Nonlinear Random Coefficients

Arthur Lewbel Boston College

original Feb. 2016, revised September 2016

Abstract

Standard random coefficients models are either linear in regressors, or equal a transformation of a linear index of regressors (e.g., random coefficient logit models). In contrast, this paper shows identification of, and consistent estimators for, general nonlinear random coefficients models with unknown parameters. For example, we consider a model that includes interaction terms in regressors and nonlinear transformations of regressors, where each regressor has a random coefficient, and the joint distribution of the random coefficients is unknown.

This packet contains a preliminary, incomplete draft of new work on Nonlinear Random Coefficients, and a closely related forthcoming JPE paper.

JEL codes: C14 C21 Keywords: unobserved heterogeneity, nonseparable errors, random utility parameters, random coefficients,

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

Standard random coefficients models are either linear in regressors, or equal a transformation of a linear index of regressors (e.g., random coefficient logit models). In contrast, this paper shows identification of, and consistent estimators for, nonlinear random coefficients models with unknown parameters.

Suppose an observed variable Y depends on a vector of observed regressors $X = (X_1, ..., X_K)$, and on a set of unobserved errors $U = (U_1, ..., U_K)$. Assume independent, identically distributed observations of X, Y. Consider a general nonlinear random coefficients model given by

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

where G is a known, nonlinear function, and θ is a parameter vector of unknown constants. The distribution of U is also unknown. Nonlinear relationships are ubiquitous in statistics, so

it is natural to extend linear random coefficients models like $Y = \sum_{k=1}^{K} X_k U_k$ to allow for nonlinearity. The basic idea of equation (1) is that, just as in the linear random coefficients model, each regressor X_k has a random coefficient U_k, but now Y is no longer constrained to being a linear function of X. The nonlinear terms are what may give rise to the additional parameters θ that would need to be estimated.

This paper provides conditions that suffice for identification of θ and the distribution of U. Consistent estimators of θ and of moments of U are provided. To illustrate the application of these results, we consider the nonlinear random coefficients model

$$Y = g\left(\theta_1 \left(X_1 U_1\right)\left(X_2 U_2\right) + \theta_2 \ln\left(X_2 U_2\right) + \theta_3 X_1 U_1 + \theta_4\right) \tag{2}$$

where $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$ and g is any known, strictly monotonic function. This example is chosen not for its behavioral significance, but rather to illustrate in a single model how our identification methodology deals with multiple types of nonlinearities, by including not just a linear term $X_1 U_1$, but also a transformation function g, an interaction term between $X_1 U_1$ and $X_2 U_2$, and a nonlinear transformation of $X_2 U_2$.

Since the joint distribution of U is unknown, identification of this distribution along with θ parameters is extremely challenging. Even in models without unknown parameters θ, very few results exist on nonparametric identification of the distribution of U, indeed, almost all existing results impose extremely strong restrictions such as linearity in X (see below for details).

The obvious approach to this problem would be to consider the conditional distribution function of Y given X, $F_{Y|X}\left(y \mid x\right)$, which equals

$$F_{Y|X}\left(y \mid x\right) = \int_{U \in supp(U)} g\left(x_1 U_1, ..., x_K U_K, \theta\right) dF_U\left(U\right) dx_1 dx_2$$

For example, one might consider sieve maximum likelihood estimation based on this equation, by replacing the unknown distribution function of U, $F_U\left(U\right)$, with a set of sieve basis functions. The problem with this approach is that it assumes identification, that is, it assumes a priori that there exists a unique function F_U and parameter vector θ that satisfies this integral equation. A similar problem would arise if we based estimation on the characteristic function of Y given X. The present paper addresses this problem by providing conditions that ensure identification. Essentially, these are restrictions on g that suffice to ensure a unique F_U and θ. This is accomplished by devising easier to solve alternative expressions involving F_U and θ.

Equation (1) does not include the usual additive error term that appears in linear random coefficients models. That is, we are not adding an additional error U_0 to the right of equations (2) or (1). In an extension we show how such an additional term could be identified, but under the restriction (which was also assumed by Beran and Hall 1992) that this added U_0 is independent of the remaining U_k random coefficients.

One example of a model in the form of equation (1) follows from Barten (1964). Assume consumers have utility functions of the form $V\left(Q_1/U_1,...Q_K/U_K, \theta\right)$, where Q_k is quantity the consumer purchases of commodity k. Let X_k be the price of commodity k divided by the total expenditures of the consumer. Maximizing utility under the standard linear budget constraint $Q'X = 1$ then yields demand functions in the form of equation (1) where for a given good k, $Y = Q_k X_k$ is the budget share of good k, G is the budget share demand function for good

k, and the elements of U are known as Barten scales. These Barten scales embody variation in tastes and preferences across consumers. Barten scales, specified as deterministic functions of observable taste shifters, have been widely used in empirical consumer demand models. See, e.g., the survey by Lewbel (1997) and references therein. Results here and in Lewbel and Pendakur (2015) allow for random variation in Barten scales, corresponding to unobserved preference heterogeneity across consumers, taking U to be random utility parameters.

Other examples could involve production function estimation, where Y is the quantity produced of some good, each X_k is the quantity of input or factor of production k (e.g., labor, materials, capital), and U_k is a measure of the quality of input k, which may vary across firms. Matzkin (1994) gives examples of production functions with a single random component of this form, but there is no economic reason to impose the restriction that only one of the factors of production would have an unobserved quality level.

Yet another potential application is to generalize the random coefficients multinomial logit model, as is widely used in economics via the BLP (Berry, Levinsohn, and Pakes (1995) model. This model writes market shares Y as logit transforms of $\sum_{k=1}^{K} X_k U_k$ where covariates X_k include prices as well as characteristics of products and of consumers. A linear random coefficients term $\sum_{k=1}^{K} X_k U_k$ corresponds to the utility associated with each choice, but it is well recognized that linearity is imposed solely for convenience in estimation, since there is no fundamental economic reason why utility should be linear. See, e.g., Berry and Haile (2009) for nonlinear examples.

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 linear simultaneous systems of random coefficients, including Masten (2015) and Hoderlein, Holzmann, and Meister (2015), 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) and Fox and Gandhi (2013). In all of these models Y is either linear in X or is a transformation of a linear index in X, i.e., specifications like $Y = G\left(X_1 U_1 + X_2 U_2\right)$.

This paper substantially extends results in Lewbel and Pendakur (2015) by allowing for unknown parameters in the specification of the regression function (in addition to unknown distributions for the random coefficients), and considering general regression models rather than focusing on additive models. Lewbel and Pendakur (2015), who consider identification of a model like equation (1) without θ. In particular, their results include a special case of Theorem 1 below. However, they focus attention on additive models like $Y = g_1\left(X_1 U_1\right) + g_2\left(X_2 U_2\right)$, and they also do not provide any associated semiparametric estimators. Another related result is Matzkin (2003) who shows identification for nonlinear random coefficients, but imposes additivity, independence among all the elements of U, and additional restrictions on the additive components. Also related is Hoderlein, Nesheim, and Simoni (2011), who provide a high level condition they call T-completeness that suffices for nonparametric identification of a vector of random parameters within a known function. They provide some examples where T-completeness can be shown to hold, but these are very restrictive classes of models such as when the random parameter distributions are in the exponential family, or are parameterizable by a single scalar. Our results can be interpreted as a method of showing how T-completeness

could be satisfied in a much larger class of models.

2 Identification

Let Y be a dependent variable and let X be a vector of covariates $X = (X_1, ..., X_K)$. For reasons that will be clear later, it is more convenient for us to write the random coefficients model as

$$Y = G\left(X_1/V_1, ..., X_K/V_K, \theta_0\right) \tag{3}$$

instead of the more usual equivalent form $G\left(X_1U_1, ..., X_KU_K, \theta_0\right)$. Let $F_v\left(v\right)$ denote the joint distribution of the random coefficients V. We begin with a generalization of results in Lewbel and Pendakur (2015).

ASSUMPTION A1: $\{X_i, Y_i\}$ for $i = 1, ..., n$ are independent and identically distributed with joint distribution $F\left(Y, X\right)$. Let θ_0 be the unknown true value of an unknown J vector of parameters $\theta = (\theta_1, ..., \theta_J)$. Equation (3) holds for some known continuous function G and some unobserved vector of random coefficients V, where $V \perp X$.

ASSUMPTION A2: $supp\left(V\right)$ is rectangular. $supp\left(X\right)$ is rectangular, includes the origin, and,
$supp\left(X_k/V_k, ...X_K/V_K \mid V = v\right) = supp\left(X\right)$ for all $v \in supp\left(V\right)$.

Since $V \perp X$, one way that Assumption A2 would be satisfied is if, for each $k = 1, ..., K$ either $supp\left(X_k\right) = \mathbb{R}$ and $supp\left(V_k\right)$ does not include zero, or $supp\left(X_k\right) = \mathbb{R}_+^K$ and V_k is strictly positive. Note that neither X nor V is assumed to be continuously distributed. Here V can be continuous, discrete, continuous with mass points, etc. We require that the support of each X_k be an interval so X cannot be discrete, but its distribution can, e.g., contain mass points.

Let X^P denote a vector consisting of a subset of elements of X, and let X^{-P} denote the vector of elements of X that are not contained in X^P. So a choice of X^P corresponds to a choice of partition P of the elements of X. Let $P \in \mathcal{P}$ denote a set of such partitions of X. Let $K^P \in \{1, ..., K\}$ denote the number of elements in X^P. In a simplifying abuse of notation, for any choice of $P \in \mathcal{P}$ let the elements of x be reordered so $X^P = \left(X_1, ...X_{K^P}\right)$ and $X^{-P} = \left(X_{K^P}, ...X_K\right)$. Let $G^P\left(s^P, \theta\right)$ denote the function G where the elements corresponding to X^P are evaluated at the corresponding elements of $s = \left(s_1, ..., s_{K^P}\right)$, and the elements corresponding to X^{-P} are set equal to zero. So, e.g., if the model were $G\left(X_1/V_1, ..., X_K/V_K, \theta\right) = X_1/V_1 + \left(X_1/V_1\right)^2\theta_1 + X_2/V_2$ with $X^P = (X_1)$, then we would have $X^{-P} = (X_2)$ and $G^P\left(s^P, \theta\right) = s_1 + s_1^2\theta_1$.

Let $j \in J$ index a set of functions h_j. Let $t \in T$ index a set of complex valued vectors $t = (t_1, ..., t_K)$ where each $t_k = t_{k0} + it_{k1}$. For a given function G, a given partition $P \in \mathcal{P}$, a given $j \in J$ and a given $t \in T$, define the function κ_{pjt} by

$$\kappa_{Ptj}\left(\theta\right) = \int_{s^P \in supp\left(X^P\right)} h_j\left(G^P\left(s^P, \theta\right)\right) s_1^{t_1-1}s_2^{t_2-1}...s_{K^P}^{t_{K^P}-1}ds_1ds_2...ds_{K^P} \tag{4}$$

ASSUMPTION A3: Given G, for a given partition $P \in \mathcal{P}$, a given $t \in T$ and a given $j \in J$ and, we can find a continuous function h_j such that
$\int_{s^P \in supp(X^P)} h_j\left(\left|G^P\left(s^P,\theta\right)\right|\right) s_1^{t_1-1} s_2^{t_2-1} ... s_{K^P}^{t_{K^P}-1} ds_1 ds_2 ... ds_{K^P}$ exists, the integral in equation (4) is convergent and has $\kappa_{Ptj}(\theta) \neq 0$ for all θ in a neighborhood of θ_0.

Assumption A3 assumes we can find a functions h_j that makes $h_j\left(G^P\left(s^P,\theta\right)\right) s_1^{t_1-1} s_2^{t_2-1} ... s_{K^P}^{t_{K^P}-1}$ integrable, and requires that the definite integral in equation (4) is finite and nonzero. This assumption imposes restrictions on G, but these restrictions are mitigated by the fact that the function h_j, the partitions p, and the constants t are freely chosen, based on knowing G. Since G is assumed to be known, in theory one can directly verify if any candidate h_j function satisfies Assumption A3 or not. We give examples of h_j functions later, but note for now that we will want to choose each h_j so that the function $h_j\left(G^P\left(s^P,\theta\right)\right)$ behaves like a thin tailed density, which would then make $\kappa_{Ptj}(\theta)$ behave like a finite, nonzero distribution moment. Note that for any given j, t, and partition P, the function $\kappa_{Ptj}(\theta)$ is a known function, since h_j and G are known functions. By its definition, $\kappa_{Ptj}(\theta)$ does not depend on the distribution of the random coefficients, so the only thing that is unknown about $\kappa_{Ptj}(\theta)$ is the true value of θ.

Define $\lambda_{Ptj}(\theta)$ and μ_{Pt} by

$$\lambda_{Ptj} = \int_{x^P \in supp(X^P)} E\left[h_j(Y) \mid X^P = x^P, X^{-P} = 0\right] x_1^{t_1-1} x_2^{t_2-1} ... x_{K^P}^{t_{K^P}-1} dx_1 dx_2 ... dx_{K^P} \tag{5}$$

$$\mu_{Pt} = E\left(V_1^{t_1} V_2^{t_2} ... V_{K^P}^{t_{K^P}}\right) \tag{6}$$

THEOREM 1: Let Assumptions A1, A2, and A3 hold. Then $\mu_{Pt} = \lambda_{Ptj}/\kappa_{Ptj}(\theta_0)$.

Note that λ_{Ptj} is identified by construction (and could, e.g., be estimated as the integral of a nonparametric regression). Lewbel and Pendakur (2015) proved a less general version of Theorem 1, and used it to identify moments μ_{Pt} assuming that the function G was known, and so in particular did not include any parameters θ. If no θ were present as in their case, then equation (6) could be used to identify μ_{Pt}, because then κ_{Ptj} would be known by construction. We now state some Corollaries of Theorem 1 that we will then use to identify θ.

COROLLARY 1: Let Assumptions A1, A2, and A3 hold for some nonempty set of partitions $\mathcal{P}$, some nonempty set T of complex vectors t and some set J that has at least two elements, j_1 and j_2. Let $S_1(\theta)$ be the vector of elements $\left(\lambda_{Ptj_1}/\kappa_{Ptj_1}(\theta)\right) - \left(\lambda_{Ptj_2}/\kappa_{Ptj_2}(\theta)\right)$ for all $P \in \mathcal{P}$, $t \in T$, $j_1 \in J$ and $j_2 \in J$. Then $S_1(\theta_0) = 0$.

COROLLARY 2: Let Assumptions A1, A2, and A3 hold for some nonempty set of partitions $\mathcal{P}$, some nonempty set T of complex vectors t and some non empty set J. Assume λ_{Ptj} and $\kappa_{Ptj}(\theta)$ are differentiable in j. Let $S_2(\theta)$ be the vector of elements $\kappa_{Ptj}(\theta)\left(\partial\lambda_{Ptj}/\partial j\right) - \left(\partial\kappa_{Ptj}(\theta)/\partial j\right)\lambda_{Ptj}$ for all $P \in \mathcal{P}$, $t \in T$, and $j \in J$. Then $S_2(\theta_0) = 0$.

COROLLARY 3: Assumptions A1, A2, and A3 hold for the vector $t = 0$, for some nonempty set of partitions $\widetilde{\mathcal{P}}$, and some nonempty set $\widetilde{J}$. Let $S_3(\theta)$ be the vector of elements $\lambda_{P0j} - \kappa_{P0j}(\theta)$ for all $P \in \widetilde{\mathcal{P}}$ and $j \in \widetilde{J}$. Then $S_3(\theta_0) = 0$.

COROLLARY 4: Let $S(\theta)$ be the vector of the union of elements of $S_1(\theta)$ if Corollary 1 holds, $S_2(\theta)$ if Corollary 2 holds, and $S_3(\theta)$ if Corollary 3 holds. If the rank of $\partial S(\theta)/\partial\theta$ equals the number of elements of θ for all θ in some neighborhood of θ_0, or if $S(\theta)$ is continuous and $S(\theta) = 0$ has only a finite number of solutions, then θ_0 is locally identified. If $S(\theta) = 0$ has only one solution, then θ_0 is point identified.

The only unknowns in the expression $\mu_{Pt} = \lambda_{Ptj}/\kappa_{Ptj}(\theta_0)$ from Theorem 1 are μ_{Pt} and θ_0. The key to Corollaries 1 and 2 is that μ_{Pt} does not depend on j. We use this fact to construct vectors $S_1(\theta_0)$ and $S_2(\theta_0)$ that eliminate μ_{Pt}, and so depend only on identified functions λ and κ and the unknown vector θ_0. In Corollary 2 we eliminate μ_{Pt} by differencing out μ_{Pt} for two different values of j, while in Corollary 3 we remove μ_{Pt} by taking the derivative of it with respect to j. Corollary 3 gives us additional equations by using the fact that for $t = 0$, μ_{Pt} is known to equal one. Corollary 4 then gathers up all these equations, and provides sufficient conditions that can be checked for identification. Note that local identification is equivalent to point identification among values of θ in a neighborhood of the true θ_0. see, e.g., Sargan (1983). Modern treatments of local identification include Chen, Chernozhukov, Lee, and Newey (2014) and Lewbel (2016).

A limiting case of Theorem 1 is when the set X^P is empty. In this case there is nothing to integrate and, letting h_j in this case be the identity function, we have the equation $E(Y \mid X = 0) - G(0, \theta_0) = 0$ which we can include as one more element of $S(\theta)$. More generally, as we will show later, use of different h_j functions and of different partitions $\mathcal{P}$ are convenient for isolating and estimating different components of θ.

Given identification of θ, we can then use Theorem 1 to identify any desired moments μ_{Pt} of the random coefficients distribution. For the case where the random coefficients are strictly positive, we can also use Theorem 1 to identify the characteristic function of $\ln U$, since in that case we can identify $\mu_t = E\left(V_1^{t_1} V_2^{t_2} ... V_K^{t_K}\right) = E\left[\exp\left(-\left(it_{11}\ln U_1 + ... + it_{K1}\ln U_K\right)\right)\right]$ where each $t_k = it_{k1}$ for any vector of real constants $(t_{11}, ..., t_{K1})$.

Alternatively, in empirical practice, it is common to parameterize the joint distribution of random coefficients. Given a parametric distribution for U, we may in general reparameterize in terms of some low order moments μ_{Pt} of V. Similarly, if U is assumed to be discretely distributed with a finite number of mass points, then its distribution would generally be recoverable from a finite number of moments μ_{Pt}. Identification of these μ_{Pt} then suffices for identification of the distribution of the random coefficients.

