

Individual heterogeneity and average welfare

**Jerry Hausman
Whitney Newey**

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP34/13

Individual Heterogeneity and Average Welfare*

Jerry A. Hausman
MIT

Whitney K. Newey
MIT

May, 2011
Revised, July 2013

Abstract

Individual heterogeneity is an important source of variation in demand. Allowing for general heterogeneity is needed for correct welfare comparisons. We consider general heterogeneous demand where preferences and linear budget sets are statistically independent. We find that the dimension of heterogeneity and the individual demand functions are not identified. We also find that the exact consumer surplus of a price change, averaged across individuals, is not identified, motivating bounds analysis. We use bounds on income effects to derive relatively simple bounds on the average surplus, including for discrete/continuous choice. We also sketch an approach to bounding surplus that does not use income effect bounds. We apply the results with income effect bounds to gasoline demand. We find little sensitivity to the income effect bounds in this application.

JEL Classification: C10, C14, C51, C54.

Keywords: Consumer surplus, deadweight loss, identification, quantiles.

*The National Science Foundation (Grant SES 1132399) provided financial support for this paper. Helpful comments were provided by D. Acemoglu, R. Blundell, C. Manski, A. Pakes and participants in seminars at Bilkent, Harvard, MIT, Northwestern, and UCL. Capable research assistance was provided by M. Ding, S. Hong, J. Lee, Y. Narita, C. Palmer, and T. Pham.

1 Introduction

Unobserved individual heterogeneity is thought to be a large source of variation in empirical demand equations. Often r-squareds are found to be quite low in cross-section and panel data applications, suggesting that much variation in demand is due to unobserved heterogeneity. Given the potential magnitude of such heterogeneity it is important that it be allowed for in a flexible way.

Demand functions could vary across individuals in general ways. For example, it seems reasonable to suppose that price and income elasticities are not confined to a one dimensional curve as they vary across individuals, meaning that heterogeneity is multi-dimensional. Demand might also arise from combined discrete and continuous choice, where heterogeneity has different effects on discrete and continuous choices. Furthermore, the dimension of heterogeneity is not identified from cross-section data, as we show. For these reasons it seems important to allow for general heterogeneity in demand analysis. In this paper we do so.

Independence of preferences and budget sets is useful for identification in cross-section data with general heterogeneity. This independence corresponds to prices and incomes being independent of preferences as well as observations coming from different markets. We impose independence, conditional on covariates and/or control functions.

Exact consumer surplus quantifies the welfare effect of price changes, including the deadweight loss of taxes. The surplus averaged over individuals is a common welfare measure. We show that average surplus is not identified. Nonidentification motivates a bounds approach. We use bounds on income effects to derive relatively simple bounds on the average of surplus across individuals, including for discrete/continuous choice models. It may be also be possible to obtain bounds on average surplus without income effect bounds, but this appears to be substantially more complicated. We sketch one approach to finding such bounds, but leave its execution to future research.

We apply the exact consumer surplus results to gasoline demand, using data from the 2001 U.S. National Household Transportation Survey. We find little sensitivity to the

income effect bounds in this application. We show that this is true more generally with small expenditure on a good. We also show how to allow for covariates and some forms of endogeneity, and incorporate these features in the application. We give confidence intervals for an identified set of average surplus values and find that the intervals are quite tight in the application.

Choice models with general heterogeneity independent of budget sets have previously been considered. In their analysis of nonlinear taxes, Burtless and Hausman (1978) pioneered models with multi dimensional heterogeneity by allowing for the income effect to vary over individuals. McFadden and Richter (1991) allowed for general heterogeneity in a revealed preference framework. Lewbel (2001) gave conditions for the average demand to satisfy the restrictions of utility maximization. Hoderlein and Stoye (2009) showed how to impose the weak axiom of revealed preference. Dette, Hoderlein, and Neumeyer (2011) proposed tests of downward sloping compensated demands. Kitamura and Stoye (2012) gave tests of the revealed preference hypothesis. Blomquist and Newey (2002) derived the form of average demand with nonlinear budget sets; see also Blomquist and Newey (2011). Our paper differs from the revealed preference work in imposing smoothness of demands. The hope is that smoothness will lead to more accurate nonparametric demand estimation, as in other nonparametric settings.

Blundell, Horowitz, and Parey (2012), Hoderlein and Vanhems (2010), and Blundell, Kristensen, and Matzkin (2011) have considered identification and estimation of welfare measures when demand depends continuously on a single unobserved variable. Blundell, Kristensen and Matzkin (2011) impose revealed preference restrictions on demand functions in that setting. Recently Lewbel and Pendakur (2013) have considered restricted multivariate heterogeneity. Our purpose is to go beyond and consider general heterogeneity.

The results of this paper build on Hausman and Newey (1995). The focus of that previous paper was to apply the consumer surplus calculation to a conditional expectation. As noted there, if the conditional expectation corresponds to the demand function for an individual then one can interpret the results as consumer surplus for an individual. Here

we allow for general heterogeneity but rely on our previous work for inference about the estimators.

2 Heterogeneity and Demand

We consider demand models where the form of heterogeneity is completely unrestricted. For simplicity we will describe these models for two goods. The bounds with income effects are straightforward to extend to multiple goods. Let q denote the quantity of a good, p its price relative to that of a numeraire good a , y the individual income level relative to the numeraire price, and η a vector of unobserved disturbances representing individual heterogeneity. We think of each value of η as corresponding to a single consumer with demand function $q(p, y, \eta)$ that is obtained as the solution to

$$\max_{q \geq 0, a \geq 0} U(q, a, \eta) \text{ s.t. } pq + a = y, \quad (2.1)$$

where $U(q, a, \eta)$ is a utility function for individual η . We impose no restrictions on the way that η affects the utility function U and hence $q(p, y, \eta)$ is completely unrestricted as a function of η . That is, demand functions are allowed to vary across individuals in any way at all.

We consider identification and estimation under independence of η and $x = (p, y)$ where the heterogeneity takes an unrestricted form. We need some restriction on the joint distribution of x and η to identify economically interesting objects, such as price and income effects. Without such restrictions economic effects may be confounded with individual heterogeneity. Independence is a natural starting condition to impose in models like this, that are nonlinear, nonseparable, and nonparametric. We also show later how this assumption can be relaxed with covariates and/or control functions.

ASSUMPTION 1: *The data (q_i, x_i) , $(i = 1, \dots, n)$ are identically distributed and satisfy $q_i = q(x_i, \eta_i)$ where x_i and η_i are statistically independent.*

This condition encompasses a statistical version of a fundamental hypothesis of consumer demand, that preferences do not vary with prices. It also encompasses the hy-

pothesis that the individual is small relative to the market of observation, so that market prices do not vary with individual preferences. This hypothesis is consistent with data where different observations come from different markets.

We also impose smoothness and the Slutsky condition on demands:

ASSUMPTION 2: *For each η the demand function $q(x, \eta)$ is continuously differentiable in $x \geq 0$ and $\partial q(x, \eta)/\partial p + q(x, \eta)\partial q(x, \eta)/\partial y \leq 0$.*

The model is the set of pairs $(q(\cdot, \cdot), F)$ of a function $q(x, \eta)$ of x and a vector η and a distribution function F for η such that Assumptions 1 and 2 are satisfied. This model is a smooth version of the stochastic revealed preference model of McFadden and Richter (1991). Assumption 1 imposes independence between the budget set and preferences, with preferences here represented by η . Also, as is well known from the results on integrability of demand, Hurwicz and Uzawa (1971), the Slutsky condition in Assumption 2 is equivalent to the axioms of revealed preference. What distinguishes this model from McFadden and Richter's (1991) is the smoothness of the demand function in Assumption 2. As in other nonparametric estimation, the hope in imposing smoothness is that it is likely to be satisfied in applications and nonparametric estimators are more precise with smoothness. We drop parts of this requirement for discrete/continuous choice models.

Turning now to identification, we adapt a standard framework to our setting, see Hsiao (1983). Here a structure is a demand function and heterogeneity distribution pair (q, F) , where for notational convenience we suppress the arguments of the demand function q . Two structures are observationally equivalent if they give the same conditional distribution of demand given prices and income, over the support of prices and income in the data.

DEFINITION 1: *(q, F) and (\tilde{q}, \tilde{F}) are observationally equivalent if and only if for all r ,*

$$\Pr\left(\int \mathbf{1}(q(x_i, \eta) \leq r)F(d\eta) = \int \mathbf{1}(\tilde{q}(x_i, \tilde{\eta}) \leq r)\tilde{F}(d\tilde{\eta})\right) = 1.$$

We consider identification of a function $\delta(q, F)$ of the structure $(q(\cdot, \cdot), F)$, that is a map from the demand function and the distribution of heterogeneity into another set. The identified set will be the set of values of this function for all structures that are observationally equivalent.

DEFINITION 2: *The identified set for δ corresponding to (q, F) is*

$$\Lambda(q, F) = \{\delta(\tilde{q}, \tilde{F}) : (q, F) \text{ and } (\tilde{q}, \tilde{F}) \text{ are observationally equivalent}\} \quad (2.2)$$

A functional is identified if $\Lambda(q, F)$ is a singleton, sometimes now referred to as point identified. An example of an identified functional is the average demand at an observed value of x , given by

$$\delta(q, F)(x) = \int q(x, \eta)F(d\eta) = E[q(x_i, \eta_i)|x_i = x],$$

where the second equality follows by independence of x_i and η_i in the model. Identification follows because this functional is an explicit function of the conditional distribution of quantity demanded conditional on x . This functional may be of interest for predicting the effect of price or income changes on the average consumption of q over the range of the data. We will discuss below how it could be estimated subject to the restrictions of the demand model. Of course δ is also the conditional expectation of quantity given x , $\delta(x) = E[q_i|x_i = x]$, so could be estimated by nonparametric regression. For example, Blomquist and Newey (2002, 2011) have done this for labor supply where the above model is generalized to allow for a nonlinear budget set. One could consider the average demand outside the range of the data on x . That functional is not identified, but the identified set may be restricted by imposition of positive demands and the Slutsky condition.

