Semi-Nonparametric Estimation of a Nonseparable Demand Function under Shape Restrictions

Richard Blundell
(UCL & IFS)

Joel L. Horowitz
(Northwestern)

Matthias Parey
(Essex & IFS)

Northwestern, November 5, 2010

Introduction

- This paper is about semi-nonparametric quantile estimation of
  - a demand function
  - the welfare consequences of price changes.
- We motivate and illustrate our methods with an application to the demand for gasoline in the U.S.
  - Real gasoline prices in the U.S. have more than doubled over the past 10 years
  - Interest in differential price effects across the income distribution.
- Simple parametric demand functions impose strong restrictions on price-income interactions.
- This motivates the use of nonparametric methods to estimate demand.

Motivation

- The nonparametric estimate of the demand function is noisy due to random sampling errors.
  - The estimated function is volatile, nonmonotonic, and inconsistent with economic theory.
- One way of overcoming this problem is to impose further structure on the demand function
  - For example, assume that the demand function has a single-index structure.
  - But there is no guarantee from economic theory that demand has such a structure.
  - Demand estimation using a misspecified model may give seriously misleading results.

Motivation (cont.)

- This paper takes a different approach
- We impose structure on the demand function by using a shape restriction from economic theory.
- Specifically, we impose the Slutsky restriction of consumer theory on the demand function
- This approach yields well-behaved estimates of the demand function and deadweight losses.
- Our method
  - maintains the flexibility of nonparametric estimation
  - generates estimates consistent with economic theory
  - Avoids using essentially arbitrary and possibly incorrect restrictions to stabilize the estimates.

Unobserved heterogeneity and quantile estimation

- In previous work, we have investigated gasoline demand, focussing on the conditional mean (Blundell, Horowitz, Parey, 2010).
- If unobserved heterogeneity enters in a non-separable manner,
  - the conditional mean represents an average across unobservables, which may be difficult to interpret
  - under suitable restrictions, quantile estimation allows to recover demand at a specific point in the distribution of unobservables.
- This motivates our interest in a quantile estimator.

- Quantile regression also allows us to study differential effects of price changes and Deadweight Loss across the distribution of unobservables
  - compare heavy users with moderate or light users.

Outline

Description of data
Semi-nonparametric estimates
Semi-nonparametric estimation subject to Slutsky restrictions
Estimation of deadweight loss
Conclusions

Data

- Data are from the 2001 National Household Travel Survey (NHTS).
- This is a household-level survey that was conducted by telephone and complemented by travel diaries and odometer readings.
- The nonparametric estimates are shown below for the three income groups whose midpoints in 2001 dollars are $42,500, $57,500 and $72,500.
  - These correspond to quartiles of the income distribution
- To minimize heterogeneity, we restrict the analysis to households with a white respondent, two or more adults, at least one child under age 16, and at least one driver.
- We take vehicle ownership as given and do not investigate how changes in gasoline prices affect vehicle purchases or ownership.
  - See Bento, Goulder, Henry, Jacobsen, and von Haefen (2005)
- The resulting sample contains 5,254 observations.
- A parametric log-log model, at the median, gives
  - a price elasticity of -0.36, and
  - an income elasticity of 0.34.

The Semi-Nonparametric Model

The demand function is

$$Q = g(P, Y, X, U)$$

To ensure identification, we assume that

- U is statistically independent of (P, Y, X).
- g is monotone increasing in its 4th argument.
- Given these assumption, we can also assume without loss of generality that $U \sim U[0; 1]$.

Under these assumptions, the α quantile of Q conditional on (P, Y, X) is

$$\mathcal{Q}_\alpha(Q|P, Y, X) = g(P, Y, X, \alpha) \equiv G_\alpha(P, Y, X)$$

Thus, the underlying demand function, evaluated at a specific value of the unobservable, can be recovered via quantile estimation.

In contrast, the conditional mean is

$$\begin{aligned} E\left(Q|P=p, Y=y, X=x\right) &= \int g(p,y,x,u)\, F(u)\, du \\ &= m(p,y,x) \end{aligned}$$

Imposing the constraint at a specific value of $U = \alpha$ is attractive because economic theory informs us about $g(\cdot)$ rather than $m(\cdot)$.

The Semi-Nonparametric Model

We can write

$$Q = G_\alpha(P, Y, X) + V_\alpha$$

where V_α is a random variable whose α quantile conditional on (P, Y, X) is zero. To mitigate the curse of dimensionality, we assume that G_α has a partially linear structure:

$$\mathcal{Q}_\alpha(Q|P, Y, X) = H_\alpha(P, Y) + \beta_\alpha X$$

where β_α is a vector of constant coefficients. Thus, the estimated model for each α is

$$Q = \hat{H}_\alpha(P, Y) + \hat{\beta}_\alpha X + V_\alpha$$

where H_α is nonparametric.

