Nonparametric Welfare Analysis

Citation: Hausman, Jerry A., and Whitney K. Newey. "Nonparametric Welfare Analysis." Annual Review of Economics, vol. 9, no. 1, Aug. 2017, pp. 521–46.

As Published: http://dx.doi.org/10.1146/ANNUREV-ECONOMICS-080315-015107

Publisher: Annual Reviews

Persistent URL: http://hdl.handle.net/1721.1/114158

Version: Original manuscript: author's manuscript prior to formal peer review

Nonparametric Welfare Analysis*

Jerry A. Hausman
MIT

Whitney K. Newey
MIT

August 2016

Abstract

Exact consumers surplus and deadweight loss are the most widely used welfare and economic efficiency measures. These measures can be computed from demand functions in straightforward ways. Nonparametric estimation can be used to estimate the welfare measures. In doing so it seems important to account correctly for unobserved heteroeneity given the high degree of unexplained demand variation often found in applications. This paper surveys work on nonparametric welfare analysis, focusing on that which allows for general heterogeneity in demand as in Hausman and Newey (2015).

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. R. Blundell and Y. Muhkin provided helpful comments. Capable research assistance was provided by M. Hardy.

1 Introduction

Exact consumers surplus and deadweight loss are the most widely used welfare and economic efficiency measures in areas of economics such as public finance. These measures can be computed from demand functions in straightforward ways. This makes welfare measures very useful for applications where quantities, prices, incomes are available. It is now possible to use nonparametric or semiparametric estimators of demand functions to estimate welfare measures, thus avoiding functional form restrictions that used to be common in empirical demand analysis.

It seems important to account for individual heterogeneity in the estimation of welfare measures. Often r-squareds are found to be low in cross-section and panel demand data, leaving open the possibility that much variation in demand is due to unobserved heterogeneity. The potential magnitude of heterogeneity suggests that allowing for heterogeneity in applications could have a significant impact.

This paper reviews the work on nonparametric welfare analysis, focusing on recent work that takes explicit account of unobserved heterogeneity. Most of our attention is given to demand models with general, multi-dimensional heterogeneity as considered in Hausman and Newey (2016). These models allow demand functions to vary across individuals in general ways. For example, it seems reasonable to suppose that price and income effects 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. Also whether demand depends monotonically on a scalar heterogeneity term is not identified from single or repeated cross-section data, as shown by Hausman and Newey (2016). Welfare measures are sensitive to assumptions about heterogeneity, motivating the focus on general heterogeneity in this paper. We also discuss recent work on welfare and demand analysis with restrictive forms of heterogeneity.

Unobserved heterogeneity in demand means that surplus will vary over individuals in unobserved ways. That means we can at best hope to learn something about the distribution of welfare effects in the population. The expected value of surplus across individuals is a common welfare measure. The distribution of surplus may also be of interest. Hausman and Newey (2016) showed that for continuous demand average surplus is generally not identified and hence neither is the distribution of surplus. Nonidentification motivates the bounds approach in Hausman and Newey (2016). They show that known bounds on income effects can be used to construct bounds on average surplus in a straightforward way. Those bounds only require the expected value of the demand function across individuals. With two goods they also show how to construct bounds on average surplus based only on utility maximization, i.e. that do not require known bounds on income effects. The approach to general bounds with two goods should also be extendable to multiple goods.

We emphasize that the bounds average over unobserved hetereogeneity while holding income and observable characteristics fixed. The variation of bounds over income and observable characteristics can be used to evaluate policy impacts on different groups of individuals. Comparisons across groups with observable differences is often relevant for welfare analysis. In this paper we compare surplus for gasoline demand across different income levels using the same data as Hausman and Newey (2016). We find that deadweight loss is quite flat as a function of income though there is some evidence that deadweight loss is largest at smaller income values and tends to decline with income. We find that the equivalent variation tends to increase with income, and hence so do gasoline taxes.

To apply these result to estimate welfare effects from data something must be assumed about how individual heterogeneity varies with prices and incomes. Most of the empirical application of nonparametric welfare analysis is based on independence of preferences and budget sets, possibly conditioned on covariates and control functions. Under independence, average demand is the conditional expectation of quantity, which can be estimated by nonparametric, or semiparametric methods in cross section data while allowing for measurement error in quantity demanded. The distribution of demand can be also estimated in analogous ways, though without allowance for measurement error.

These estimates can be used to estimate surplus bounds.

Independence of preferences and the budget set, possibly conditioned on covariates and control functions, is an essential assumption with multi dimensional heterogeneity at this point in the development of nonparametric welfare analysis. Without independence it is not known how to do nonparametric welfare analysis with multi-dimensional heterogeneity. Independence can be tested by comparing results with and without the use of control functions. This was done informally in Hausman and Newey (2016) where it was found that using a control function to control for price endogeneity did not have much effect. A formal Hausman test could also be constructed by statistical comparison of bounds with and without a control function. When demand is monotonic in scalar heterogeneity and there is an instrument independent of heterogeneity, nonparametric instrumental variable estimation as in Chernozhukov, Imbens, and Newey (2007) could be used to estimate the demand function. Blundell, Horowitz, and Parey (2016) take this approach while imposing the Slutzky condition on the demand function.

While Hausman and Newey (2016) find a non-point identification result for surplus with continuous demand, Bhattacharya (2015) demonstrates point identification in the situation of discrete choice. The result follows because in the special case of discrete choice the change in the Marshallian (uncompensated) consumer surplus equals the average Hicksian (compensated) equivalent variation even if income effects are not constant. Bhattacharya (2015) demonstrates that his result does not hold for ordered discrete choice where the price remains constant over units. Thus, his results demonstrate how the special situation of purchasing a single unit of a good allows for point identification which does not hold if an individual chooses the number of units to purchase.

Turning now to an account of the literature, Hausman (1981) showed how surplus could be obtained from the demand function and used that insight to solve for surplus for some widely used parametric specifications. Vartia (1983) gave an ordinary differential equation for surplus in terms of the demand function and proposed some algorithms for solving it numerically. Hausman and Newey (1995) suggested solving that equation using demand estimators based on nonparametric regression and gave asymptotic inference results, including functional expansions, for series and kernel estimators. Vanhems (2006) gives further results on asymptotic properties of kernel estimators.

With heterogeneity in demand one could ask what is being estimated by a nonparametric regression like that in Hausman and Newey (1995). The regression will give a demand function for a particular consumer under some restrictive conditions discussed in Hausman and Newey (1995) and reviewed below. However, the literature also explored implications of more general forms of heterogeneity. Brown and Walker (1989) showed that the residual generally is heteroskedastic, which is important for inference. Lewbel (2001) explored the properties of the conditional expectation of demand with general heterogeneity. He found that the interpretation of elasticity calculation as applying to the average consumer holds true. However, the interpretation does not hold for utility and welfare measures unless the unobserved heterogeneity does not affect the marginal utility of income. Within the context of a random coefficients specification (and more generally), it will not be the case that this restriction holds. Thus, as in Hausman and Newey (2016), following from Gorman (1961), only under special conditions on how income enters demand functions in a quasi-homothetic manner and its interaction with heterogeneity, will estimated demand function be interpretable as arising from utility maximization, which allows for exact welfare measures to be calculated. However, the necessary restriction for this outcome are inconsistent with typical finding in applied econometric estimation. These results suggest the potential importance of allowing for heterogeneity in nonparametric welfare analysis and in choice analysis more generally.

The previous literature considered some forms of heterogeneity. In their analysis of labor supply with nonlinear taxes, Burtless and Hausman (1978) allowed the income effect to vary over individuals. Blomquist and Newey (2002) allowed for nonparametric, scalar, and monotonic unobserved heterogeneity with nonlinear taxes and their results turn out to be valid with general heterogeneity, see Blomquist, Kumar, Liang, Newey (2016). More recently Dette, Hoderlein, and Neumeyer (2016) showed that with general heterogeneity the quantile of any linear combination of goods must satisfy a Slutzky condition, implying that the quantile of any one good is a demand function as only the price of that good varies.

Specific kinds of nonparametric heterogeneity have recently been considered for welfare analysis. Blundell, Horowitz, and Parey (2016) and Hoderlein and Vanhems (2013) consider two goods and demand that is monotonic in scalar heterogeneity. Lewbel and Pendakur (2016) have considered a random coefficients demand model, with some restrictions on the distribution of coefficients that make it relatively easy to estimate.