3 Estimation

The method of identification provided in this paper is largely constructive. Here we briefly discuss how estimators could be constructed based on the identification. These estimators all take the form of two step extremum estimators with either a parametric or nonparametric first

step. We do not provide associated limiting distribution theory, because it would consist of entirely standard applications of results in, e.g., Newey and McFadden (1994) based on kernel estimators, or sieve estimators could be used as in Chen (2007).

Based on Corollaries 1 and 4, a possible estimator for θ would be

$$\widehat{\theta} = \arg\min \sum_{P \in \mathcal{P}, t \in T} \sum_{j_1 \in J} \sum_{j_2 \in J} \left[\left(\widehat{\lambda}_{Ptj_1} / \kappa_{Ptj_1}(\theta) \right) - \left(\widehat{\lambda}_{Ptj_2} / \kappa_{Ptj_2}(\theta) \right) \right]^2 \tag{7}$$

given a consistent estimators $\widehat{\lambda}_{Ptj}$ of each λ_{Ptj}. The moments of Corollaries 2 or 3 could also be included by adding $\sum_{P \in \mathcal{P}, t \in T} \sum_{j \in J} \left[\kappa_{Ptj}(\theta) \left(\partial \widehat{\lambda}_{Ptj} / \partial j \right) - \left(\partial \kappa_{Ptj}(\theta) / \partial j \right) \widehat{\lambda}_{Ptj} \right]$ or $\sum_{P \in \widetilde{\mathcal{P}}} \sum_{j \in \widetilde{J}} \left[\widehat{\lambda}_{j0} - \kappa_{j0}(\theta) \right]^2$ to the minimization in equation (7). Equivalently, we are estimating $\widehat{\theta}$ by minimizing a quadratic form in $\widehat{S}(\theta)$, where each element of $\widehat{S}(\theta)$ equals the corresponding element of $S(\theta)$ with $\widehat{\lambda}$ replacing λ.

One choice of estimator $\widehat{\lambda}_{Ptj}$ in these expressions could be obtained by substituting a Nadayara-Watson kernel regression estimate of $E\left[h_j(Y) \mid X^P = x^P, X^{-P} = 0 \right]$ into equation (5). Consistency of $\widehat{\theta}$ would follow from the identification argument and the uniform consistency of each $\widehat{\lambda}_{Ptj}$.

Given estimates $\widehat{\theta}$, consistently estimated random coefficient moments $\widehat{\mu}_t$ for each $t \in T$ could be obtained by

$$\widehat{\mu}_t = \sum_{j \in J} w_{jt} \widehat{\lambda}_{jt} / \kappa_{jt}\left(\widehat{\theta} \right)$$

where weights w_{jt} (possibly non uniform to increase efficiency) are chosen by the econometrician, such that $\sum_{j \in J} w_j$.

Suppose identification only requires the partition $X^P = X$. For example, this will hold automatically if $K = 1$. In that case we can rewrite equation (5) as

$$\lambda_{Ptj} = E\left(\frac{h_j(Y) x_1^{t_1 - 1} x_2^{t_2 - 1} \ldots x_K^{t_K - 1}}{f_x(X)} \right) \tag{8}$$

where $f_x(X)$ is the probability density function of X. We may then construct a consistent estimator $\widehat{\lambda}_{Ptj}$ as the sample average $h_j(Y) x_1^{t_1 - 1} x_2^{t_2 - 1} \ldots x_K^{t_K - 1} / \widehat{f}_x(X_i)$ where $\widehat{f}_x(X)$ is a Rosenblatt-Parzen kernel density estimator. If desired, equation (7) could be replaced with weighted sums or a more general quadratic form, with weights chosen to increase efficiency of the resulting estimator. Estimation can be further simplified if $f_x(X)$, the density function of X, is finitely parameterized as, say $f_x(X, \alpha)$. In this case, the entire problem can be recast as an ordinary GMM estimator of the parameters α, θ, where moments consist of the score function of maximum likelihood estimation of α, that is,

$$E\left(\frac{\partial \ln f_x(X, \alpha)}{\partial \alpha} \right) = 0$$

along with the moments based on the above and Corollary 1 of

$$E\left(\frac{h_{j_1 t}(Y) x_1^{t_1 - 1} x_2^{t_2 - 1} \ldots x_K^{t_K - 1}}{f_x(X, \alpha) \kappa_{j_1 t}(\theta)} - \frac{h_{j_2 t}(Y) x_1^{t_1 - 1} x_2^{t_2 - 1} \ldots x_K^{t_K - 1}}{f_x(X, \alpha) \kappa_{j_2 t}(\theta)} \right) = 0 \text{ for } t \in T,\ j_1 \in J, \text{ and } j_2 \in J$$

We could at the same time also estimate moments μ_t by including

$$E\left(\mu_t - \frac{h_{jt}(Y)\, x_1^{t_1-1} x_2^{t_2-1} ... x_K^{t_K-1}}{f_x(X,\alpha)\,\kappa_{jt}(\theta)}\right) = 0 \text{ for } t \in T,\ j \in J.$$

as additional moments in the GMM estimation.

4 Example

Consider identification of the example of equation (2) from the introduction. We first note that setting $\theta_3 = 1$ and $\theta_4 = 0$ are free normalizations, since if θ_3 and θ_4 took on any other values we could without loss of generality replace U_1 with U_1^*/θ_3, replace U_2 with $U_2^* e^{-\theta_4}$ and replace θ_1 with $\theta_1^* = \theta_1\theta_3 e^{\theta_4}$ to get the model $Y = g\left(X_1U_1^* + \left(X_1U_1^*\right)\left(X_2U_2^*\right)\theta_1^* + \ln\left(X_2U_2^*\right)\theta_2\right)$, which is observationally equivalent to equation (2) with $\theta_0 = 1$ and $\theta_4 = 0$. So what remains to identify is θ_1, θ_2, and the distribution of (U_1, U_2) in the model

$$Y = g\left(X_1U_1 + (X_1U_1)(X_2U_2)\theta_1 + \ln(X_2U_2)\theta_2\right) \tag{9}$$

where g is known and strictly monotonic, We will now show identification of this model, using the Corollaries of Theorem 1. We choose a specific class of h_j functions and partitions X^P to formally prove identification of each element of θ. This is mainly done for analytic tractibility of the derived results. In practice one would not generally perform algebraic calculations like those given below, but would instead simply choose a thin tailed class of functions h_j and a variety of values of j and of t to obtain a long vector of equations $S(\theta) = 0$, and verify (perhaps numerically) that multiple solutions for this expression do not exist.

Assume that $X_k \geq 0$ and that is strictly positive and bounded. Nonnegativity is is required for the X_2U_2 term since we are taking the logarithm of that term, but these restriction are more generally sensible in most of the applications listed in the introduction, where the regressors X_k would be objects like income or factors of production. In particular, these economic rationality imposes these restrictions in the empirical application of Lewbel and Pendakur (2015), where each X_k is a price divided by income and each U_k is a Barten (1964) scale, which is a type of strictly positive random utility parameter.

As noted earlier, to satisfy the assumptions of Theorem 1, one should choose an h_j function that corresponds to a thin tailed density over x. For this application, a convenient choice of h_j (one that yields simple expressions for θ) is

$$h_j(y) = \widetilde{h}_j\left(g^{-1}(Y)\right) = \frac{\exp\left(-jg^{-1}(y)\right)}{\left(\exp\left(-jg^{-1}(y)\right)+1\right)^2/j}.$$

here $\widetilde{h}_j$ is the logistic probability density function, which is evaluated at the inverse of the g function.

We first show identification of θ_2. We will now use the partition P where $X^P = (X_2)$ and let $t_2 = 0$. This gives

$$\lambda_{P0j} = \int_0^\infty E\left[h_j(Y) \mid X_1 = 0, X_2 = x_2\right] x_2^{-1} dx_2$$

which corresponds to

$$\kappa_{P0j}(\theta) = \int_0^\infty h_j\left[g(\theta_2 \ln s_2)\right] s_2^{-1} ds_2 = \int_0^\infty \widetilde{h}_j(\theta_2 \ln s_2)\, s_2^{-1} ds_2 = \int_0^\infty \frac{j \exp(j\theta_2 \ln s_2)}{(\exp(j\theta_2 \ln s_2)+1)^2} s_2^{-1} ds_2$$

Assume $\theta_2 \geq 0$ (a similar analysis is possible if θ_2 is negative. Do the change of variables $q = \theta_2 \ln s_2$. Then

$$\kappa_{P0j}(\theta) = \int_{-\infty}^\infty \frac{j e^{-jq}}{\left(e^{-jq}+1\right)^2} \frac{1}{\theta_2} dq = \frac{1}{\theta_2}$$

It then follows from Corollary 3 that $\lambda_{P0j} = \kappa_{P0j}(\theta)$, so θ_2 is identified by $\theta_2 = 1/\lambda_{P0j}$ for any choice of $j > 0$.

This particular model is simple enough that θ_2 might have alternatively been identified by other means. For example, if g were assumed to be differentiable, then θ_2 could have been recovered from $\partial E\left[h_j(Y) \mid X_1 = 0, X_2 = x_2\right]/\partial x_2$ evaluated at $x_2 = 0$. Still, the above analysis shows identification using this paper's methods, specifically Corollary 3. It would also be possible to use Corollaries 1 or 2 to construct additional moments for θ_2 that could be used for estimation or as an alternative demonstrations of identification, most easily by keeping the same choice of $h_j(y)$ function, but with different choices for t_2.

Now consider identification of θ_1, which for convenience we will assume is nonnegative. We now use the partition P where $X^P = X$ and let $t_1 = 1$ and $t_2 = 1$. This gives

$$\lambda_{Ptj} = \int_0^\infty \int_0^\infty E\left[h_j(Y) \mid X_1 = x_1, X_2 = x_2\right] dx_1 dx_2$$

and

$$\kappa_{Ptj}(\theta) = \int_0^\infty \int_0^\infty \widetilde{h}_j\left[(s_1 + \theta_1 s_1 s_2 + \theta_2 \ln s_2)\right] ds_1 ds_2$$

Now do a change in variables in $\kappa_{Ptj}(\theta)$, changing s_1 for $r = s_1 + \theta_1 s_1 s_2 + \theta_2 \ln s_2 = \theta_2 \ln s_2 + (1 + s_2\theta_1) s_1$ so $dr = (1 + s_2\theta_1) ds_1$. Recalling that we assumed $\theta_1 > 0$ (an analogous derivation applies if θ_1 is negative), the change of variables gives

$$\begin{aligned}
\kappa_{jt}(\theta) &= \int_0^\infty \frac{1}{(1+s_2\theta_1)} \left(\int_{\theta_2 \ln s_2}^\infty \widetilde{h}_j(r)\, dr \right) ds_2 \\
&= \int_0^\infty \frac{\left(1 - \widetilde{H}_j(\theta_2 \ln s_2)\right)}{(1+s_2\theta_1)} ds_2
\end{aligned}$$

where $\widetilde{H}_j(\widetilde{y}) = \int_{-\infty}^{\widetilde{y}} \widetilde{h}_j(\widetilde{y})\, d\widetilde{y} = 1/(1 + \exp(-j\widetilde{y}))$, so

$$\begin{aligned}
\kappa_{jt}(\theta) &= \int_0^\infty \frac{\exp(-j\theta_2 \ln s_2)}{1 + \exp(-j\theta_2 \ln s_2)} \frac{1}{(1+s_2\theta_1)} ds_2 \\
&= \int_0^\infty \frac{1}{\left(1 + s_2^{j\theta_2}\right)(1+s_2\theta_1)} ds_2.
\end{aligned}$$

We will apply Corollary 3, which requires evaluating the derivative of the above expression for $\kappa_{jt}(\theta)$ with respect to j. This derivative is

$$\frac{\partial \kappa_{jt}(\theta)}{\partial j} = -\theta_2 \int_0^\infty \frac{s_2^{j\theta_2} \ln s_2}{\left(1 + s_2^{j\theta_2}\right)^2 (1 + s_2\theta_1)} ds_2$$

In general these integrals are difficult to evaluate analytically. However, we are free to choose any positive value for j, and these integrals become managable when we can choose j to make $j\theta_2$ be an integer. We can choose j this way because θ_2 has already been identified. In particular, letting $j = j_1 = 1/\theta_2$ we obtain

$$\kappa_{j_1 t}(\theta) = \int_0^\infty \frac{1}{(1 + s_2)(1 + s_2\theta_1)} ds_2 = \frac{\ln(\theta_1)}{\theta_1 - 1}$$

and

$$\frac{\partial \kappa_{j_1 t}(\theta)}{\partial j} = -\theta_2 \int_0^\infty \frac{s_2 \ln s_2}{(1 + s_2)^2 (1 + s_2\theta_1)} ds_2 = -\frac{\theta_2}{2} \left(\frac{\ln(\theta_1)}{\theta_1 - 1}\right)^2$$

It then follows from Corollary 3 that

$$\frac{\partial \lambda_{j_1 t}}{\partial j} / \lambda_{j_1 t} = -\frac{\theta_2}{2} \frac{\ln(\theta_1)}{\theta_1 - 1}$$

which is a strictly monotonic function of θ_1, and can therefore be uniquely solved for θ_1, thereby identifying θ_1. We could obtain additional or alternative analytically tractible moments for identification of θ_1 using Corollary 2, by letting $j_2 = c/\theta_2$ for integers c and evaluating $\kappa_{j_1 t}(\theta) / \kappa_{j_2 t}(\theta) = \lambda_{j_1 t} / \lambda_{j_2 t}$. More generally, equations like these could be applied with arbitrary values of j, by evaluating the resulting integrals numerically.

Finally, given that both θ_1 and θ_2 are identified, the joint distribution of (U_1, U_2) is identified as described at the end of the Identification section.

5 Extension: including an independent additive error

Assume that instead of $Y = G(X_1 U_1, ..., X_K U_K, \theta)$, we had the model $Y = G(X_1 U_1, ..., X_K U_K, \theta) + U_0$. Assume $U_0 \perp U_1, ..., U_K, X_1, ..., X_K$. This restriction was also assumed in one of the original papers on identification and estimation of random coefficients, Beran and Hall (1992), in the standard random coefficient framework where G is linear. As noted in the introduction, this is a strong assumption, but is justified in some contexts, e.g., in the model of Lewbel and Pendakur (2015), $U_1, ..., U_K$ are structural random utility parameters, and U_0 is measurement error in Y, which is plausibly independent of both structural heterogeneity parameters $U_1, ..., U_K$ and covariates $X_1, ..., X_K$. Assume also that the characteristic function of U_0 is everywhere nonzero, and without loss of generality, assume the median of U_0 equals zero (if not then we can just add a constant to the function G to make the median of U_0 equals zero).

Let $F(Y \mid X)$ denote the conditional distribution function of Y given $X = (X_1, ..., X_K)$. Then $F(Y \mid 0)$ equals and therefore identifies the marginal distribution of $U_0^* = U_0 + c$ where

the unknown constant $c = G(0, \theta)$. It therefore follows that the marginal distribution of U_0 is identified, because it equals the distribution of $U_0^* - med\left(U_0^*\right)$. Let $Y^* = Y - U_0$. The distribution function F is known, and the distribution of U_0 is known and is independent of X, and so by taking a conditional deconvolution Y with respect to $-U_0$, we identify the conditional distribution of Y^* given X. All of our previous identification result can then be applied replacing Y with Y^*.

By the above logic, including the additive error U_0 does not interfere with identification of the rest of the model. However, inclusion of this additive error could have profound implications for rates of convergence and other aspects of estimation. In particular, if U_0 is a thin tailed density, then the rate of convergence of the estimated distribution of Y^* given X could be very slow, and estimation of all other parameters might be correspondingly affected.

6 Proof of Theorem 1

$$\lambda_{Ptj} = \int_{x^P \in supp\left(X^P\right)} E\left[h_j(Y) \mid X^P = x^P, X^{-P} = 0\right] x_1^{t_1-1} x_2^{t_2-1} ... x_{K^P}^{t_{K^P}-1} dx_1 dx_2 ... dx_{K^P} \tag{10}$$

For a given partition P and associated ordering of the elements of X and U, let $\Omega_X^P = supp\left(X^P\right)$, $\Omega_U^P = supp\left(U^P\right)$, and let $F_U^P\left(u^P\right)$ be the joint distribution of U^P evaluated at the value u^P. By the definition of λ_{Ptj} we have

$$\begin{aligned}
\lambda_{Ptj} &= \int_{x^P \in \Omega_X^P} \int_{u \in \Omega_U^P} h_j\left(G\left(x_1 u_1, ..., x_{K^P} u_{K^P}, 0, ..., 0\right), \theta_0\right) dF_U^P\left(u^P\right) x_1^{t_1-1} x_2^{t_2-1} ... x_{K^P}^{t_{K^P}-1} dx_1 dx_2 ... dx_{K^P} \\
&= \int_{u \in \Omega_U^P} \int_{x^P \in \Omega_X^P} h_j\left(G\left(x_1 u_1, ..., x_{K^P} u_{K^P}, 0, ..., 0\right), \theta_0\right) x_1^{t_1-1} x_2^{t_2-1} ... x_{K^P}^{t_{K^P}-1} dx_1 dx_2 ... dx_{K^P} dF_U^P\left(u^P\right)
\end{aligned}$$

where the second equality follows from Fubini's theorem. Do a change of variables on the inner integral, letting $s_k = x_k u_k$ for $k = 1, ..., K$ to get

$$\begin{aligned}
\lambda_{Ptj} &= \int_{u \in \Omega_U^P} \int_{s^P \in supp\left(X_1 U_1, ..., X_{K^P} U_{K^P} \mid U = u\right)} h_j\left(G\left(s^P, 0, ..., 0\right), \theta_0\right) s_1^{t_1-1} s_2^{t_2-1} ... s_{K^P}^{t_{K^P}-1} u_1^{-t_1} u_2^{-t_2} ... u_{K^P}^{-t_{K^P}} \\
&\qquad\qquad ds_1 ds_2 ... ds_{K^P} dF_U^P\left(u^P\right) \\
&= \int_{u \in \Omega_U^P} \int_{s^P \in \Omega_X^P} h_j\left(G\left(s^P, 0, ..., 0\right), \theta_0\right) s_1^{t_1-1} s_2^{t_2-1} ... s_{K^P}^{t_{K^P}-1} u_1^{-t_1} u_2^{-t_2} ... u_{K^P}^{-t_{K^P}} ds_1 ds_2 ... ds_{K^P} dF_U^P\left(u^P\right) \\
&= \int_{u \in \Omega_U^P} \int_{s^P \in \Omega_X^P} h_j\left(G\left(s^P, 0, ..., 0\right), \theta_0\right) s_1^{t_1-1} s_2^{t_2-1} ... s_{K^P}^{t_{K^P}-1} ds_1 ds_2 ... ds_{K^P} u_1^{-t_1} u_2^{-t_2} ... u_{K^P}^{-t_{K^P}} dF_U^P\left(u^P\right) \\
&= \int_{u \in \Omega_U^P} \kappa_{Ptj}(\theta_0) u_1^{-t_1} u_2^{-t_2} ... u_{K^P}^{-t_{K^P}} dF_U^P\left(u^P\right) = \kappa_{Ptj}(\theta_0) \int_{u \in \Omega_U^P} u_1^{-t_1} u_2^{-t_2} ... u_{K^P}^{-t_{K^P}} dF_U^P\left(u^P\right) \\
&= \kappa_{Ptj}(\theta_0) E\left(U_1^{-t_1} U_2^{-t_2} ... U_K^{-t_K}\right) = \kappa_t(\theta_0) \mu_t
\end{aligned}$$

where the second equality above uses Assumption A2 regarding support conditions.