Semi-Nonparametric Estimation

- We estimate H_α using B-splines:

$$H_\alpha(P, Y) = \sum_{m_1} \sum_{m_2} c_{m_1, m_2;\, \alpha}\, B^p_{m_1}(P) B^y_{m_2}(Y)$$

  where B^p and B^y (with indices m_1 and m_2) are spline functions following Powell (1981), and $c_{m_1, m_2;\, \alpha}$ is the matrix of coefficients.
- Denote the data by $\{Q_i, P_i, Y_i, X_i : i = 1, ..., n\}$.

The estimator is defined in the following optimization problem:

$$\min_{\{c_{m_1, m_2;\, \alpha}\}, \beta_\alpha} \sum_{i=1}^{n} \rho_\alpha\left(Q_i \;-\; H_\alpha(P_i, Y_i) \;-\beta_\alpha X_i\right)$$

where $\rho_\alpha(V) = (\alpha - \mathbb{1}\{V < 0\})\, V$ is the check function.

Semi-Nonparametric Estimation (cont.)

Implementation:

- We use cubic splines
- To avoid influence of low-density areas, we restrict attention to a price-income rectangle of interest:
  - price between $1.22 and $1.47
  - income between $35,000 and $100,000.
- Set of covariates contained in X: Public transit availability, urbanity indicators, population density indicators.
- For each quantile of interest, the number of knots is obtained by cross-validation.

Outline

Imposing the Slutsky Condition

- We impose the Slutsky condition on the nonparametric estimate of the conditional quantile function.
- The condition is

$$\frac{\partial G_\alpha(P,Y,X)}{\partial p} + G_\alpha(P,Y,X)\,\frac{\partial G_\alpha(P,Y,X)}{\partial y} \;\leq\; 0.$$

- The constrained estimator is obtained by solving the problem

$$\min_{\{c_{m_1,m_2;\,\alpha}\},\beta_\alpha} \sum_{i=1}^{n} \rho_\alpha\left(Q_i \;-\; \hat{G}^C_\alpha(P_i,Y_i,X_i)\right)$$

subject to

$$\frac{\partial \hat{G}^C_\alpha(P,Y,X)}{\partial p} + \hat{G}^C_\alpha(P,Y,X)\frac{\partial \hat{G}^C_\alpha(P,Y,X)}{\partial y} \leq 0$$

for all (P,Y,X). This problem has uncountably many constraints.

Implementation

- We replace the continuum of constraints by a discrete set, thereby solving:

$$\min_{\{c_{m_1,m_2;\,\alpha}\},\beta_\alpha} \sum_{i=1}^{n} \rho_\alpha \Big(Q_i \; - \; \hat{G}_\alpha^C(P_i, Y_i, X_i) \Big)$$

subject to

$$\frac{\partial \hat{G}_\alpha^C(p_j, y_j, x_j)}{\partial p} + \hat{G}_\alpha^C(p_j, y_j, x_j) \frac{\partial \hat{G}_\alpha^C(p_j, y_j, x_j)}{\partial y} \leq 0, \quad j = 1, ..., J,$$

where $\{p_j, y_j, x_j : j = 1, ..., J\}$ is a grid of points.

- To implement this, we use a method of Koenker & Ng (2004) for quantile regression subject to (linear) inequality constraints.
  - We linearize the restriction by holding $G_\alpha(P, Y, X)$ in the income effect constant at the unconstrained solution.
  - We then iterate this procedure, using the constrained estimate for $G_\alpha(P, Y, X)$ previously obtained.

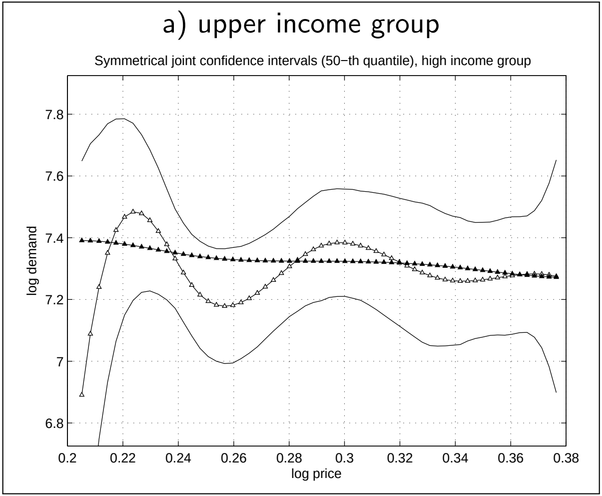
Role of partially-linear covariates in imposing the constraint

- Given the partially linear structure, the choice of the x_j, the X where we impose the constraint, may matter.
- In principle, this can make the restriction more or less binding, but we do not exploit this.
- We choose x_j as follows: At each grid point, we find the household in the data closest to that (p_j, y_j) combination.
- We set x_j equal to the covariates for that household.