Another strand of the literature is about revealed stochastic preference. This work can be thought of as demand analysis with unrestricted heterogeneity and possibly multivalued demand. McFadden (2005) characterized the restrictions on the distribution of demand implied by those models. Hoderlein and Stoye (2014) showed how to bound the proportion of consumers that satisfy the weak axiom of revealed preference. Kitamura and Stoye (2012) gave tests of the revealed stochastic preference restrictions. For two goods Blundell, Kristensen, and Matzkin (2014) developed methods for predicting the distribution of demand outside the range of the data while imposing the revealed stochastic preference restrictions, as we explain below. Quite recently, Cosaert and Demuynck (2014) derived bounds on surplus and demand prediction using the weak axiom of revealed preference while allowing for general heterogeneity.

In Section 2 of the paper we begin by discussing consumer surplus and deadweight loss for a single consumer. Section 3 gives an account of nonparametric estimation like that considered in Hausman and Newey (1995). Section 4 introduces general heterogeneity in demand. Section 5 discusses some useful results for the case of two goods. Section 6 reviews work on identification of demand and surplus with general heterogeneity. Section 7 explains how bounds for average surplus can be constructed using bounds on income effects. Section 8 describes recent work on bounds with two goods that does not require bounds on income effects. Section 9 outlines the assumptions that are important for applying the theoretical results to data. Section 10 reviews existing work with two goods and/or restricted forms of heterogeneity. Section 11 gives an empirical application to gasoline demand, showing how surplus and deadweight loss bounds vary with income level. Section 12 offers some conclusions.

2 Welfare Analysis for a Single Consumer

We will first review welfare analysis for a single consumer and then consider heterogeneity. We begin with describing choice for a consumer. Let q denote the quantity of a vector of goods, a the quantity of a numeraire good, p the price vector for q relative to a, and y the individual income level relative to the numeraire price. Also let $x = (p^T, y)^T$, where throughout we adopt the notational convention that vectors are column vectors. The demand function $q(x)$ will denote the consumer's choice of q for given prices and income x. In what follows we will be assuming that we observe the choices of individuals for given prices and incomes, so we focus on demand functions as the empirically relevant object. We follow much of the existing welfare analysis literature in this approach.

Demand $q(x)$ will result from maximizing a utility function $U(q, a)$ that is monotonic increasing in q and a and strictly increasing in at least one argument, subject to the budget constraint that expenditure on goods cannot exceed income. That is

$$q(x) = \arg \max_{q \geq 0, a \geq 0} U(q, a) \text{ s.t. } p^T q + a \leq y.$$

We will assume throughout that demand is single valued. Strict quasi-concavity of the utility function is sufficient for single valued demand and is necessary over all positive prices and incomes.

Utility maximization imposes restrictions on the demand function as a function of prices and income. Assuming that the demand function is continuously differentiable and restricting attention to positive prices and incomes utility maximization implies that

$$\partial q(x)/\partial p + q(x)[\partial q(x)/\partial y]^T \textit{ is symmetric and negative semi-definite,} \qquad (2.1)$$

where we adopt the usual Jacobian notation with the ij^{th} element of $\partial q(x)/\partial p$ being $\partial q_i(x)/\partial p_j$. By Hurwicz and Uzawa (1971), this condition and $p^T q(x) + a = y$ are also sufficient for existence of a utility function, with $q(x)$ maximizing the utility function subject to the budget constraint. In this sense, formulating a model with demand functions satisfying the Slutzky symmetry and negative definiteness condition in equation (2.1) is equivalent to formulating a model based on utility maximization. In what follows we work with demand functions satisfying the Slutzky condition as the primitive underlying economic objects that are empirically relevant.

Our goal is to quantify the welfare effects of price changes. We follow standard practice in using the expenditure function to do so. The expenditure function is given by

$$e(p,u)=\min_{q\geq 0,a\geq 0}\{p^Tq+a,\ s.t.\ U(q,a)\geq u\}.$$

Changes in the expenditure function as price changes measure the welfare effects of price changes. Let p^0 denote a vector of initial prices and p^1 a vector of final prices. Also let

$$u^j=\max_{q\geq 0,a\geq 0}U(q,a)\text{ s.t. }(p^j)^Tq+a\leq y,\ j=0,1,$$

denote the utility at prices p^0 and p^1. A money metric for the change in utility between prices p^0 and p^1 is given by the equivalent variation

$$EV(p^0,p^1,y)=e(p^0,u^0)-e(p^0,u^1)=y-e(p^0,u^1).$$

Also equivalent variation can be viewed as the effect on expenditure of varying prices while holding utility fixed at u^1, since

$$EV(p^0,p^1,y)=y-e(p^0,u^1)=e(p^1,u^1)-e(p^0,u^1).$$

The corresponding deadweight loss measure is the equivalent variation minus the tax receipts,

$$DWL(p^0,p^1,y)=y-e(p^0,u^1)-\Delta p^Tq(p^1,y),\ \Delta p=p^1-p^0.$$

Compensating variation can also be used as a money measure of the welfare effects of price changes. We focus on equivalent variation because it can be used to compare welfare across different price changes while compensating variation cannot and because deadweight loss is more complicated for compensating variation.

The key to empirical welfare analysis is that equivalent variation and deadweight loss can be computed from the demand functions, allowing us to estimate welfare measures from individuals' observed choices. Let $\{p(t)\}_{t=0}^1$ be a continuously differentiable price path from $p(0) = p^0$ to $p(1) = p^1$. Let

$$s(t) = y - e(p(t), u^1) \tag{2.2}$$

be the equivalent variation for a price change from $p(t)$ to p^1. Let $h(p, u)$ denote the compensated demand, i.e. $h(p, u) = \arg\min_q p^T q + a$ s.t. $U(q, a) \geq u$. Differentiating equation (2.2) with respect to t, applying the chain rule and Shephard's Lemma, and using $h(p, u) = q(p, e(p, u))$,

$$\begin{aligned} \frac{ds(t)}{dt} &= -\frac{de(p(t), u^1)}{dt} = -\frac{\partial e(p(t), u^1)}{\partial p}^T \frac{dp(t)}{dt} = -h(p(t), u^1)^T \frac{dp(t)}{dt} \\ &= -q(p(t), e(p(t), u^1))^T \frac{dp(t)}{dt} = -q(p(t), y - s(t))^T \frac{dp(t)}{dt}, \end{aligned} \tag{2.3}$$

where the last equality follows by the definition of $s(t)$. Notice how $s(t)$ compensates income so that the individual remains on the same indifference curve as t varies. The resulting equation is an ordinary, nonlinear differential equation with final condition $s(1) = 0$; see Vartia (1983). The solution $s(0)$ to this differential equation at $t = 0$ is the equivalent variation for a price change from p^0 to p^1.

The solution to this differential equation will not depend on the particular path $\{p(t)\}_{t=0}^1$ as long as the matrix $\partial q(x)/\partial p + q(x)[\partial q(x)/\partial y]^T$ is symmetric. By the integrability arguments in Hurwicz and Uzawa (1971) this symmetry condition implies that there exists a function $\tilde{e}(p)$ such that

$$-q(p(t), y - s(t))^T \frac{dp(t)}{dt} = \frac{d\tilde{e}(p(t))}{dt}.$$

It then follows from $p(0) = p^0$ and $p(1) = p^1$ that $s(0) = \tilde{e}(p^1) - \tilde{e}(p^0)$, which does not depend on the path. Given this invariance of equivalent variation to the price path we are free to pick whatever path is convenient. One convenient path is the convex combination of p^0 and p^1, where $p(t) = tp^1 + (1-t)p^0 = p^0 + t\Delta p$ for $\Delta p = p^1 - p^0$. The key differential equation then becomes

$$\frac{ds(t)}{dt} = -q(p^0 + t\Delta p, y - s(t))^T \Delta p, \quad s(1) = 0. \tag{2.4}$$

Computation of equivalent variation from this differential equation is straightforward. Given a demand function $q(p, y)$ any of a variety of numerical methods can be used.

Vartia (1983) discussed some methods. There are many fast methods that are available in modern software packages.

When only the price of one good is changing the surplus will depend only on the demand function of that good. To see this suppose that only the price p_1 of the first good is changing and that the prices of the other goods are fixed at $\bar{p}_2$. Then we can take the price path to be $p(t) = (p_1^0 + t\Delta p_1, \bar{p}_2)$, so that $\Delta p = (\Delta p_1, 0^T)^T$. The differential equation then becomes

$$\frac{ds(t)}{dt} = -q_1(p_1^0 + t\Delta p_1, \bar{p}_2, y - s(t))\Delta p_1, \quad s(1) = 0, \tag{2.5}$$

which depends only on the demand function for the first good. This result can be thought of as an implication of path independence of the surplus, and hence of symmetry, that allows us to pick a path where the differential equation only depends on the first good. It was shown by Hausman (1981).