A functional is not identified if $\Lambda(q, F)$ is bigger than a single point, sometimes now referred to as set identified. An interesting example of an unidentified (or set identified) functional is the dimension $\dim(\eta)$ of the heterogeneity. To see nonidentification, consider the conditional quantile $Q_q(\tau|x)$, $0 < \tau < 1$, of q_i given $x = x_i$. An important observation

due to Dette, Hoderlein, Neumeyer (2011) is that $Q_q(\tau|x)$ satisfies the Slutsky condition for every τ , under certain regularity conditions, and so is a demand function for each τ . We refer to these as the quantile demands. By standard arguments it then follows that $(Q_q(\cdot|\cdot), \tilde{F})$ is a demand specification that is observationally equivalent to (q, F) , where \tilde{F} is the uniform distribution on $(0, 1)$. Furthermore, the dimension of $\tilde{\eta}$ is 1, so that if $\dim(\eta) > 1$ we have nonidentification of the dimension.

A precise result on the observational equivalence of quantile and actual demands follows.

THEOREM 1: *Suppose that $\eta = (u, \varepsilon)$ for scalar ε , $q(x, \eta) = q(x, u, \varepsilon)$ is continuously differentiable in ε , there is $C > 0$ with $\partial q(x, u, \varepsilon)/\partial \varepsilon \geq 1/C$, $\|\partial q(x, \eta)/\partial x\| \leq C$ everywhere, ε is continuously distributed conditional on u , with conditional pdf $f(\varepsilon|u)$ that is bounded and continuous in ε . If Assumptions 1 and 2 are satisfied, then (q, F) and (\tilde{q}, \tilde{F}) are observationally equivalent demand specifications, where $0 < \tilde{\eta} < 1$, $\tilde{q}(x, \tilde{\eta}) = Q_q(\tilde{\eta}|x)$, and \tilde{F} is the uniform distribution on $(0, 1)$.*

Nonidentification of $\dim(\eta)$ is a corollary of this result:

COROLLARY 2: *$\dim(\eta)$ is not identified at any (q, F) satisfying the hypotheses of Theorem 1 with $\dim(\eta) > 1$.*

A specification (q, F) where the dimension is not identified is a linear random coefficients model where

$$q_i = \eta_{1i} + \eta_{2i}p_i + \eta_{3i}y_i, (\eta_{1i}, \eta_{2i}, \eta_{3i}) \text{ independent of } (p_i, y_i), \Pr(q_i \geq 0, \eta_{2i} + q_i\eta_{3i} \leq 0) = 1.$$

Under the conditions of Theorem 1 the quantile demand is observationally equivalent to this random parameter model. We have $\dim(\eta) = 3$ for the random coefficients while $\dim(\eta) = 1$ for the observationally equivalent quantile demand. Thus, $\dim(\eta)$ is not identified for this linear random coefficients specification.

It is also clear that the demand function is not identified when $\dim(\eta) > 1$, since the demand depends on η .

COROLLARY 3: $q(x, \eta)$ is not identified at any (q, F) satisfying the hypotheses of Theorem 1 with $\dim(\eta) > 1$.

Another example of an unidentified function is average exact consumer surplus, that we consider in the next Section.

3 Average Surplus

We turn now to consumer surplus. We first consider exact consumer surplus for a single individual. We focus on equivalent variation though a similar analysis could be carried out for compensating variation. Let $S(p, y, \eta)$ be the equivalent variation for a price change from p to p_1 for an individual η with income y . As discussed in Hausman and Newey (1995), Shephard's Lemma implies that $S(p, y, \eta)$ solves the following differential equation,

$$\frac{\partial S(p, y, \eta)}{\partial p} = -q(p, y - S(p, y, \eta), \eta), S(p_1, y, \eta) = 0. \quad (3.3)$$

Like the demand function, $S(p, y, \eta)$ is not identified from cross-section data. To see this, note that the mapping from the demand function to $S(p, y, \eta)$ is one-to-one, with the solution to equation (3.3) giving S as a function of q and $q(p, y, \eta) = -\partial S(p, y + S(p, y, \eta), \eta) / \partial p$. Thus, $S(p, y, \eta)$ will not be identified because $q(p, y, \eta)$ is not identified.

COROLLARY 4: *Equivalent variation is not identified at any (q, F) satisfying the hypotheses of Theorem 1 with $\dim(\eta) > 1$.*

One interesting functional is the average, or expected consumer surplus. It is given by

$$\bar{S}(p, y) = E[S(p, y, \eta_i)] = \int S(p, y, \eta) F(d\eta).$$

Average surplus is often used in public economics as a welfare measure. It corresponds to a social welfare function that is the average of the expenditure function across consumers. One could consider other social welfare functions. We leave that to future work.

In general, average consumer surplus is not identified. To show this it suffices to consider a linear, random coefficients model with an income coefficient taking on two values.

THEOREM 5: *In a specification $q(x, \eta) = \eta_1 + \beta p + \eta_2 y$ with η_1 continuously distributed and η_2 having a two point mixture distribution, the average consumer surplus $\bar{S}(p, y)$ is not identified.*

This result is proved by showing that the average surplus for quantile demand is different than the average surplus for the true demand function in an example. Because it helps to quantify the misspecification that can result from quantile demands we briefly describe the example. In the example $\eta_1 \sim U(0, 1)$ and η_2 is distributed independently of η_1 with $\Pr(\eta_2 = 1/3) = \Pr(\eta_2 = 2/3) = 1/2$. For $\beta = -1$, $p = .1$, $p_1 = .2$, and $y = 3/4$ we calculate the ratio of average surplus for quantile demands to true average surplus to be .903. Thus, in this example the average surplus for quantile demands is only about 90 percent of the true average surplus.

A key to this result is that the income effect η_2 varies across individuals. Average surplus is identified if the demand function is restricted to have a constant income effect that does not vary across individuals, corresponding to the Gorman (1961) polar form (GPF). GPF demand functions satisfy

$$q(p, y, \eta) = a(p, \eta) + b(p)y. \quad (3.4)$$

Here the average demand will be $\bar{q}(p, y) = \bar{a}(p) + b(p)y$ where $\bar{a}(p) = \int a(p, \eta)F(d\eta)$. Also, the differential equation for the individual equivalent variation $S(p, y, \eta)$ is given by

$$\frac{\partial S(p, y, \eta)}{\partial p} = -[a(p, \eta) + b(p)\{y - S(p, y, \eta)\}].$$

Because this equation is linear in functions of η , integrating both sides with respect to η and interchanging the order of differentiation and integration gives

$$\frac{\partial \bar{S}(p, y)}{\partial p} = -[\bar{a}(p) + b(p)\{y - \bar{S}(p, y)\}] = -\bar{q}(p, y - \bar{S}(p, y)), \bar{S}(p_1, y) = 0. \quad (3.5)$$

Thus, for the GPF, average surplus can be obtained by solving the differential equation for equivalent variation using the average demand.

THEOREM 6: *If $q(x, \eta)$ is restricted to satisfy $q(x, \eta) = a(p, \eta) + b(p)y$ for functions $a(p, \eta)$ and $b(p)$, Assumptions 1 and 2 are satisfied, and $E[|a(p, \eta_i)|] < \infty$ then $\bar{S}(p, y)$ is identified as the solution to equation (3.5).*

Identification of average consumer surplus for GPF demand is consistent with the well known aggregation results of Gorman (1961), who showed that the GPF is necessary and sufficient for the average of the demand functions to be the demand for a consumer with utility equal to the average of the individual utilities. Indeed the preceding discussion is just a simple demonstration of a partial dual result, that the GPF is sufficient for the expenditure function for average (or aggregate) demand to be the average (or aggregate) of individual expenditure functions when income is independent of preferences. It also is interesting to note that average surplus for the quantile demand is the same as average surplus for the true demand when demand is GPF.

The GPF is too restrictive for most applications. However related ideas can be used to bound average welfare for other demands. The approach is based on restricting the range of income effects. The idea is that if we have known income effect bounds then the actual demand, on the right hand side of equation (3.3) will be bounded above and below by a linear in y demand. Since the upper and lower bounds are linear in y , a similar calculation as for the GPF shows that average demand can be used to obtain upper and lower bounds for average surplus.

The following result makes this idea precise. For any constant C define $\bar{S}_C(p, y)$ as the solution to the differential equation

$$\frac{\partial \bar{S}_C(p, y)}{\partial p} = -\bar{q}(p, y) + C\bar{S}_C(p, y), \bar{S}_C(p^1, y) = 0. \quad (3.6)$$

THEOREM 7: *If $q(p, y, \eta)$ is continuously differentiable in y and there are constants b and B such that $b \leq \partial q(p, y, \eta) / \partial y \leq B$ then for all $p \leq p^1$,*

$$\bar{S}_B(p, y) \leq \bar{S}(p, y) \leq \bar{S}_b(p, y).$$

This result shows that income effect bounds lead to corresponding bounds on expected consumer surplus that can be computed from average demand. Thus, under Assumption 1 and bounds on the income effect, the conditional expectation $E[q|p, y] = \bar{q}(p, y)$ can be used to identify bounds on the average surplus via equation (3.6) and Theorem 7. In this way a nonparametric regression of q on (p, y) can be used to obtain bounds on average consumer surplus. It is interesting that the bounds come from the conditional expectation of demand and not the log of demand or some other nonlinear function of demand.

The key ingredients for this result are the bounds on the income effect. Prior restrictions may provide bounds. The lower bound is $b = 0$ for a normal good and the upper bound is $B = 0$ for an inferior good. The data can also be informative about the bounds. Let $Q_q(\tau|x)$ denote quantile demand, as discussed above. By Hoderlein and Mammen (2007),

$$\frac{\partial Q_q(\tau|x)}{\partial y} = E\left[\frac{\partial q(x, \eta)}{\partial y} \mid q(x, \eta) = Q_q(\tau|x)\right].$$

Thus, as τ varies the quantile derivative traces out the average of income effects over the set of η values given by $q(x, \eta) = Q_q(\tau|x)$. The bounds b and B must be below and above, respectively, all of these quantile derivatives. The bounds are not in general identified because the demand function is not identified.

