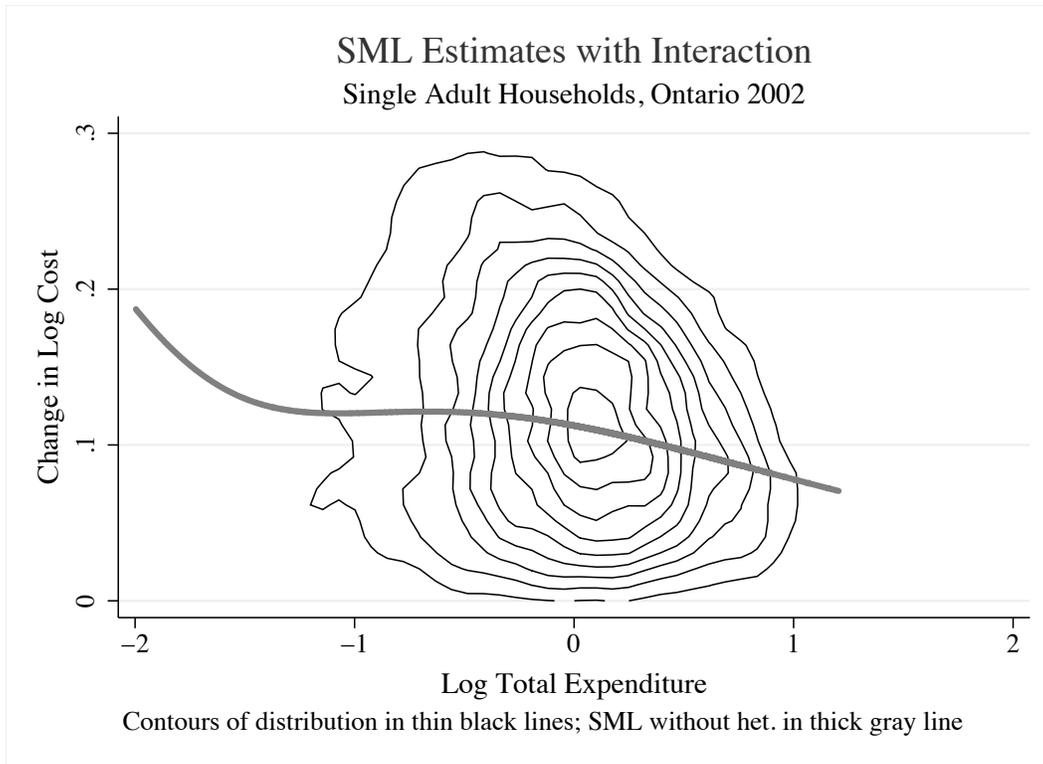
Figure 8: Distribution of Consumer Surplus

SML Estimates with Interaction

Single Adult Households, Ontario 2002

Contours of distribution in thin black lines; SML without het. in thick gray line