The differential equation becomes linear, with an explicit solution, when there are constant income effects over the range of $(p(t)^T, y - s(t))^T$. In fact if only the price of one good is changing then there is an explicit solution when only the income effect for that good is constant. Suppose that $\partial q_1(p_1^0 + t\Delta p_1, \bar{p}_2, y - s(t))/\partial y = \lambda$ for each $t \in [0,1]$. Then the differential equation becomes

$$\frac{ds(t)}{dt} = -q_1(p_1^0 + t\Delta p_1, \bar{p}_2, y)\Delta p_1 + \lambda \Delta p_1 s(t), \quad s(1) = 0. \tag{2.6}$$

This linear differential equation has an explicit solution

$$s(0) = \Delta p_1 \int_0^1 q_1(p_1^0 + t\Delta p_1, \bar{p}_2, y) \exp(-t\lambda\Delta p_1) dt = \int_{p_1^0}^{p_1^1} q_1(p_1, \bar{p}_2, \bar{y}, \eta) \exp(-\lambda(p_1 - p_1^0)) dp_1.$$

We discuss this result because it is useful for identification and bounds when there is unobserved heterogeneity. Beyond that it may not have much interest because constant income effects seems a strong assumption for practice and the numerical calculation of surplus from equation (2.4) is straightforward without an explicit solution.

Marshallian surplus solves equation (2.5) while replacing $s(t)$ in the demand function with zero, i.e. while not compensating income to remain on the same indifference curve.

Marshallian surplus s_M is given by

$$s_M = \int_0^1 q(p^0 + t\Delta p, y)^T \Delta p dt.$$

This surplus measure ignores income effects. Ignoring income effects can lead to a poor approximation to deadweight loss, see Hausman (1981). For this reason we focus on exact surplus in our analysis, though as known from Willig (1976), Marshallian surplus provides a useful upper bound for equivalent variation for a normal good.

3 Nonparametric Estimation for a Single Demand Function

Nonparametric estimation of surplus for a single demand function is straightforward. The idea is to replace the true demand function with a nonparametric estimator and then solve the differential equation for surplus numerically. Let $\hat{q}(x)$ be a nonparametric estimator of the demand function. Plugging this estimator in the differential equation (2.5) leads to an estimator $\hat{s} = \hat{s}(0)$ obtained as the solution at $t = 0$ to

$$\frac{d\hat{s}(t)}{dt} = -\hat{q}(p^0 + t\Delta p, y - s(t))^T \Delta p, \quad \hat{s}(1) = 0.$$

The estimators of equivalent variation and deadweight loss are then given by

$$\widehat{EV} = \hat{s}, \widehat{DWL} = \hat{s} - \hat{q}(p^1, y)^T \Delta p.$$

Computation of $\hat{s}$ is straightforward using a variety of ordinary differential equation solvers found in various computer packages. That computation will just require calculating the nonparametric estimator $\hat{q}(p, y)$ at various values of p and y. That calculation is simple to do using a variety of possible nonparametric estimators of the demand function, such as series, locally linear, or kernel estimators. One could also use partially linear or index specifications that allow for covariates, and then fix those covariates at specific values when calculating $\hat{q}(p, y)$. Allowing for covariates amounts to allowing for observed heterogeneity in the demand function. For brevity we will not catalog the various possible estimators one could use. We emphasize that all that is needed is calculation of the estimator $\hat{q}(p, y)$ at various values of p and y.

Confidence intervals for the true EV and DWL may be obtained using the bootstrap. For cross-section data $z_1, ..., z_n$ with mutually independent observations z_i, a bootstrap sample could be constructed by sampling n observations with replacement from the original data. That is, if the data are n independent observations $z_1, ..., z_n$ then a bootstrap sample $z_1^b, ..., z_n^b$ could be constructed by drawing z_i^b, $(i = 1, ..., n)$, independently from the distribution which puts probability weight $1/n$ on each z_i. Let $\hat{q}^b(x)$ be the demand estimator obtained from the bootstrap sample and $(\widehat{EV})^b$ and $(\widehat{DWL})^b$ be computed from $\hat{q}^b(x)$. Multiple simulated estimates can then be constructed by repeating this procedure B times to get $(\widehat{EV})^1, ..., (\widehat{EV})^B$. A confidence interval can then be formed in the usual way using the standard deviation of these bootstrap draws as the standard error.

Large sample confidence intervals can also be constructed using analytical standard errors rather than bootstrap ones. For series estimators these can be obtained by treating the series estimator as if it were least squares for a correctly specified model and applying the delta method, see Newey (1997). For kernel estimators the delta method of Newey (1994) can be used. It should be straightforward to extend that approach to locally linear estimators. Hausman and Newey (1995) show how to construct analytical standard errors for surplus and deadweight loss for series and kernel estimators. Constructing such analytical standard errors does require extensive derivations and calculation of various derivatives. The bootstrap avoids all that and so is attractive in substituting computing time for researcher's time.

Series estimators of the demand function may be computationally convenient because of the many times $\hat{q}(p, y)$ needs to be computed for estimation and bootstrap inference. For a series estimate calculation of $\hat{q}(p, y)$ only requires calculating a linear combination of relatively few approximating functions while locally linear and kernel estimators requires summing across all observations in the data set. Of course if computational time is not a concern then the savings from using a series estimate will not be important.

As an example we consider the partially linear model series estimator from Hausman and Newey (1995). Let $m(x) = (m_1(x), ..., m_J(x))^T$ denote approximating functions such as powers or splines of the log of components of p and y and let w denote a vector covariates. Let $\hat{\beta}$ and $\hat{\gamma}$ be the coefficients obtained from regressing $\ln q_i$ on $m(x_i)$ and covariate observations w_i. Let $\bar{w}$ be some chosen value for the covariates. The estimator of Hausman and Newey (1995) is

$$\hat{q}(x) = \exp(m(x)^T \hat{\beta} + \bar{w}^T \hat{\gamma}).$$

One could apply the methods we have described to estimate the equivalent variation for this function. Hausman and Newey (1995) do so for gasoline demand.

In this example the function $q(x)$ being estimated by $\hat{q}(x)$ corresponds to a partially linear specification where

$$q(x) = \exp(E[\ln q_i | x_i = x, w_i = \bar{w}]), E[\ln q_i | x_i, w_i] = r_0(x_i) + w_i^T \gamma_0.$$

Treating $q(x)$ as a demand function ignores the residual $\varepsilon_i = q_i - E[\ln q_i | x_i, w_i]$. Much of applied welfare analysis had followed the same practice until recently. One can ignore the residual if it is all measurement error but not if it contains individual heterogeneity. When ε_i contains heterogeneity, it may still be possible to interpret $q(x)$ as the demand function of a consumer. Suppose that $\varepsilon_i = \varepsilon(x_i, w_i, \eta_i)$ for some function $\varepsilon(x, w, \eta)$ of prices, income, covariates, and a vector η of taste variables that is independent of prices, income, and covariates. In general $\varepsilon(x, w, \eta)$ will depend on p and y, as shown by Brown and Walker (1989). Nevertheless, if $\varepsilon(x, w, \eta)$ is identically zero for some value of η then $q(x)$ can be interpreted as the demand function for that value of η. As an example suppose the conditional mean $E[\ln q_i | x_i = x, w_i = w]$ also equals the conditional median. Considering the conditional mean of $\ln q_i$ rather than q_i makes this seem more plausible because the log transformation can help make the distribution of demand more symmetric. Then if η is scalar and $\varepsilon(x, w, \eta)$ is monotonic increasing in η we will have $\varepsilon(x, w, \eta) = 0$ at the median of η_i, as further discussed below.

In general one would not expect that one could interpret $q(x)$ as the demand function for an individual, see Lewbel (2001). Also, even if $q(x)$ is the demand function for an individual one might want to consider surplus measures that account for individual heterogeneity. In the following Sections we do so.

