Nonlinear Random Coefficients

Arthur Lewbel
Boston College

2016

Notes on Contents

Talk is based on two papers:

Lewbel, A., (2016) "Nonlinear Random Coefficients," Working paper in progress.

Lewbel, A., and K. Pendakur, (2016) "Unobserved Preference Heterogeneity in Demand Using Generalized Random Coefficients," forthcoming, Journal of Political Economy.

Introduction

Standard Linear Random Coefficients are

$$Y = \sum_{k=1}^{K} X_k U_k + U_0$$

for regressors $X = (X_1, ..., X_K)$ and unobserved errors (random coefficients) $U = (U_0, U_1, ..., U_K)$.

Standard Assumptions: i) IID observations of Y, X. ii) U is independent of X. iii) X continuous.

Popular extension: $Y = g\left(\sum_{k=1}^{K} X_k U_k + U_0\right)$ for known g, e.g., discrete choice models like BLP.

Typical applications assume $F_U(U)$ is normal.

But for above models linear in X and linear in U, is known can nonparametrically identify and estimate $F_U(U)$.

This paper: Consider Nonlinear Random Coefficients:

$$Y = G\left(X_1 U_1, ..., X_K U_K, \theta\right) + U_0$$

Identify parameter vector θ and nonparametric joint distribution F_U with known G.

Caveat: will add the strong restriction that U_0 is independent of $\left(U_1, ..., U_K\right)$, or that U_0 is not present. Price we pay for general nonlinearity.

EXAMPLE: Let $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$ and g is any known, strictly monotonic function.

$$Y = g\left[\theta_1 X_1 U_1 X_2 U_2 + \theta_2 \ln\left(X_2 U_2\right) + \theta_3 X_1 U_1 + \theta_4\right] \tag{1}$$

Not chosen for behavioral meaning. Illustrates multiple types of nonlinearities: a transformation function g, an interaction term between $X_1 U_1$ and $X_2 U_2$, a linear term $X_1 U_1$, and a nonlinear transformation of $X_2 U_2$.

Motivating Examples

1. Additive indirect utility functions with unobserved preference heterogeneity - random Barten (1964) scales. Yields budget share $Y = G(X_1 U_1, X_2 U_2)$ where X_j are prices divided by total expenditures. Also add an independent U_0 corresponding to measurement error in Y.

2. Production function $Y = G(X_1 U_1, ..., X_K U_K)$. $Y =$ output. $X_k =$ quantity of input or factor of production k (e.g., labor, capital). $U_k =$ unobserved quality of input k, or technology. Varies across firms. Generalizes Matzkin (1994) who considers a single random component.

Allow for nonlinear utility in random coefficient discrete choice models like Berry, Levinsohn, and Pakes (1995). X_k can be prices, income, characteristics. BLP assumes utility of each choice is linear in X, making market shares $Y = g\left(\sum_{k=1}^{K} X_k U_k\right)$. for logistic g. But No economic rationale exists for utility linear in X.

Outline

1. Literature Review.

2. Identify F_U in $Y = G(X_1 U_1, ..., X_K U_K)$ for known G.

3. Identify θ and F_U in $Y = G(X_1 U_1, ..., X_K U_K, \theta)$

4. Do example: $Y = g[\theta_1 X_1 U_1 X_2 U_2 + \theta_2 \ln(X_2 U_2) + \theta_3 X_1 U_1 + \theta_4]$

5. Extensions: append U_0, discrete choice
$D = I[V + G(X_1 U_1, ..., X_K U_K) + U_0 \geq 0]$

6. Empirical application: Random Barten Scales (preference heterogeneity) in Energy Demand.

Literature Review

Nonparametric identification and estimation of random coefficients: Beran and Hall (1992), Beran, Feuerverger, and Hall, (1996) and Hoderlein, Klemelae, and Mammen (2010).

Recent generalizations include linear systems of equations with random coefficients: Masten (2015), Hoderlein, Holzmann, and Meister (2015); random coefficient linear index models in binary choice: 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).

Matzkin (2003) in an appendix gives some generic identifying conditions for additive models with unobserved heterogeneity. Hoderlein, Nesheim, and Simoni (2011) give high level conditions for identification and estimation of parametric models containing a vector of random parameters.

Structural unobserved heterogeneity: Heckman and Singer (1984) and Lewbel (2001). Recent general nonseparable identification and estimation: Chesher (2003), Altonji and Matzkin (2005), Hoderlein, and Mammen (2007), Matzkin (2007a, 2008), and Imbens and Newey (2009).

Preference heterogeneity in continuous demand systems: Engel (1895), Sydenstricker and King (1921), Rothbarth (1943), Prais and Houthakker (1955), Barten (1964), Pollak and Wales (1981), Ray (1992), Brown and Walker (1989), McFadden and Richter (1991) Hildenbrand (1994), Lewbel (1997, 2001, 2007, 2008), Comon and Calvet (2003), McFadden (2004) Beckert (2006) Matzkin (2007, 2010), Beckert and Blundell (2008), Blundell, Kristensen and Matzkin (2011), Blundell and Matzkin (2011), Hoderlein and Stoye (2014), and Kitamura and Stoye (2014).