Allowing the income effect bounds to depend on p can lead to tighter surplus bounds. Suppose that there are functions $b(p)$ and $B(p)$ such that

$$b(p) \leq \frac{\partial q(x, \eta)}{\partial y} \leq B(p).$$

In this case the upper and lower bounds could be obtained from the solution $\bar{S}_C(p, y)$ to

$$\frac{\partial \bar{S}_C(p, y)}{\partial p} = -\bar{q}(p, y) + C(p)\bar{S}_C(p, y), \bar{S}_C(p^1, y) = 0,$$

where $C(p) = b(p)$ and $C(p) = B(p)$ respectively. The corresponding bounds will be tighter than those for constant bounds $b = \inf_p b(p)$ and $B = \sup_p B(p)$, respectively.

The surplus bounds will not be very sensitive to the income effect bounds when expenditure on q is small. This result is related to the Hotelling (1938) result that when

expenditure is small approximate consumers surplus is typically close to actual consumers surplus. To see this note that the solution to the linear differential equation (3.6) is given by

$$\bar{S}_C(p, y) = \int_p^{p^1} \bar{q}(\tilde{p}, y) e^{C(p-\tilde{p})} d\tilde{p}.$$

Differentiating this with respect to C gives

$$\begin{aligned} \frac{\partial \bar{S}_C(p, y)}{\partial C} &= \int_p^{p^1} \bar{q}(\tilde{p}, y) (p - \tilde{p}) e^{C(p-\tilde{p})} d\tilde{p} \\ &= \int_p^{p^1} [\bar{q}(\tilde{p}, y) \tilde{p}] \left(\frac{p}{\tilde{p}} - 1 \right) e^{C(p-\tilde{p})} d\tilde{p}. \end{aligned}$$

It follows that the magnitude of the derivative increases with $q(\tilde{p}, y)\tilde{p}$. Therefore the higher the expenditure on the good the more sensitive will be the surplus bounds to the income effect bounds.

It is straightforward to carry out a similar analysis for deadweight loss. For an individual the deadweight loss of a tax $p^1 - p^0$ is

$$DWL(p, y, \eta) = S(p, y, \eta) - (p^1 - p)q(p^1, y, \eta).$$

Integrating the deadweight loss over η gives the average

$$\overline{DWL}(p, y) = \bar{S}(p, y) - (p^1 - p)\bar{q}(p^1, y).$$

Upper and lower bounds for average deadweight loss are then

$$\overline{DWL}_b(p, y) = \bar{S}_b(p, y) - (p^1 - p)\bar{q}(p^1, y), \overline{DWL}_B(p, y) = \bar{S}_B(p, y) - (p^1 - p)\bar{q}(p^1, y).$$

These bounds are obtained by simply subtracting the average tax revenue term $(p^1 - p)\bar{q}(p^1, y)$ from the surplus bounds.

There is a way to check the validity of the bounds using quantile demands. It turns out that the integral over quantiles of the consumer surplus must lie between the bounds if the income effect is bounded. To describe this result let $S^\tau(p, y)$ be the exact consumer surplus for the quantile demand given by the solution to

$$\frac{\partial S^\tau(p, y)}{\partial p} = -Q_q(\tau|p, y - S^\tau(p, y)), S^\tau(p_1, y) = 0.$$

The average surplus under scalar heterogeneity would be the integral $\int_0^1 S^\tau(p, y) d\tau$ of $S^\tau(p, y)$ over a uniform distribution for τ . The following result shows that this object is always between the bounds.

THEOREM 8: *Under the conditions of Theorem 1, for all $p \leq p^1$*

$$\bar{S}_B(p, y) \leq \int_0^1 S^\tau(p, y) d\tau \leq \bar{S}_b(p, y).$$

Also, if η is a scalar, $q(p, y, \eta)$ is monotonic in η and η is continuously distributed then

$$\int_0^1 S^\tau(p, y) d\tau = \int S(p, y, \eta) F(d\eta).$$

Thus the validity of the bounds on income effects can be checked by comparing the bounds with what would be average surplus with scalar heterogeneity.

The average surplus bounds given above are based on income effect bounds. One could also construct average surplus bounds based only on utility maximization. Consider the set of functions $\tilde{q}(x, \eta)$ and CDF's \tilde{F} for η in the set

$$\tilde{R} = \left\{ (\tilde{q}, \tilde{F}) : F(r|x) = \int \mathbf{1}(\tilde{q}(x, \eta) \leq r) \tilde{F}(d\eta), \frac{\partial \tilde{q}(x, \eta)}{\partial p} + \tilde{q}(x, \eta) \frac{\partial \tilde{q}(x, \eta)}{\partial y} \leq 0 \right\}, \quad (3.7)$$

where the first equality and the second inequality hold for all r, x in the support of x_i , and η in the support of \tilde{F} . One could approximate \tilde{R} using a series approximation, with $\tilde{q}(x, \eta) \approx F^{-1}(\eta_0|x) + \sum_{j=1}^J \eta_j m_j(x)$, where $m_j(x)$ are approximating functions, the marginal distribution of η_0 is $U(0, 1)$, and the distribution of $\eta = (\eta_1, \dots, \eta_J)$ is a discrete mixture that is independent of η_0 . Many, say L , values for η could be drawn at random, only keeping those that satisfy the second condition in equation (3.7) over a grid of x points. The distribution of η over the resulting values could be taken to be independent of η_0 and a discrete mixture with $\rho_\ell = \Pr(\eta = \eta^\ell)$. One could impose the first condition in equation (3.7) by requiring equality of a nonparametric estimator of the conditional CDF $F(r|x)$ with that implied by the model. Let $\hat{F}(r|x)$ be a nonparametric estimator that satisfies the Slutsky conditions (the conditional CDF satisfies the Slutsky condition

by the inverse function theorem). Then equality of this estimator with that implied by the model is

$$\hat{F}(r|x) = \sum_{\ell=1}^L \rho_{\ell} \hat{F}(r - \sum_{j=1}^J \eta_j^{\ell} m_j(x)|x).$$

Specifying that this holds over a grid of r and x points then leads to a set of equality restrictions on ρ_{ℓ} , with L being taken large in order to model general heterogeneity. Bounds on average exact consumer surplus (or average demand) could then be estimated by linear programming. Average surplus and average demand are linear combinations of ρ_{ℓ} and so bounds could be estimated by maximizing and minimizing those linear combinations subject to the above constraints, $\rho_{\ell} \geq 0$, and $\sum_{\ell=1}^L \rho_{\ell} = 1$. This approach is like the revealed preference approaches except that it imposes the Slutsky condition on smooth demands rather than revealed preference inequalities. The hope is that using smooth approximations will improve nonparametric estimation as it does in other nonparametric settings. We intend to investigate this approach in future work.

4 Discrete and Continuous Choice

Discrete and continuous choice models are important in applications. For instance, gasoline demand could be modeled as gasoline purchases that are made jointly with the purchase of automobiles. In those models the heterogeneity can influence the discrete choices as well as the demand for a particular commodity, e.g. see Dubin and McFadden (1984) and Hausman (1985). Multiple sources of heterogeneity are an integral part of these models, with separate disturbances for discrete and continuous choices. The general heterogeneity we consider allows for such multi dimensional heterogeneity.

We first consider the individual choice problem and the associated expenditure function. We adopt the framework of Hausman (1985), extending previous results to the expenditure function. Suppose that the agent is choosing among J discrete choices in addition to choosing q . The consumer choice problem is

$$\max_{j,q,a} U(q, a, j, \eta) \text{ s.t. } pq + a + r_j \leq y \tag{4.8}$$

where r_j is the usage price of choice j . Here we assume that for each η and j the function $U(q, a, j, \eta)$ is strictly quasi-concave (preferences are strictly convex) and satisfies local nonsatiation.

To describe the demand function, let $V_j(p, \tilde{y}, \eta) = \max_{q,a} \{U(q, a, j, \eta) \text{ s.t. } pq + a \leq \tilde{y}\}$ denote the indirect utility function associated with the j^{th} discrete choice and $q_j(p, \tilde{y}, \eta) = -[\partial V_j(p, \tilde{y}, \eta) / \partial \tilde{y}]^{-1} \partial V_j(p, \tilde{y}, \eta) / \partial p$ the demand function. The utility maximizing choice of the discrete good will be $\arg \max_j V_j(p, y - r_j, \eta)$. When there is a unique discrete choice j (depending on p, y, r_1, \dots, r_J , and η) that maximizes utility, i.e. where $V_j(p, y - r_j, \eta) > V_k(p, y - r_k, \eta)$ for all $k \neq j$, the demand $q(p, y, r_1, \dots, r_J, \eta)$ will be

$$q(p, y, r_1, \dots, r_J, \eta) = q_j(p, y - r_j, \eta).$$

When there are multiple values of the discrete choice that maximize utility the demand will generally be a correspondence, containing one point for each value of j that maximizes utility.

In what follows we will assume that $(V_1(p, y - r_1, \eta), \dots, V_J(p, y - r_J, \eta))$ is continuously distributed so that the probability of ties is zero. Nevertheless the case with ties is important for us. Surplus is calculated by integrating the demand function as price changes while income is compensated to keep utility constant. As compensated income changes ties may occur and the demand for q may jump. With gasoline demand, compensated income changes could result in a choice of car with different gas mileage, leading to a jump. Such jumps must be accounted for in the bounds analysis.

In what follows we will group the usage prices with individual heterogeneity η , so that usage prices are treated as unobserved by the econometrician. Assumption 1 then requires that the income y and price p are independent of the usage prices. This seems a reasonable assumption for some applications to cross-section data, such as gasoline demand. There Assumption 1 states that car purchase prices are independent of gasoline prices in the cross-section. If the prices of different types of cars do not vary in the cross-section then independence would automatically hold.