4 Unobserved Individual Heterogeneity

We will allow for unobserved individual heterogeneity by letting the demand function depend on a vector of unobserved disturbances η of unknown dimension. One might think of each value of η as corresponding to a consumer though we do allow η to be continuously distributed. Similarly as before we are implicitly assuming that the utility function is strictly quasi-concave, only now we are making that assumption for each individual. Also, the Slutzky restrictions on the demand function are now assumed to hold for each individual. We summarize these restrictions in the following condition

$$\begin{aligned}&\textit{For each } \eta \textit{ the function } q(x,\eta) \textit{ is continuously differentiable in } x \textit{ at all } x \\ &\textit{with strictly positive prices and income, } \partial q(x,\eta)/\partial p + q(x,\eta)[\partial q(x,\eta)/\partial y]^T \\ &\textit{is symmetric and negative semi-definite for all } x \textit{ in } \chi, \text{ and } p^T q(x,\eta) + a(x,\eta) = y.\end{aligned} \tag{4.7}$$

The set χ is the set of prices and income over which the Slutzky condition is assumed to hold. It may be larger than the set of data on prices and income in order to use utility maximization to make predictions outside the range of data. In what follows we take demand functions satisfying this condition as primitive elements of the model. We also need technical conditions in order to make probability statements using these demand functions. These technical conditions are found in McFadden (2005) and the Appendix to Hausman and Newey (2016).

We follow the existing literature in modeling heterogeneity as corresponding to a distribution of demand for given prices and income x. Here we do this by letting η have a CDF G. Let r denote a possible value of quantity demanded. The CDF $F(r|x,q,G)$ of quantity when prices and income equal x for all individuals is given by

$$F(r|x,q,G) = \int 1(q(x,\eta) \leq r) G(d\eta). \tag{4.8}$$

The model we consider is one with a CDF for this form for $q(x,\eta)$ satisfying equation (4.7) and some distribution G of η.

This model is a random utility model (RUM) of the kind considered by McFadden (2005, see also McFadden and Richter, 1991). The model here specializes the RUM to single valued demands that are smooth in prices and income. Single valued, smooth demand specifications are often used in applications. In particular, smoothness has often proven useful in applications of nonparametric models and we expect it will here. We consider identification and estimation of surplus and deadweight loss in this RUM.

Viewing demand as a stochastic process indexed by x helps explain identification and other aspects of demand analysis with heterogeneity. Here $q(x,\eta)$ is a function of x for each η, that varies stochastically with η, i.e. $q(x,\eta)$ is a stochastic process. In this way the pair (q,G) can be thought of as a demand process. In the language of stochastic processes the distribution of $q(x,\eta)$ for fixed x is a marginal distribution, while the distribution of $(q(x^1,\eta),...,q(x^K,\eta))^T$ for some set $\{x^1,...,x^K\}$ of prices and income is a joint distribution. In our notation the marginal CDF of this stochastic process is $F(r|x,q,G)$. Thus, the thing being modeled in this paper is the marginal distribution of the demand process. We focus on the marginal distribution because that is what is identified in cross-section data where x is independent of η.

With individual heterogeneity there will be a distribution of surplus and deadweight loss that corresponds to the distribution of demand functions. For a price change from p^0 to p^1 and income at $\bar{y}$ let $S(\eta)$ denote the equivalent variation corresponding to the demand function $q(x,\eta)$ and $D(\eta)=S(\eta)-q(p^1,\bar{y},\eta)^T\Delta p$, the associated deadweight loss. As previously discussed $S(\eta)$ is the solution $s(0,\eta)$ at $t=0$ to the ordinary differential equation

$$\frac{ds(t,\eta)}{dt}=-q(p^0+t\Delta p,\bar{y}-s(t,\eta),\eta)^T\Delta p, \qquad s(1,\eta)=0. \tag{4.9}$$

The distribution of surplus and deadweight loss we consider will be the distribution of $S(\eta)$ and $D(\eta)$ that are implied by G.

Objects that we will focus on and that are of common interest are the average surplus $\bar{S}$ and deadweight loss $\bar{D}$ across individuals, given by

$$\bar{S}=\int S(\eta)G(d\eta), \bar{D}=\int D(\eta)G(d\eta).$$

As is known from Hicks (1939), when $\bar{S}$ is positive it is possible to redistribute income so that individuals are better off under p^0 than under p^1. Also, weighted averages of $\bar{S}$ over different $\bar{y}$ values can provide measures of social welfare when income and heterogeneity are independent in the population. We illustrate this use of average surplus in the gasoline demand application to follow.

Average surplus depends only on average demand $\bar{q}(p,y) = \int q(x,\eta)G(d\eta)$ when income effects are constant across individuals, prices, and income. When the price of only one good is changing and the income effect of that good is constant then average surplus depends only on the average demand for that good. To see this result consider a price change of just the first good. Suppose that the income effect for that good is constant with $\partial q_1(p_1, \bar{p}_2, y, \eta)/\partial y = \lambda$ over $p_1 \in [p_1^0, p_1^1]$, $y \in [\bar{y} - S(\eta), \bar{y}]$, and η. Then $S(\eta)$ is the solution at $t = 0$ to

$$\frac{ds(t,\eta)}{dt} = -[q_1(p_1^0 + t\Delta p_1, \bar{p}_2, \bar{y}, \eta) - \lambda s(t,\eta)]\Delta p_1, \quad s\,(1,\eta) = 0. \tag{4.10}$$

This is a linear differential equation with explicit solution

$$S(\eta) = \Delta p_1 \int_0^1 q_1(p_1^0 + t\Delta p_1, \bar{p}_2, \bar{y}, \eta)\exp(-t\lambda\Delta p_1)dt = \int_{p_1^0}^{p_1^1} q_1(p_1, \bar{p}_2, \bar{y}, \eta)\exp(-\lambda(p_1 - p_1^0))dp_1.$$

Taking expectations under the integral gives

$$\bar{S} = \int_{p_1^0}^{p_1^1} \bar{q}_1(p_1, \bar{p}_2, \bar{y})\exp(-\lambda(p_1 - p_1^0))dp_1.$$

This can also be represented as the solution at $t = 0$ to

$$\frac{d\bar{s}(t)}{dt} = -\bar{q}(p^0 + t\Delta p, \bar{y} - \bar{s}(t), \eta)^T \Delta p, \quad \bar{s}(1) = 0. \tag{4.11}$$

Comparing equation (4.11) with (4.9) we see that, if only the price of one good is changing and the income effect that good is constant then, average surplus solves the same differential equation as individual surplus, with average demand replacing individual demand. This result generalizes to multiple price changes where the income effects are constant for all goods with changing prices.

Obtaining average surplus from average demand is consistent with the well known aggregation results of Gorman (1961), who showed that constant income effects are necessary and sufficient for demand aggregation. The preceding discussion is a demonstration of a partial dual result, that when the price of one good is changing and the income effect is constant for that good then surplus for average demand is the average of surplus. McFadden (2004) derived and used this result in the case where income effects are constant for all goods. Hausman and Newey (2016) showed that it is sufficient that the income effect only be constant for the goods with prices that vary between p^0 and p^1.

5 Heterogenous Demand with Two Goods

The case with two goods has some special features that are important. One feature is that there are simple, intuitive restrictions on the distribution of demand that are equivalent to utility maximization with general heterogeneity. In addition there are a number of recent papers about modeling and estimating heterogenous demand for two goods. For these reasons it seems appropriate to devote some attention to the two good case.

Much of the revealed stochastic preference literature is concerned with deriving restrictions on $F(r|x,q,G)$ as a function of r and x that are necessary and sufficient for a RUM. McFadden (2005) provides a set of inequalities that are necessary and sufficient for the RUM with continuous demands. With two goods there is a simple, alternative characterization in terms of quantiles that is useful in the identification analysis to follow. The characterization is that each quantile is a demand function, or equivalently for smooth demands that each quantile satisfies the Slutzky condition that compensated demand is downward sloping. With two goods the Slutzky condition and the budget constraint are necessary and sufficient for a function to be a demand function.