Unobserved Preference Heterogeneity in Demand Using Generalized Random Coefficients

Arthur Lewbel and Krishna Pendakur
Boston College and Simon Fraser University

original Sept. 2011, revised Nov. 2015

Abstract

We prove a new identification theorem showing nonparametric identification of the joint distribution of random coefficients in general nonlinear and additive models. This differs from existing random coefficients models by not imposing a linear index structure for the regressors. We then model unobserved preference heterogeneity in consumer demand as utility functions with random Barten scales. These Barten scales appear as random coefficients in nonlinear demand equations. Using Canadian data, we compare estimated energy demand functions with and without random Barten scales. We find that unobserved preference heterogeneity substantially affects the estimated consumer surplus costs of an energy tax.

JEL codes: C14 D12 D13 C21 Keywords: unobserved heterogeneity, nonseparable errors, random utility parameters, random coefficients, equivalence scales, consumer surplus, welfare calculations. The authors wish to thank Jinyong Hahn, Richard Blundell, James Heckman, Xavier d'Haultfoeuille, Stefan Hoderlein, Rosa Matzkin, and anonymous referees for helpful discussions and suggestions. Earlier versions of this paper were circulated under the name "Generalized Random Coefficients With Equivalence Scale Applications."

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). In this paper we propose an analogous way to introduce unobserved preference heterogeneity in nonlinear continuous demand systems. This includes a new identification theorem showing how the joint distribution of random coefficients in general nonlinear and additive models can be nonparametrically identified. This is in sharp contrast to almost all existing random coefficients theorems, which, like BLP, assume a linear index structure for the regressors.

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. Moreover, we find that this variation in energy price elasticities is particularly large among poorer households. 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 equivalent 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. We call these, "random Barten scales."

Allowing for random 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 like 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 our demand application, generalized random coefficients on prices are the random Barten scales.

In this paper we first provide some identification theorems, showing that the joint distribution of random coefficients can be nonparametrically identified in nonlinear, and in additive 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.

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 more than twice as large in our model compared to an analogous 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 set

of unobserved errors $U_0, U_1, ..., U_K$ that are (possibly after conditioning on other covariates Z) independent of X. We propose a generalized random coefficients model given by

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

for some function G. We focus mainly on results for the special case of equation (1) where G takes the additive model form

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

and the functions $G_1, ..., G_K$ are unknown.

In these models the vector $U = (U_1, ..., U_K)$ represents unobserved heterogeneity in the dependence of Y on X, while U_0, if present, represents measurement error or other independent variation in Y. We provide conditions under which the joint distribution of the vector U is nonparametrically identified. If present, U_0 is assumed independent of these other errors and has a marginal distribution that is also nonparametrically identified.

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 random Barten scale. All previous empirical implementations of Barten scales have exactly these forms, 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 allow the Barten scales to be random, depending on both observable and unobservable characteristics. We show that the joint distribution of these random Barten scales can be nonparametrically identified, under low level regularity conditions.

One of our identification theorems shows that if G is known, then under some conditions the joint distribution of the elements of U is nonparametrically identified. We also provide a theorem giving conditions under which, in equation (2), each function G_k can be nonparametrically identified. Combining both theorems then allows us to simultaneously nonparametrically identify the joint distribution of U and nonparametrically identify each G_k function. Combining both theorems also provides additional restrictions that we argue might be exploited to further generalize the model (specifically, by possibly relaxing the additivity assumption with some interaction terms).

Imposing the additivity of equation (2) 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 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 a large increase in the price of, or taxes on, energy goods. We then list some conclusions. One important finding is that an energy tax may impact inequality and welfare through a previously unrecognized channel, which is the variation in consumer surplus effects for households that have the same observed budgets and characteristics. We find that this variation is particularly large for poor households.

Following the conclusions is an appendix that contains proofs of Theorems. We also have an online supplemental appendix. The online supplemental appendix contains an extensive set of additional empirical analyses verifying the robustness of our estimated results to a wide variety of alternative model specifications, including adding complexity to the utility function specification, relaxing the parametric structure on preference and error distributions, and dealing with potential endogeneity of regressors. The online supplemental appendix also includes a Monte Carlo analysis of our estimated model, and contains some additional technical results of lesser importance.

2 Literature Review

We use generalized random coefficients to represent price scales in consumer demand models. There is a long history of using such scales to empirically model observed sources of preference heterogeneity. See, e.g., Rothbarth (1943), Prais and Houthakker (1955), Barten (1964), Pollak and Wales (1981) and Jorgenson, Lau, and Stoker (1982), and see Lewbel (1997) for a survey. Barten (1964) type price scales (hereafter: *Barten scales*) take the form of multiplying 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 Barten 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 (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. As an alternative to modeling unobserved heterogeneity, Hausman and Newey (2014) provide bounds on average consumer surplus.

In contrast, our model follows the prior consumer demand literature by including preference heterogeneity in the form of Barten 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 of 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), McFadden and Richter (1991), Lewbel (2001), McFadden (2005), Beckert (2006), Matzkin (2007b), Beckert and Blundell (2008), Hoderlein and Stoye, (2014), and Kitamura and Stoye (2014). 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 linear simultaneous systems of random coefficients, including Masten (2015) and Hoderlein, Holzmann, and Meister (2015), 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).

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

Fox and Gandhi (2013) provide general conditions for identification of random utility parameters in multinomial choice problems, including linear index models with random coefficients, and models analogous to Berry and Haile (2009) that exploit Lewbel (2000) type special regressors. They note that the only general sufficient condition known for one of their identifying assumptions is utility functions that are real analytic functions.

A related result to ours is Hoderlein, Nesheim, and Simoni (2011), who provide a high level condition they call T-completeness that suffices for nonparametric identification of a vector of random parameters within a known function. They provide some examples where T-completeness can be shown to hold, such as when error distributions are in the exponential family, or are parameterizable by a single scalar. Our model when G is known is a special case of their general setup, and so our theorem proving identification for this model provides a new framework where T-completeness could be satisfied. More generally, one goal of our analysis is to provide relatively low level conditions that serve to identify our model, instead of high level, difficult to verify conditions as in Fox and Gandhi (2013), or like T-completeness.

Perhaps the result that comes closest to our identification theorem is Matzkin (2003), which

in an appendix describes sufficient conditions for identification of a general class of additive models with unobserved heterogeneity. The biggest difference between our results and Matzkin (2003) is that we identify the joint distribution of U, while Matzkin assumes the elements of U are mutually independent. However, even our model when $K = 1$ (the case where there is no joint distribution to be identified) does not satisfy her identification assumptions and so even in that case our Theorem is new and cannot be derived from her results. Our online supplemental appendix contains details regarding these differences.

3 Generalized Random Coefficient Model Identification

In this section we first provide, in Theorem 1, assumptions under which the joint distribution of random coefficients can be identified in the general model of equation (1) when G is known. We then provide, in Theorem 2, separate assumptions under which each function G_k (and the marginal distributions of each random coefficient U_k) can be identified in additive models given by equation (2). We then combine both theorems to nonparametrically identify both the joint distribution of random coefficients and the functions G_k.

Later sections provide the connections between these theorems and our Barten scales model of demand. However, we note upfront that in our empirical application X is positive (though not bounded away from zero), so it is relevant that our identification theorems allow for zero being on the boundary of the closure of the support of X.

3.1 Identification of the Distribution of Generalized Random Coefficients

Let $\widetilde{Y}$ be a dependent variable, X is a vector of covariates $(X_1, ..., X_K)$ which will have random coefficients $(U_1, ..., .U_K)$. Let Z be a vector of additional covariates that may affect the distribution of these random coefficients. For now think of $\widetilde{Y}$ as being observable along with X and Z, although later we will generalize to having the dependent variable be $Y = \widetilde{Y} + U_0$. 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.

ASSUMPTION A1: The conditional distribution $F_{\widetilde{Y}|X,Z}(\widetilde{y} \mid x, z)$ and the marginal distribution $F_Z(z)$ are identified. $\widetilde{Y} = G(X_1U_1, ..., X_KU_K)$ for some continuous function G and some unobserved random coefficients $U = (U_1, ..., .U_K)$, where $U \perp X \mid Z$.

Assumption A1 first assumes identification of the distributions of observables, which would in general follow from a sample of observations of $\widetilde{Y}$, X, Z with sample size going to infinity. We will first consider identification of $F_{U|Z}(U \mid Z)$, the joint distribution of the random coefficients (conditional on Z), assuming the function G is known. Later we will provide nonparametric identification for a class of G functions. Standard random coefficients is the special case in which the known function G is just the sum of the X_kU_k terms.

In our empirical application, the random coefficients U_k will represent unobserved taste heterogeneity, and Z could then be a vector of observable characteristics that also affect tastes. Another possible role for Z is to serve as control function residuals, which then allows the

random coefficients U_k to be unconditionally correlated with X, so elements of X can 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.

The vector Z can be empty, so all the results we provide 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: The distribution of $\left(U_1^{-1}, ..., U_K^{-1} \mid Z\right)$ is identified from its integer moments. $supp\,(X)$ is rectangular, $supp\,(X \mid Z) = supp\,(X)$, and the closure of $supp\,(X \mid Z)$ equals the closure of $supp\,(U_1 X_1, ..., U_K X_K \mid Z, U)$

We will identify $F_{U|Z}$ by identifying moments of the distribution of $\left(U_1^{-1}, ..., U_K^{-1} \mid Z\right)$. Necessary and sufficient conditions for integer moments to identify a distribution are known, and are weaker than the conditions needed for existence of a moment generating function. See, e.g., Assumption 7 of Fox, Kim, Ryan, and Bajari (2012). In our empirical application it will be reasonable to assume that each U_k is bounded and bounded away from zero, which is sufficient and stronger than necessary.

Given $U \perp X \mid Z$, the support condition in Assumption A2 could be satisfied in a few different ways. In particular, the support condition holds if the closure of $supp\,(X \mid Z) = \mathbb{R}_+^K$ and $supp\,(U \mid Z) \subseteq \mathbb{R}_+^K$, or if $supp\,(X \mid Z) = \mathbb{R}^K$ and U has any support that excludes the origin. In our application we satisfy the first of these conditions by assuming X (prices scaled by total expenditures) can take on any positive value and by noting that Barten scales U must be positive.

Let $t = (t_1, ..., t_K)$ denote a K vector of positive integers. For a given function h and vector t, define κ_t by

$$\kappa_t = \int_{supp(X)} h\left[G\left(s_1, ...s_K\right), t\right] s_1^{t_1-1} s_2^{t_2-1} ... s_K^{t_K-1} ds_1 ds_2 ... ds_K \tag{3}$$

ASSUMPTION A3: Given G, for any K vector of positive integers t we can find a continuous function h such that $\int_{supp(X)} \left|h\left[G\left(s_1, ...s_K\right), t\right] s_1^{t_1-1} s_2^{t_2-1} ... s_K^{t_K-1}\right| ds_1 ds_2 ... ds_K$ exists, the integral defining κ_t is convergent, and $\kappa_t \neq 0$.

Assumption A3 assumes integrability of the function being integrated in equation (3), and requires that the integral is finite and nonzero. This assumption imposes restrictions on G, but these restrictions are mitigated by the fact that the function h is freely chosen, based on knowing both G and t. The classes of general G functions that satisfy Assumption A3 are difficult to characterize, so Lemma 1 below provides a sufficient condition that is relevant for our application.

LEMMA 1: Assume $supp(X) = \mathbb{R}_+^K$, $G(X_1, ...X_K) = \sum_{k=1}^K G_k(X_k)$, and there exist positive constants $c_1, ..., c_K$ such that $G_k(X_k) \geq c_k X_k$ for $k = 1, ..., K$. Then Assumption A3 holds.

Proofs are in the appendix. Assumption A3 is most readily satisfied by taking h to be a positive function that is thin tailed in its first argument, as in the proof of Lemma 1 where h as a function of G is the exponential density function. As is also illustrated by Lemma 1, the function h is not required to depend on t, but it is allowed to do so.

There are examples of G functions for which identification of the joint distribution of U_1 and U_2 is clearly not possible. An example is $G(X_1, X_2) = \ln(X_1) + \ln(X_2)$, since in this case U_1 and U_2 appear only in the form $\ln(U_1) + \ln(U_2)$ with no way to separately identify each. Lemma 2 in the appendix proves that this function violates Assumption A3, so the assumed existence of a function h does limit the class of allowable G functions.

THEOREM 1: Let Assumptions A1, A2, and A3 hold. If the function G is known or identified, then the joint distribution function $F_{U|Z}(U_1,...,U_K \mid Z)$ is identified.

The proof is in the appendix. To provide an idea of how Theorem 1 works, consider the case of $K = 1$ and observe that, by a change of variables argument, $E\left(\int h(G(UX)) X^{t-1} dX\right) = E\left(\int h(G(s)) s^{t-1} U^{-t} ds\right)$. The left side of this equation can be estimated, while the right equals a known function times $E\left(U^{-t}\right)$. We can in this way identify any moment of U^{-1} and thereby identify the distribution of U.

Providing some more detail, define $\lambda_t(Z)$ by

$$\lambda_t(Z) = \int_{X \in supp(X)} E\left[h\left(\widetilde{Y}, t\right) \mid X_1, X_2...X_K, Z\right] X_1^{t_1-1} X_2^{t_2-1} ... X_K^{t_K-1} dX_1 dX_2 ... dX_K \quad (4)$$

$\lambda_t(Z)$ is an integral of a known conditional expectation, and so is identified. The proof of Theorem 1 works by showing that the moment $E\left(U_1^{-t_1} U_2^{-t_2} ... U_K^{-t_K} \mid Z\right)$ is identified by equaling the ratio of identified objects $\lambda_t(Z)/\kappa_t$. Identification of these moments for any K vector of integers t then implies identification of the distribution of $U_1^{-1} U_2^{-1} ... U_K^{-1} \mid Z$ by Assumption A2, and hence identification of $F_{U|Z}(U_1,...,U_K \mid Z)$. We could have instead tried to directly identify $F_{U|Z}$ by working with negative values of t, but Assumption A3 would then be more difficult to satisfy than with positive integers t. More speculatively, we might instead try to identify the characteristic function of $F_{U|Z}$ by replacing t with the square root of minus one times t for real vectors t, but in that case dealing with boundary issues associated with the change of variables step in the proof becomes more complicated.

A special case of Theorem 1 is when $G(s_1, ...s_K) = s_1 + ... + s_K$, corresponding to the standard linear random coefficients model. Nonparametric identification of the linear model under varying conditions is established by Beran and Hall (1992), Beran and Millar (1994), Beran, Feuerverger, and Hall (1996), and Hoderlein, Klemelae, and Mammen (2010), using assumptions and methods that differ substantially from Theorem 1. The latter authors give key sufficient identifying assumptions in the linear model as being U independent of X as we assumed, and that the distribution of U have square integrable derivatives of sufficiently high order. Generally, identification of the distribution linear random coefficients has been based on

either assuming large support for X or assuming thin tails for U, (see, e.g., Masten 2013 for a summary), while Theorem 1 for a general G function requires both support and tail restrictions as in Assumption A2. The linear G model automatically satisfies our Assumption A3 by being a special case of Lemma 1.

Although Theorem 1 imposes large support restrictions on the regressors, it does not depend on identification at infinity or other types of "thin set" identification arguments (see Khan and Tamer 2011). The support restrictions in Theorem 1 are only used to ensure that boundary terms equal zero in the change of variables step in the proof of the Theorem.

3.2 Additive Model Identification

Theorem 1 identified the joint distribution $F_{U|Z}(U_1,...,U_K \mid Z)$, assuming the function G is identified. We now provide a theorem that shows nonparametric identification for a class of G functions. We will then combine both Theorems. Some assumptions of Theorem 2 duplicate those of Theorem 1; we write the assumptions this way so that each Theorem can stand alone. 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.

ASSUMPTION A4: The conditional distribution $F_{\widetilde{Y}|X,Z}(y \mid x,z)$ and the marginal distribution $F_Z(z)$ are identified. $Y = \sum_{k=1}^{K} G_k(X_k U_k) + U_0$ for some unknown functions G_k and some unobserved random coefficients U_0 and $U = (U_1, ..., .U_K)$, where $(U_0, U_1, ..., U_K) \perp X \mid Z$ and $U \perp U_0 \mid Z$. Either U_0 has a nonvanishing characteristic function (conditional on Z)[1] or U_0 is identically zero. $supp(U_0) \subseteq supp(Y)$.

ASSUMPTION A5: For $k \in \{1, ..., K\}$, $(U_k, X_k) \mid Z$ is continuously distributed, and for every $r \in supp(X_k U_k)$ there exists an $x_k \in supp(X_k)$ such that $f_{U_k}\left(x_k^{-1} r\right) \neq 0$. X has rectangular support and $\{0, e_1, ..., e_K\}$ is a subset of the closure of $supp(X)$.

ASSUMPTION A6: For $k \in \{1, ..., K\}$, G_k is a strictly monotonically increasing, differentiable 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.

As before, Assumption A4 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 $k = 1, ..., K$. Assumption A4 also imposes conditional independence and support requirements on U, X and Z. These are standard assumptions for random coefficients models, except for the assumption that the additive error U_0 is conditionally independent of the other random coefficients. Beran and Hall (1992) assumed independence of U_0, but later linear G models do not impose this restriction. Although independence of U_0 is a strong assumption, we show it's plausible in our empirical application.

[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 A5 assumes that the regressors and random coefficients are continuously distributed. Assumption A5 also calls for some support restrictions, but these are all much milder than the support restrictions that were required by Theorem 1. A key feature of this assumption is that the closure of the support of each element of X includes zero.