Additive separability and nonparametric additive regression: Gorman (1976), Blackorby, Primont, and Russell (1978), Hastie and Tibshirani (1990), Linton (2000), and Wood (2006).

Joint Distribution Identification

Drop U_0 for now. Later extensions bring U_0 back, and allow for control function type endogeneity.

First consider identification of the joint distribution $F_U(U)$ for $U = (U_1, ..., U_K)$ with known G, so

$$Y = G(X_1 U_1, ..., X_K U_K)$$

ASSUMPTION A1: $F_{Y|X}(y \mid x)$ is identified (e.g., could have IID observations of Y, X). G is continuous. U is independent of X.

Continuity of X and U is not required. X cannot be discrete, but its distribution can, e.g., contain mass points. U can be continuous, discrete, continuous with mass points, etc.

Side Note: Why not look at the conditional distribution function or characteristic function?

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

If this integral equation has a unique solution for F_U given known G, then F_U is identified.

If identified, estimators could be based on this equation (or the conditional characteristic function).

The identification problem: find restrictions on G that suffice to ensure a unique F_U.

Accomplished by devising easier to solve alternative expressions for (features of) F_U.

$$Y = G\left(X_1 U_1, ..., X_K U_K\right)$$

ASSUMPTION A2: $supp\left(X\right)$ is rectangular. The closure of $supp\left(X\right)$ equals the closure of $supp\left(U_1 X_1, ..., U_K X_K \mid U\right)$. The Moment Generating Function of $\left(U_1^{-1}, ..., U_K^{-1}\right)$ exists.

Sufficient for Assumption A2 is $supp\left(X\right) = \mathbb{R}_+^K$ and $supp\left(U\right) \subseteq \mathbb{R}_+^K$. Alternatively, could have $supp\left(X\right) = \mathbb{R}^K$ and U has any support, but the density of U must shrink quickly to zero as any element of U goes to zero.

We identify F_U by identifying moments of the distribution of $\left(U_1^{-1}, ..., U_K^{-1}\right)$. Necessary and sufficient conditions for integer moments to identify a distribution are known. Can replace existence of the MGF with, e.g., Assumption 7 of Fox, Kim, Ryan, and Bajari (2012).

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

$$\kappa_t = \int_{supp(X)} h\left[G\left(s_1, ...s_K\right), t\right] s_1^{t_1 - 1} s_2^{t_2 - 1} ... s_K^{t_K - 1} ds_1 ds_2 ... ds_K \qquad (2)$$

ASSUMPTION A3: Given G, for any K vector of positive integers t we can find a nonnegative, bounded function h such that $h\left[G\left(s_1, ...s_K\right), t\right] s_1^{t_1 - 1} s_2^{t_2 - 1} ... s_K^{t_K - 1}$ is absolutely integrable in s, and κ_t is convergent and nonzero.

Restricts G, but note h is chosen knowing G and t.

When does an h exist? How to find it? If G grows relatively quickly in its arguments, then h should look like a thin tailed density.

Can show an h exists for any additive G that grows faster than linearly:

LEMMA 1: If $supp\left(X\right) = \mathbb{R}_{+}^{K}$, $Y = \sum_{k=1}^{K} G_k\left(U_k X_k\right)$, and there exist positive constants c_k such that $G_k\left(s_k\right) \geq c_k s_k$ for $k = 1, ..., K$. Then Assumption A3 holds.

PROOF of Lemma 1: Let $h\left(G, t\right) = e^{-\rho G}$ for any $\rho > 0$. Then

$$\begin{aligned} \kappa_t &= \prod_{k=1}^{K} \int_0^{\infty} e^{-\rho G_k(s_k)} s_k^{t_k - 1} ds_k \leq \prod_{k=1}^{K} \int_0^{\infty} e^{-\rho c_k s_k} s_k^{t_k - 1} ds_k \\ &= \prod_{k=1}^{K} \left(\rho c_k\right)^{-t_k} \int_0^{\infty} e^{-r_k} r_k^{t_k - 1} dr_k = \prod_{k=1}^{K} \left(\rho c_k\right)^{-t_k} \Gamma\left(t_k\right) \end{aligned}$$

which is finite and positive, because the gamma function $\Gamma\left(t_k\right)$ is finite and positive.

An example that is not identified, an h does not exist:

If $G(X_1, X_2) = \ln(X_1) + \ln(X_2)$
then $G(U_1 X_1, U_2 X_2) = [\ln(U_1) + \ln(U_2)] + \ln(X_1) + \ln(X_2)$.
$F_U(U)$ can't be identified because can't separate U_1 from U_2

Lemma 2: An h does not exist for $G(X_1, X_2) = \ln(X_1) + \ln(X_2)$.
PROOF of Lemma 2: For any function h, change variables replacing s_2 with $r = s_1 s_2$ to get

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

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

Main Identification Theorem

ASSUMPTION A1: $Y = G(X_1 U_1, ..., X_K U_K)$. $F_{Y|X}(y \mid x)$ is identified (e.g., could have IID observations of Y, X). G is continuous. $U \perp X$. U is independent of X.