It is interesting to note that the components of η that affect the discrete choice may

be different than the components that affect the continuous choice. For example, this is the case in Dubin and McFadden (1984). A discrete/continuous choice model thus leads naturally to multiple sources of heterogeneity in q .

Turning to welfare analysis, let $e(p, u, \eta)$ denote the expenditure function in this discrete/continuous choice setting, defined as

$$e(p, u, \eta) = \min\{y \text{ s.t. } \max_{j,q,a}\{U(q, a, j, \eta) \text{ s.t. } pq + a + r_j \leq y\} \geq u\}.$$

As usual it is the minimum value of income that allows individual η to attain utility level u . The exact equivalent variation $S(p, y, \eta)$ for a price change from p to p^1 with income y for individual η is defined as before, as $S(p, y, \eta) = y - e(p, u^1, \eta)$. The next result gives conditions for $S(p, y, \eta)$ to satisfy the same differential equation as in the continuous case.

THEOREM 9: *If for each j and η the utility $U(q, a, j, \eta)$ is strictly quasi-concave and satisfies local nonsatiation then at any p, y , and η such that there is j with $V_j(p, y - S(p, y, \eta) - r_j, \eta) > V_k(p, y - S(p, y, \eta) - r_k, \eta)$ for all $k \neq j$, it follows that $S(p, y, \eta)$ is differentiable and*

$$\frac{\partial S(p, y, \eta)}{\partial p} = -q(p, y - S(p, y, \eta), \eta) = -q_j(p, y - S(p, y, \eta) - r_j, \eta).$$

Thus when there are no ties the surplus is differentiable and satisfies the same differential equation as in the continuous case. Also, by standard arguments the surplus will be continuous in p and y . As long as ties only occur at a finite number of points it then follows that surplus is the solution to the differential equation in between the jump points that is connected together in a continuous function across the jump points.

The discontinuity of individual demand does affect the bounds for average consumer surplus. The previous bounds depend on demand derivatives. With jumps we construct bounds that are based on limits on the size of the jump and on the proportion of individuals whose demand would jump as income is compensated along with the price change. For that purpose we make use of a demand decomposition into continuous and jump components.

Assumption 3: *There are $b, B > 0$ such that for each η there is $D(\eta)$ such that for $p \in [p^0, p^1]$ and $0 \leq s \leq S(p^0, y, \eta)$,*

$$q(p, y, \eta) = \dot{q}(p, y, \eta) + d(p, y, \eta), bs \leq -\dot{q}(p, y - s, \eta) + \dot{q}(p, y, \eta) \leq Bs, |d(p, y, \eta)| \leq D(\eta).$$

Here we assume that the demand function can be decomposed into a jump component $d(p, y, \eta)$ and a Lipschitz continuous component $\dot{q}(p, y, \eta)$, with lower and upper bounds b and B , respectively, on how much \dot{q} may vary with y . The constants b and B are bounds on the income effect, for the continuous part. The term $D(\eta)$ is an individual specific bound on the jump. It will be zero for individuals whose demand function does not jump as income is compensated up to the surplus amount $S(p^0, y, \eta)$. For example, for gasoline demand it will be zero for individuals who would not change car types over the range of income being compensated.

To describe bounds on average surplus that allow for jumps, let $\bar{S}_\ell(p, y)$ and $\bar{S}_u(p, y)$ be the solutions to the differential equations

$$\begin{aligned} \frac{\partial \bar{S}_u(p, y)}{\partial p} &= -\bar{q}(p, y) + 2E[D(\eta)] + b \cdot \bar{S}_u(p, y), \bar{S}_u(p^1, y) = 0, \\ \frac{\partial \bar{S}_\ell(p, y)}{\partial p} &= -\bar{q}(p, y) - 2E[D(\eta)] + B \cdot \bar{S}_\ell(p, y), \bar{S}_\ell(p^1, y) = 0. \end{aligned}$$

THEOREM 10: *If Assumptions 1, 2, and 3 are satisfied then $\bar{S}_\ell(p, y) \leq \bar{S}(p, y) \leq \bar{S}_u(p, y)$.*

These bounds adjust for the possible presence of discontinuity in individual demands by adding $2E[D(\eta)]$ to $-\bar{q}(p, y)$ in the equation for the upper bound and subtracting the same term in the equation for the lower bound. This adjustment will be small when the largest possible jump is small or when the proportion of individuals with a discontinuity is small.

5 Covariates and Control Functions

In some settings it may be useful to account for the presence of covariates and/or to control for endogeneity. In this Section we describe how this can be done. Covariates

can be allowed for by assuming that the utility function depends on a vector of functions $v(w, \delta)$ of covariates w and parameters δ . For example, a linear index would be $v(w, \delta) = w_1 + w_2\delta$, with the usual scale and location normalization imposed. We will allow the utility function to depend on $v(w, \delta)$ in a completely general way. In demand these covariates might be demographic variables that represent observed components of the utility.

If the utility can depend on $v(w, \delta)$ in any way at all then so can demand. In keeping with a nonparametric specification, we will take demand to be

$$q(x, v(w, \delta_0), \eta),$$

where δ_0 denotes the true value of the parameter. Assuming that η is distributed independently of $p, y, v(w, \delta)$, the conditional expectation of demand will be

$$E[q(x, v(w, \delta_0), \eta)|p, y, v(w, \delta_0)] = \bar{q}(p, y, v(w, \delta_0)).$$

This is a semiparametric specification for the conditional mean where the conditional mean depends on p, y , and the index $v(w, \delta)$. It can be estimated by semiparametric index regression.

With covariates all that we have done before applies conditionally on the covariates $v(w, \delta)$. Bounds on expected consumer surplus can be obtained for some value of the estimated index $v(w, \hat{\delta})$. Calculations exactly like those before lead to bounds that are conditional on values of the covariates.

Endogeneity can also be allowed for if there is a control variable. A control variable ξ is an estimable variable such that $x = (p, y)$ and η are independent of each other conditional on ξ . In that case $\int q(p, y, \eta)F(d\eta) = \int E[q|p, y, \xi]F(d\xi)$ when the support of ξ conditional on x does not vary with x ; see Blundell and Powell (2003).

6 Estimation and Welfare Analysis of Gasoline Demand

In this section we estimate average consumer surplus and deadweight loss from changes in the gasoline tax in the US while allowing for unrestricted multidimensional individual heterogeneity. We use data from the 2001 U.S. National Household Transportation Survey (NHTS). This survey is conducted every 5-8 years by the Federal Highway Administration. The survey is designed to be a nationally representative cross section which captures 24-hour travel behavior of randomly-selected households. Data collected includes detailed trip data and household characteristics such as income, age, and number of drivers. We restrict our estimation sample to households with either one or two gasoline-powered cars, vans, SUVs and pickup trucks. We exclude Alaska and Hawaii. We use daily gasoline consumption, monthly state gasoline prices, and annual household income. We have 8,908 observations. Summary statistics are given in Table 1. Note that the mean price of gasoline was \$1.33 per gallon with the mean number of drivers in a household equal to 2.04.

To estimate the average gasoline demand we estimate up to a 4th degree polynomial with interaction and predetermined variables along with price and income for household i :

$$\bar{q}(p, \widehat{y}, w) = \sum_{j,k,\ell=1}^4 \hat{\beta}_{j,k,\ell} (\ln p)^j (\ln y)^k (v(w, \hat{\delta}))^\ell \quad (6.9)$$

We estimate equation (6.9) taking the price of gasoline as predetermined assuming a world market for gasoline. We also allow for the gasoline price to be jointly endogenous using state tax rates as instruments and also distance of the state from the Gulf of Mexico, as in Blundell, Horowitz and Pairey (2012). Here we take a control function approach where in the first stage we use the instruments z_i , along with household income, and the predetermined variables w_i . We then take the estimated residuals from this first stage $\hat{\xi}_i$ and use them as a control function in equation (6.9), constructing

$$E[q_i | \widehat{p}, \widehat{y}, w, \xi] = \sum_{j,k,\ell,m=1}^4 \tilde{\beta}_{j,k,\ell} (\ln p)^j (\ln y)^k (w' \delta)^\ell (\hat{\xi})^m, \quad (6.10)$$

where $\tilde{\beta}_{j,k,\ell,m}$ are the coefficients from the regression of q_i on log price, income, the covariates index, and the first stage residual. The average demand is then estimated as

$$\bar{q}(p, \widehat{y}, w) = \frac{1}{n} \sum_{i=1}^n \sum_{j,k,\ell,m=1}^4 \hat{\beta}_{j,k,\ell} (\ln p)^j (\ln y)^k (w' \delta)^\ell (\hat{\xi}_i)^m \quad (6.11)$$

In Figure 1 we plot the OLS estimates taking gasoline price to be predetermined. Note that it is generally downward sloping except at low prices. In Figure 2 we estimate the demand function using the control function approach and find it to be better behaved. In Table 2 we consider the estimated price elasticities for OLS and IV. We see that the estimated price elasticity has the incorrect sign for the 75th quantile for three out of the four specifications, while the IV estimates all have the correct sign. However, the IV estimates are somewhat large except perhaps for the 3rd and 4th order specification. In Table 3 we see that the estimated income elasticities for both OLS and IV are quite similar and also similar to previous estimates, e.g. Hausman and Newey (1995).

We now set bounds on the income effect. We estimate the income derivatives from a local linear quantile regression of log of gasoline demand on log price and log income. We assume that gasoline is a normal good and so choose the lower bound b to be 0.0. We set the upper bound B to be 20 times the derivative from the 0.9 quantile income effect which is equal to 0.0197 and is quite small. We assume that the previously discussed jump term from changing car type is negligible, consistent with few individuals changing car type and/or small changes in gas consumption, as income is compensated.

In Figure 3 we graph the bounds on the change in the average equivalent variation for a price increase from the stated price on the lower axis to \$1.40 per gallon. We use the estimates from the 3rd order power series, with a control function, evaluated at median income. Note that the lower bound and upper bound estimates are almost the same and it is difficult to distinguish between them. This result follows from the small income derivative and the small share of gasoline expenditure in overall household expenditure. The results demonstrate that although the welfare function is not point identified, in this type of situation the upper and lower bound estimates are very similar.