To see why a demand model implies that the quantiles satisfy the budget constraint and Slutzky condition, let $Q(\tau|x) = \inf\{r : F(r|x,q,G) \geq \tau\}$ denote the τ^{th} conditional quantile corresponding to $F(r|x,q,G)$, where $0 < \tau < 1$ and we drop dependence of Q on q and G for notational convenience. Note that $Q(\tau|x)$ is the τ^{th} quantile of $q(x,\eta)$ so that the budget constraint $pQ(\tau|x) \leq y$ is satisfied by $pq(x,\eta) \leq y$ for all η. Hoderlein and Mammen (2007) gave a useful result on the derivatives of the quantile that can be used to show the Slutzky condition. The Hoderlein and Mammen (2007) result has been used and verified by Chernozhukov, Fernandez-Val, Hoderlein, and Newey (2015) and others. This characterization says that under certain regularity conditions

$$\frac{\partial Q(\tau|x)}{\partial x} = E[\frac{\partial q(x,\eta)}{\partial x}|q(x,\eta) = Q(\tau|x)].$$

It then follows that

$$\begin{aligned}\frac{\partial Q(\tau|x)}{\partial p} + Q(\tau|x)\frac{\partial Q(\tau|x)}{\partial y} &= E[\frac{\partial q(x,\eta)}{\partial p}|q(x,\eta) = Q(\tau|x)] + Q(\tau|x)E[\frac{\partial q(x,\eta)}{\partial y}|q(x,\eta) = Q(\tau|x)] \\ &= E[\frac{\partial q(x,\eta)}{\partial p} + q(x,\eta)\frac{\partial q(x,\eta)}{\partial y}|q(x,\eta) = Q(\tau|x)] \leq 0,\end{aligned}$$

where the inequality follows from the Slutzky condition for the demand function $q(x,\eta)$. That is, the quantile satisfies the Slutzky condition

$$\frac{\partial Q(\tau|x)}{\partial p} + Q(\tau|x)\frac{\partial Q(\tau|x)}{\partial y} \leq 0.$$

Thus each quantile is a demand function. This result was shown by Dette, Hoderlein, and Neumeyer (2016), who have made it the basis of testing the negative definiteness part of the Slutzky conditions. Some regularity conditions are required for this result, e.g. as given in Assumption A2 of Hausman and Newey (2016).

To see why a quantile satisfying the Slutzky condition and budget constraint $pQ(\tau|x) \leq y$ is sufficient for a demand model, let $\tilde{\eta}$ denote a scalar random variable that is independent of x and consider

$$\tilde{q}(x,\tilde{\eta}) = Q(\tilde{\eta}|x), \tilde{\eta} \sim U(0,1).$$

This is a demand model because $\tilde{q}(x,\tilde{\eta})$ satisfies the Slutzky condition and budget constraint for all $\tilde{\eta} \in (0,1)$. Furthermore, it is well known that when the τ argument in the quantile is replaced by a $U(0,1)$ random variable the resulting random variable has the distribution corresponding to that quantile function. Then, for $F(r|x,q,G)$ the distribution from which the quantile is obtained,

$$\int_0^1 1(Q(\tilde{\eta}|x) \leq r)d\tilde{\eta} = F(r|x,q,G).$$

Thus when the quantile satisfies the Slutzky condition and the budget constraint there is a demand model with scalar (uniform) heterogeneity that gives the same conditional distribution of quantity given x as the quantile. We refer to this model where the quantile is the demand function as quantile demand. This result is pointed out in Hausman and Newey (2016). Thus we see that for two goods and single valued smooth demands the revealed, stochastic preference conditions are that each quantile is a demand function.

Even though each quantile is a demand function it is not, in general, the demand function for particular consumers. That would only be correct if demand is monotonic in scalar heterogeneity. With multidimensonal heterogeneity we do not necessarily follow the same consumers as we change prices and income. The derivatives of the quantile are averages of derivatives of demand over those individuals at the quantile. The individuals at the quantile generally change with price and income. Furthermore, in cross section data we are not able to distinguish scalar heterogeneity from multivariate heterogeneity, as discussed below. Thus there is no way to tell from cross section data whether a quantile function can be interpreted as a demand function for particular consumers.

The conditional CDF of q given x also satisfies a Slutzky like condition. By the inverse function theorem, the quantile satisfies the Slutzky condition if and only if

$$\frac{\partial F(r|x,q,G)}{dp} + r\frac{\partial F(r|x,q,G)}{dy} \geq 0.$$

As with the quantile, the CDF satisfying this Slutzky condition is necessary and sufficient for a demand model, under the regularity conditions for existence of derivatives and for the inverse function theorem. This result is pointed out in Blomquist, Kumar, Liang, and Newey (2015).

These characterizations have important empirical and theoretical implications. In applications where we are estimating the demand for one of two goods, imposing the Slutzky condition and budget constraint on the quantiles imposes all the restrictions of utility maximization. Thus, empirical analysis based on such quantile or distribution estimates uses all those restrictions. Blundell, Kristensen and Matzkin (2014) is an example of this approach. In addition one can construct estimates of the conditional mean of demand that impose all the restrictions of utility maximization using the Slutzky condition for the CDF. Blomquist et. al. (2015) show how to do this. A theoretical implication of quantile demand with two goods is that the quantile demand model is observationally equivalent to the true model. This implication is useful in the identification analysis for consumer surplus.

6 Identification

We consider identification of objects of interest when we know the marginal CDF $F(r|x,q,G)$ of the demand process over a set $\bar{\chi}$ of prices and income. This corresponds to cross section data, where we only observe one price and income for each individual. If more than one value of x were observed for each individual, as in panel data, then one could identify some joint distributions of demand at different values of x. We touch on this topic below.

We adapt a standard framework to our setting, as in Hsiao (1983), by specifying that a structure is a demand function and heterogeneity distribution pair (q,G), where for notational convenience we suppress the arguments of q and G.

Definition 1: *(q,G) and $(\tilde{q},\tilde{G})$ are observationally equivalent if and only if for all r and $x \in \bar{\chi}$,*

$$F(r|x,q,G) = F(r|x,\tilde{q},\tilde{G}).$$

The set $\bar{\chi}$ will correspond to the set of x that is observed. We allow $\bar{\chi}$ to differ from the χ where the Slutzky conditions are imposed in order to allow the Slutzky conditions to be imposed outside the range of the data. We consider identification of an object $\delta(q,G)$ that is a function of the structure (q,G). Here $\delta(q,G)$ is a map from the demand function and the distribution of heterogeneity into some set. The identified set for δ we consider 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_0,G_0) is $\Lambda(q_0,G_0) = \{\delta(\tilde{q},\tilde{G}) : (q_0,G_0)$ and $(\tilde{q},\tilde{G})$ are observationally equivalent$\}$.*

The (q_0,G_0) in this definition can be thought of as the true values of the demand function and heterogeneity distribution. The identified set $\Lambda(q_0,G_0)$ is the set of δ that is consistent with the distribution of demand $F(r|x, q_0, G_0)$ implied by the true values. The set $\Lambda(q_0, G_0)$ is nonempty since it always includes the true value $\delta(q_0, G_0)$. The set $\Lambda(q_0, G_0)$ is sharp, given only knowledge of $F(r|x, q_0, G_0)$, because it consists exactly of those δ that correspond to some $(\tilde{q}, \tilde{G})$ that generates the same distribution of demand as the true values. In other words, sharpness of $\Lambda(q_0, G_0)$ holds automatically here because we are explicitly formulating the identified set in terms of all the restrictions on the distribution of demand that are implied by the model, and we are assuming that the distribution of demand is all we know.

The view of demand as a stochastic process indexed by x helps explain identification. As previously noted the marginal CDF of this stochastic process is $F(r|x, q, G)$ in our notation. Thus, two demand processes will be observationally equivalent if and only if they have the same marginal distribution.

One interesting and useful result is that objects $\delta(q, G)$ that depend only on the marginal distribution of the demand process are point identified, because they are the same for all observationally equivalent structures. For example, average demand $\bar{q}(x) = \int q(x, \eta)G(d\eta) = \int rF(dr|x, q, G)$ is identified, as are functionals of it, such as the bounds below.

Joint distributions of the demand process, such as the joint distribution of $(q(\tilde{x}, \eta), q(\bar{x}, \eta))^T$ for two different values of x, will not be identified. We will show this result for certain demand processes below and the intuition is straightforward. Intuitively, joint distributions are not identified from marginal distributions. Because joint distributions are not identified, distributions and averages of objects that depend on varying x for given η will not be identified. As was shown rigorously by Hausman and Newey (2016), such nonidentified objects will include average surplus, which depends on varying both price and income for a given η.

It will generally be impossible to identify demand functions for individuals from the marginal distribution of demand. Again the intuition is straightforward, with individual demands not identified because we only observe one price and income for each individual. More formally, we can think of the ability to identify individual demands as

$q_0(\tilde{x}, \eta)$ being perfectly predictable for each $\tilde{x}$ if we know $q_0(\bar{x}, \eta)$ for some $\bar{x}$, i.e. as $Var(\tilde{q}_0(\tilde{x}, \tilde{\eta})|\tilde{q}_0(\bar{x}, \tilde{\eta})) = 0$ for any $(\tilde{q}, \tilde{G})$ that is observationally equivalent to the truth. This is a property of the joint distribution of the demand process, and so is not identified from the marginal distribution of the demand process.