ASSUMPTION A2: $supp(X)$ is rectangular. The closure of $supp(X)$ equals the closure of $supp(U_1 X_1, ..., U_K X_K \mid U)$. The Moment Generating Function of $\left(U_1^{-1}, ..., U_K^{-1}\right)$ exists.

ASSUMPTION A3: Given G, for any K vector of positive integers t we can find a nonnegative, bounded function h such that

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

is absolutely integrable, convergent, and nonzero.

THEOREM 1: Let Assumptions A1, A2, and A3 hold. If G is known or identified, then $F_U(U_1, ..., U_K)$ is identified.

Proof sketch for $K = 2$. Define the identified term

$$\lambda_t = \int_{X \in supp(X)} E\left[h(Y,t) \mid X_1, X_2\right] X_1^{t_1-1} X_2^{t_2-1} dX_1 dX_2 =$$

$$\int_{X \in supp(X)} \int_{U \in supp(U)} h\left(G\left(X_1 U_1, X_2 U_2\right), t\right) dF\left(U_1, U_2\right) X_1^{t_1-1} X_2^{t_2-1} dX_1 dX_2$$

$$\int_{U \in supp(U)} \int_{X \in supp(X)} h\left(G\left(X_1 U_1, X_2 U_2\right), t\right) X_1^{t_1-1} X_2^{t_2-1} dX_1 dX_2 dF\left(U_1, U_2\right)$$

Change variables on the inner integral, letting $s_k = X_k U_k$:

$$\int_{U \in supp(U)} \int_{s \in supp(X_1 U_1, X_2 U_2 | U)} h\left(G\left(s_1, s_2\right), t\right) s_1^{t_1-1} s_2^{t_2-1} U_1^{-t_1} U_2^{-t_2} ds_1 ds_2 dF\left(\right.$$

$$\int_{U \in supp(U)} \int_{s \in supp(X)} h\left(G\left(s_1, s_2\right), t\right) s_1^{t_1-1} s_2^{t_2-1} ds_1 ds_2 U_1^{-t_1} U_2^{-t_2} dF\left(U_1, U_2\right)$$

$$\int_{U \in supp(U)} \kappa_t U_1^{-t_1} U_2^{-t_2} dF\left(U_1, U_2\right) = \kappa_t E\left(U_1^{-t_1} U_2^{-t_2}\right)$$

So $E\left(U_1^{-t_1} U_2^{-t_2}\right) = \lambda_t / \kappa_t$ identifies the moments of $\left(U_1^{-1}, U_2^{-1}\right)$, which by existence of MGF identifies $F_U\left(U_1, U_2\right)$.

EXTENSION 1: Satisfying Assumption A3 when some elements of t equal zero is difficult. Can instead partition X into two subvectors X^P and X^{-P}, and apply theorem conditioning on $X^{-P} = 0$ instead of integrating over X^{-P}.

Example: With $X = (X_1, X_2)$ let $X_2 = 0$ to get $E\left(U_1^{-t_1}\right) = \lambda_{Pt}/\kappa_{Pt}$

EXTENSION 2: Replace $h(Y, t)$ with $h_j(Y, t)$ for different h_j functions.

Let $\mu_t = E\left(U_1^{-t_1} U_2^{-t_2} ... U_2^{-t_K}\right)$. For each t (including partition P) and each j, will now get

$$\lambda_{Ptj}/\kappa_{Ptj} = \mu_t$$

Can apply with multiple j to get multiple expressions for μ_t.

EXTENSION 3: Unknown parameter vector θ. Model is now:

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

for known G, unknown vector θ.

Given a vector $t \in T$, partition $P \in \mathcal{P}$, and function h_j for $j \in J$, construct

$$\begin{aligned}
\lambda_{Ptj} &= \int_{x^P \in supp(X^P)} E\left[h_j\left(Y\right) \mid X^P = x^P, X^{-P} = 0, t\right] \\
&\qquad x_1^{t_1 - 1} x_2^{t_2 - 1} ... x_{K^P}^{t_{K^P} - 1} dx_1 dx_2 ... dx_{K^P} \\
\kappa_{Ptj}\left(\theta\right) &= \int_{s^P \in supp(X^P)} h_j\left(G^P\left(s^P, \theta\right), t\right) \\
&\qquad s_1^{t_1 - 1} s_2^{t_2 - 1} ... s_{K^P}^{t_{K^P} - 1} ds_1 ds_2 ... ds_{K^P}
\end{aligned}$$

By Theorem

$$\frac{\lambda_{Ptj_1}}{\kappa_{Ptj_1}\left(\theta\right)} = E\left(U_1^{-t_1} U_2^{-t_2} ... U_2^{-t_K}\right)$$

right side only depends on $F_U\left(U\right)$ and t, not on θ.