In Figure 4 we graph the bounds on deadweight loss for a price increase from the price on the lower axis to \$1.40. Again we use the 3rd order power series control function estimates evaluated at the median income. Again, the lower and upper bound estimates are very similar and difficult to distinguish except for very low gasoline prices. Since deadweight loss is a second order calculation compared to the first order calculation of equivalent variation, e.g. Hausman (1981), the closeness of the bounding estimates allows for policy evaluation, even in the absence of point identification.

We now estimate confidence sets for our estimated bounds, since both equivalent variation and deadweight loss are only set identified instead of being point identified. We construct confidence sets that cover the true identification region with probability asymptotically equal to 95%. Let the estimated set identification regions be given by $\hat{\Theta} = [\hat{\theta}_\ell, \hat{\theta}_u]$ and the true identification region be given by $\Theta_0 = [\theta_{0\ell}, \theta_{0u}]$. Let the joint asymptotic variance matrix of the bounds be Σ . It will follow from Hausman and Newey (1995) that the bounds are joint asymptotically normal. We estimate Σ by treating the model as if it were parametric and applying the delta method, as works for series estimates, as in Newey (1997). Here we use the delta method on the estimated equation (6.11) and the derivatives of the upper and lower bounds. The derivatives of the upper and lower bounds with respect to β are straightforward to calculate since the equivalent variation and deadweight loss are linear in these parameters with the derivative with respect to the coefficient β of a price term given by

$$\frac{\partial \hat{S}_B(p, y)}{\partial \beta} = e^{Bp} \int_p^{p^1} (\ln(\tilde{p}))^i e^{-B\tilde{p}} d\tilde{p}$$

We compute the other elements in the surplus expression in a similar way using numerical integration. We compute the derivatives for the deadweight loss.

$$\frac{\partial \widehat{DWL}_B}{\partial \beta} = \frac{\partial \hat{S}_B(p, y)}{\partial \beta} - (p^1 - p)(\ln(p^1))^i.$$

We then form an estimate $\hat{\Sigma}$ of the the joint asymptotic variance matrix of the upper and lower bounds in the usual way.

We use the Beresteanu and Molinari (2008) procedure to construct critical values. The results are given in Table 4 for the equivalent variation estimates with the estimate

standard errors in parenthesis and the 95% confidence intervals given in brackets. Concentrating on the 3rd order estimates which we plotted before we see that the estimated standard errors are quite small at both the lower bound and the upper bound and the 95th percentile confidence interval goes from \$13.72 to \$16.24 which is small enough for reliable policy analysis. In Table 5 we give the standard errors and bounds and the DWL estimates. Here we find that the standard errors are reasonably small but the estimate confidence intervals are sufficiently large to make their use in policy analysis problematic.

We have used our bounds approach to estimate household gasoline demand functions allowing for unrestricted heterogeneity. While the welfare measures are not point identified, using our bounds approach we find that the lower and upper bound estimates are very close to each other given the relatively small income elasticity and share of gasoline consumption in overall household expenditure.

7 Appendix

Proof of Theorem 1: Let $F(e|u) = \Pr(\varepsilon \leq e|u) = \int_{-\infty}^e f_{\varepsilon}(t|u)dt$. Then by the fundamental theorem of calculus, $F(e|\eta)$ is differentiable in e with derivative $f(e|u)$ that is continuous in e . Let $q^{-1}(x, u, y)$ denote the inverse function of $q(x, u, \varepsilon)$ as a function of ε . Then

$$\Pr(q(x, \eta_i) \leq y) = E[1(q(x, \eta_i) \leq y)] = E[E[1(\varepsilon_i \leq q^{-1}(x, u_i, y))|u_i]] = E[F(q^{-1}(x, u_i, y)|u_i)].$$

By the inverse function theorem $q^{-1}(x, u, y)$ is continuously differentiable in x and y , with

$$\partial q^{-1}(x, u, y)/\partial y = [\partial q(x, u, q^{-1}(x, u, y))/\partial \varepsilon]^{-1}, \partial q^{-1}(x, u, y)/\partial x = -\frac{\partial q(x, u, q^{-1}(x, u, y))/\partial x}{\partial q(x, u, q^{-1}(x, u, y))/\partial \varepsilon}.$$

By Assumption 2 both $\partial q^{-1}(x, u, y)/\partial y$ and $\partial q^{-1}(x, u, y)/\partial x$ are continuous in y and x and bounded. Let q_x denote the random variable $q(x, \eta)$, $f_{q_x}(y)$ be its marginal pdf at y , and $f_{q_x}(y|u)$ its conditional pdf at y given $u_i = u$. Then by the chain rule and a change of variables from ε to $y = q(x, \eta)$ conditional on u ,

$$\frac{\partial F(q^{-1}(x, u, y)|u)}{\partial y} = f(q^{-1}(x, u, y)|u)\partial q^{-1}(x, u, y)/\partial y = f_{q_x}(y|u),$$

$$\frac{\partial F(q^{-1}(x, u, y)|u)}{\partial x} = f(q^{-1}(x, u, y)|u)\partial q^{-1}(x, u, y)/\partial x = -f_{q_x}(y|u) \frac{\partial q(x, \eta)}{\partial x} \Big|_{q(x, \eta)=y}.$$

Also, these partial derivatives are all continuous and bounded in y and x . Then by standard results we can differentiate inside the integral and integrate over the distribution of u , so that $\Pr(q(x, \eta_i) \leq y)$ is continuously differentiable in x and y with

$$\frac{\partial \Pr(q(x, \eta_i) \leq y)}{\partial y} = f_{q_x}(y) > 0, \quad \frac{\partial \Pr(q(x, \eta_i) \leq y)}{\partial x} = -f_{q_x}(y) E\left[\frac{\partial q(x, \eta_i)}{\partial x} \Big| q(x, \eta_i) = y\right].$$

Note that $Q_q(\tau|x)$ is the inverse function of $\Pr(q(x, \eta_i) \leq y)$, so it then follows by the inverse function theorem that $Q_q(\tau|x)$ is continuously differentiable in x with

$$\frac{\partial Q_q(\tau|x)}{\partial x} = E\left[\frac{\partial q(x, \eta_i)}{\partial x} \Big| q(x, \eta_i) = Q_q(\tau|x)\right],$$

the conclusion from Hoderlein and Mammen (2007).

Since $q(x, \eta)$ is a demand function it satisfies downward sloping compensated demand, i.e. $\partial q(x, \eta_i)/\partial p + q(x, \eta_i)\partial q(x, \eta_i)/\partial p \leq 0$ for all η_i . Therefore, following Dette et al. (2011), we have

$$\begin{aligned} \frac{\partial Q_q(\tau|x)}{\partial p} + Q_q(\tau|x) \frac{\partial Q_q(\tau|x)}{\partial x} &= E\left[\frac{\partial q(x, \eta_i)}{\partial x} + Q_q(\tau|x) \frac{\partial q(x, \eta_i)}{\partial x} \Big| q(x, \eta_i) = Q_q(\tau|x)\right] \\ &= E\left[\frac{\partial q(x, \eta_i)}{\partial x} + q(x, \eta_i) \frac{\partial q(x, \eta_i)}{\partial x} \Big| q(x, \eta_i) = Q_q(\tau|x)\right] \leq 0. \end{aligned}$$

Therefore, $Q_q(\tau|x)$ is a demand function for each τ . Since $\Pr(q(x, \eta_i) \leq y)$ is differentiable in y with positive derivative, $Q_q(\tau|x)$ is strictly monotonic and continuous in τ for $\tau \in (0, 1)$. For $0 < \tilde{\eta} < 1$, let $\tilde{q}(x, \tilde{\eta}) = Q_q(\tilde{\eta}|x)$. It follows from above that $\tilde{q}(x, \tilde{\eta})$ satisfies the Slutsky condition and hence is a demand function. Furthermore, for \tilde{F} the uniform distribution on $(0, 1)$ it follows by standard arguments that (\tilde{q}, \tilde{F}) is observationally equivalent to (q, F) . *Q.E.D.*

Proof of Corollary 2: Let $\delta(q, F) = \dim(\eta)$. In the specification (q, F) we have $\delta(q, F)$. Let (\tilde{q}, \tilde{F}) be the specification in the conclusion of Theorem 1. Then $\delta(\tilde{q}, \tilde{F}) = 1 \neq \delta(q, F)$. *Q.E.D.*

Proof of Corollary 3: For (\tilde{q}, \tilde{F}) as in the conclusion of Theorem 1 the functions q and \tilde{q} are different, because they have different domains. *Q.E.D.*

Proof of Corollary 4: Follows by Corollary 3 and equivalent variation being a one-to-one mapping of the demand function, as shown in the body of the paper. *Q.E.D.*

Proof of Theorem 5: Consider a specification $q(x, \eta) = \eta_1 + \beta p + \eta_2 y$ where β is constant, $\eta_1 \sim U[0, 1]$, and η_2 is independent of η_1 and has a two point distribution, with $\Pr(\eta_2 = 1/3) = \Pr(\eta_2 = 2/3) = 1/2$. Throughout we impose on β , p , and y that $q(x, \eta) \geq 0$ and that the Slutsky condition $\beta + q(x, \eta)\eta_2 \leq 0$ is satisfied. With this $q(x, \eta)$ eq. (3.3) has the solution

$$S(p, y, \eta) = \left(\eta_2^{-1} q(p, y, \eta) + \beta \eta_2^{-2} \right) - \left(\eta_2^{-1} q(p_1, y, \eta) + \beta \eta_2^{-2} \right) \exp(\eta_2(p - p_1)).$$

By $q(p, y, \eta) = q(p, y, \eta_1, \eta_2)$ linear in η_1 it follows that $S(p, y, \eta)$ is linear in η_1 . Then by independence of η_1 and η_2 , integrating first over η_1 and then summing over values for η_2 gives true average equivalent variation

$$\begin{aligned} \bar{S}(p, y) &= \int S(p, y, \eta) F(d\eta) = \int S(p, y, E[\eta_1], \eta_2) F(d\eta_2) \\ &= [S(p, y, 1/2, 1/3) + S(p, y, 1/2, 2/3)]/2. \end{aligned}$$