Imposing the Slutsky Condition (cont.)

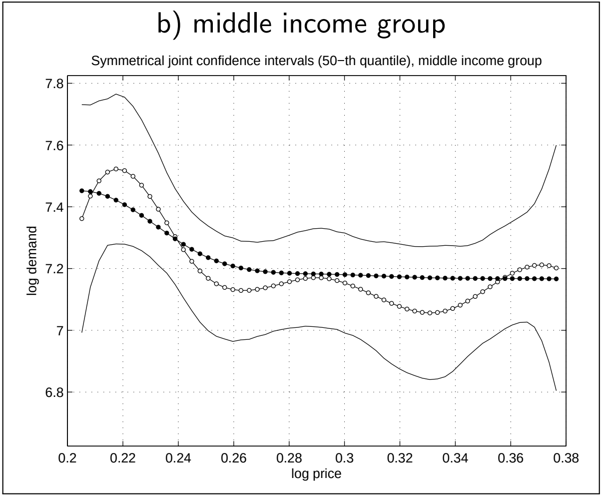
- We use the bootstrap for inference under the assumption that the Slutsky constraint does not bind in the population.
- The constrained estimates are
  - Consistent with economic theory.
  - Contained in a 90% confidence band around the unconstrained ones; consistent with random sampling error interpretation.
  - The middle income group is more sensitive to price than are the two outer groups.

Figure: Demand estimates and simultaneous confidence intervals (at $\alpha = 0.5$)



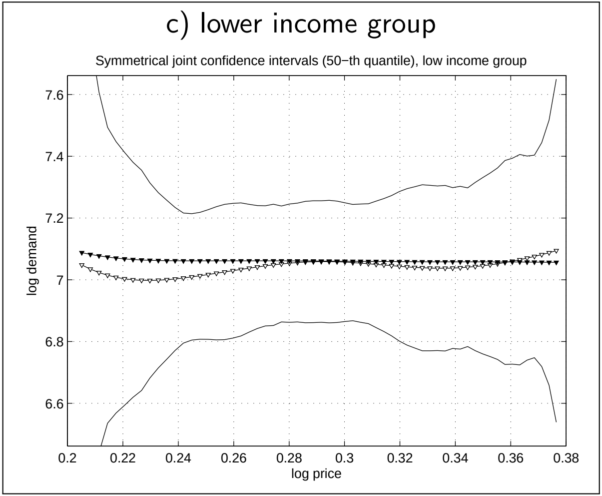
Note: Confidence intervals shown refer to bootstrapped symmetrical, simultaneous confidence intervals with a confidence level of 10%, based on 1,999 replications.

Figure: Demand estimates and simultaneous confidence intervals (at $\alpha = 0.5$)



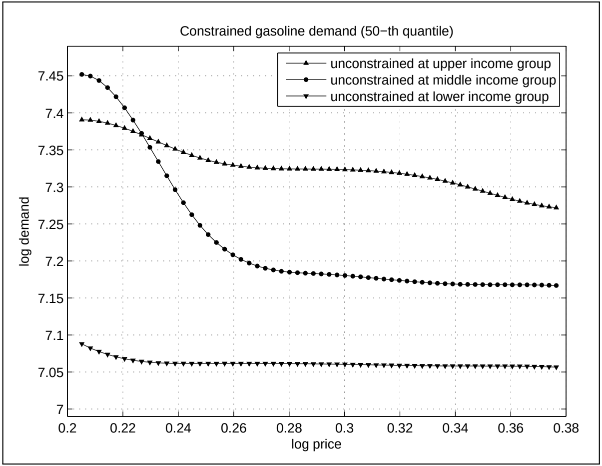
Note: Confidence intervals shown refer to bootstrapped symmetrical, simultaneous confidence intervals with a confidence level of 10%, based on 1,999 replications.