$$\text{Have} \quad \frac{\lambda_{Ptj_1}}{\kappa_{Ptj_1}(\theta)} = E\left(U_1^{-t_1} U_2^{-t_2} \ldots U_2^{-t_K}\right)$$

Therefore, available equations for identifying θ include:

1. $\frac{\lambda_{Ptj_1}}{\kappa_{Ptj_1}(\theta)} = \frac{\lambda_{Ptj_2}}{\kappa_{Ptj_2}(\theta)}$ for all $P \in \mathcal{P}$, $t \in T$, $j_1 \in J$, and $j_2 \in J$

2. $\kappa_{Ptj}(\theta)\left(\frac{\partial \lambda_{Ptj}}{\partial j}\right) = \left(\frac{\partial \kappa_{Ptj}(\theta)}{\partial j}\right)\lambda_{Ptj}$ for all $P \in \mathcal{P}$, $t \in T$, and $j \in$

3. $\lambda_{P0j} = \kappa_{P0j}(\theta)$ for all $P \in \mathcal{P}$ and $j \in J$. Here $t = 0$.

First use any combination of these to identify θ, then apply Theorem 1 to identify F_U.

Notes on Possible Estimators

1. Construct $\kappa_{Ptj}(\theta)$ for each choice of j, P, t.
2. Replace $E\left[h_j(Y) \mid X^P = x^P, X^{-P} = 0, t\right]$ with nonparametric regression in λ_{Ptj} to get $\widehat{\lambda}_{Ptj}$.
3. Use minimum distance estimate θ, e.g.,

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

4. Given $\widehat{\theta}$, estimate random coefficient moments μ_t of the U distribution for each $t \in T$ using

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

with weights w_{jt} chosen such that $\sum_{j \in J} w_j$.

These are essentially semiparametric two step estimators with nonparametric first step.

Alternatives: If partitions of X are not needed, can rewrite the λ_{Ptj} equation as

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

So the integral of a nonparametric regression can be replaced by estimation of the density of X, $f_x\left(X\right)$, and a simple average.

If the density of X is finitely parameterized, and we only want a finite number of moments μ_t of the U distribution, then all of the estimation steps can be combined into an ordinary GMM.

Alternative to all of the above might be sieve maximum likelihood. Assume U is continuous, write the likelihood function for the model $Y = G\left(X_1U_1, ..., X_KU_K, \theta\right)$ in terms of the density of U, and approximate the density with basis functions (e.g. mixtures of normals or Hermite polynomial expansions).

Example

$$Y = g\left[\theta_1 X_1 U_1 X_2 U_2 + \theta_2 \ln\left(X_2 U_2\right) + \theta_3 X_1 U_1 + \theta_4\right]$$

To satisfy assumptions, assume g is known and monotonic, $supp\left(X\right) = \mathbb{R}_+^K$, $supp\left(U\right) \subseteq \mathbb{R}_+^K$, and $\theta_1 > 0$.

Wlog, can choose scale normalizations for U_1 and U_2 to make $\theta_3 = 1$ and $\theta_4 = 0$.

Goal: identify θ_1, θ_2 and joint distribution of U_1 and U_2.

For this application, a convenient h_j (one that yields simple expressions for θ) is

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

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

Model: $$Y = g\left[\theta_1 X_1 U_1 X_2 U_2 + \theta_2 \ln\left(X_2 U_2\right) + X_1 U_1\right]$$

First identify θ_2. Choose P where $X_1 = 0$ so t_1 drops out and $X^P = (X_2)$. Let $t_2 = 0$. This gives

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

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

Do the change of variables $q = \theta_2 \ln s_2$. Then

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

Now $\lambda_{P0j} = \kappa_{P0j}\left(\theta\right) = 1/\theta_2$ so identified by $\theta_2 = 1/\lambda_{P0j}$.

Model: $Y = g\left[\theta_1 X_1 U_1 X_2 U_2 + \theta_2 \ln\left(X_2 U_2\right) + X_1 U_1\right]$

Next identify θ_1. Now use partition $X^P = X$, and let $t_1 = t_2 = 1$. This gives

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

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

Now do change in variables replace s_1 with $r = s_1 + \theta_1 s_1 s_2 + \theta_2 \ln s_2$ to get

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

Let $j = 1/\theta_2$ to get

$$\kappa_{jt}\left(\theta\right) = \frac{\ln\left(\theta_1\right)}{\theta_1 - 1} \quad \text{and} \quad \frac{\partial \kappa_{j_1 t}\left(\theta\right)}{\partial j} = -\frac{\theta_2}{2}\left(\frac{\ln\left(\theta_1\right)}{\theta_1 - 1}\right)^2$$

Plugging into $\kappa_{Ptj}\left(\theta\right)\left(\partial\lambda_{Ptj}/\partial j\right) - \left(\partial\kappa_{Ptj}\left(\theta\right)/\partial j\right)\lambda_{Ptj}$, can uniquely solve for θ_1.