We compare this with the average equivalent variation for the quantile demand. Let $F_U(u)$ denote the CDF for $U(0, 1)$. The CDF of $q(x, \eta)$ is given by

$$\Pr(q(x, \eta) \leq q|x) = [F_U(q - \beta p - y/3) + F_U(q - \beta p - 2y/3)]/2.$$

Define

$$Q^1(\tilde{\eta}|x) = q(p, y, 2\tilde{\eta}, 1/3), Q^2(\tilde{\eta}|x) = q(p, y, \tilde{\eta}, 1/2), Q^3(\tilde{\eta}|x) = q(p, y, 2\tilde{\eta} - 1, 2/3).$$

Then the conditional quantile function is given by

$$Q(\tilde{\eta}|x) = \begin{cases} Q^1(\tilde{\eta}|x), & 0 \leq \tilde{\eta} \leq y/6, \\ Q^2(\tilde{\eta}|x), & y/6 < \tilde{\eta} < 1 - y/6, \\ Q^3(\tilde{\eta}|x), & 1 - y/6 \leq \tilde{\eta} \leq 1. \end{cases}$$

Let $S_j(p, y, \tilde{\eta})$ denote $S(p, y, \eta)$ with $Q^j(\tilde{\eta}|x)$ replacing $q(p, y, \eta)$. Average exact consumer surplus for the quantile demand is then given by

$$\begin{aligned}\bar{S}(p, y) &= \int_0^{y/6} S_1(p, y, \tilde{\eta})d\tilde{\eta} + \int_{y/6}^{1-y/6} S_2(p, y, \tilde{\eta})d\tilde{\eta} + \int_{1-y/6}^1 S_3(p, y, \tilde{\eta})d\tilde{\eta} \\ &= (y/6)S(p, y, y/6, 1/3) + (1 - y/3)S(p, y, 1/2, 1/2) + (y/6)S(p, y, 1 - y/6, 2/3).\end{aligned}$$

We restrict the value of y and the range for p so that demand is positive and the Slutsky condition is satisfied for all η_1, η_2 in their respective supports. By the $q(x, \eta)$ monotonic in η_1 and η_2 separately, these conditions are, respectively,

$$\beta p + y/3 \geq 0, \beta + (1 + \beta p + 2/3y)(2/3) \leq 0.$$

Choosing $\beta = -1$ as a further example, these conditions become $p \leq y/3, p \geq 2y/3 - 1/2$. Choose $y = 3/4$, so the constraints are $p \leq 1/4, p \geq 0$. We compare the true average surplus with the quantile demand average surplus for $q(x, \eta) = \eta_1 + \beta p + \eta_2 y, p = .1, p_1 = .2, y = 3/4, \beta = -1, \eta_1$ distributed $U(0, 1)$, and η_2 a two point distribution with support $\{1/3, 2/3\}$ and probability of each point equal to $1/2$. For this specification the ratio of average surplus for quantile demands to true average surplus is .90279. Thus, in this example the average surplus for quantile demands is only about 90 percent of the true average surplus.

The calculations here are clearly continuous in all the variables involved, so this difference will remain for small modifications of these parameters. Thus, we have provided an example where average surplus is not identified because there are two observationally equivalent models with average surplus that differs by about 10 percent, and average surplus remains unidentified for small variations in the parameters. *Q.E.D.*

Proof of Theorem 6: Given in text. *Q.E.D.*

Proof of Theorem 7: Note first that $S(p, y, \eta)$ is positive because $q(p, y, \eta)$ is obtained as the solution to eq. (2.1). Then expanding the right-hand side of equation (3.3) around $S = 0$ it follows that

$$\frac{\partial S(p, y, \eta)}{\partial p} = -q(p, y, \eta) + \frac{\partial q(p, y - \dot{S}(p, y, \eta))}{\partial y} S(p, y, \eta) \leq -q(p, y, \eta) + BS(p, y, \eta),$$

where $0 \leq \dot{S}(p, y, \eta) \leq S(p, y, \eta)$ is an intermediate value. Let $S_B(p, y, \eta)$ be the solution to

$$\frac{\partial S_B(p, y, \eta)}{\partial p} = -q(p, y, \eta) + BS_B(p, y, \eta), S_B(p^1, y, \eta) = 0. \quad (7.12)$$

For notational convenience, suppress the y and η arguments, and let $S_B(p) = S_B(p, y, \eta)$ and $S(p) = S(p, y, \eta)$, etc.. By equation (7.12) it follows that for all $p < p^1$,

$$\frac{d}{dp}(e^{-B \cdot p} S_B(p)) = -B \cdot e^{-B \cdot p} S_B(p) + e^{-B \cdot p} \frac{dS_B(p)}{dp} = -e^{-B \cdot p} q(p).$$

Then by the above inequality for $dS(p)/dp$ we have

$$\begin{aligned} \frac{d}{dp}(e^{-B \cdot p} S(p)) &= -B \cdot e^{-B \cdot p} S(p) + e^{-B \cdot p} \frac{dS(p)}{dp} \leq -B \cdot e^{-B \cdot p} S(p) + e^{-B \cdot p} [-q(p) + BS(p)] \\ &= -e^{-B \cdot p} q(p) = \frac{d}{dp}(e^{-B \cdot p} S_B(p)). \end{aligned}$$

Consider now $G(p) = e^{-B \cdot p} [S_B(p) - S(p)]$. Note that $G(p^1) = 0$ and for $p < p^1$ we have $dG(p)/dp \geq 0$. Therefore, for $p < p^1$ we have $G(p) \leq 0$, implying that

$$S_B(p) \leq S(p).$$

It follows similarly that

$$S(p) \leq S_b(p).$$

Adding back the y and η notation, we have

$$S_B(p, y, \eta) \leq S(p, y, \eta) \leq S_b(p, y, \eta).$$

Taking expectations gives

$$E[S_B(p, y, \eta)] \leq \bar{S}(p, y) \leq E[S_b(p, y, \eta)].$$

Furthermore, it follows as in the discussion preceding Theorem 1 that

$$E[S_B(p, y, \eta)] = \bar{S}_B(p, y), E[S_b(p, y, \eta)] = \bar{S}_b(p, y),$$

giving the conclusion. *Q.E.D.*

Proof of Theorem 8: Integrating this over a uniform on $[0, 1]$, and making the change of variables $\eta = F_\eta^{-1}(\tau) = Q_\eta(\tau)$ gives

$$\int_0^1 S^\tau(p, y) d\tau = \int S^{F_\eta(\eta)}(p, y) F(d\eta).$$

By $Q_\eta(F_\eta(\eta)) = \eta$ it follows that $S^{F_\eta(\eta)}(p, y)$ is the consumer surplus for individual η , giving the second conclusion.

Similarly to the proof of Theorem 7, define $S_C^\tau(p, y)$ as the solution to

$$\frac{\partial S_C^\tau(p, y)}{\partial p} = -Q_q(\tau|p, y) + C S_C^\tau(p, y), S_C^\tau(p^1, y) = 0.$$

As in the proof of Theorem 7, the solution to this linear differential equation is $S_C^\tau(p, y) = \int_p^{p^1} Q_q(\tau|\tilde{p}, y) e^{C(p-\tilde{p})} d\tilde{p}$. Integrating this, and interchanging the order of integration, gives

$$\int_0^1 S_C^\tau(p, y) d\tau = \int_p^{p^1} \left[\int_0^1 Q_q(\tau|\tilde{p}, y) d\tau \right] e^{C(p-\tilde{p})} d\tilde{p} = \bar{S}_C(p, y),$$

where the second equality follows by $E[q|p, y] = \int_0^1 Q_q(\tau|p, y) d\tau$.

Next, as in the proof of Theorem 7, an expansion of the differential equation for $S^\tau(y, p)$ gives

$$\frac{\partial S^\tau(p, y)}{\partial p} = -Q_q(\tau|p, y) + \frac{\partial Q_q(\tau|p, y - \ddot{S}^\tau(p, y))}{\partial y} S^\tau(p, y),$$

where $\ddot{S}^\tau(p, y)$ is an intermediate value. Furthermore, by Hoderlein and Mammen (2007) and $\partial q(p, y, \eta)/\partial y \leq B$ for all p, y , and η ,

$$\frac{\partial Q_q(\tau|p, y - \ddot{S}^\tau(p, y))}{\partial y} = E\left[\frac{\partial q}{\partial y}(p, y - \ddot{S}^\tau(p, y), \eta) | q(p, y, \eta) = Q_q(\tau|p, y)\right] \leq B.$$

Plugging this in, it follows that

$$\frac{\partial S^\tau(p, y)}{\partial p} \leq -Q_q(\tau|p, y) + B S^\tau(p, y).$$

The right hand side of this equation is identical with that for $S_B^\tau(p, y)$, so by a similar argument as in the proof of Theorem 7, we have $S^\tau(p, y) \geq S_B^\tau(p, y)$. It follows analogously that $S^\tau(p, y) \leq S_b^\tau(p, y)$, so that

$$S_B^\tau(p, y) \leq S^\tau(p, y) \leq S_b^\tau(p, y).$$

Integrating both sides then gives

$$\bar{S}_B(p, y) = \int_0^1 S_B^\tau(p, y) d\tau \leq \int_0^1 S^\tau(p, y) d\tau \leq \int_0^1 S_b^\tau(p, y) d\tau = \bar{S}_b(p, y).$$

Thus we find that, under the general heterogeneity conditions of Theorem 1, the integral of the surplus for the τ^{th} quantile lies between the bounds for average surplus. *Q.E.D.*

Before proving Theorem 9 we give a useful Lemma. For notational convenience suppress the η argument and let q be a vector that includes a and p be the absolute price vector. Let $e_j(p, u) = \min\{p'q : U(q, j) \geq u\}$ be the expenditure function for the utility function $U(q, j)$, ($j = 1, \dots, J$)'.