For two goods the quantile demand characterization of utility maximization provides a key to the proof of nonidentification of average surplus. As discussed in the previous section $Q(\tau|x)$ will be a demand function and $Q(\tilde{\eta}|x)$ for $\tilde{\eta} \sim U(0, 1)$ gives the same conditional distribution of quantity as true demand. Thus the quantile demand is observationally equivalent to true demand. The joint distribution of the quantile process can differ from the true one. For example, the true demand may have $Var(q(\tilde{x}, \eta)|q(\bar{x}, \eta)) > 0$ but $Q(\tilde{\eta}|x)$ will be one-to-one in $\tilde{\eta}$ for each x so $Var(Q(\tilde{\eta}|\tilde{x})|Q(\tilde{\eta}|\bar{x})) = 0$.

Some intuition for the nonidentification of average surplus is provided by a demand specification that is a random coefficients linear model, where

$$q_0(x, \eta) = \eta_1 + \eta_2 p + \eta_3 y,$$

and η_3 varies across individuals. This demand process is a familiar specification. Quantile demand will be observationally equivalent to this true demand. Thus, there is no way to distinguish nonparametrically this true, linear, varying coefficients process from quantile demand. Also, true average surplus will generally be different than average surplus for quantile demand. Intuitively, the true demand is linear in income y but quantile demand will generally be nonlinear in y because of varying η_3. The nonlinearities in income of quantile demand lead to average surplus for quantile demand being different than average surplus for the true demand. This is the basis of the nonidentification result for average surplus shown in Hausman and Newey (2016).

In panel data we could have multiple observations for a single individual. In that case it should be possible to test for whether $q(x, \eta)$ is monotonic in scalar η. Also, panel data could be used to tighten the bounds for surplus. In the limit as the number of observations per individual gets large it should be possible to identify individual surplus. Note though that panel data will only be helpful in these ways if there are some restrictions on how the demand function varies over time for a given individual, e.g. that the demand function is the same in each time period. One might want to let demand functions differ over time for a given individual to better fit the data. If the demand function in each time period is allowed to be completely different then panel data does not help identify more than the marginal distribution of the demand process.

7 Income Effect Bounds

Known bounds on income effects can be used to bound average surplus and deadweight loss using average demand. The idea is to extend the result that constant income effects allow computation of average surplus from average demand, to identify bounds on surplus from average demand. To describe the result, for any constant B let

$$\bar{S}_B = \int_0^1 [\bar{q}(p^0 + t\Delta p, \bar{y})^T \Delta p] e^{-Bt} dt \tag{7.12}$$

be the solution $\bar{s}_B(t)$ at $t = 0$ to the linear differential equation

$$\frac{d\bar{s}_B(t)}{dt} = -\bar{q}(p^0 + t\Delta p, \bar{y})^T \Delta p + B\bar{s}_B(t), \quad \bar{s}_B(1) = 0. \tag{7.13}$$

From Section 2 we see that $\bar{S}_B$ would be the average surplus if just the price of the first good were changing and the demand for the first good had a constant income effect $\partial q_1(p(t), y, \eta)/\partial y = B/\Delta p_1$.

If i) $q(p(t), \bar{y} - s, \eta)^T \Delta p \geq 0$ for $s \in [0, S(\eta)]$, *ii) there are constants* $\underline{B}$ *and* $\overline{B}$ *such that* $\underline{B} \leq [\partial q(x, \eta)/\partial y]^T \Delta p \leq \overline{B}$ *for all* $x \in \chi$; *iii) all prices in* $p(t)$ *are bounded away from zero then*

$$\bar{S}_{\overline{B}} \leq \bar{S} \leq \bar{S}_{\underline{B}}, \bar{S}_{\overline{B}} - \bar{q}(p^1, \bar{y})^T \Delta p \leq \bar{D} \leq \bar{S}_{\underline{B}} - \bar{q}(p^1, \bar{y})^T \Delta p.$$

Condition i) is a restriction on the price path that is automatically satisfied when only the price of the first good is changing and $p_1^1 > p_1^0$. Also, the bounds in the conclusion are satisfied under weaker conditions than bounded income effects, as discussed in the Appendix to Hausman and Newey (2016).

The key ingredient for these average surplus bounds are bounds on the income effect $[\partial q(x,\eta)/\partial y]^T \Delta p$. Economics can deliver such bounds. Consider again, and for the rest of this Section, a price change in the first good, where $\underline{B}$ and $\overline{B}$ are bounds on $\Delta p_1 \partial q_1(x,\eta)/\partial y$ and $\Delta p_1 > 0$. If q_1 is a normal good then the income effect is nonnegative, so we can take $\underline{B} = 0$. Then an upper bound for average equivalent variation and deadweight loss can be obtained from Marshallian surplus for average demand, that is

$$\bar{S} \leq \bar{S}_M = \int_0^1 \left[\bar{q}(p^0 + t\Delta p, \bar{y})^T \Delta p\right] dt, \bar{D} \leq \bar{S}_M - \bar{q}(p^1, \bar{y})^T \Delta p.$$

The upper bound on average deadweight loss could be useful for policy purposes, e.g. to proceed with a tax if average public benefits (e.g. environmental benefits) exceed average deadweight loss and the appropriate separability conditions are satisfied.

Economics can also deliver upper bounds on income effects. If no more than a fraction π of additional income is spent on q_1 then

$$\partial q_1(x,\eta)/\partial y \leq \pi/p_1 \leq \pi/p_1^0,$$

so that $\bar{B} = \Delta p_1 \pi/p_1^0 = \pi(p_1^1/p_1^0 - 1)$ is an upper bound on $[\partial q(x,\eta)/\partial y]^T \Delta p$. For example, in the gasoline demand application below we are quite certain that only a small fraction of any increase in income is spent on gasoline, making our choice of $\bar{B}$ very credible. The Slutzky condition also can limit the size of income effects relative to price effects. In the next Section we consider bounds based on the Slutzky condition.

The quantiles of the demand distribution are informative about income effects. Let $Q_1(\tau|x)$ denote the conditional quantile of the first good, where we continue to suppress dependence on q and G. By Hoderlein and Mammen (2007),

$$\frac{\partial Q_{{}_1}(\tau|x)}{\partial y} = E[\frac{\partial q_1(x,\eta)}{\partial y}|q_1(x,\eta) = Q_{{}_1}(\tau|x)],$$

where η is a random variable with distribution G. Note that constancy of the income effect will also imply constancy of $\partial Q_{{}_1}(\tau|x)/\partial y$ as τ varies. Thus, if $\partial Q_{{}_1}(\tau|x)/\partial y$ varies with τ the income effect for the first good is heterogenous. Also, a necessary condition for $\underline{B}$ and $\overline{B}$ to bound $\Delta p_1 \partial q_1(x,\eta)/\partial y$ is $\underline{B} \leq \Delta p_1 \partial Q_{{}_1}(\tau|x)/\partial y \leq \overline{B}$. This result can be used to guide the choice of bounds on income effects. For example, one might choose an upper bound that is much larger than derivatives of many quantiles, as we do in the gasoline application to follow. Note though that this approach does not serve to identify the bounds, because we cannot tell from the quantile derivative how the income effect varies over η with $q_1(x,\eta) = Q_1(\tau|x)$.

The conditional quantile is also informative about the surplus bounds. Let S^τ be the exact surplus obtained by treating $Q_1(\tau|x)$ as if it were a demand function, obtained as the solution $s^\tau(0)$ at $t = 0$ to the differential equation

$$\frac{ds^\tau(t)}{dt} = -Q_1(\tau|p_1^0 + t\Delta p_1, \bar{p}_2, \bar{y} - s^\tau(t))\Delta p_1, \qquad s^\tau\,(1) = 0.$$

With two goods and scalar heterogeneity, the average surplus would be $\int_0^1 S^\tau d\tau$. It turns out that $\int_0^1 S^\tau d\tau$ is between the surplus bounds in general. Hausman and Newey (2016) showed that

$$\bar{S}_{\overline{B}} \le \int_0^1 S^\tau d\tau \le \bar{S}_{\underline{B}}.$$

Surplus bounds are relatively insensitive to income effect bounds when a small proportion of income is spent on the good. This result is related to the Hotelling (1938) result that when expenditure is small approximate consumer surplus is typically close to actual consumer surplus. Differentiate equation (7.12) with respect to B to obtain

$$\begin{aligned} \bar{y}^{-1}\frac{\partial \bar{S}_B}{\partial B} &= -\bar{y}^{-1}\int_0^1 [\bar{q}_1(p_1^0 + t\Delta p_1, \bar{p}_2, \bar{y})\Delta p_1] t e^{-Bt} dt \\ &= -\int_{p_1^0}^{p_1^1} [\bar{q}_1(p_1, \bar{p}_2, \bar{y})p_1/\bar{y}](\frac{1 - p_1^0/p_1}{\Delta p_1}) \exp(-B\frac{p_1 - p_1^0}{\Delta p_1}) dp_1. \end{aligned}$$

In this way the bounds are less sensitive to B when share $\bar{q}_1(p_1, \bar{p}_2, \bar{y})p_1/\bar{y}$ of income spent on the first good is smaller.