Additive Model Identification

Additive Model: $Y = c + \sum_{k=1}^{K} G_k \left(X_k U_k \right)$

Both the joint distribution of random coefficients $F_U \left(U \right)$ and the G_k functions are unknown, need to be nonparametrically identified.

Maintain assumptions A1, A2 and A3. Recall $G_k \left(s_k \right) \geq c_k s_k$ suffices for A3.

ASSUMPTION A4: U and X are continuously distributed. Each G_k is strictly monotonically increasing, wlog normalize $G_k \left(0 \right) = 0$, $G_k \left(1 \right) = 1$.

Extensions

1. $Y = G(X_1 U_1, ..., X_K U_K, \theta) + U_0 = \widetilde{Y} + U_0$ for unobserved $\widetilde{Y}$
Assume U_0 independent of U_1,...,U_K, and U_0 has nonvanishing characteristic function. WLOG let $G(0) = 0$.
$F_{Y|X}(y \mid 0) = F_{U_0}(y)$ identifies F_{U_0}. Deconvolution of $Y|X$ with U_0 identifies $F_{\widetilde{Y}|X}$. Can then proceed as before.

2. Discrete choice. Assume for unobserved Y:
$D = I[Y - V \geq 0] = I[G(X_1 U_1, ..., X_K U_K) - V + U_0 \geq 0]$
V is a special regressor (Lewbel 1998, 2000, 2015): linear, continuous, large support, independent of U. $E(1 - D \mid V, X) = F_{Y|X}(V \mid X)$ identifies $F_{Y|X}$. Can then proceed as before.

3. Can replace $F_U(U)$ with $F_U(U \mid Z)$, let all assumptions hold conditional on covariates Z, observable characteristics. Allows for observable preference heterogeneity and/or control function type endogeneity.

Barten Scales

Utility function $S\left(Q_1/U_1, ..., Q_J/U_J\right)$
$Q_1, ..., Q_J$ are quantities of goods consumed
$U_1, ..., U_J$ are Barten (1964) scales, reference values one.

Example: A couple rides together in their car 50% of the time. For quantity of gasoline Q_j, they get utility as if the quantity bought was $Q_j * 1.5 = Q_j/U_j$ where Barten scale $U_j = 2/3$.

If they did not share car at all, would have $U_j = 1/2$. If they shared all the time $U_j = 1$.

Barten scales can also reflect preference heterogeneity. If I need to eat more than you to get the same utility from $j =$ food, then I have a larger value of U_j in my utility function you have in yours.

Barten Scales

Let $W_j^* = Q_j P_j / M$ be the good j budget share and $X_j = P_j / M$. If max utility function $S(Q_1 / U_1, ..., Q_J / U_J)$ given $\sum_{j=1}^{J} Q_j P_j = M$, get Marshallian demands (in budget share form):

$W_j^* = \omega_j (U_1 X_1, ..., U_J X_J)$ for each good j.

Traditional Barten scales: $U_j = \alpha(Z, \theta)$, Z are observable household characteristics (age, family size, etc.), estimate parameters θ.

This paper's Barten scales: U_j are random utility parameters, reflecting unobserved preference heterogeneity. Each U_j has a conditional pdf $= f_j(U_j \mid Z)$.

The functional form of $\omega_j(X_1, ..., X_J)$ depends only on the functional form of $S(Q_1, ..., Q_J)$, so $U_1, ..., U_J$ can vary independently of $X_1, ..., X_J$.

Empirical Demand Model Specification

Since identified, could consider nonparametric sieve estimation.

Due to sample size and curse of dimensionality, will instead do MLE with 'sieve inspired' model specification.

Specify indirect utility $V^{-1} = h_1(U_1 X_1) + h_2(U_2 X_2)$ where

$$h_k(X_k) = \int_{\ln X_k} \left(\beta_{k0} + \beta_{k1} e^{r} + \beta_{k2} e^{2r} + ... + \beta_{kS} e^{Sr} \right)^2 dr$$

Yielding Marshallian budget shares proportional to polynomials. Almost all standard demand models are proportional to polynomials. See, e.g., Lewbel (2008) and references therein.

Empirical Barten Scale Specification

Include observable taste shifters Z. Could identify nonparametric $F_U(U \mid Z)$, but to reduce dimensionality, let $U_k = \alpha_k(Z)\widetilde{U}_k$ where $\alpha_k(Z) = \exp\left(\theta'_{1k}Z + Z'\theta_{2k}Z\right)$ is a traditional deterministic Barten scale.

Remaining unobserved random component of the Barten scales $\widetilde{U} = \left(\widetilde{U}_1, \widetilde{U}_2\right)$ is specified as (trimmed) bivariate log normal density

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

The density is trimmed at $I = \pm 3\sigma$, since Theorem 2 needs MGF.

Density f_0 of error U_0 is mean zero normal with variance σ_0^2.

Estimator

We later also consider hermite polynomial sieve expansion densities for the unobservables.