LEMMA A1: *If the Assumptions of Theorem 9 are satisfied then $e(p, u) = \min_j\{e_j(p, u) + r_j\}$.*

Proof: Define $\bar{e}(p, u) = \min_j\{e_j(p, u) + r_j\}$. By the definition of $\bar{e}(p, u)$ it follows that $\bar{e}(p, u) = e_{j^*}(p, u) + r_{j^*}$ for some j^* that need not be unique. By the definition of $e_{j^*}(p, u)$ and standard results there is q^* such that $U(q^*, j^*) \geq u$ and $p'q^* = e_{j^*}(p, u)$, so $p'q^* + r_{j^*} = \bar{e}(p, u)$. Since $U(q^*, j^*) \geq u$ and $p'q^* + r_{j^*} \leq \bar{e}(p, u)$, it follows that

$$\max_{q, j} \{U(q, j) \text{ s.t. } p'q + r_j \leq \bar{e}(p, u)\} \geq U(q^*, j^*) \geq u.$$

It follows that $e(p, u) \leq \bar{e}(p, u)$. Next, consider any $\bar{y} < \bar{e}(p, u)$. Then by the definition of $\bar{e}(p, u)$ we have $\bar{y} - r_j < e_j(p, u)$ for all $j \in \{1, \dots, J\}$. Since $e_j(p, u)$ is the expenditure function it follows that $\max_q \{U(q, j) \text{ s.t. } p'q \leq \bar{y} - r_j\} < u$ for every j , and so $\max_q \{U(q, j) \text{ s.t. } p'q \leq \bar{y} - r_j\} < u$. It follows that $\bar{y} < e(p, u)$. Since this is true for every $\bar{y} < \bar{e}(p, u)$ it follows that $\bar{e}(p, u) = e(p, u)$. *Q.E.D.*

Proof of Theorem 9: By definition, $e(p, u^1) = y - S(p, y)$ and $V(p, e(p, u)) \geq u$. Then by definition of V , Lemma A1, and $V_k(p, e_k(p, u)) = u$ for every k , there is j such that

$$V(p, e(p, u)) = V_j(p, e(p, u) - r_j) \leq V_j(p, e_j(p, u)) = u$$

Therefore we have $V(p, e(p, u)) = u$. Similarly we have $e(p, V(p, y)) \leq y$ by the definitions and there is j such that

$$e(p, V(p, y)) = e_j(p, V(p, y)) + r_j \geq e_j(p, V_j(p, y - r_j)) + r_j = y,$$

so that $e(p, V(p, y)) = y$.

Now consider j such that $V(p, e(p, u^1)) = V_j(p, e(p, u^1) - r_j)$. For any $k \neq j$ it follows by duality that $V_k(p, e_k(p, u^1)) = u^1$. Therefore, we have

$$V_k(p, e_k(p, u^1)) = u^1 = V(p, e(p, u^1)) = V_j(p, e(p, u^1) - r_j) > V_k(p, e(p, u^1) - r_k).$$

By $V_k(p, y)$ monotonically increasing in y , it follows that $e_k(p, u^1) > e(p, u^1) - r_k$. Since this is true for any $k \neq j$ we have

$$e_k(p, u^1) + r_k > e(p, u^1) = e_j(p, u^1) + r_j.$$

Since each $e_k(p, u^1)$ is continuous in p , this inequality continues to hold in a neighborhood of p . Therefore, by $e_j(p, u^1)$ differentiable and by Shephard's lemma, on that neighborhood $S(p, y) = y - e(p, u^1)$ is differentiable and

$$\begin{aligned} -\frac{\partial S(p, y)}{\partial p} &= \frac{\partial e(p, u^1)}{\partial p} = \frac{\partial e_j(p, u^1)}{\partial p} = h_j(p, u^1) = q_j(p, e_j(p, u^1)) \\ &= q_j(p, e(p, u^1) - r_j) = q_j(p, y - S(p, y) - r_j) = q(p, y - S(p, y)), \end{aligned}$$

where the last equality follows by the $q(p, y) = q_j(p, y - r_j)$ when $V_j(p, y - r_j) > V_k(p, y - r_k)$ for all $k \neq j$. *Q.E.D.*

Proof of Theorem 10: For notational convenience, suppress the η argument. Note first that $0 \leq S(p, y) \leq S(p^0, y)$ by $p \geq p^0$. Then by Assumption 3,

$$\begin{aligned} \frac{\partial S(p, y)}{\partial p} &= -\dot{q}(p, y - S(p, y)) + d(p, y - S(p, y)) \leq -\dot{q}(p, y, \eta) + b \cdot S(p, y) + D \\ &= -q(p, y) + d(p, y) + b \cdot S(p, y) + D \leq -q(p, y) + 2D + b \cdot S(p, y). \end{aligned}$$

Then the first conclusion follows as in the proof of Theorem 7 while the second conclusion follows similarly. *Q.E.D.*

8 References

Beresteanu, A. and F. Molinari (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica* 76, 763-814.

Blomquist, S. and W.K. Newey (2002): "Nonparametric Estimation with Nonlinear Budget Sets," *Econometrica* 70, 2455-2480.

Blomquist, S. and W.K. Newey (2011): "Individual Heterogeneity and Taxable Income," working paper, MIT.

Blundell, R. and J.L. Powell (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," in M. Dewatripont, L.P. Hansen and S.J. Turnovsky, eds. *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress, Vol. II*, Cambridge: Cambridge University Press.

Blundell, R., D. Kristensen, and R. Matzkin (2011): "Bounding Quantile Demand Functions Using Revealed Preference Inequalities," CEMMAP working paper 21/11.

Blundell, R., J. Horowitz, and M. Parey (2012): "Measuring the Price Responsiveness of Gasoline Demand: Economic Shape Restrictions and Nonparametric Demand Estimation," *Quantitative Economics* 3, 29-51.

Burtless, G. and J.A. Hausman (1978): "The Effect of Taxation on Labor Supply: Evaluating the Gary Negative Income Tax Experiments," *Journal of Political Economy* 86, 1103-1130.

Dette, H., S. Hoderlein, and N. Neumeyer (2011): "Testing Multivariate Economic Restrictions Using Quantiles: The Example of Slutsky Negative Semidefiniteness," CEMMAP Working Paper CWP14/11.

Dubin, J.A. and D.L. McFadden (1984): "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," *Econometrica* 52, 345-362.

Gorman, W. M. (1961): "On a Class of Preference Fields," *Metroeconomica* 13, 53-56.

Hausman, J.A. (1985): "The Econometrics of Nonlinear Budget Sets," *Econometrica* 53: 1255-1282.

Hausman, J.A. (1981): "Exact Consumer Surplus and Deadweight Loss," *American*

Economic Review 71: 662-676.

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

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

Hoderlein, S. and J. Stoye (2009): "Revealed Preference in a Heterogeneous Population," Discussion paper, Boston College.

Hoderlein, S. and A. Vanhems (2010): "Welfare Analysis Using Nonseparable Models," working paper.

Hotelling, H. (1938): "The General Welfare in Relation to Problems of Taxation and of Rail-way and Utility Rates," *Econometrica* 6, 242-269.

Hsiao, C. (1983): "Identification," in the *Handbook of Econometrics, Vol. I*, ed. by Z. Griliches and M. Intriligator, 223-283, Amsterdam: North-Holland.

Hurwicz, J. and H. Uzawa (1971): "On the Integrability of Demand Functions," in Chipman, J. ed., *Preferences, Utility and Demand*, New York: Harcourt Brace Jovanavich.

Kitamura, Y. and J. Stoye (2012) "Nonparametric Analysis of Random Utility Models," Discussion paper, Cornell University.

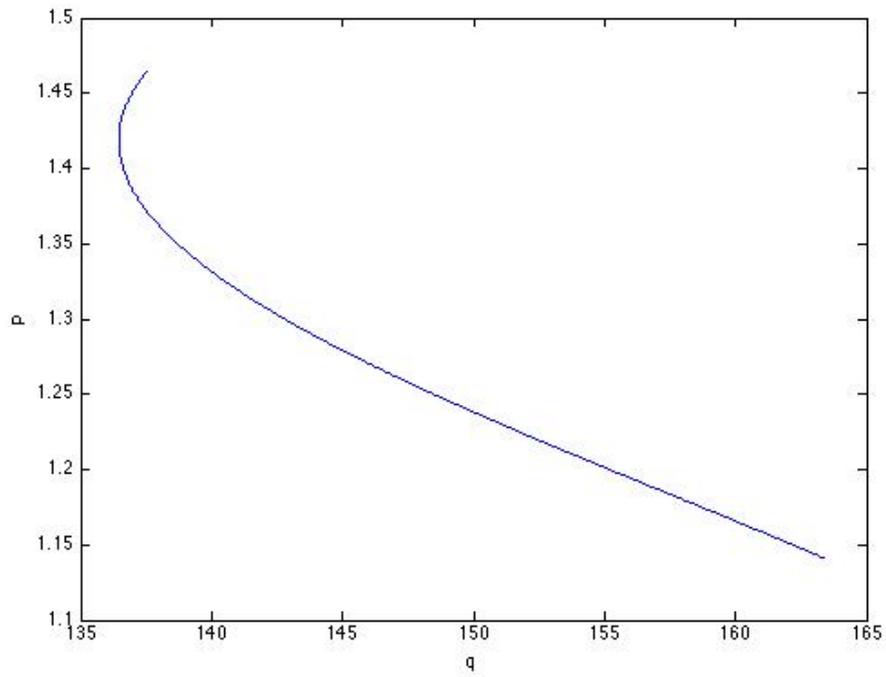
Lewbel, A. (2001): "Demand Systems With and Without Errors," *American Economic Review* 91, 611-618.

Lewbel, A. and K. Pendakur (2013): "Unobserved Preference Heterogeneity in Demand Using Generalized Random Coefficients," working paper.