The role of average demand in these bounds has implication for econometric practice. Average demand is the expectation of quantity demanded and not log-quantity or some other function of quantity. Thus, for estimating the bounds we need to estimate the conditional expectation of quantity. In practice it has often been the case that the some nonlinear function of quantity has been used in the specification in an effort to fit the data, e.g. Hausman and Newey (1995). What we find here is that quantity itself should be used for estimating the bounds. Share equations have also been used in the specification of demand models. That is alright because shares are linear in quantity.

8 General Bounds with Two Goods

The surplus and deadweights loss bounds based on income effects are computationally straightforward but depend on knowing bounds on income effects. We can also bound surplus using just utility maximization, i.e. using only the Slutzky condition and the budget constraint. The goal here is to estimate the largest and smallest surplus that are consistent with the Slutzky condition and with the distribution of the data. Hausman and Newey (2016) suggested doing this for two goods by using a discrete mixture expansion around quantile demand.

That approach uses a flexible demand specification that is a series expansion with random coefficients around quantile demand. To describe this specification let $m_j(x)$, $j = 1, ..., J$ be approximating functions such as power series or splines and $m(x) = (m_1(x), ..., m_J(x))^T$. Let $\breve{\eta} = (\breve{\eta}_1, ..., \breve{\eta}_J)^T$ be random coefficients for these approximating functions. We will assume that the vector $\breve{\eta}$ is discretely distributed with L points of support $\{\breve{\eta}^1, ..., \breve{\eta}^L\}$. Let $\tilde{\eta}$ be a scalar and $Q(\tau|x)$ be the conditional quantile of quantity given x. Consider a demand specification where $\eta = (\tilde{\eta}, \breve{\eta}^T)^T$, $\tilde{\eta} \sim U(0,1)$, $\breve{\eta}$ has a discrete distribution conditional on $\tilde{\eta}$ with points of support in $\{\breve{\eta}^1, ..., \breve{\eta}^L\}$, and

$$\tilde{q}(x, \eta) = Q(\tilde{\eta}|x) \exp(m(x)^T \breve{\eta}).$$

This will be a demand model of the kind are considering as long as the function $\tilde{q}(x, \tilde{\eta}, \breve{\eta})$ satisfies the budget constraint and the Slutzky condition for all $x \in \chi$, $\tilde{\eta} \in (0,1)$, and $\breve{\eta} \in \{\breve{\eta}^1, ..., \breve{\eta}^L\}$. For computation purposes we impose these conditions on a grid of x values that lie in χ. We do this by drawing candidates for support points $\breve{\eta}^\ell$ randomly from a distribution and then only keeping those such that $q(x, \tilde{\eta}, \breve{\eta}^\ell)$ satisfy the budget and Slutzky conditions for x in this grid and $\tilde{\eta}$ on a grid of values in $(0,1)$.

We also let the mixture probabilities for $\breve{\eta}$ vary with $\tilde{\eta}$ in a flexible way. We do this by taking those mixture probabilities to be convex combinations of fixed probabilities where the convex combination varies with $\tilde{\eta}$ in a flexible way. To describe this approach let $\varphi_{\varkappa}(\tilde{\eta})$, $(\varkappa = 1, ..., \Upsilon)$ be a partition of unity, satisfying $\varphi_{\varkappa}(\tilde{\eta}) \geq 0$ and $\sum_{\varkappa=1}^{\Upsilon} \varphi_{\varkappa}(\tilde{\eta}) = 1$. For example, we could choose $\varphi_{\varkappa}(\tilde{\eta})$ to be B-splines. Let $\rho_{\ell}^{\varkappa} \geq 0$ be probabilities satisfying $\sum_{\ell=1}^{L} \rho_{\ell}^{\varkappa} = 1$. We take the conditional distribution of $\breve{\eta}$ given $\tilde{\eta}$ to be $\Pr(\breve{\eta} = \breve{\eta}^{\ell}|\tilde{\eta}) = \sum_{\varkappa=1}^{\Upsilon} \varphi_{\varkappa}(\tilde{\eta})\rho_{\ell}^{\varkappa}$. This is a flexible specification of the discrete distribution of $\breve{\eta}$ conditional on $\tilde{\eta}$.

As J grows so that any function of x can be approximated, as L grows and the support $\{\breve{\eta}^1, ..., \breve{\eta}^L\}$ becomes richer, and as Υ grows so the distribution of $\breve{\eta}$ given $\tilde{\eta}$ becomes more flexible this demand specification should be able to approximate any demand process. Consequently the maximum and minimum surplus for this demand specification should be close to bounds for surplus over all demand processes.

This demand process is computationally convenient for imposing the constraints implied by the data distribution, because the CDF for this process is linear in the probabilities of the points of support for $\breve{\eta}$. Let $F(r|x) = Q^{-1}(r|x)$ be the CDF corresponding to the quantile $Q(\tau|x)$ that we assume to be invertible in τ. Define

$$\Psi_{\ell}^{\varkappa}(r, x) = \int_{0}^{F(r \cdot \exp(-m(x)^T \breve{\eta}^{\ell})|x)} \varphi_{\varkappa}(\tilde{\eta}) d\tilde{\eta}.$$

Integrating over $\tilde{\eta}$ gives

$$\begin{aligned} F(r|x, \tilde{q}, \tilde{G}) &= \Pr(\tilde{q}(x, \eta) \leq r) = E[E[1(\tilde{q}(x, \eta) \leq r)|\tilde{\eta}, x]|x] \\ &= E[E[1(Q(\tilde{\eta}|x) \leq r \exp(-m(x)^T \breve{\eta}))|\tilde{\eta}, x]|x] \\ &= E[E[1(\tilde{\eta} \leq F(r \exp(-m(x)^T \breve{\eta})|x)|\tilde{\eta}, x]|x] \\ &= E[\sum_{\ell=1}^{L} \sum_{\varkappa=1}^{\Upsilon} \varphi_{\varkappa}(\tilde{\eta}) \rho_{\ell}^{\varkappa} 1(\tilde{\eta} \leq F(r \exp(-m(x)^T \breve{\eta}^{\ell})|x)|x] = \sum_{\ell=1}^{L} \sum_{\varkappa=1}^{\Upsilon} \rho_{\ell}^{\varkappa} \Psi_{\ell}^{\varkappa}(r, x). \end{aligned}$$

Here we see that the demand distribution for the model we have specified is linear in the probabilities $\rho_{\ell}^{\varkappa}$.

An important feature of this demand model is that it includes the quantile demand as long as 0 is one of the elements of the support set $\{\breve{\eta}^1, ..., \breve{\eta}^L\}$. In that case this model will equal quantile demand when $\Pr(\breve{\eta} = 0) = 1$. Thus this model can be thought of as allowing multiplicative variations around quantile demand through the term $m(x)^T\breve{\eta}$.

The distribution implied by the true model imposes constraints on the probabilities $\rho_\ell^\varkappa$. For computational purposes we consider imposing a subset of these constraints on a grid of M values for quantity prices and income, (r_m, x_m), $(m = 1, ..., M)$. Let $\Gamma = \{(r_1, x_1), ..., (r_M, x_M)\}$ denote the grid points. The constraints take the form

$$F(r_m|x_m) = \sum_{\ell=1}^{L}\sum_{\varkappa=1}^{\Upsilon} \rho_\ell^\varkappa \Psi_\ell^\varkappa(r_m, x_m), (r_m, x_m) \in \Gamma, \rho_\ell^\varkappa \geq 0, \sum_{\ell=1}^{L} \rho_\ell^\varkappa = 1.$$

As M grows with J, L, and Υ the constraints will approximately impose all the restrictions of the data distribution. A convenient feature of these constraints for computation is that they are linear in the probabilities $\rho_\ell^\varkappa$.