Assuming n iid households the resulting likelihood function is

$$\sum_{i=1}^{n} \ln f_{W_1|X_1,X_2,Z}\left(w_{1i} \mid x_{1i}, x_{2i}, z_i; \alpha, \beta\right).$$

where in Model 1 (standard deterministic Barten, $\widetilde{u} = 1$)

$$f_{W_1|X_1,X_2,Z}\left(w_1 \mid x_1, x_2, z; \alpha, \beta\right)$$
$$= \exp\left(\frac{-1}{2\sigma_0^2}\left[\lambda\left(W_1\right) - \ln\left(\left(\frac{\beta_{10} + \sum_{s=1}^{S} \beta_{1s}\left(\alpha_1\left(z\right)x_1\right)^s}{1 + \sum_{s=1}^{S} \beta_{2s}\left(\alpha_2\left(z\right)x_2\right)^s}\right)^2\right)\right]^2\right)$$

While in Model 2 (random Barten scales, $\widetilde{u}$ density $f_{\ln \widetilde{U}}$ is trimmed log normal) the likelihood function is

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

where

$$\begin{aligned} & f_{W_1|X_1,X_2,Z}\left(w_1 \mid x_1, x_2, z; \alpha, \beta, \sigma, \rho\right) \\ = \; & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{f_{\ln \widetilde{U}}\left(\widetilde{u}_1, \widetilde{u}_2, \sigma, \rho\right)}{(2\pi)^{1/2}\sigma_0} \\ & \exp\left(\frac{-1}{2\sigma_0^2}\left[\lambda\left(W_1\right) - \ln\left(\left(\frac{\beta_{10} + \sum_{s=1}^{S}\beta_{1s}\left(\widetilde{u}_1\alpha_1(z)\,x_1\right)^s}{1 + \sum_{s=1}^{S}\beta_{2s}\left(\widetilde{u}_2\alpha_2(z)\,x_2\right)^s}\right)^2\right)\right]^2\right) \\ & \qquad d\ln\widetilde{u}_1\, d\ln\widetilde{u}_2 \end{aligned}$$

Note numerical integration over the $\widetilde{U}_1, \widetilde{U}_2$ distribution.

Data

1997 to 2008 Canadian Survey of Household Spending, urban working age singles, trimmed.

M = total nondurable expenditures: sum of household spending on food, clothing, health care, alcohol and tobacco, public transportation, private transportation operation, and personal care, plus the energy goods fuel oil, electricity, natural gas and gasoline.

W_1 = energy share of total nondurable expenditures. P_1 and P_2 are household specific within group budget share weighted Stone Indices of energy and non-energy goods, respectively, normalised to 1 in Ontario in 2002.

Z = characteristics: dummy for female; age (by 5 year age groups); calendar year; dummy for in Quebec; Number of days requiring heating and cooling in each province in each year (normalized as z-scores); dummy for renter; dummy for more than 10% of gross income from government transfers.

Data - continued

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	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
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

Empirical Example - Energy Demand Estimates

$S = 3$ third order polynomial in each $\ln X_j$. Model 1 deterministic Barten, Model 2 random Barten. Model 2 has generally smaller standard errors, roughly analogous to how generalized least squares lowers standard errors by modeling the heteroskedasticity.

Table 2: Estimated Parameters - part 1

	Model 1		Model 2	
		llf=−10043.1		llf=−9706.9
Parameter	Estimate	*Std Err*	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*

Table 2: Estimated Parameters - part 2

Parameter		Model 1		Model 2	
		Estimate	*Std Err*	Estimate	*Std Err*
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*

Barten summary terms. Note $\ln \alpha_j(z)$ is deterministic component, σ_j is standard deviation of random component $\widetilde{U}_j$, $\ln U_j = \ln \widetilde{U}_j + \ln \alpha_j(z)$.

Table 2: Estimated Parameters - part 3

		Model 1		Model 2	
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	

Log Barten Scale Distributions

Next two slides show contour plots of estimated joint density of log Barten Scales.

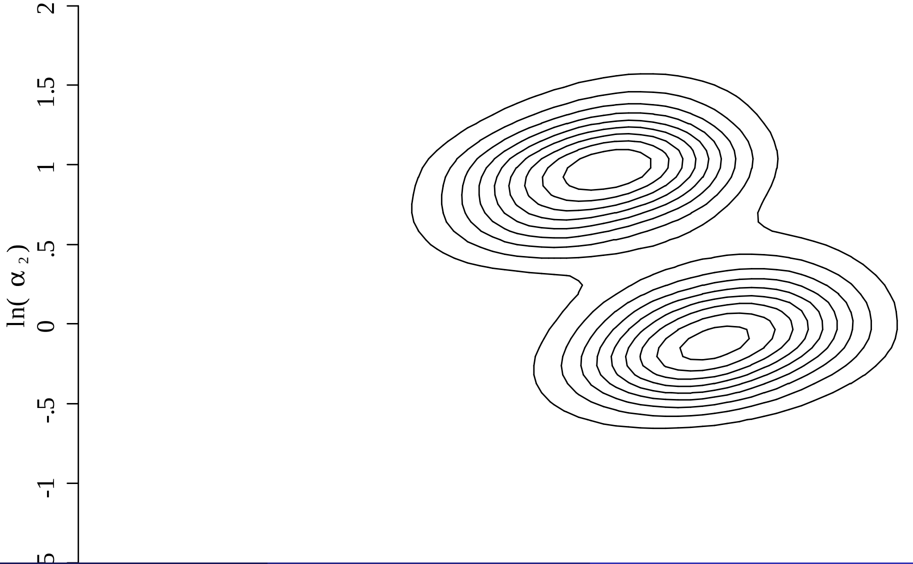
First is Model 1 Joint density of $\ln \alpha_1(Z)$, $\ln \alpha_2(Z)$. These are traditional deterministic log Barten scales.