McFadden, D., and M. Richter (1991): "Stochastic Rationality and Revealed Stochastic Preference," in J. Chipman, D. McFadden, and M. Richter (eds.) *Preferences, Uncertainty and Optimality: Essays in Honour of Leonid Hurwicz*. Boulder, Co.: Westview Press.

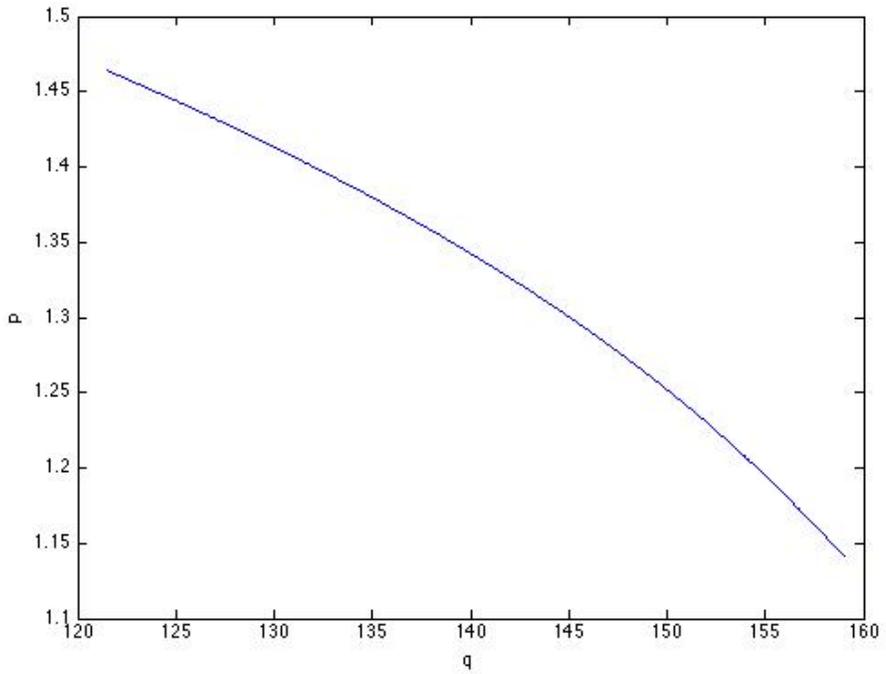
Newey, W.K. (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics* 79, 147-168.

Figure 1. Estimated Demand: OLS



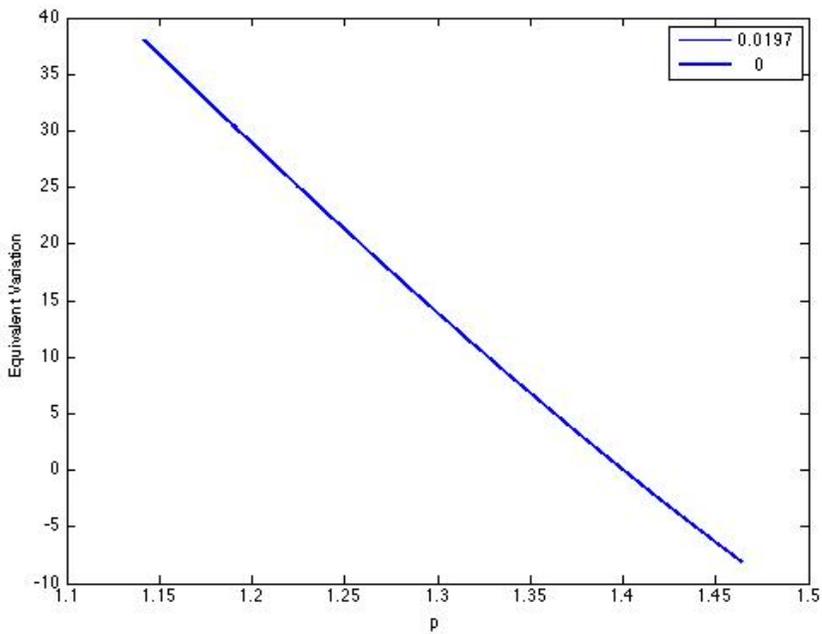
Notes: Demand estimated from 3rd order series regression evaluated at median income.

Figure 2. Estimated Demand: Control Function



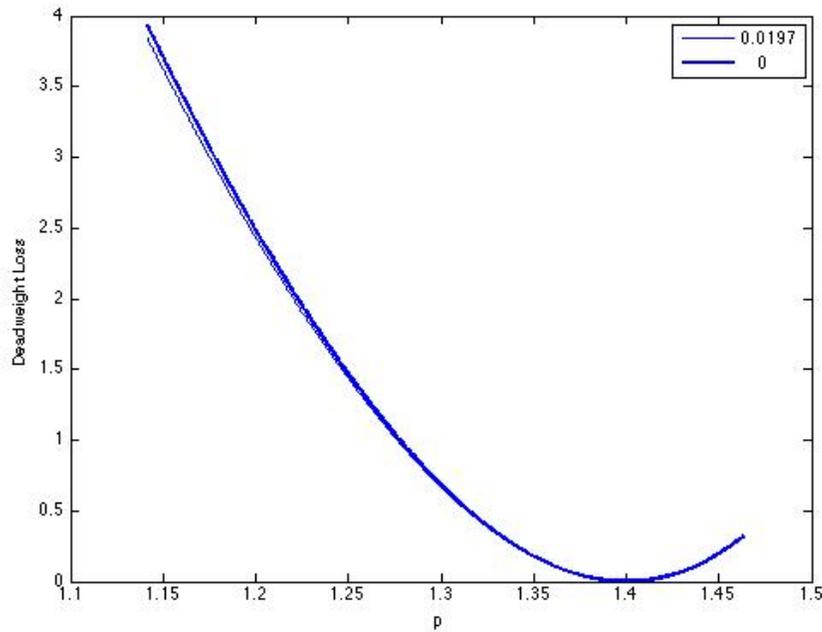
Notes: Demand estimated from 3rd order power series control function regression evaluated at median income.

Figure 3. Equivalent Variation Bounds



Notes: Graph shows change in equivalent variation for a price increase from p to \$1.40, evaluated at upper and lower bounds of income derivative and at median income and estimated from 3rd order power series control function estimated demand.

Figure 4. Deadweight Loss Bounds



Notes: Graph shows change in deadweight loss for a price increase from p to \$1.40, evaluated at upper and lower bounds of income derivative and at median income and estimated from 3rd order power series control function estimated demand.

Table 1. Summary Statistics

Variable	Mean	Median	Std Dev	Min	Max
price (\$)	1.33	1.32	0.08	1.14	1.46
quantity (gallons)	4.90	2.65	7.53	0.01	195.52
income (1,000 \$)	62.19	47.5	47.47	2.08	170.72
number of drivers	2.04	2	0.78	1	7
public transit availability	0.24	0	0.42	0	1
Observations	8,908				

Table 2. Estimated Price Elasticities

Quantiles	OLS Estimates			Control Function Estimates		
	0.25	0.5	0.75	0.25	0.5	0.75
Order 1	-0.698 (0.254)	-0.656 (0.244)	-0.631 (0.243)	-1.111 (0.282)	-1.060 (0.280)	-1.043 (0.291)
Order 2	-1.597 (0.469)	-0.798 (0.283)	0.069 (0.509)	-1.675 (0.476)	-1.350 (0.320)	-1.111 (0.657)
Order 3	-1.214 (0.753)	-0.798 (0.570)	0.271 (0.721)	-1.037 (0.860)	-1.102 (0.670)	-0.872 (0.952)
Order 4	-0.713 (0.877)	-0.583 (0.623)	0.140 (0.801)	-0.389 (1.158)	-0.588 (0.707)	-0.853 (1.032)

Table 3. Estimated Income Elasticities

Quantiles	OLS Estimates			Control Function Estimates		
	0.25	0.5	0.75	0.25	0.5	0.75
Order 1	0.168 (0.022)	0.157 (0.019)	0.151 (0.017)	0.167 (0.022)	0.159 (0.019)	0.157 (0.018)
Order 2	0.221 (0.032)	0.244 (0.025)	0.261 (0.031)	0.210 (0.032)	0.233 (0.025)	0.262 (0.034)
Order 3	0.217 (0.057)	0.221 (0.040)	0.236 (0.039)	0.220 (0.059)	0.221 (0.041)	0.256 (0.043)
Order 4	0.266 (0.067)	0.305 (0.054)	0.275 (0.047)	0.267 (0.074)	0.294 (0.058)	0.277 (0.052)

Table 4. Bounds on Equivalent Variation Estimates

	From \$1.20 to 1.30		From \$1.20 to 1.40	
	Lower Bound	Upper Bound	Lower Bound	Upper Bound
Order 1	16.777	16.794	32.281	32.343
	0.349	0.349	0.502	0.501
	[16.104, 17.468]		[31.355, 33.270]	
Order 2	15.147		28.829	28.884
	0.443	0.443	0.753	0.751
	[14.275, 16.005]		[27.405, 30.309]	
Order 3	14.972	14.987	28.845	28.900
	0.647	0.646	0.884	0.882
	[13.715, 16.244]		[27.163, 30.583]	
Order 4	14.625	14.639	28.546	28.601
	0.660	0.659	0.924	0.922
	[13.340, 15.925]		[26.788, 30.360]	

Table 5. Bounds on Deadweight Loss Estimates

	From \$1.20 to 1.30		From \$1.20 to 1.40	
	Lower Bound	Upper Bound	Lower Bound	Upper Bound
Order 1	0.646	0.663	2.467	2.529
	0.175	0.175	0.669	0.668
	[0.319, 0.990]		[1.219, 3.776]	
Order 2	0.735		2.821	2.876
	0.233	0.233	0.728	0.727
	[0.277, 1.178]		[1.452, 4.245]	
Order 3	0.485	0.500	2.434	2.489
	0.393	0.393	1.04	1.039
	[-0.272, 1.257]		[0.447, 4.475]	
Order 4	0.321	0.335	1.827	1.882
	0.498	0.497	1.137	1.135
	[-0.641, 1.298]		[-0.343, 4.052]	