Figure: Demand estimates and simultaneous confidence intervals (at $\alpha = 0.5$)



Note: Confidence intervals shown refer to bootstrapped symmetrical, simultaneous confidence intervals with a confidence level of 10%, based on 1,999 replications.

Figure: Summary constrained demand estimates (at $\alpha = 0.5$)



Do different income groups have different price derivatives?

- Interest in testing whether the price derivatives between two given income levels is equal.
- Null hypothesis:

$$\left.\frac{\partial G_\alpha(P,Y,X)}{\partial p}\right|_{Y=y^0} - \left.\frac{\partial G_\alpha(P,Y,X)}{\partial p}\right|_{Y=y^1} = 0$$

  for all p in some interval.
- Given our specification for G_α, this leads to:

$$\sum_{m_1}\sum_{m_2} \frac{\partial B^p_{m_1}(p)}{\partial p}\, c_{m_1,m_2;\alpha}\, \left(B^y_{m_2}(y^1) - B^y_{m_2}(y^0)\right) = 0$$

- This implies a linear restriction on the vector of coefficients.
- We can test this using a Wald test.

Differential price effects – preliminary results

Comparison with high income group ($y_1 = \$72,500$) for unconstrained median estimates:

income (y_0)	p-value
middle income group (\$57,500)	0.0442
lower income group (\$42,500)	0.4143

Note: Cubic splines with 5 knots in price and 2 knots in income dimension. Unconstrained variance-covariance matrix is bootstrapped (999 replications).

- Evidence of differential price effects across the income distribution.

Outline

Estimating Deadweight Loss (DWL)

- Let $e(p)$ denote the expenditure function at price p and some reference utility level (keeping X implicit).
- The DWL of a tax that changes the price from p^0 to p^1 is

$$L(p^0, p^1) = e(p^1) - e(p^0) - (p^1 - p^0)\, G_\alpha \left[p^1, e(p^1), X\right].$$

- We estimate $L(p^0, p^1)$ by replacing e and g with consistent estimates.
- The estimator of e, $\hat{e}$, is obtained by numerical solution of the differential equation

$$\frac{d\hat{e}(t)}{dt} = \hat{G}_\alpha \left[p(t), \hat{e}(t), X\right] \frac{dp(t)}{dt},$$

where $[p(t), \hat{e}(t)]$ $(0 \leq t \leq 1)$ is a price-(estimated) expenditure path.

	DWL (as % of tax paid)		
	Semi-nonparametric		Parametric
	unconstrained	constrained	log-log
Income	(1)	(2)	(3)
	Panel A: 75th-percentile		
$72,500	-1.23 %	3.63 %	2.02 %
$57,500	4.42 %	8.35 %	2.01 %
$42,500	-13.12 %	0.34 %	1.99 %
	Panel B: 50th-percentile		
$72,500	2.67 %	5.63 %	3.06 %
$57,500	0.61 %	6.37 %	3.04 %
$42,500	-5.45 %	0.54 %	3.02 %
	Panel C: 25th-percentile		
$72,500	3.56 %	6.34 %	6.17 %
$57,500	23.31 %	13.85 %	6.16 %
$42,500	52.65 %	45.73 %	6.14 %

- Some unconstrained estimates of DWL have the incorrect sign
- The nonparametric estimates show a strong dependence of DWL on income.
- Parametric estimates are very different from the nonparametric ones and do not show the income dependence of DWL measures.

Conclusions

- Simple parametric models of demand functions can be misspecified and, consequently, yield misleading estimates of price sensitivity and DWL.
- Nonparametric estimates eliminate the risk of specification error but can be poorly behaved (e.g., non-monotonic) due to random sampling errors.
- This paper has shown that these problems can be overcome by constraining nonparametric estimates to satisfy the Slutsky condition of economic theory.
  - This stabilizes the nonparametric estimates without the need for essentially arbitrary parametric or semiparametric restrictions that have no theoretical basis.
- We have illustrated this approach by estimating a gasoline demand function.
  - Fully nonparametric estimates are nonmonotonic and violate consumer theory
  - The constrained estimates are well-behaved and reveal features not found with parametric model

Conclusions (2)

- We have also estimated deadweight losses of some tax increases
  - The unconstrained nonparametric estimates sometimes have incorrect signs.
  - The constrained nonparametric estimates have correct signs and are very different from those obtained from a simple parametric model.
  - The constrained nonparametric estimates reveal patterns of income dependence that are not present in the parametric estimates
- Further research:
  - Estimate Slutsky-constrained models in which prices may be endogenous, with the distance variable we have previously used to instrument for P.
  - Estimate systems of demand functions for several goods