Second is Model 2 Joint density of our random Barten scales $\ln U_1$, $\ln U_2$ where, $\ln U_j = \ln \alpha_j(z) + \ln \widetilde{U}_j$.

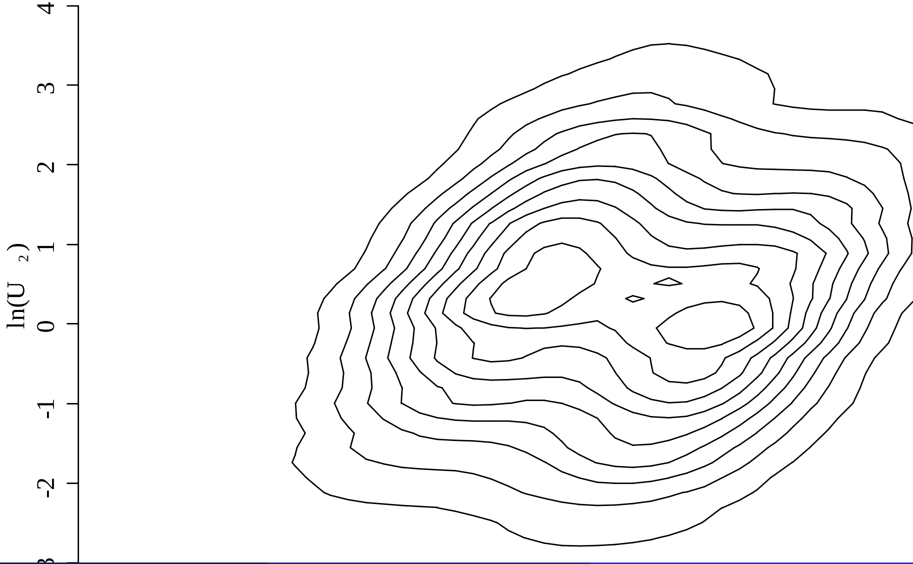
Two modes correspond to separate mean energy expenditures of renters vs owners.

In Model 2, after controlling for renter vs owners, $var(\ln U_2)$ is a little higher than $var(\ln U_1)$, correlation about 0.7

Estimated Distribution of ln(α 1),ln(α 2): Mod



Estimated Distribution of ln(U $_1$),ln(U $_2$): Mod

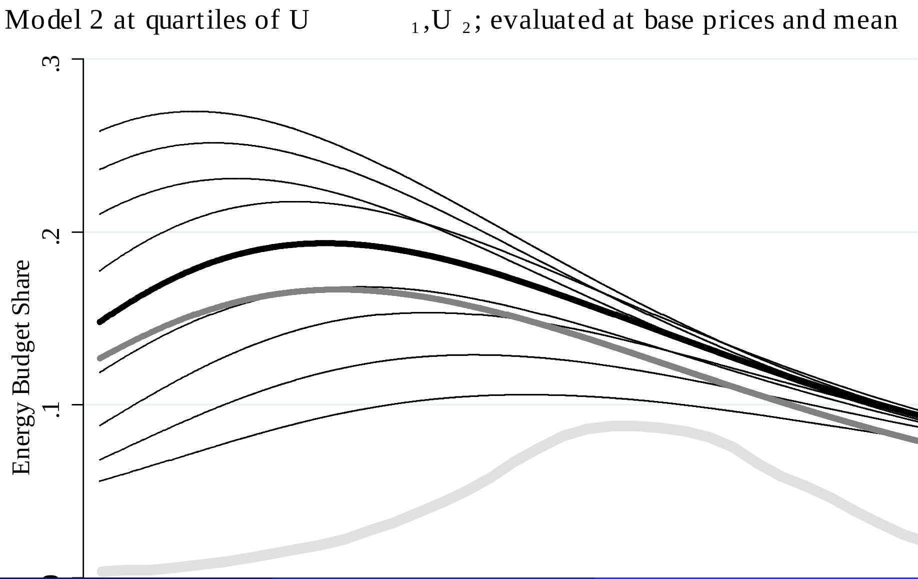


Budget Shares - Engel Curves

Engel curve: energy budget share W_1 as a function of log total expenditures, $\ln M$, evaluated at $P_1 = P_2 = 1$, at quartiles of the distributions of U_1, U_2. For comparison, model 1 is gray line and model 2 without random $\widetilde{U}_1$ and $\widetilde{U}_2$ is thick black line.