Assumption A6 requires each G_k function to be smooth and monotonic. This facilitates identifying the distribution of unobservables U_k that lie inside each G_k. In our application, economic theory will imply this monotonicity. The normalizations in Assumption A6 are all free normalizations, that is, they are made without loss of generality. This is 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 2 below, though in empirical applications alternative normalizations may be more natural, e.g., choosing location to make $E(U_0) = 0$.

THEOREM 2: Let Assumptions A4, A5, and A6 hold. Then the functions $G_1, G_2, ..., G_K$ and the distributions $F_{U_0|Z}$, $F_{U_1|Z}$,...$F_{U_K|Z}$ are all nonparametrically identified.[2]

The proof is in the appendix. As noted in the literature review, Theorem 2 is similar to, but is not a direct corollary of, results in Matzkin (2003). In an online spplemental appendix, we summarize the ways in which Theorem 2 differs from Matzkin (2003). One obvious (though not the only) difference is that Matzkin assumes the elements of U are mutually independent, either conditionally or unconditionally, and we do not.

The proof of Theorem 2 depends on an identification at zero argument, i.e., thin set identification (see Khan and Tamer 2011). This is undesirable because it means that most of the observable population distribution is not used for identification, and as a result often implies slow rates of convergence for corresponding nonparametric estimators. However, two arguments help mitigate these concerns. First, additional pieces of information can be constructed and used for estimation or testing. For example, given a G_k and $F_{U_k|Z}$ function identified by Theorem 1, an additional equality that these functions satisfy for all values of x (not just in the neighborhood of zero) is

$$\frac{\partial E(Y \mid X = x, Z = z)}{\partial x_k} = \int_{u \in supp(U_k|Z=z)} \left[\partial G_k(x_k u)/\partial x_k\right] dF_{U_k|Z}(u \mid z) \tag{5}$$

Equation (5) provides additional restrictions on the G_k and $F_{U_k|Z}$ that might be exploited for estimation, testing, or alternative identification arguments. The restrictions provided by equation (5) apply over the whole support of the data, not just on a thin set.

A small extension to Theorem 2 is the following.

[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. 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 $\| X \| \leq c$ and taking limits as $c \to 0$.

COROLLARY 1: Let $Y = G(X_1U_1, ..., X_KU_K) + U_0$ for any function G that includes $X_kU_ke_k$ in its domain, for $k = 1, ..., K$. Then

i) There exists functions $G_1, ..., G_K$, and $\widetilde{G}$ such that

$$Y = \widetilde{G}(X_1U_1, ..., X_KU_K) + \sum_{k=1}^{K} G_k(X_kU_k) + U_0 \tag{6}$$

where the function $\widetilde{G}(X_1U_1, ..., X_KU_K)$ equals zero when all but one of its elements equal zero, and

ii). Theorem 2 holds replacing $Y = \sum_{k=1}^{K} G_k(X_kU_k) + U_0$ in Assumption A4 with equation (6).

In Corollary 1, the function $\widetilde{G}$ is not identified, so the main points of this corollary are first that any function G can be decomposed into an additive part $\sum_{k=1}^{K} G_k$ and an interactions part $\widetilde{G}$, and second that the presence of the interaction function $\widetilde{G}$ does not interfere with identification of the G_k and $F_{U_k|Z}$ functions using Theorem 2. Corollary 1 would then be useful in contexts where $\widetilde{G}$ is known or can be identified by other means.

3.3 Full Model Identification

Here we combine Theorems 1 and 2 to identify all the unknown functions in equation (2).

COROLLARY 2: Let $Y = \sum_{k=1}^{K} G_k(X_kU_k) + U_0$ and let $G(X_1U_1, ..., X_KU_K) = \sum_{k=1}^{K} G_k(X_kU_k)$. Let Assumptions A1, A2, A3, A4, A5, and A6 hold. Then the functions $G_1, G_2, ..., G_K$ and the joint distribution function $F_{U|Z}(U_1,...,U_K \mid Z)$ are identified.

Corollary 2 shows identification of the model, but also may provide additional restrictions implied by the model that might usable either for alternative identification strategies, or to obtain identification under more general conditions. In particular, both Theorems 1 and 2 identify the functions $F_{U_k|Z}$ for $k = 1, ..., K$, and so restrictions on the functions $G_1, G_2, ..., G_K$ and $F_{U_1|Z}, ..., F_{U_K|Z}$ are obtained by equating the construction of the functions $F_{U_k|Z}$ from each of the two theorems for each k. Equation (5) provides more such restrictions. Still more restrictions might be obtained by applying Theorem 2 using different h functions, and it may be possible to use that variation to identify the G_k functions without the use of Theorem 2 at all. Similarly, the additional equations that are obtainable using multiple h functions might also be used to identify richer models, such as those containing interaction terms like the function $\widetilde{G}$ in Corollary 1. We defer discussion of these conjectures to the online supplemental appendix.

Theorems 1 and 2, and hence our model identification result of Corollary 2, is constructive. However, designing an estimator based on mimicing the steps of these identification arguments would likely be both inefficient and difficult to implement. Inefficiency is likely because Theorem 2 uses thin set identification, and Theorem 1 provides equations based on specific choices of the function h, and it is hard to see how one might choose the function h to maximize efficiency. Indeed, different h functions might be optimal for each moment and each function to be estimated. Also, Theorem 1 identifies moments of $U_1^{-1}, ...U_K^{-1}$, so an inversion would be needed to directly obtain the distribution function of U. Finally, sequentially applying Theorem 2 to

estimate $G_1, G_2, ..., G_K$ and Theorem 1 to estimate F_U would ignore the additional identifying information discussed in the previous paragraph. In our empirical application we parameterize these unknown functions and use maximum likelihood estimation, though one could consider extending our estimator to allow for nonparametric specification of these functions using sieve maximum likelihood. To faciliate such an extension, we provide series expansions that could serve as basis functions for a possible sieve estimator.

4 Barten Scales in Utility Functions

Let a "consumer" refer to an individual or household that maximizes a single well behaved utility function. Let Q_k denote the quantity purchased of a good k, and let $S(Q, U)$ denote the direct utility function over the bundle of goods $Q = (Q_1, ..., Q_K)$ of a consumer having a vector of preference heterogeneity parameters $U = (U_1, ..., U_K)$. Assume S is continuous, non-decreasing, twice differentiable in Q and quasi-concave in Q. Define the reference consumer to be a consumer that has heterogeneity parameters U normalized to equal one, and let $\overline{S}(Q_1, ..., Q_K)$ denote the direct utility function of a reference consumer. Each consumer chooses quantities to maximize utility subject to the standard linear budget constraint $\sum_{k=1}^{K} P_k Q_k = M$ where P_k is the price of good k and M is the total amount of money the consumer spends on this bundle of goods. Let $W_k^* = Q_k P_k / M$ be the share of the money budget M that is spent on good k (called the budget share of good k). Write the Marshallian budget share functions that result from maximizing the reference utility function $\overline{S}$ subject to the budget constraint as $W_k^* = \omega_k(P_1/M, ..., P_K/M)$. Let $V(P_1/M, ..., P_K/M)$ denote the indirect utility function corresponding to $\overline{S}$, obtained by substituting $Q_k = \omega_k(P_1/M, ..., P_K/M)\, M/P_k$ into $\overline{S}(Q_1, ..., Q_K)$ for $k = 1, ..., K$.

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 $S(Q_1, .., Q_K; \alpha_{z1}, ..., \alpha_{zK}) = \overline{S}(Q_1/\alpha_{z1}, ..., Q_K/\alpha_{zK})$, where the Barten scales $\alpha_{z1}, ..., \alpha_{zK}$ are positive functions of observable household attributes z, 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_{zk} is, the greater the economies of scale to consumption of good k 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 $S(Q_1, .., Q_K; U_1, ..., U_K) = \overline{S}(Q_1/U_1, ..., Q_K/U_K)$, where $U_1, ..., U_K$ are positive random utility parameters embodying preference heterogeneity (both observed and unobserved) across consumers. More formally, we could write each random Barten scale as $U_{zk}(z)$, since for each good k, the distribution function that U_{zk} is drawn from could depend on observable household attributes z. Barten's original model is then the special case where the distribution of each U_{zk} is degenerate with a single mass point at α_{zk}.

Define normalised prices $X_k = P_k/M$ for each good k and rewrite the budget constraint as $\sum_{k=1}^{K} X_k Q_k = 1$. Now $\overline{S}(Q_1, ..., Q_K)$ and $V(X_1, ..., X_K)$ are the direct and indirect utility functions of the reference consumer, and $\omega_k(X_1, ..., X_K)$ 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 $W_k^* = \omega_k(U_1X_1, ..., U_KX_K)$ for each good k if and only if the consumer's direct and indirect utility function equal, up to an arbitrary monotonic transformation, $\overline{S}(Q_1/U_1, ..., Q_K/U_K)$ and $V(U_1X_1, ..., U_KX_K)$, respectively. Also, given a specification of reference indirect utility $V(X_1, ..., X_K)$, the corresponding Barten scaled demand functions can be obtained by the logarithmic form of Roy's identity:

$$W_k^* = \omega_k(U_1X_1, ..., U_KX_K) = \frac{\partial V(U_1X_1, ..., U_KX_K)}{\partial \ln X_k} \Big/ \left(\sum_{\ell=1}^{K} \frac{\partial V(U_1X_1, ..., U_KX_K)}{\partial \ln X_\ell} \right) \tag{7}$$

Notice that the functional form of each ω_k only depends on the functional form of $\overline{S}$ or equivalently of $\overline{V}$, so $U_1, ...U_K$ can vary independently of $X_1, ..., X_K$ 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_k in place of deterministic functions α_{hk} 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 Indirectly Additively Separable Utility

In our empirical application, we let ω_1 be the budget share of a single good of interest, energy, and we let ω_2 denote the budget share of all other goods, corresponding to the general Barten scaled model with $K = 2$. This case only requires estimating a single equation for ω_1, since the equation for ω_2 is automatically determined by construction as $\omega_2 = 1 - \omega_1$. If we had $K > 2$, then we would have $K-1$ separate equations to estimate, and we would have further restrictions to impose because the same Barten scales, with the same joint distribution $F_{U|Z}(U_1,...,U_K \mid Z)$,

would appear in each equation.

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 K Barten scales $(U_1, ..., U_K)$ and K demand equations, Matzkin's identification method for systems of equations cannot be applied here because there are actually only $K-1$ distinct demand functions ω_1,...,ω_{K-1}, with the remaining demand function ω_K determined by the adding up constraint that $\sum_{k=1}^{K}\omega_k = 1$. We therefore have more random parameters than distinct equations in the system.

The decomposition of total consumption into $K = 2$ goods is often done in empirical work when one wishes to focus on the welfare effects of a price change 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 (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 $K = 2$ goods, our model is $W_1^* = \omega_1(U_1X_1, U_2X_2)$ and $W_2^* = 1 - W_1^*$, and with $K = 2$ we can rewrite equation (7) as

$$\lambda\left(W_1^*\right) = \ln\left(\frac{\partial V(U_1X_1, U_2X_2)}{\partial \ln X_1}\right) - \ln\left(\frac{\partial V(U_1X_1, U_2X_2)}{\partial \ln X_2}\right) \tag{8}$$

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

Due to the constraints of Slutsky symmetry, imposing additivity directly on the Marshallian budget share function $\omega_1(X_1, X_2)$ would result in extreme restrictions on behavior. See, e.g., Blackorby, Primont, and Russell (1978). So we instead impose additivity on the logit transformation of $\omega_1(X_1, X_2)$ (later this will be relaxed to allow for interaction terms), thereby assuming demands have the additive form

$$\lambda(W_1) = \lambda\left[\omega_1(U_1X_1, U_2X_2)\right] + U_0 = g_1(U_1X_1) + g_2(U_2X_2) + U_0 \tag{9}$$

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 for good 1 is given by

$$W_1^* = \omega_1(U_1X_1, U_2X_2) = \left(1 + e^{-g_1(U_1X_1) - g_2(U_2X_2)}\right)^{-1} \tag{10}$$

Use of the logit transformation here, and assumed additivity in logit transformed budget shares, has as far as we know not been considered before in the estimation of continuous demand functions. However, this logit transformed model has a number of advantages. First, $\lambda(W_1)$ has support on the whole real line, so the measurement error U_0 has unrestricted support, instead of a support that necessarily depends on covariates. Second, with this transform no constraints need to be placed on the range of values the nonparametric functions g_1 and g_2 take on. Third, unlike all other semiparametric or nonparametric applications of the Hausman (1981) consumer

surplus type methodology (such as those cited above), a closed form expression for the indirect utility function that gives rise Marshallian demands (10) and hence (9) exists, and is given by Theorem 3.

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

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

for some 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

$$h_1(U_1X_1) + h_2(U_2X_2) = \int e^{g_1(U_1x_1)} d\ln x_1 + \int e^{-g_2(U_2x_2)} d\ln x_2 \tag{11}$$

and

$$g_1(U_1X_1) + g_2(U_2X_2) = \ln\left(\frac{\partial h_1(U_1X_1)}{\partial \ln X_1}\right) - \ln\left(\frac{\partial h_2(U_2X_2)}{\partial \ln X_2}\right) \tag{12}$$

Also, the functions $h_1(U_1P_1/M)$ and $h_2(U_2P_2/M)$ are each nonincreasing, and their sum is strictly increasing in M and quasiconvex in P_1,P_2, and M.

The proof is in the Appendix. The function H in Theorem 3 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.[3] We can therefore just write the indirect utility function in Theorem 3 as

$$V(U_1X_1, U_2X_2)^{-1} = h_1(U_1X_1) + h_2(U_2X_2). \tag{13}$$

which takes H to be the reciprocal function (this is a convenient normalization since we later take h_1 and h_2 to be increasing functions, and utility must be nondecreasing in total expenditures).

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 H, $V(X_1, X_2) = H[h_1(X_1) + h_2(X_2)]$ 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 (10) 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.

4.2 Random Barten Scales: Identification

From equation (9) we have the demand model

$$\lambda(W_1) = g_1(U_1X_1) + g_2(U_2X_2) + U_0 \tag{14}$$

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

Identification of this model can be obtained by Corollary 2, letting $Y = \lambda(W_1)$ and $G_k = g_k$. A condition that suffices to make the monotonicity of Assumption A6 hold is that the goods not be Giffen goods.[4] Having good 1 not be Giffen guarantees monotonicity of g_1, and similarly the restriction that good 2 is not Giffen means that ω_2 is monotonic in X_2, which by the adding up constraint $\omega_1 + \omega_2 = 1$ implies monotonicity of g_2. As discussed earlier, Assumption A3 is satisfied for a wide range of possible g_k functions, e.g., those satisfying Lemma 1.

Next consider Assumptions A1, A2, and A4. U_0 is assumed to be measurement error in Y, not a taste parameter, and hence independent of the other variables. As discussed earlier, Barten scales are traditionally modeled as deterministic functions of demographic characteristics and other factors that affect demand, so in our extension to random Barten scales we take Z to be demographic characteristics, and other taste shifters for energy such as weather. Continuity of each X_k and U_k is straightforward. U_1 and U_2 are preference parameters, and it is common to assume that tastes are determined independently of regressors in partial equilibrium analyses. In our application W_1 will be energy demand by Canadian households and the regressors X_k are prices divided by budgets. One might therefore be concerned about correlations between U_k and X_k caused by endogeneity of prices, however, Canadian households comprise a very small fraction of world energy demand, and so the likely effect of U_k on energy prices should be very small. In our online supplemental appendix, we verify that endogeneity, if any is present due to this or other sources (such as potential measurement error in total expenditures), appears too small empirically to significantly change our results. Economic theory also bounds the support of U_1 and U_2 in the positive orthant, assuming maximizing utility does not imply zero consumption of either energy or other goods.

Each X_k is by construction nonnegative so to satisfy Assumptions A2 and A5 we assume the support of each X_k is $(0, \infty)$, the closure of which includes zero. We are therefore assuming that in theory prices can be very low and/or total expenditures can be arbitrarily large. In our empirical application we observe a very wide range of total expenditure levels (from poor to wealthy individuals), and substantial relative price variation. Further, in our empirical specification we also exploit variation in other regressors Z to aid identification (see the next subsection for details). An additional mitigating factor is that our identification results do not depend solely on thin sets, given the additional restrictions that were discussed in section 3.3. We also provide a Monte Carlo study (in the online supplemental appendix) with design given by our actual estimated model and using our actual X_1, X_2 and Z data. This Monte Carlo shows good performance of our estimator for sample sizes in the range of our empirical application.

4.3 Random Barten Scales: Specification and Estimation

In our main results, we present estimates of a parametric model wherein the demand function (9) given by the difference in the logs of squared cubic functions of X_1 and X_2, the unobserved preference heterogeneity parameters follow a truncated bivariate log-normal distribution, and the normalized prices X_k are taken to be exogenous. We then consider consumer surplus and

[4] 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. A rare example is Jensen and Miller (2008), who show that some grains may have been Giffen goods for extremely poor households in rural China.

social welfare analysis with this parametric specification. In our online supplemental appendix, we provide a variety of robustness checks and tests of model adequacy. In particular, we: (1) allow for the demand function to depend on higher-order squared polynomials of X_k; (2) allow for interaction terms in X_k in the demand function; (3) allow for more general distributions of unobserved preference heterogeneity and of measurement error; (4) allow for heteroskedasticity of the measurement error term U_0; (5) implement a Monte Carlo analysis with design based on our actual estimated model and data; and (6) allow for possible endogeneity in prices (which are the numerators of X_k) via estimation with control functions and exogenous supply shifters. The results are that our major findings regarding demand, inference, consumer surplus, and social welfare analysis survive these robustness checks.

Marshallian budget shares are commonly modeled as equal to, or proportional to, polynomials, almost always of third or lower order in terms of flexibility. See, e.g., Lewbel (2008) and references therein. We therefore specify the functions g_1 and g_2 in (9)as squared third-order polynomials

$$g_1(X_1) = \ln\left[\left(\beta_{10} + \beta_{11}X_1 + \beta_{12}X_1^2 + \beta_{13}X_1^3\right)^2\right] \tag{15}$$

$$g_2(X_2) = -\ln\left[\left(\beta_{20} + \beta_{21}X_2 + \beta_{22}X_2^2 + \beta_{23}X_2^3\right)^2\right] \tag{16}$$

with constants β_{ks} for $k = 1, 2$ and $s = 0, .., 3$. We square these polynomials, analogous to Gallant and Nychka (1987), to ensure that the resulting demand functions will not entail taking logs of a negative number. This specification also has the advantage that we can analytically evaluate the integrals that define the corresponding indirect utility function in Theorem 3. Specifically, by equation (11) we get $V^{-1} = h_1(U_1X_1) + h_2(U_2X_2)$ where

$$h_k(X_k) = \int_{\ln X_k} \left(\beta_{k0} + \beta_{k1}e^r + \beta_{k2}e^{2r} + \beta_{k3}e^{3r}\right)^2 dr \tag{17}$$

As noted earlier, it is both unusual and convenient to have closed form expressions for utility functions corresponding to arbitrary demand function components like these.