Density of $\ln M$ also shown; upward sloping portion of Engel curves are only in the lower tail of $\ln M$ distribution.

Estimated Budget Shares, Models 1 and 2



Price effects, Consumer Surplus

We have a closed form expression for indirect utility. Therefore can compute cost of living consumer surplus without Vartia (1984) type approximations. Would otherwise need a numeric differential equation solution for every value U_1, U_2 can take on.

To show price effects clearly, consider a large price change: a 50% increase in the price of energy at $P_1 = P_2 = 1$ (approximating the effect of a $300 per ton CO2 tax,see, e.g., Rhodes and Jaccard 2014). The cost-of-living impact, $\pi(U_1, U_2, M)$ is defined as the solution to

$$V\left(\frac{U_1 P_1}{M}, \frac{U_2 P_2}{M}\right) = V\left(\frac{1.5 U_1 P_1}{\pi M}, \frac{U_2 P_2}{\pi M}\right).$$

Next slide shows joint density (contour plot) of π and $\ln M$, variation from U_1, U_2. $\overline{W}_1 = 0.146$, so 50% energy tax without substitution effects would increase costs by 7.3%. Most mass is below .073 horizontal line from substitution effects. Low M households have higher mean and variance of harm.

Distribution of Log Cost of Living Impacts, Model 2

ven base prices, 50% increase in Energy Price, and estimated

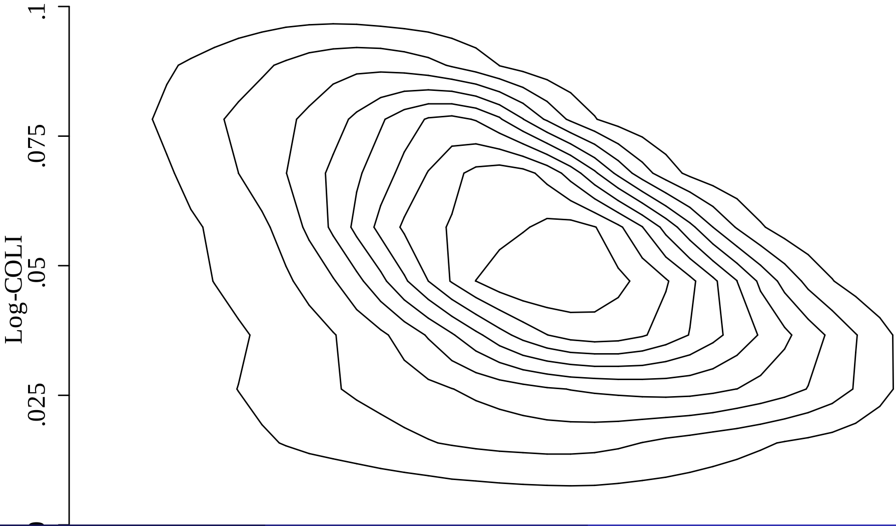


Table 4: Cost of Living Impacts: 50% Energy Price Increase

Per Cent Increase		Model 1		Model 2	
$\pi - 1$, per cent		Estimate	*Std Err*	Estimate	*Std Err*
$\alpha_j = \overline{\alpha}_j, \widetilde{U}_j = 1$	Mean	5.34	*0.22*	5.66	*0.17*
	Std Dev	1.26	*0.06*	1.30	*0.05*
$\alpha_j, \widetilde{U}_j = 1$	Mean	**5.31**	*0.24*	5.64	*0.17*
	Std Dev	**1.85**	*0.21*	1.69	*0.08*
$\alpha_j, \widetilde{U}_j$	Mean			**5.37**	*0.20*
	Std Dev			**4.31**	*0.46*

If had no substitution effects the mean effect above would be 7.3%.

Comparing first 2 and second 2 rows shows allowing for observed heterogeneity in U has little effect on COLI $\pi - 1$.

Comparing models 1 and 2 shows allowing for **un**observed heterogeneity in U has little effect on mean COLI but more than doubles(!) its standard deviation, from 1.85 to 4.31. As previous graph shows, wider variation particularly impacts the poor.

Conclusions

Have shown identification of generalized random coefficients models.

$$Y = G\left(X_1 U_1, ..., X_K U_K, \theta\right) + U_0$$

Potential applications:
Production functions with unobserved qualities U_k of inputs X_k
Discrete choice and BLP type models without artificial restriction of linearity in covariates
Polynomial instead of linear random coefficients.

Empirical application: Extend existing observed heterogeneity in demand model (Barten scales) to unobserved heterogeneity - highly relevant for distribution of welfare effects of an energy tax.

Ongoing work: Estimation asymptotics, characterizations of feasible g functions, multiple equation systems.