We impose the normalizations $\beta_{20} = 1$ and $E(U_0) = 0$, which are free normalizations that take the place of the normalizations of the g functions described in Theorem 2. These observationally equivalent normalizations are used in place of the ones used to prove Theorem 2, because they're more natural and easier to impose in our particular application.

Applying Theorem 3 and substituting (15) and (16) into (9) gives:

$$\lambda(W_1) = \ln\left[\left(\beta_{10} + \beta_{11}X_1 + \beta_{12}X_1^2 + \beta_{13}X_1^3\right)^2\right] - \ln\left[\left(\beta_{20} + \beta_{21}X_2 + \beta_{22}X_2^2 + \beta_{23}X_2^3\right)^2\right] + U_0. \tag{18}$$

We next need to specify the distributions of U_0, U_1 and U_2. The distribution of $F_{U|Z}(U_1, U_2 \mid Z)$ for a vector of observed demographic characteristics Z is in theory nonparametrically identified. But to reduce the dimensionality of the model, instead of letting the dependence of U on Z be entirely unrestricted, we assume each Barten scale takes the form

$$U_k = \alpha_k(Z)\,\widetilde{U}_k, \tag{19}$$

for $k = 1, 2$ and where each function $\alpha_k(Z)$ is a traditional deterministic Barten scale, and the remaining random variation given by $\widetilde{U}_k$ in each Barten scale is assumed to be independent of the covariates X_k, Z. We model $\ln[\alpha_k(Z)]$ as linear in a vector of demographic characteristics Z. This index does not include a constant term, because the scaling of $\alpha_k(Z)$ is freely absorbed into the β_{ks} parameters. An additional advantage of this specification is that we now have the support of each $\alpha_k(Z) X_k$ instead of just the support of X_k to help identify the distribution of $\widetilde{U}_k$. Thus, we exploit variation in Z to aid in the identification of the distribution of $\widetilde{U}_k$.

We specify f_0, the density of U_0, as a mean zero normal with variance σ_0^2. We specify the joint distribution of the random component of the Barten scales, $\widetilde{U} = (\widetilde{U}_1, \widetilde{U}_2)$, to be a truncated (trimmed) bivariate log-normal. Specifically, before truncation, the density of $\ln \widetilde{U}$ is

$$f_{\ln \widetilde{U}}\left(\widetilde{U}_1, \widetilde{U}_2, \sigma_1, \sigma_2, \rho\right) = \frac{1}{2\pi\sigma_1\sigma_2\left(1-\rho^2\right)^{1/2}} \exp\left(\frac{\left(\frac{\ln \widetilde{U}_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{\ln \widetilde{U}_2}{\sigma_2}\right)\left(\frac{\ln \widetilde{U}_1}{\sigma_1}\right) + \left(\frac{\ln \widetilde{U}_2}{\sigma_2}\right)^2}{-2\left(1-\rho^2\right)}\right). \tag{20}$$

We truncate the distribution to make the support of $\ln \widetilde{U}$ equal the box defined by $\pm 3\sigma_1$, $\pm 3\sigma_2$. Bounding the support of $\widetilde{U}$ in this way satisfies the assumptions of Theorem 1 that $\widetilde{U}$ and hence U be bounded away from zero and ensures that a moment generating function exists. Our estimation method uses numerical integration, and in our the empirical application we implement this distribution by integrating $\ln \widetilde{U}$ over its bounded support with a stepsize of $0.06\sigma_k$, yielding a 10,000 point grid for the numerical integration.

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

$$\begin{aligned} & f_{W_1|X_1,X_2,Z}(w_1 \mid x_1, x_2, z; \alpha, \beta, \sigma, \rho) \\ = & \int_0^\infty \int_0^\infty f_0\left[\lambda(W_1) - \lambda\left[\omega_1\left(\alpha_1(z)\widetilde{u}_1 x_1, \alpha_2(z)\widetilde{u}_2 x_2, \beta\right)\right], \sigma_0\right] f_{\widetilde{U}}\left(\widetilde{u}_1, \widetilde{u}_2, \sigma_1, \sigma_2, \rho\right) \partial\widetilde{u}_1 \partial\widetilde{u}_2. \end{aligned} \tag{21}$$

Substituting in the logit transformation λ, the demand function (18), the barten scale functions (19), the joint log-normal distribution (20) (for U_1, U_2) and the normal distribution (for U_0) into the conditional density function (21) gives

$$\begin{aligned} & f_{W_1|X_1,X_2,Z}(w_1 \mid x_1, x_2, z; \alpha, \beta, \sigma, \rho) = \\ & \int_{-\infty}^\infty \int_{-\infty}^\infty \exp\left(\frac{-1}{2\sigma_0^2}\left[\ln\left(\frac{W_1}{1-W_1}\right) - \ln\left(\left(\frac{\sum_{s=0}^3 \beta_{1s}\left(\widetilde{u}_1\alpha_1(z)x_1\right)^s}{\sum_{s=0}^3 \beta_{2s}\left(\widetilde{u}_2\alpha_2(z)x_2\right)^s}\right)^2\right)\right]^2\right) \frac{f_{\ln \widetilde{U}}\left(\widetilde{u}_1, \widetilde{u}_2, \sigma_1, \sigma_2, \rho\right)}{(2\pi)^{1/2}\sigma_0} \partial \ln \widetilde{u}_1 \end{aligned} \tag{22}$$

Assuming N independently, identically distributed observations $w_{1i}, x_{1i}, x_{2i}, z_i$ of consumers i, estimation proceeds by searching over parameters α, β, σ, and ρ to maximize the log likelihood function

$$\sum_{i=1}^N \ln f_{W_1|X_1,X_2,Z}(w_{1i} \mid x_{1i}, x_{2i}, z_i; \alpha, \beta, \sigma, \rho). \tag{23}$$

5 Empirical Results

5.1 Data

We estimate a baseline parametric specification as above, and with it undertake Engel curve, cost-of-living and social welfare analyses. We estimate the model 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 25-64 residing in cities of 30,000 or more residents in provinces other than Prince Edward Island (due to data masking). We drop observations whose expenditures on energy goods are zero, and eliminate a few extreme outliers by removing those whose total nondurable expenditures are in the top or bottom percentile of the total nondurable expenditure distribution. This leaves 9971 observations for estimation.

We consider the budget share of energy goods, W_1, defined as the share of total nondurable expenditures devoted to energy goods. Total nondurable expenditures are constructed as 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 defined as fuel oil, electricity, natural gas and gasoline (reported in thousands of dollars). We include eight demographic characteristics, comprising the vector Z, as observed preference shifters: a dummy for female individuals; age of the individual (on an 8 unit integer scale for 5 year age groups with age 40 to 44 coded as 0); calendar year minus 2002; a dummy for residence in the francophone province of Quebec; Environment Canada ex poste records of the number of days requiring heating and cooling in each province in each year (normalized as z-scores from the full sample of all households in all provinces in all city sizes); an indicator that the household is a renter (spending more than $100 on rent in the year); and an indicator that the household received more than 10% of its gross income from government transfers. These demographic characteristics equal zero for the reference consumer (whose utility function is $\overline{S}$ and indirect utility function is V): a single male aged 40-44 with less than 10% transfer income living in owned accomodation outside Quebec in 2002 with average heating and cooling days.

Table 1: Summary Statistics				
9971 Observations	mean	std dev	min	max
logit energy share, Y	-1.949	0.766	-7.140	1.005
energy share, W_1	0.146	0.085	0.001	0.732
nondurable expenditure, M	15.661	7.104	2.064	41.245
energy price, P_1	1.039	0.230	0.426	1.896
non-energy price, P_2	0.965	0.075	0.755	1.284
energy normalized price, X_1	0.082	0.049	0.015	0.570
non-energy normalized price, X_2	0.077	0.045	0.020	0.476
female indicator	0.482	0.500	0.000	1.000
age group-4	0.549	2.262	-3.000	4.000
year-2002	0.363	3.339	-5.000	6.000
Quebec resident	0.168	0.374	0.000	1.000
heat days, normalized	-0.102	0.990	-2.507	2.253
cooling days, normalized	0.014	1.007	-1.729	4.013
renter indicator	0.512	0.500	0.000	1.000
transfer income indicator	0.184	0.387	0.000	1.000

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 (a free normalization) in Ontario in 2002. To account for the impact on prices of individual variation in compositional differences of these aggregate commodities, we follow the methodology of Lewbel (1989) and Hoderlein and Mihaleva (2008) in constructing 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. This construction has the feature of further increasing relative price variation across households. The budget, M, is equal to the total nondurable expenditures of the household. The normalized prices X_k are then given by $X_k = P_k/M$. Finally, the regressand, Y, is the logit transformation of the energy budget share, so $Y = \lambda(W_1)$. Table 1 gives summary statistics for these budget shares, expenditures, prices, normalised prices and demographic preference (Barten scale) shifters.

5.2 Parameter Estimates

Our main analyses are based on two models. The first, Model 1, imposes the restriction that $\widetilde{U}_1 = \widetilde{U}_2 = 1$ (so $\widetilde{U}$ is degenerate), and simplifies equation (21) to

$$f_{W_1|X_1,X_2,Z}(w_1 \mid x_1, x_2, z; \alpha, \beta, \sigma_0) = \frac{\exp\left(\frac{-1}{2\sigma_0^2}\left[\ln\left(\frac{W_1}{1-W_1}\right) - \ln\left(\left(\frac{\sum_{s=0}^{3}\beta_{1s}(\alpha_1(z)x_1)^s}{\sum_{s=0}^{3}\beta_{2s}(\alpha_2(z)x_2)^s}\right)^2\right)\right]^2\right)}{(2\pi)^{1/2}\sigma_0}. \tag{24}$$

This is just a traditional deterministic Barten scale model, having $U_k = \alpha_k(z)$, estimated using our general functional form for energy demand. Model 1 is then compared to our real specification, Model 2, which is equation (22) with the distribution of $\widetilde{U}$ given by equation (20), and

therefore contains our random Barten scales $U_k = \alpha_k(z)\,\widetilde{U}_k$. Both models are estimated using maximum likelihood in Stata, with likelihood functions given by substituting equation (24) or (22) into equation (23). Estimated coefficients are given in Table 2 below.

As noted earlier, in our online supplemental appendix we provide an extensive set of analyses to verify the robustness of our baseline empirical results (given in Table 2). A brief summary of these results is that, while some departures from our baseline Model 2 are statistically significant, none result in big changes in our economic analyses or conclusions, indicating that our results are robust to many different possible sources of misspecification and estimation imprecision.

Our model imposes the equality constraints of Slutsky symmetry and homogeneity. However, we do not impose the inequality constraints that the g functions be monotonic or that the Slutsky matrix be negative semidefinite (concavity).[5] Despite not imposing these conditions, we find that our Model 1 estimates satisfy monotonicity throughout our observed data, and that our Model 2 estimates satisfy monotonicity at 97.7% of the data points in our sample. Similarly, our Model 1 estimates satisfy negative semidefiniteness at 99.6% of the data points in our sample, and our Model 2 estimates satisfy it at 99.9% of the data points in our sample.

Model 2 has three more parameters than Model 1. They are σ_1, σ_2, ρ, the standard deviations and correlation coefficient of the bivariate normal distribution of $\ln U_1, \ln U_2$. The likelihood ratio test statistic for the restriction that these parameters are all zero is 672, so the parameters that allow for random Barten scales parameters are highly jointly significant. One can see in Table 2 that they are also individually highly significant.

Figures 1 and 2 show the estimated joint distribution of $\ln \alpha_1(z)$ and $\ln \alpha_2(z)$ (logged deterministic portion of the Barten scales) in Model 1 and Model 2, respectively. Summary statistics for these distributions are provided in the bottom panel of Table 2. In both Models the estimated distributions of $\ln \alpha_1(z), \ln \alpha_2(z)$ are bimodal. The two modes are driven almost entirely by the renter variable; conditioning on just renters or just owners produces unimodal distributions. In Canada, most renters do not pay for their own home heating or electricity (this is included in rents and doesn't depend on usage), causing a ceteris paribus reduction in their energy shares relative to home owners.

Looking at the bottom of Table 2, we see in both models that the standard deviation of $\ln \alpha_2(z)$ is much larger than that of $\ln \alpha_1(z)$. This indicates that heterogeneity in preferences due to observables is larger for non-energy than for energy goods. This is not surprising; it just says that people vary more (based on observable characteristics) in their taste for non-energy goods than in their taste for energy goods. Unconditionally, $\ln \alpha_1(z), \ln \alpha_2(z)$ are slightly negatively correlated in both models. However, conditional on rental tenure, the deterministic components of Barten scales are strongly positively correlated.

[5] In a parametric setting like ours, failing to impose inequality constraints on estimation that are satisfied by the true model does not affect standard limiting distribution theory, assuming that the true parameter values do not lie on the boundary of the parameter space.

Table 2: Estimated Parameters

Parameter		Model 1 llf=−10043.1 Estimate	*Std Err*	Model 2 llf=−9706.9 Estimate	*Std Err*
β_{10}		0.145	*0.010*	0.185	*0.007*
β_{11}		8.113	*0.487*	7.623	*0.287*
β_{12}		-37.563	*2.924*	-32.871	*2.147*
β_{13}		51.576	*5.650*	40.630	*4.390*
β_{21}		2.484	*0.568*	1.805	*0.266*
β_{22}		-1.743	*0.663*	1.053	*0.314*
β_{23}		0.152	*0.141*	-0.996	*0.139*
α_1	female	-0.214	*0.031*	-0.228	*0.015*
	agegp	0.002	*0.009*	0.013	*0.004*
	time	-0.013	*0.004*	-0.003	*0.002*
	PQ	0.085	*0.043*	0.043	*0.021*
	heat	0.036	*0.016*	0.026	*0.008*
	cool	-0.062	*0.015*	-0.035	*0.007*
	renter	-0.292	*0.058*	-0.440	*0.026*
	social	0.034	*0.038*	0.054	*0.020*
α_2	female	-0.130	*0.076*	-0.117	*0.010*
	agegp	-0.068	*0.023*	-0.038	*0.002*
	time	0.018	*0.010*	0.044	*0.001*
	PQ	0.402	*0.100*	0.217	*0.017*
	heat	0.015	*0.040*	-0.021	*0.008*
	cool	-0.077	*0.043*	-0.014	*0.006*
	renter	0.943	*0.155*	0.605	*0.008*
	social	-0.085	*0.091*	-0.110	*0.011*
σ_0		0.663	*0.005*	0.469	*0.009*
σ_1				0.165	*0.036*
σ_2				1.336	*0.011*
ρ				0.883	*0.100*
std dev	$\ln(\alpha_1)$	0.197		0.252	
	$\ln(\alpha_2)$	0.568		0.380	
correlation	$\ln(\alpha_1), \ln(\alpha_2)$	-0.479		-0.700	
(all obs)	$\ln U_1, \ln U_2$			0.293	
correlation	$\ln(\alpha_1), \ln(\alpha_2)$	0.426		0.105	
(renter=0)	$\ln U_1, \ln U_2$			0.699	
correlation	$\ln(\alpha_1), \ln(\alpha_2)$	0.420		0.087	
(renter=1)	$\ln U_1, \ln U_2$			0.691	

In Model 1 the log Barten scales equal $\ln \alpha_k(z)$, but in Model 2 the log Barten scales are given by $\ln U_k = \ln \alpha_k(z) + \ln \widetilde{U}_k$. The components $\ln \alpha_k(z)$ and $\ln \widetilde{U}_k$ are, respectively, the observed deterministic and unobserved random components of these Barten scales. Thus the variance and correlations of the $\ln \widetilde{U}_k$ terms in Model 2 are directly comparable to the corre-

sponding statistics of the $\ln \alpha_k(z)$ terms. The estimated parameters of the distribution of $\ln \widetilde{U}_k$ show some similar features to that of $\ln \alpha_k$. The standard deviation $\ln \widetilde{U}_2$ (equal to σ_2) is much larger than that of $\ln \widetilde{U}_1$ (equal to σ_1), so both observed and unobserved components of Barten scales vary more across consumers for nonenergy goods than for energy goods. Likewise, $\ln \widetilde{U}_1$ and $\ln \widetilde{U}_2$ are positively correlated (like $\ln \alpha_1$ and $\ln \alpha_2$ after conditioning on the rental dummy). That the estimated effects of the unobserved random components of Barten scales have similar patterns to the estimated effects of observed preference shifters of Barten scales is a reasuring indicator of the sensibility and reasonableness of our model.

Overall, we find that unobserved preference heterogeneity is about as important as observed preference heterogeneity in driving variation in Barten scales. The estimated joint distribution of $\ln U_1, \ln U_2$, summing the effects of $\ln \alpha_k(z)$ and $\ln \widetilde{U}_k$, is shown in Figure 3. Comparing Figures 1 and 3 shows that accounting for unobserved heterogeneity substantially increases the estimated total heterogeneity in tastes across individuals.

The unobserved preference heterogeneity terms $\widetilde{U}_k$ partly pick up unobserved variation that would otherwise be subsumed by the non behavioral error term U_0, making the estimated standard deviation of U_0 fall from 0.666 in Model 1 to 0.469 in Model 2. But more significantly, U_1 and U_2 also pick up a substantial portion of what would otherwise be unexplained heteroskedasticity in demand. In the online supplemental appendix, when we consider heteroskedastic U_0, we find that allowing for unobserved preference heterogeneity via our random Barten scales accounts for a great deal of variation that would otherwise have been falsely attributed to heteroskedastic measurement error.

Accounting for unobserved heterogeneity appears to yield precision benefits as well. Appropriately modeling the heteroskedasticity driven by unobserved preference heterogeneity should increase precision in parameter estimates, just as correctly specified generalized least squares estimation usually reduces standard errors relative to ordinary least squares estimation in heteroskedastic regression models. Empirically, we do see an improvement in estimation precision, comparing across the columns in Table 2. The parameter estimates in Model 2 generally have standard errors about 20 per cent to 50 per cent smaller than those of Model 1.

Table 3 gives summary statistics on predicted values of the logit transformed budget share Y and of the budget share itself, W_1. We give estimates for Model 1 evaluated at the observed data in the left panel, and for Model 2 evaluated at the observed data with unobserved preference heterogeneity parameters "turned off" ($\widetilde{U}_1 = \widetilde{U}_2 = 1$) in the middle panel. In the rightmost panel, we present estimates simulated at the observed data with the estimated distribution of unobserved preference heterogeneity parameters $\widetilde{U}_1, \widetilde{U}_2$.

Since Model 1 has just a single additively separable error term, the average prediction from Model 1 including variation from all regressors exactly equals the mean of the observed Y (-1.949). The predicted average mean of Y in Model 2 is somewhat larger for the cases where we don't account for unobserved preference heterogeneity $\widetilde{U}_k$, with Y averaging about -1.78. But, when we account for unobserved preference heterogeneity (in the rightmost panel), the average prediction of Model 2 predicts is -1.997 which is very close to the mean of the observed Y. The standard deviation of Model 1 predictions of W_1 is 0.044, while that of Model 2 is 0.065. Comparing these predictions to Table 1 shows that, for both Y and W_1, Model 2 gives closer predictions to the actual empirical standard deviation of these variables than does Model 1.

Table 3: Estimated Energy Budget Shares						
	Model 1		Model 2 ($\widetilde{U}_k = 1$)		Model 2	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Logit Budget Shares Y	-1.949	0.387	-1.775	0.371	-1.997	0.896
Budget Shares W_1	0.131	0.044	0.151	0.047	0.137	0.065

Figure 4 shows estimated Engel curves from the models, showing W_1 as a function of $\ln M$. These are estimated demand functions evaluated at the Ontario 2002 prices $P_1 = P_2 = 1$ and at average demographics $\overline{\alpha}_k$. Model 2 implies a different Engel curve for every value that $\widetilde{U}_1$ and $\widetilde{U}_2$ can take on. The single Engel curve for Model 1 is shown as a thick gray line, while that for Model 2 evaluated at $\widetilde{U}_k = 1$ is shown as a thick black line. To illustrate the range of Engel curves implied by our model, we also evaluate Model 2 at each quartile of the distribution of $\widetilde{U}_1$ paired with each quartile of the distribution of $\widetilde{U}_2$, for a total of nine pairs of values. This yields eight additional Engel curves, which are depicted by thin gray lines in Figure 4. Finally, we show the estimated marginal density of $\ln M$ (divided by 10 to fit in the graph) as a thick light gray line at the bottom of Figure 4.

On average, richer consumers tend to spend a smaller fraction of their budget on energy goods than poorer consumers. This can be seen in the mostly downward slope of the Engel curves in Figure 4. Comparing these curves to the depicted density function of $\ln M$ shows that only a small fraction of all consumers, the poorest ones, are on the upward sloping parts of these curves.

A striking feature of Figure 4 is that for each given value of M, there is substantial variation in the level and slope of Engel curves, due entirely to variation in $\widetilde{U}_k$. To reduce clutter we did not include standard error bars on this graph, but the differences between these estimated Engel curves are statistically significant. For example, at the mean value of $\ln M$ ($\ln M = 2.64$), the top Engel curve displayed is that of the top quartile of both $\widetilde{U}_1$ and $\widetilde{U}_2$, and the bottom Engel curve is that of the bottom quartile of both $\widetilde{U}_1$ and $\widetilde{U}_2$. The estimated levels of these Engel curves at $\ln M = 2.64$ are 0.166 and 0.104, respectively, with standard errors of 0.002 and 0.008, repectively. We find that variation in the random components $\widetilde{U}_k$ of Barten scales, corresponding to unobserved variation in tastes across consumers, yields significant differences in both the levels and slopes of the estimated Engel curves.

Figure 4 showed the effects on W_1 of just the random component of the Barten scales at different budget levels M. In contrast, Figure 5 illustrates the total effect of Barten scales on W_1. Specifically, Figure 5 shows a contour plot of the joint distribution of W_1 and M predicted by Model 2, evaluated at Ontario 2002 prices $P_1 = P_2 = 1$ and observed demographics z. The vertical variation in this graph therefore shows the estimated variation in W_1 due to Barten scales (both observed and unobserved components) at different M levels. This wide variation in tastes will have important implications for welfare analyses below.

Taken together, all of the above results show that Model 2's inclusion of random Barten scales accounts for more and richer variation in observed behaviour than does Model 1. This is due to the fact that budget shares are highly variable and heteroskedastic, and Model 1 treats this variance and heteroskedasticity entirely as meaningless variation in a nonbehavioral error term, while Model 2 captures much of this variation in a behaviorally sensible, structural way, via random Barten scales.

5.3 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.[6] Using equation (17), we have a closed form expression for indirect utility.[7] We can therefore numerically invert the indirect utility function (17) to obtain the exact cost of living impact (consumer surplus) of a price change. Without Theorem 3, we could instead use numerical approximation (instead of an exact solution) such as in Vartia (1984), or we could 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.

Recall the indirect utility function defined over normalized prices X_k and Barten scales U_k is $V(U_1X_1, U_2X_2) = V(U_1P_1/M, U_2P_2/M)$. For an individual facing initial prices $\overline{P}_1$, $\overline{P}_2$, having total expenditures M, and having 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

$$V\left(\frac{U_1\overline{P}_1}{M}, \frac{U_2\overline{P}_2}{M}\right) = V\left(\frac{U_1P_1}{\pi M}, \frac{U_2P_2}{\pi M}\right).$$

Here π is the cost-of-living index giving the proportionate change in costs M needed to compensate for the price change from $\overline{P}_1$, $\overline{P}_2$ to P_1, P_2.

To show price effects clearly, we consider a large price change: a 50% increase in the price of energy. This price increase is chosen to approximate the effect of a \$300 per ton CO2 tax (see, e.g., Rhodes and Jaccard 2014)[8]. We solve for the π function given the initial price vector $\overline{P}_1 = \overline{P}_2 = 1$ and the new price vector $P_1 = 1.5$, $P_2 = 1$. Figure 6 shows the resulting estimated joint distribution (contour plot) of $\ln \pi$ and $\ln M$ from Model 2 evaluated at the observed demographics Z and budgets M. This plot is constructed by calculating π for each observation in the data, with draws from the estimated distribution of $\widetilde{U}_1$, $\widetilde{U}_2$, and using observed values of the preference shifters z.

Table 4 gives summary statistics (means and standard deviations) of these distributions for Model 1 and Model 2. Analogous to Table 3, to assess the contribution of variation in π due to observed and unobserved preference heterogeneity, we also calculate the π distribution imposing $\widetilde{U}_k = 1$ for Model 2. For ease of presentation this Table reports percent changes in cost of living, that is, $100(\pi - 1)$. Standard errors for these statistics that account for the sampling variability of the parameter estimates (estimated via simulation) are provided in *italics*.

[6] 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.

[7] The integral in equation (17) is readily evaluated, e.g., for our squared cubic functions we have $\int \left(\beta_{k0} + \beta_{k1}e^r + \beta_{k2}e^{2r} + \beta_{k3}e^{3r}\right)^2 dr =$ $\frac{1}{6}\beta_{k3}^2 e^{6r} + \frac{2}{5}\beta_{k2}\beta_{k3}e^{5r} + \frac{1}{4}\left(2\beta_{k1}\beta_{k3} + \beta_{k2}^2\right)e^{4r} + \frac{2}{3}\left(\beta_{k0}\beta_{k3} + \beta_{k1}\beta_{k2}\right)e^{3r} + \frac{1}{2}\left(2\beta_{k0}\beta_{k2} + \beta_{k1}^2\right)e^{2r} + 2\beta_{k0}\beta_{k1}e^r + r\beta_{k0}^2$. Each $h_k(U_kX_k)$ function is given by substituting $r = \ln(U_kX_k)$ into this expression.

[8] The Canadian province of British Columbia has a CO2 tax. It charges 6.7 cent/liter of gasoline for 30\$/ton. A 67 cent/liter CO2 tax is about half as large as the pump price of gasoline in 2002 (the base year for this analysis).

Table 4: Cost of Living Impacts: 50% Energy Price Increase							
Per Cent Increase		Model 1		Model 2 ($\widetilde{U}_k = 1$)		Model 2	
$\pi - 1$, per cent		Estimate	*Std Err*	Estimate	*Std Err*	Estimate	*Std Err*
$\alpha_k, \widetilde{U}_k = 1$	Mean	5.31	*0.24*	5.64	*0.17*	5.37	*0.20*
	Std Dev	1.85	*0.21*	1.69	*0.08*	4.31	*0.46*

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, essentially by ignoring substitution effects (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 0.146, 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 7.3 per cent. This would be the first order approximation based estimate of π.

The estimated cost-of-living impacts given in Table 4, averaging about 5.4 per cent, are much smaller than 7.3 per cent, showing substantial responses to relative prices. This difference of nearly 2 percentage points is very large relative to the standard error of estimated means in Models 1 and 2 of about 0.20 percentage points, so the hypothesis that the model estimates have a mean of 7.3 is strongly rejected at conventional levels. These results 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 the consumer surplus and welfare effects of large price changes. Moreover, one goal of an energy tax would be to reduce energy consumption (a substitution effect), so it's important to account for the impact on welfare of this reduction.

Models 1 and 2, with or without variation in α_k or $\widetilde{U}_k$, deliver similar estimates of the mean effects of the energy tax on cost of living. However, the inclusion of the random Barten scale components $\widetilde{U}_k$ in Model 2 more than doubles the estimated standard deviation of π across consumers. This difference is statistically significant as well as being economically large; the z-test statistic for the hypothesis that these standard deviations are the same has value of 6.1. Though less substantial economically, the difference in mean effects between the Model 2 estimates without and with unobserved preference heterogeneity (5.64% and 5.37%, respectively) is also statistically significant with a z-test statistic of 3.9.

The large estimated standard deviation of π in Model 2 (which is mostly due to unobserved variation in tastes $\widetilde{U}_k$) has substantial welfare implications. The larger is the variation in π, the larger is the variation in impacts of an energy tax. Although the average consumer would need to have their budget M increased by 5.37% to compensate for the tax, some consumers (those near the bottom of Figure 6) would only need a slight increase in their budget to be made whole, while others (those near the top of Figure 6) would be greatly harmed by the tax, needing more than a 10% increase in the budget M to compensate.

What makes this substantial variation in cost of living impacts particularly relevant economically is that it mostly impacts poorer consumers. As can be seen in Figure 6, both the mean and the variation in cost of living impacts is larger at low values of M than at high values, so those consumers who are hurt the most by the tax in percentage terms are also predominantly the poorer consumers, who can least afford the increase in costs. Not only do richer consumers spend a smaller fraction of their budget on energy goods (as seen in Figure 4), but they also

appear to have a greater ability to substitute away from energy when the relative price of energy increases.

5.4 Social Welfare Implications of a Carbon Tax

The above analysis showed the distribution across consumers of the cost of living effects of an energy tax. We now evaluate the implications of these results for aggregate welfare, based on a range of social welfare functions. To evaluate social welfare functions, we require interpersonally comparable, cardinalized measures of individual utility. 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 of V having the property that, evaluated at base prices $\overline{P}_1$, $\overline{P}_2$, the function $\widetilde{V} = M$. We therefore define cardinalized utility $\widetilde{V}$ by

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

where, by definition, the monotonic transformation function H is chosen so that

$$\widetilde{V}\left(U_1\overline{P}_1/M, U_2\overline{P}_2/M\right) = M$$

for all values of M, U_1, U_2. We let base prices be $\overline{P}_1 = \overline{P}_2 = 1$.

The money-metric function $\widetilde{V}$ gives a utility level that may be interpreted as the number of dollars that, at base prices, delivers the same level of utility that the consumer can achieve with a budget of M dollars when facing prices P_1, P_2. Attained utility depends on U_1, U_2, so the function H that yields $\widetilde{V}$ will likewise depend on these Barten scales.

The range of social welfare functions that we consider are in the Atkinson (1970) Mean-of-Order-r class, defined by

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

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. Since $\widetilde{V}_i$ is measured in dollars, so too are the welfare functions S_r.

We compute proportionate welfare losses $\triangle S_r$ equal to welfare at base prices minus welfare at new prices divided by welfare at base prices. The money metric at base prices ($\overline{P}_1 = \overline{P}_2 = 1$) equals the budget M_i of each consumer. Letting $\widetilde{V}_i$ be the money metric at M_i and new prices $P_1 = 1.5$, $P_2 = 1$, we have

$$\triangle S_r = \left(S_r(M_1 ... M_N) - S_r(\widetilde{V}_1 ... \widetilde{V}_N)\right)/S_r(M_1 ... M_N)$$

A first cut at welfare analysis is to employ a first-order approximation of the money metric. A standard approximation of the individual money-metric utility associated with a given price vector, which in our case is (1.5, 1), is given by $\widetilde{V}_i = M_i/(W_{1i} * 1.5 + (1 - W_{1i}) * 1)$. This is the Laspeyres index approximation to the money metric for consumer i. The welfare loss associated with this approximate money metric accounts for some heterogeneity across individuals

(because W_{1i} differs across individuals) but does not account for substitution responses. The arithmetic mean welfare index computed using this approximate money-metric utility shows a welfare loss of 6.30 per cent. The inequality-averse geometric and harmonic mean welfare indices show higher welfare losses of 6.74 and 7.18 per cent, respectively. This is because energy budget shares W_{1i} are negatively correlated with budgets M_i, so the welfare indices that upweight poorer households show a greater welfare loss associated with lower budgets and hence with higher energy budget shares.

These approximate social welfare loss numbers can be compared with the estimates from Models 1 and 2, provided in Table 5. Standard errors accounting for the sampling variability of the estimated parameters were generated via simulation and are shown in *italics*. In Table 5 we account for both substitution effects and heterogeneity using our models.

Table 5: Social Welfare Loss: 50% Energy Price Increase

		Model 1		Model 2 ($\widetilde{U}_k = 1$)		Model 2	
Welfare Loss, Per Cent	r	Estimate	*Std Err*	Estimate	*Std Err*	Estimate	*Std Err*
Arithmetic Mean	1	4.81	*0.19*	5.14	*0.14*	4.91	*0.17*
Geometric Mean	0	5.27	*0.26*	5.43	*0.16*	5.37	*0.22*
Harmonic Mean	-1	6.14	*0.54*	5.48	*0.17*	5.85	*0.30*

Every estimate in Table 5 is lower than its corresponding first-order approximation. For example, the welfare loss given the geometric mean welfare index is about 1.5 percentage points lower (about one-fourth lower) than the first-order approximation. This shows that accounting for substitution effects has a substantial effect on welfare. All the estimates also show welfare losses increasing with the inequality aversion of the welfare index. As discussed above, this is primarily due to the downward sloping Engel curves as seen in Figure 4.

Another feature seen in Table 5 is that the estimates based on Model 2, which account for unobserved preference heterogeneity, have smaller standard errors. For the more inequality-averse welfare measures, this improvement in precision is substantial, e.g., the estimated harmonic mean welfare loss for Model 2 has about half the standard error of that for Model 1. This is due to the fact that the Model 2 treatment of unobserved heterogeneity increases precision of estimated parameters over model 1, as discussed earlier.

Further, we find that dealing with unobserved preference heterogeneity affects both the level and pattern of estimated welfare losses. There are two ways in which the misspecification of Model 1 compared to Model 2 matters for welfare analysis. First, Model 2 Engel Curves are on average less downward sloping than those of Model 1. This means that, even without accounting for unobserved preference heterogeneity, when we consider welfare functions that are very inequality-averse, Model 1 will tend to overstate welfare losses. For example, in the upper panel of Table 5, the harmonic mean index shows a welfare loss of 6.14 per cent for Model 1 but only 5.48 per cent for Model 2.

A second difference is that Model 2 has greater variance in individual utility losses than does Model 1, and inequality-averse welfare functions will tend to penalize such variance. This effect can be seen by comparing the middle and right panels of Table 5. In the right panel we account for unobserved preference heterogeneity, which modestly increases the welfare loss for inequality-averse welfare functions (from 5.48 to 5.85 per cent).

The welfare loss in Model 2 not accounting for unobserved preference heterogeneity (middle

panel) is 0.34 percentage points higher with the harmonic versus the arithmetic mean welfare index. Because these welfare loss measures are highly positively correlated, this difference is statistically significant, with a standard error of 0.05. In the right panel of Table 5, where we add variation due to unobserved preference heterogeneity, the difference in estimated welfare loss is 0.94 percentage points, with a standard error of the difference of 0.15. Thus, appropriately accounting for unobserved preference heterogeneity dramatically increases the sensitivity of estimated welfare loss to the inequality-aversion of the welfare index.

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 impacts of price changes. In our example, the standard deviation of cost-of-living impacts due to a 50% energy tax is more than doubled by accounting for unobserved preference heterogeneity by random Barten scales. Second, we find that accounting for unobserved preference heterogeneity changes welfare analyses in a variety of ways, and in particular decreases estimated welfare loss when the welfare index is inequality-averse.

6 Conclusions

We show nonparametric identification of a generalized random coefficients model. We also provide an empirical application in which the generalized random coefficient structure arises naturally from extending an existing commonly used economic model of observed heterogeneity to a model allowing for unobserved heterogeneity. In this Barten scales application, we allow for general forms of unobserved preference heterogeneity that are shown to be important for empirically evaluating the welfare effects of potential policy interventions such a carbon tax. For example, we find that accounting for unobserved preference heterogeneity more than doubles the estimated variation in impacts of an energy tax (as measured by the standard deviation across consumers of the cost of living impact of the tax).

Accounting for unobserved preference heterogeneity via Barten scales is economically important because, by including them, we discovered that the variance in the impacts of energy prices is particularly large among poorer consumers. As a result, an energy tax may impact inequality and welfare through a previously unrecognized channel, which is the variation in energy price elasticities across households. We find that this variation in price responses is particularly large for poorer households. This can be seen in Figure 6, where the variation in cost of living impacts of an energy price increase is largest at low values of total expenditures. This figure also confirms the previously known result that mean cost of living impacts of energy price increases are also higher for the poor. Roughly, these results say that not only are energy taxes regressive on average (the previously known result), but also the degree of regressivity varies more among poorer consumers than among the rich.

This result has important implications for social welfare. Empirically, the unobserved preference heterogeneity revealed by our model strongly interacts with inequality aversion in social welfare calculations, reversing conclusions that would have been made using almost all previous demand models, which fail to account for such preference heterogeneity.

Useful areas for further work on the theory of generalized random coefficients would be finding conditions under which Theorem 1 alone could be used to identify the model with-

out the identification at zero assumptions used in Theorem 2, formalizing the extent to which additional implications of the model like those discussed in section 3.3 might be used to identify more general structures. It would also be useful to investigate how the assumptions used for identification might interact with assumptions needed for inference based on nonparametric estimators such as sieve maximum likelihood.

Our application focuses on consumers with single utility functions, that is, unitary households. A possible extension would be to consider collective household models. For example, Barten scales are used in the collective household models of Browning, Chiappori, and Lewbel (2013). It would also be useful to extend our carbon tax analyses into a general equilibrium setting, which would affect the conditional independence assumptions we used for identification of the random Barten scales.

7 Appendix A: Proofs

PROOF of Lemma 1: We have $G_k(X_k) \geq c_k X_k$. Let $h(G,t) = e^{-\rho G}$ for any $\rho > 0$. Then

$$\kappa_t = \prod_{k=1}^{K} \int_0^\infty e^{-\rho G_k(s_k)} s_k^{t_k-1} ds_k \leq \prod_{k=1}^{K} \int_0^\infty e^{-\rho c_k s_k} s_k^{t_k-1} ds_k = \prod_{k=1}^{K} (\rho c_k)^{-t_k} \int_0^\infty e^{-r_k} r_k^{t_k-1} dr_k$$

which is finite, because the gamma function $\Gamma(t_k) = \int_0^\infty e^{-r_k} r_k^{t_k-1} dr_k$ is finite. Also, $\kappa_t > 0$ because the function being integrated is strictly positive everywhere in the interior of $supp(X)$.

PROOF of Lemma 2: Let $G(X_1, X_2) = \ln(X_1) + \ln(X_2)$. Start with any choice of the function h. Do the change of variables replacing s_1 and s_2 with s_1 and $r = s_1 s_2$ to get

$$\begin{aligned} \kappa_t &= \int_0^\infty \int_0^\infty h[\ln(s_1) + \ln(s_2), t] s_1^{t_1-1} s_2^{t_2-1} ds_1 ds_2 \\ &= \int_0^\infty \int_0^\infty h[\ln(r), t] s_1^{t_1-t_2-1} r^{t_2-1} ds_1 dr \\ &\quad \left(\int_0^\infty h[\ln(r), t] r^{t_2-1} dr \right) \int_0^\infty s_1^{t_1-t_2-1} ds_1 \end{aligned}$$

and the second integral is not convergent for $t_1 > t_2 - 1$.

PROOF of Theorem 1:

Let $\Omega_X = supp(X)$, $\Omega_{X|Z} = supp(X \mid Z)$, etc. By the definition of $\lambda_t(Z)$ we have $\lambda_t(Z) =$

$$\int_{X \in s\Omega_X} \int_{U \in \Omega_{U|Z}} h(G(X_1U_1, ..., X_KU_K), t) \, dF_{U|Z}(U \mid Z) \, X_1^{t_1-1} X_2^{t_2-1} ... X_K^{t_K-1} dX_1 dX_2 ... dX_K$$

$$= \int_{U \in \Omega_{U|Z}} \int_{X \in \Omega_{X|Z}} h(G(X_1U_1, ..., X_KU_K), t) \, X_1^{t_1-1} X_2^{t_2-1} ... X_K^{t_K-1} dX_1 dX_2 ... dX_K dF_{U|Z}(U \mid Z)$$

where the second equality follows from Fubini's theorem and $supp(X) = supp(X \mid Z)$. Do a change of variables on the inner integral, letting $s_k = X_k U_k$ for $k = 1, ..., K$ to get $\lambda_t(Z) =$

$$\int_{U \in s\Omega_{U|Z}} \int_{s \in supp(X_1 U_1, ..., X_K U_K | Z, U)} h(G(s_1, ..., s), t) s_1^{t_1 - 1} s_2^{t_2 - 1} ... s_K^{t_K - 1} U_1^{-t_1} U_2^{-t_2} ... U_K^{-t_K} ds_1 ds_2 ... ds_K dF_{U|Z}(U \mid Z)$$

$$= \int_{U \in \Omega_{U|Z}} \int_{s \in \Omega_X} h(G(s_1, ..., s), t) s_1^{t_1 - 1} s_2^{t_2 - 1} ... s_K^{t_K - 1} U_1^{-t_1} U_2^{-t_2} ... U_K^{-t_K} ds_1 ds_2 ... ds_K dF_{U|Z}(U \mid Z)$$

$$= \int_{U \in \Omega_{U|Z}} \int_{s \in \Omega_X} h(G(s_1, ..., s), t) s_1^{t_1 - 1} s_2^{t_2 - 1} ... s_K^{t_K - 1} ds_1 ds_2 ... ds_K U_1^{-t_1} U_2^{-t_2} ... U_K^{-t_K} dF_{U|Z}(U \mid Z)$$

$$= \int_{U \in \Omega_{U|Z}} \kappa_t U_1^{-t_1} U_2^{-t_2} ... U_K^{-t_K} dF_{U|Z}(U \mid Z)$$

$$= \kappa_t \int_{U \in \Omega_{U|Z}} U_1^{-t_1} U_2^{-t_2} ... U_K^{-t_K} dF_{U|Z}(U \mid Z) = \kappa_t E\left(U_1^{-t_1} U_2^{-t_2} ... U_K^{-t_K} \mid Z\right)$$

where the second equality above uses $supp(U_1 X_1, ..., U_K X_K \mid Z, U) = supp(X)$. It therefore follows that moments $E\left(U_1^{-t_1} U_2^{-t_2} ... U_K^{-t_K} \mid Z\right)$ are identified for all positive integers t_1,...,t_k by equalling the ratio of identified objects $\lambda_t(Z)/\kappa_t$. To identify moments where $t_k = 0$ for one or more values of k, redefine $\lambda_t(Z)$ and κ_t setting the corresponding X_k terms equal to zero. For example, to identify moments having $t_1 = 0$, replacing equation (4) with

$$\lambda_t(Z) = \int_{(X_2, ... X_K) \in supp(X_2, ... X_K)} E\left[h\left(\widetilde{Y}, 0, t_2, ..., t_K\right) \mid X_1 = 0, X_2 ... X_K, Z\right] X_2^{t_2 - 1} ... X_K^{t_K - 1} dX_2 ... dX_K$$

and correspondingly redefining κ_t as

$$\kappa_t = \int_{(s_2, ... s_K) \in supp(X_2, ... X_K)} h[G(0, s_2, ... s_K), 0, ..., t_K] s_2^{t_2 - 1} ... s_K^{t_K - 1} ds_2 ... ds_K$$

gives $\lambda_t(Z)/\kappa_t$ equal to $E\left(U_2^{-t_2} ... U_K^{-t_K} \mid Z\right)$.

We have now shown that $E\left(U_1^{-t_1} U_2^{-t_2} ... U_K^{-t_K} \mid Z\right)$ is identified for any nonnegative integers t_1,...,t_k. It then follows from Assumption A2 that the joint distribution of $U_1^{-1} U_2^{-1} ... U_K^{-1} \mid Z$ is identified from these moments, and therefore that the joint distribution $F_{U|Z}(U_1, ..., U_K \mid Z)$ is identified.

Before proving Theorem 2, we prove a couple of lemmas.

LEMMA 3: 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 $\widetilde{y}_k = G_k(r)$. Assume H_k is differentiable. 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, for all x_k and $\widetilde{y}_k$ such that $x_k \neq 0$ and $f_{U_k}\left(x_k^{-1} H_k(\widetilde{y}_k)\right) \neq 0$,

$$H_k(\widetilde{y}_k) = sign\left(sign(x_k) \frac{\partial S_k\left(\widetilde{y}_k, x_k^{-1}\right)}{\partial x_k^{-1}}\right) \exp\left(\int_{y_0}^{\widetilde{y}} \frac{x_k \partial S_k\left(\widetilde{y}_k, x_k^{-1}\right)/\partial \widetilde{y}}{\partial S_k\left(\widetilde{y}_k, x_k^{-1}\right)/\partial x_k^{-1}} d\widetilde{y}_k\right) \tag{25}$$

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 3 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 3: For any $\widetilde{y}_k = G_k(x_k U_k)$ and any $x_k > 0$ we have

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

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

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

Together these equations say

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

So

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

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

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

and

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

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

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

so

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

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

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

So the right side of equation (25) equals $sign\left(H_k\left(\widetilde{y}_k\right)\right)\left|H_k\left(\widetilde{y}_k\right)\right| = H_k\left(\widetilde{y}_k\right)$ as claimed.

LEMMA 4: If Assumptions A4 and A5 hold, and the normalization $G_k\left(0\right) = 0$ 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 $\widetilde{Y} = \sum_{k=1}^{K} G_k\left(X_k U_k\right)$.

PROOF of Lemma 4:

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

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 $\widetilde{Y} = \sum_{k=1}^{K} G_k\left(X_k U_k\right)$.

PROOF of Theorem 2: 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 4 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 3 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: For part i), Given any function G, for $k = 1, ..., K$, define $G_k(X_k U_k) = G(X_k U_k e_k) = G(0, ..., 0, X_k U_k, 0, ..., 0)$ and define $\widetilde{G}$ by $\widetilde{G}(X_1 U_1, ..., X_K U_K) = G(X_1 U_1, ..., X_K U_K) - \sum_{k=1}^{K} G_k(X_k U_k)$. Then, by construction, part i) holds. For part ii), we have that the function $G(X_1 U_1, ..., X_K U_K)$ 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 (6) is equivalent to equation (2). For equation (2), the proof of Theorem 1 shows (for $k = 1, ..., K$) identification of $F_{U_k | Z}$ and of G_k only using $X = 0$ and $X = X_k e_k$, so these functions are also identified for equation (6).

PROOF of Corollary 2: Theorem 2 identifies the functions $G_1, G_2, ..., G_K$, and shows that the distribution of $\widetilde{Y}$ defined by $\widetilde{Y} = G(X_1 U_1, ..., X_K U_K)$ is identified. Given that the distribution of $\widetilde{Y}$ is identified, Theorem 1 can be applied to identify the joint distribution function $F_{U | Z}(U_1, ..., U_K \mid Z)$.

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 (7), given any indirect utility function V, the corresponding demand function ω_1 is given by

$$\omega_1(X_1, X_2) = \frac{\partial V(X_1, X_2) / \partial \ln X_1}{\left[\partial V(X_1, X_2) / \partial \ln X_1\right] + \left[\partial V(X_1, X_2) / \partial \ln X_2\right]} \tag{26}$$

This is just one way to write Roys identity in a demand system of two goods. Then by the definition of λ, we have that for any demand function ω_1, the corresponding indirect utility function V satisfies

$$\lambda\left[\omega_1(X_1, X_2)\right] = \ln\left(\frac{\partial V(X_1, X_2)}{\partial \ln X_1}\right) - \ln\left(\frac{\partial V(X_1, X_2)}{\partial \ln X_2}\right) \tag{27}$$

and similarly, given any V the corresponding ω_1 satisfies equation (27).

It follows immediately that, given any differentiable functions $h_1(X_1)$ and $h_2(X_2)$, if $V(X_1, X_2) = h_1(X_1) + h_2(X_2)$ then equation (27) equals

$$\lambda[\omega_1(X_1, X_2)] = \ln\left(\frac{\partial h_1(X_1)}{\partial \ln X_1}\right) - \ln\left(\frac{\partial h_2(X_2)}{\partial \ln X_2}\right) \tag{28}$$

which is in the form of equation

$$\lambda[\omega_1(X_1, X_2)] = g_1(X_1) + g_2(X_2), \tag{29}$$

showing that any additive indirect utility function generates a demand equation in the form of (29). Also, by equation (28) given the functions h_1 and h_2 we can define g_1 and g_2 by

$$g_1(X_1) + g_2(X_2) = \ln\left(\frac{\partial h_1(X_1)}{\partial \ln X_1}\right) - \ln\left(\frac{\partial h_2(X_2)}{\partial \ln X_2}\right). \tag{30}$$

To go the other direction, start by supposing that equation (29) holds for some functions g_1 and g_2. We will apply the following two component special case of Lemma 4.1 in Blackorby, Primont, and Russell (1978, p. 160): Assume a function $F(r_1, r_2)$ is twice continuously differentiable and strictly increasing in its elements. Then $F(r_1, r_2) = F_0(F_1(r_1) + F_2(r_2))$ for some functions F_0, F_1, and F_2 if and only if

$$\frac{\partial}{\partial r_1} \ln\left(\frac{\partial F(r_1, r_2)/\partial r_2}{\partial F(r_1, r_2)/\partial r_1}\right) = \psi(r_1)$$

for some function ψ. Taking the derivative of equation (27) with respect to $\ln X_1$ and using equation (29) gives

$$\frac{\partial \lambda[\omega_1(X_1, X_2)]}{\partial \ln X_1} = \frac{\partial g_1(X_1)}{\partial \ln X_1} = \ln\left(\frac{\partial V(X_1, X_2)/\partial \ln X_2}{\partial V(X_1, X_2)/\partial \ln X_1}\right).$$

Apply the Lemma with $r_k = \ln X_k$, $F(r_1, r_2) = V(X_1, X_2)$, $\psi(r_1) = \partial g_1(X_1)/\partial \ln X_1$, and $F_k(r_k) = h_k(X_k)$ to prove that $V(X_1, X_2)$ must equal $F_0(h_1(X_1) + h_2(X_2))$ for some function F_0, which by the properties of indirect utility functions must be monotonically increasing (recall also that twice differentiability was one of the assumed properties of our indirect utility functions). Finally, applying equation (26) to this indirect utility function shows that equation (30) holds, which we can integrate to obtain

$$\int e^{g_1(x_1)} d\ln x_1 + \int e^{-g_2(x_2)} d\ln x_2 = h_1(X_1) + h_2(X_2).$$

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 (12) and (11). The properties of the functions h_1 and h_2 given at the end of Theorem 2 follow from the fact that the indirect utility function $h_1(U_1P_1/M) + h_2(U_2P_2/M)$ 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.

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

Beran, R. and P. W. Millar (1994), “Minimum Distance Estimation in Random Coefficient Regression Models,” Annals of Statistics, 22, 1976–1992.

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 (2013), "Estimating Consumption Economies of Scale, Adult Equivalence Scales, and Household Bargaining power,” Review of Economic Studies, 80, 1267-1303.

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.

Fox, J.T., K. I. Kim, S. P. Ryan, and P. Bajari (2012), "The random coefficients logit model is identified," Journal of Econometrics, 166, 204-212.

Fox, J.T. and A. Gandhi, (2013), "Nonparametric Identification and Estimation of Random Coefficients in Multinomial Choice Models," Unpublished Manuscript.

Gallant, A. R. and D. W. Nychka (1987), "Semi-Nonparametric Maximum Likelihood Estimation," Econometrica, 55, 363-390.

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.

Hausman, J. A. and W. K. Newey, (2014), "Individual Heterogeneity and Average Welfare," unpublished manuscript.

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., H. Holzmann, and A. Meister (2015), "The Triangular Model with Random Coefficients," Unpublished Manuscript.

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 J. Stoye, (2014), Revealed Preferences in a Heterogeneous Population,” Review of Economics and Statistics, forthcoming

Hoderlein, S. and A. Vanhems (2011), "Welfare Analysis using Nonseparable Models," Unpublished Manuscript, Boston College.

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

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

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

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

Khan, S. and E. Tamer, (2010), "Irregular Identification, Support Conditions, and Inverse Weight Estimation," Econometrica, 78, 2021-2042.

Kitamura, Y. and J. Stoye (2014), "Nonparametric Analysis of Random Utility Models," Unpublished Manuscript

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.

Masten, M. (2015), “Random Coefficients on Endogenous Variables in Simultaneous Equations Models," Unpublished Manuscript.

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

McFadden, D. (2005), “Revealed Stochastic Preference: A Synthesis,” Economic Theory, 26, 245–264.

McFadden, D. and M. K. Richter (1991), "Stochastic Rationality and Revealed Stochastic Preference," in J. Chipman, D. McFadden, and M.K. Richter (eds.), Preferences, Uncertainty, and Optimality, Essays in Honor of Leo Hurwicz, 161-186, Westview Press: Boulder, CO.

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.

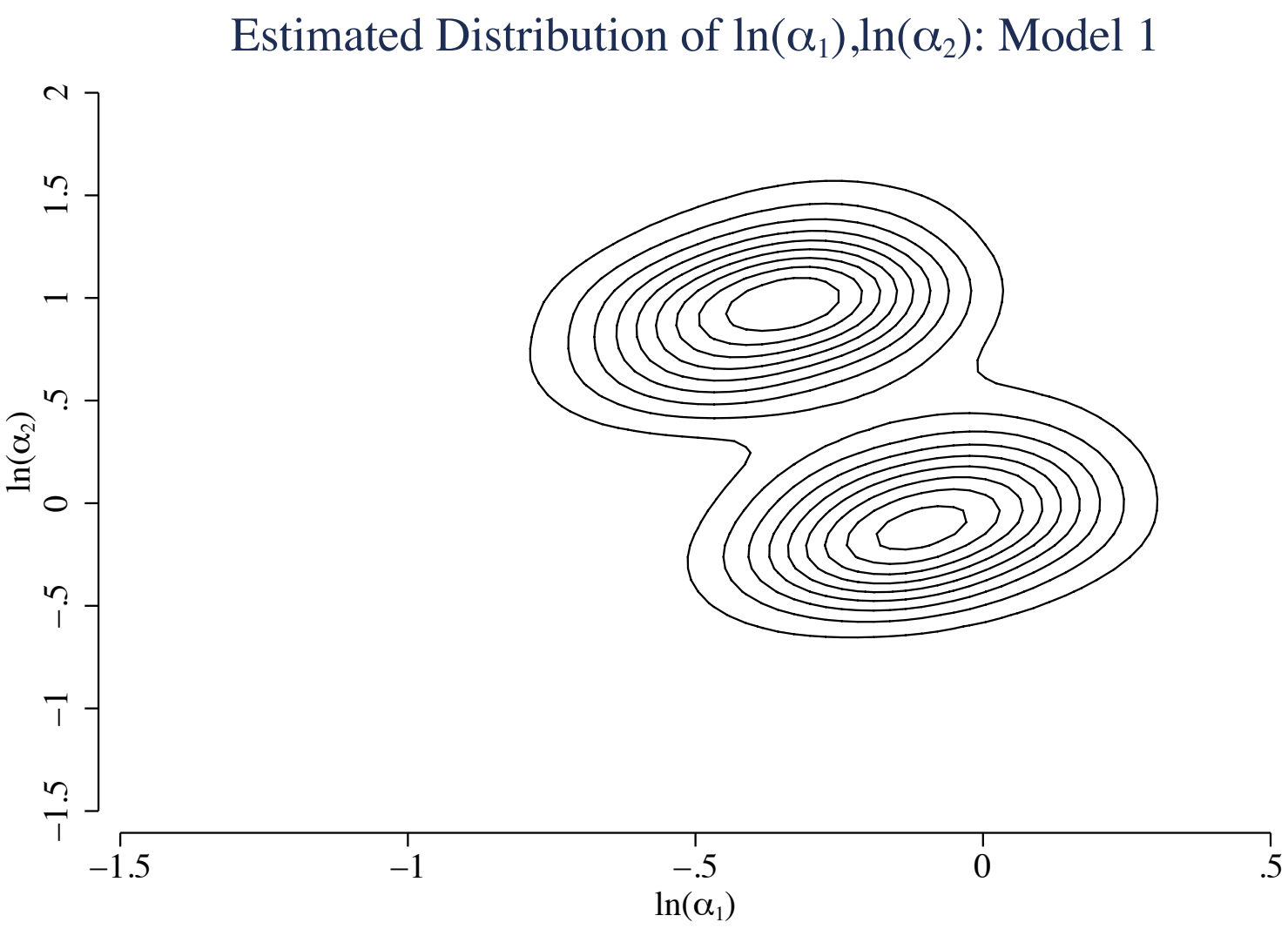
Matzkin, R. L. (2008) “Identification in Nonparametric Simultaneous Equations,” Econometrica, 76, 945–978.

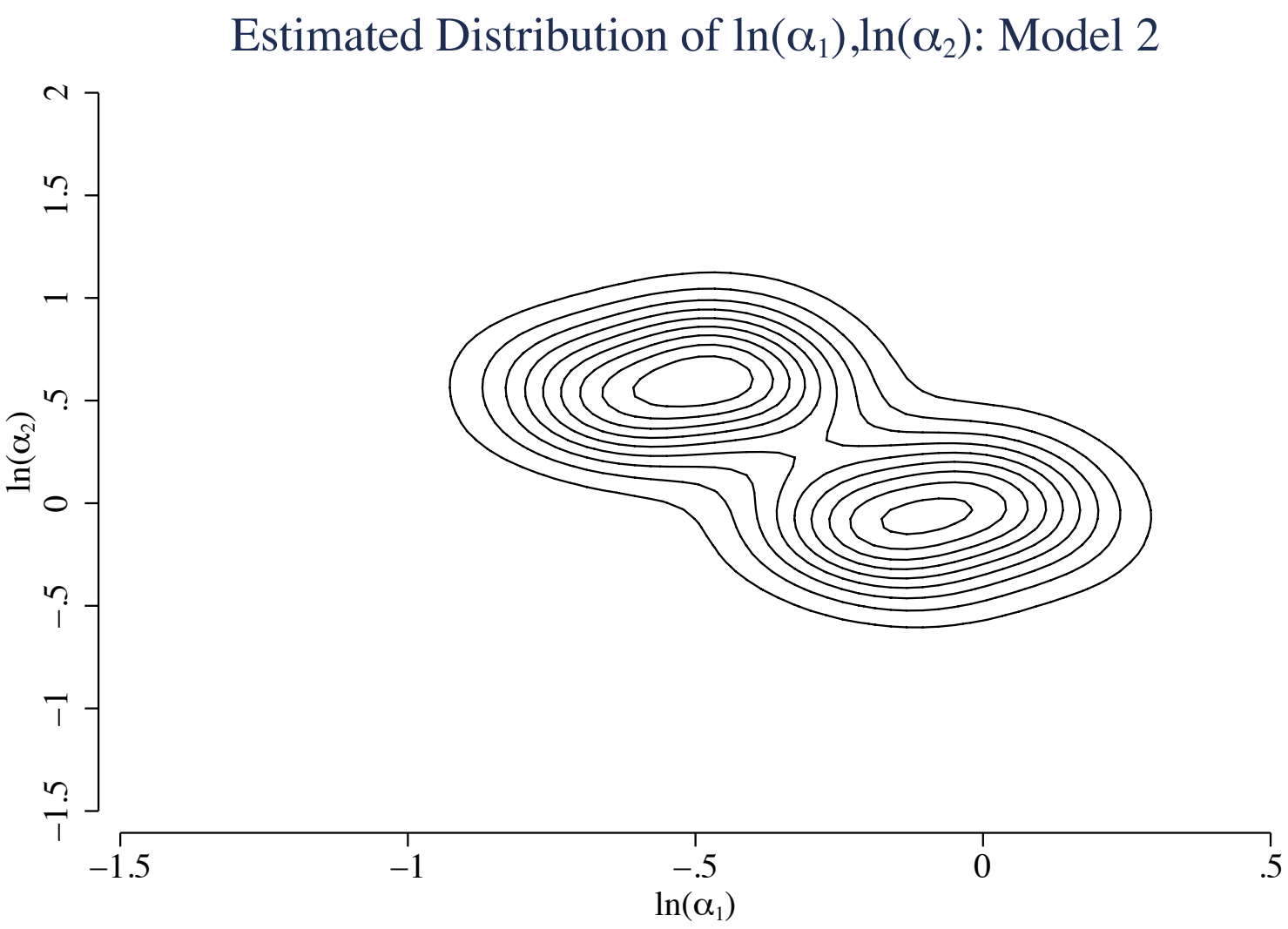
Meister, A. (2005), "Non-estimability in spite of identifiability in density deconvolution,". Mathematical Methods of Statistcs, 14, 479-487.

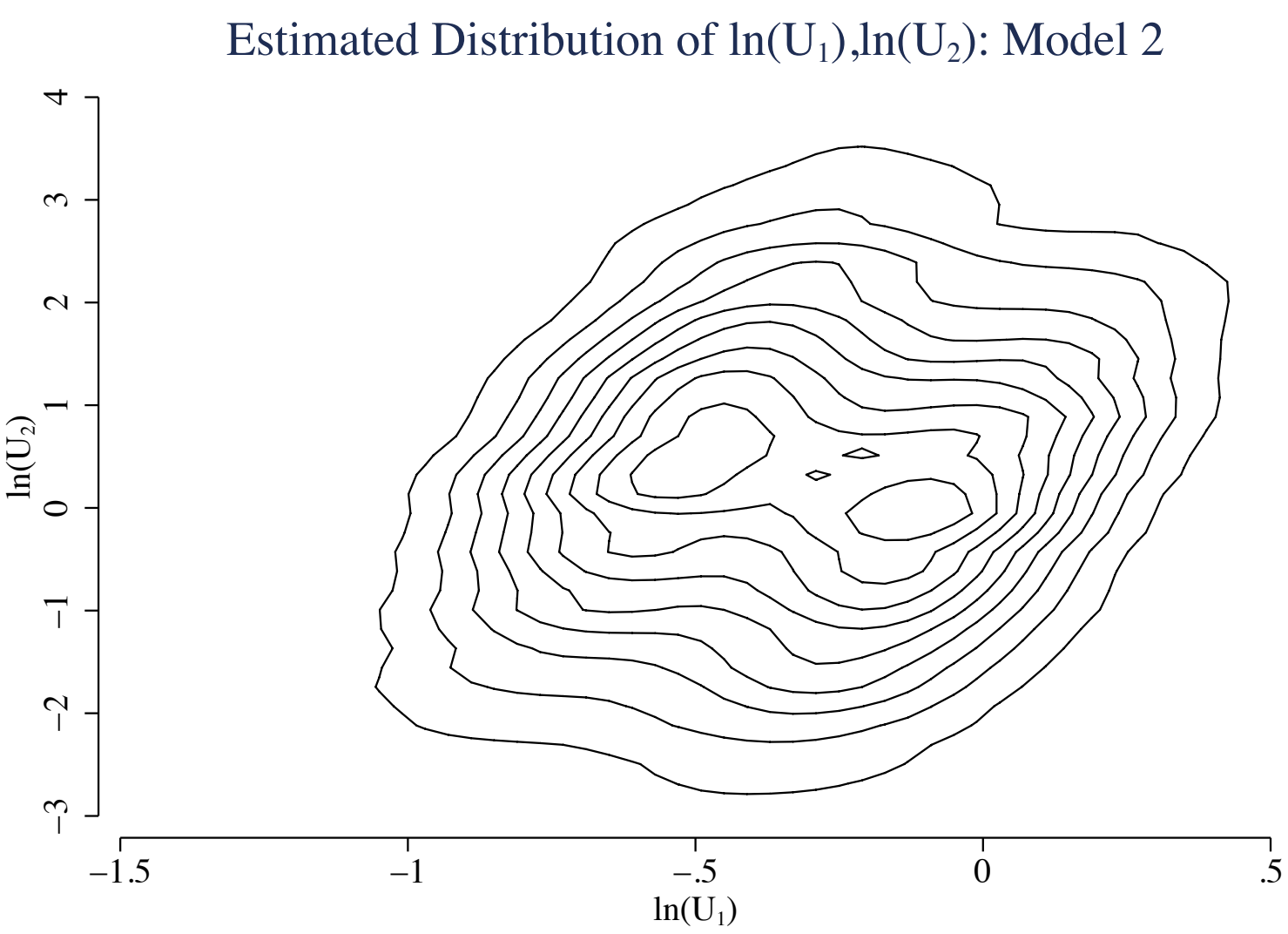
Stern, N. H, (1987), "The Theory of Optimal Commodity and Income Taxation: An Introduction," in D. M. G. Newbery and N. H. Stern, eds., The Theory of Taxation for Developing Countries, Oxford: Oxford University Press.

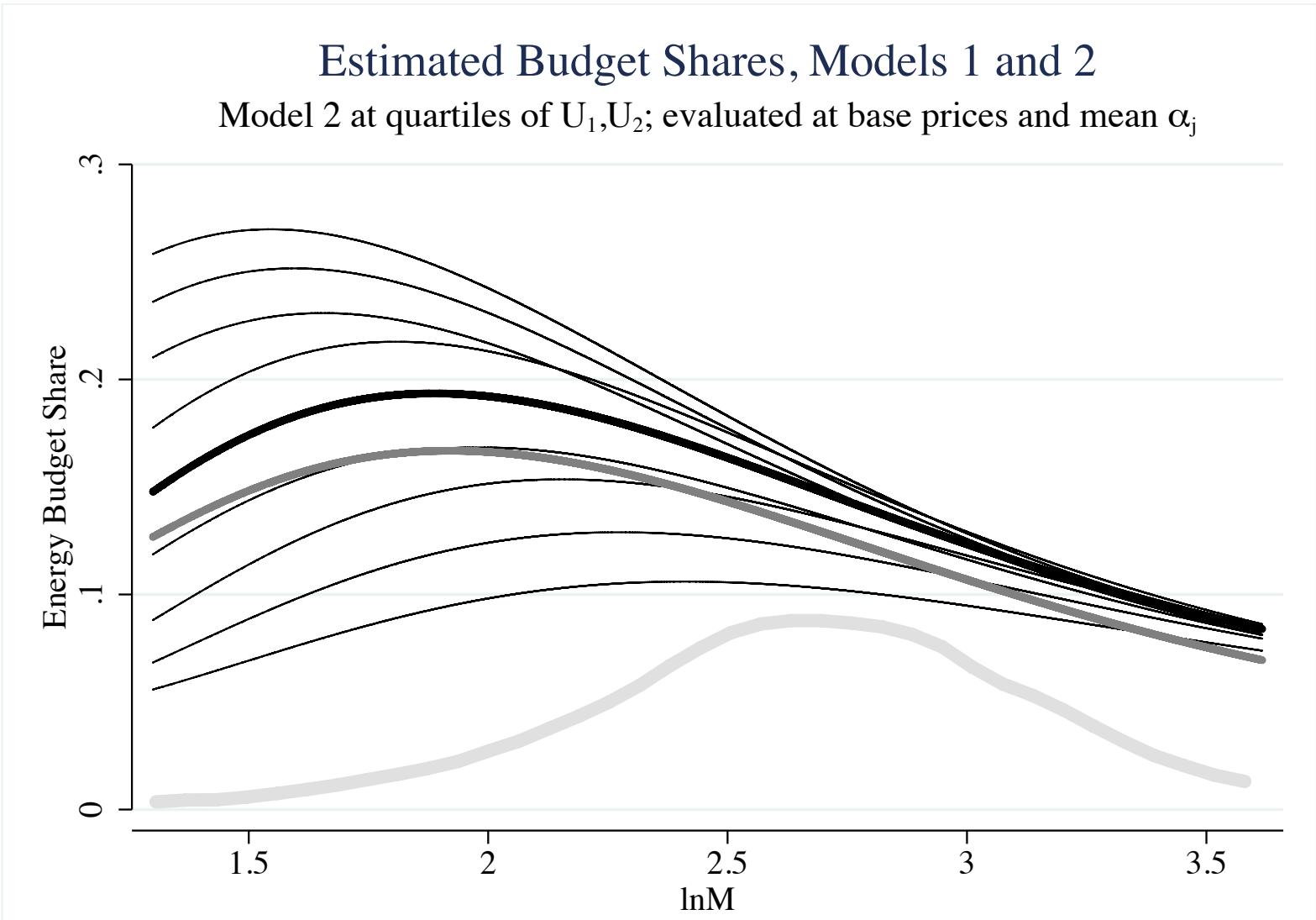
Wood, S. N. (2006). Generalized Additive Models: An Introduction with R, Chapman & Hall/CRC.

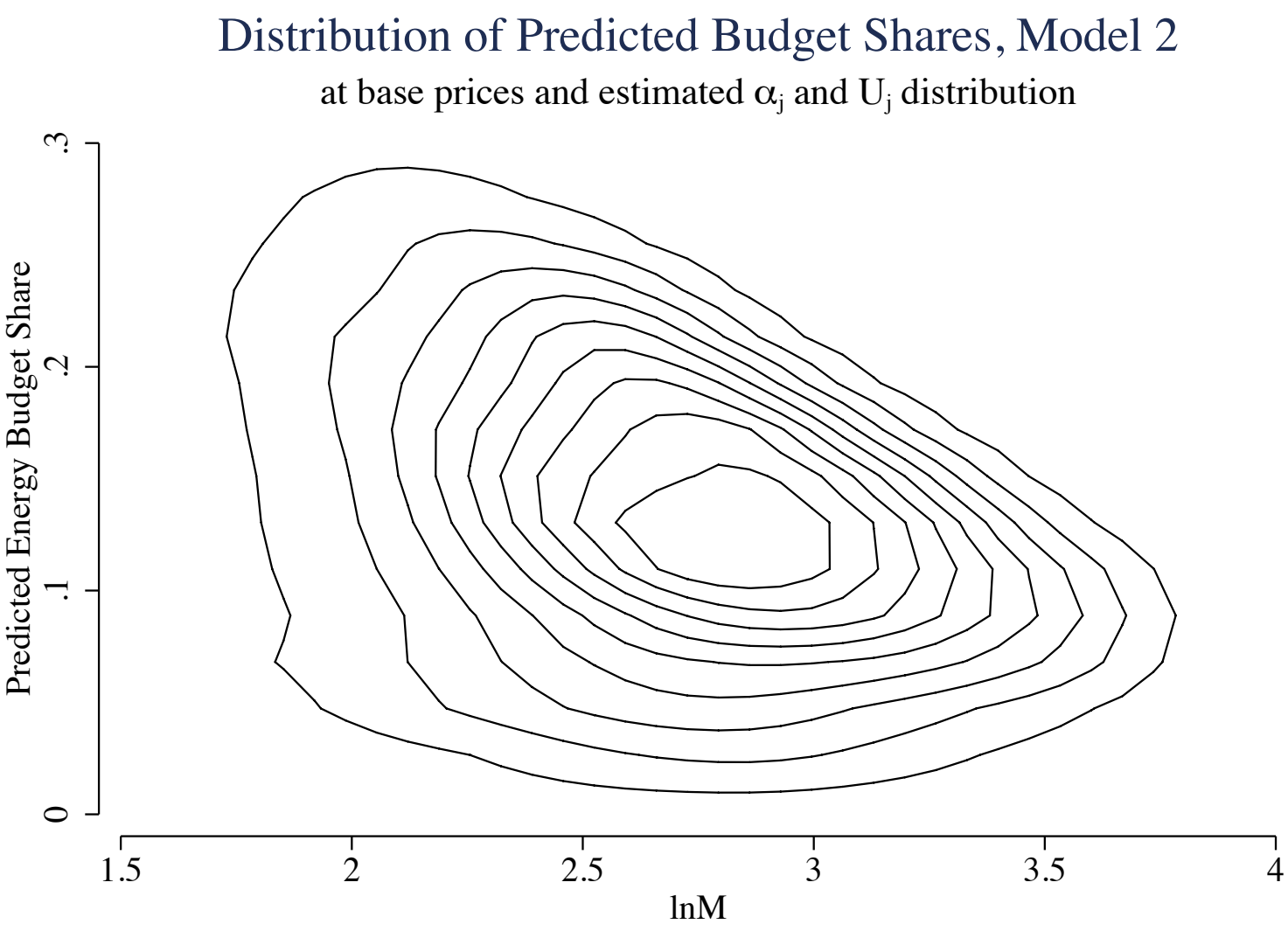
Working, H. (1943), "Statistical Laws of Family Expenditures," Journal of the American Statistical Association, 38, 43-56.


Estimated Distribution of ln(α1),ln(α2): Model 1
ln(α2)
2
1.5
1
.5
0
−.5
−1
−1.5
−1.5
−1
−.5
0
.5
ln(α1)


Estimated Distribution of ln(α1),ln(α2): Model 2
ln(α2)
2
1.5
1
.5
0
−.5
−1
−1.5
−1.5
−1
−.5
0
.5
ln(α1)


Estimated Distribution of ln(U1),ln(U2): Model 2
ln(U2)
4
3
2
1
0
−1
−2
−3
−1.5
−1
−.5
0
.5
ln(U1)


Estimated Budget Shares, Models 1 and 2
Model 2 at quartiles of U_1,U_2; evaluated at base prices and mean α_j
Energy Budget Share
0
.1
.2
.3
1.5
2
2.5
3
3.5
lnM


Distribution of Predicted Budget Shares, Model 2
at base prices and estimated α_j and U_j distribution
Predicted Energy Budget Share
0
.1
.2
.3
1.5
2
2.5
3
3.5
4
lnM

Distribution of Log Cost of Living Impacts, Model 2

Given base prices, 50% increase in Energy Price, and estimated α_j and U_j