Bounds on the average and the distribution of surplus can be constructed by maximizing and minimizing over all the mixture probabilities $\rho_\ell^\varkappa$ that satisfy the constraints. Let $\tilde{S}^\ell(\tilde{\eta})$ be the surplus for $\tilde{q}(x, \tilde{\eta}, \breve{\eta}^\ell)$ and $\bar{S}_\ell^\varkappa = \int_0^1 \varphi_\varkappa(\tilde{\eta})\tilde{S}^\ell(\tilde{\eta})d\tilde{\eta}$. We can get an approximate upper bound for average surplus by solving the linear program

$$\max_{\rho_\ell^\varkappa} \sum_{\varkappa=1}^{\Upsilon}\sum_{\ell=1}^{L} \rho_\ell^\varkappa \bar{S}_\ell^\varkappa \text{ s.t. } F(r_m|x_m) = \sum_{\ell=1}^{L}\sum_{\varkappa=1}^{\Upsilon} \rho_\ell^\varkappa \Psi_\ell^\varkappa(r_m, x_m), (r_m, x_m) \in \Gamma, \rho_\ell^\varkappa \geq 0, \sum_{\ell=1}^{L} \rho_\ell^\varkappa = 1.$$

This is a linear program so computation is straightforward. For the CDF of surplus let $F_{S\ell}^\varkappa(s) = \int_0^1 1(\tilde{S}^\ell(\tilde{\eta}) \leq s)\varphi_\varkappa(\tilde{\eta})d\tilde{\eta}$. An upper bound on the CDF of surplus is

$$\max_{\rho_\ell^\varkappa} \sum_{\varkappa=1}^{\Upsilon}\sum_{\ell=1}^{L} \rho_\ell^\varkappa F_{S\ell}^\varkappa(s) \text{ s.t. } F(r_m|x_m) = \sum_{\ell=1}^{L}\sum_{\varkappa=1}^{\Upsilon} \rho_\ell^\varkappa \Psi_\ell^\varkappa(r_m, x_m), (r_m, x_m) \in \Gamma, \rho_\ell^\varkappa \geq 0, \sum_{\ell=1}^{L} \rho_\ell^\varkappa = 1.$$

This is also a linear program where computation is straightforward.

As for other estimators of partially identified objects (e.g. Manski and Tamer, 2002), it may be important for consistent estimation to include some slackness in the constraints. For average surplus this could be accomplished using the quadratic program,

$$\max_{\rho_\ell^\varkappa} \sum_{\varkappa=1}^{\Upsilon}\sum_{\ell=1}^{L} \rho_\ell^\varkappa \bar{S}_\ell^\varkappa \text{ s.t. } \sum_{(r_m, x_m)\in\Gamma} [F(r_m|x_m) - \sum_{\ell=1}^{L}\sum_{\varkappa=1}^{\Upsilon} \rho_\ell^\varkappa \Psi_\ell^\varkappa(r_m, x_m)]^2 \leq \varepsilon, \rho_\ell^\varkappa \geq 0, \sum_{\ell=1}^{L} \rho_\ell^\varkappa = 1.$$

Here the constraints are allowed to depart from zero by some small amount $\varepsilon > 0$. This quadratic program can be solved quite easily using standard software.

This approach provides approximate surplus bounds using series approximation to the set of all demand processes that are consistent with the conditional CDF $F(r|x)$. Approximation to the true bounds depends on large J, L, Υ, and M. The choice of these tuning parameters and the corresponding approximation and inference theory are beyond the scope of this paper. Note though that these bounds are even of interest for some fixed J, L, and Υ. As long as $\breve{\eta}^{\ell} = 0$ for some ℓ the average quantile surplus will be between the bounds computed using this procedure. Thus the general bounds described here give a measure of how much surplus can vary away from the quantile surplus for other random utility specifications consistent with the data. Also, increasing J, L, or Υ only leads to wider bounds, so the results will give a lower bound on how wide the identified interval for surplus might be.

This series approximation approach provides a way of empirically implementing the RUM, i.e. of finding identified sets for objects of interest under revealed stochastic preference conditions. This approach differs from Kitamura and Stoye (2012) where revealed stochastic preference inequalities are imposed. Here we impose the Slutzky conditions on a grid and then interpolate between points using a series approximation. This approach relies on and exploits smoothness in underlying demand functions.

9 Empirical Application

The previous results are based on the average and distribution of demand for fixed price and income. These objects are identified when prices and income in the data are independent of preferences, i.e. when the data are (q_i, x_i), $(i = 1, ..., n)$ with $q_i = q_0(x_i, \eta_i)$ and x_i and η_i are statistically independent. In that case

$$E[q_i|x_i = x] = \bar{q}_0(x), \Pr(q_i \leq r|x_i = x) = F(r|x, q_0, G_0).$$

Here average demand is the conditional expectation of quantity given prices and income in the data, and similarly for the distribution of demand. The conditional expectation of quantity, and not some other function of quantity, such as the log, is special because it equals the average demand which is used in bounds based on income effects. Average demand could also be recovered from the conditional expectation of the share of income spent on q.

The conditional expectation $E[q_i|x_i = x]$ could be estimated by nonparametric regression, as we do in the gasoline demand application below. Alternatively, if there are many goods, so that nonparametric estimation is affected by the curse of dimensionality, a semiparametric or parametric estimate of the conditional expectation of quantity could be used. Those estimators could have functional form misspecification but are useful with high dimensional regressors.

Independence of η_i and x_i encompasses a statistical version of a fundamental hypothesis of consumer demand, that preferences do not vary with prices. It is also based on the individual being small relative to the market of observation, as would hold when different observations come from different markets. The independence of income from preferences has been a concern in some demand specifications where allowance is made for dynamic consumption, but is an important starting point and is commonly imposed in the gasoline demand application we consider.

Independence of η_i and x_i could be relaxed to allow for covariates. Consider an index specification where there are covariates w_i with possible value w and it is assumed that there is a vector of functions $v(w, \delta)$ that affect utility such that η_i and $(x_i^T, w_i^T)^T$ are independent. These covariates might be demographic variables that represent observed components of the utility. For example, one could use a single, linear index $v(w, \delta) = w_1 + w_2^T \delta$, with the usual scale and location normalization imposed. The demand function $q_0(x, v(w, \delta_0), \eta)$ would then depend on the index $v(w, \delta_0)$, as would the average demand

$$\bar{q}_0(x, v(w, \delta_0)) = \int q_0(x, v(w, \delta_0), \eta) G(d\eta) = E[q_i | x_i = x, v(w_i, \delta_0) = v]$$

Here average demand is equal to a partial index regression of quantity q_i on x_i and $v(w_i, \delta)$. Similar approaches to conditioning on covariates are common in demand analysis.

Endogeneity can be accounted for if there is an estimable control variable ξ such that x_i and η_i are independent conditional on ξ_i and the conditional support of ξ_i given x_i equals the marginal support of ξ_i. In that case it follows as in Blundell and Powell (2003) and Imbens and Newey (2009) that

$$\int E[q_i|x_i = x, \xi_i = \xi]F_\xi(d\xi) = \bar{q}_0(x), \int \Pr(q_i \le r|x_i = x, \xi_i = \xi)F_\xi(d\xi) = F(r|x, q_0, G_0),$$

where $F_\xi(\xi)$ is the CDF of ξ_i. Although conditions for existence of a control variable are quite strong (see Blundell and Matzkin, 2014), this approach does provide a way to allow for some forms of endogeneity.

Bounds on average surplus based on average demand are robust to measurement error in the observed quantity that preserves conditional expectations. For example if $q_i = q_i^* + \nu_i$ where q_i^* is true demand and v_i is measurement error satisfying $E[v_i|x_i, w_i, \xi_i] = 0$ then the bounds based on income effects are still valid. This is not true for the general bounds that make used of the distribution of demand.

10 Applications with Two Goods or Restricted Heterogeneity

Recently a number of papers have considered welfare analysis with two goods and scalar heterogeneity. The demand specification they have used is one where $q_i = q(x_i, \eta_i)$ for $q(x, \eta)$ monotonic in the scalar η and $x = (p, y)^T$ for a scalar p. Under independence of x_i and η_i the conditional quantiles of quantity will be demands for quantiles of η. To see this result note that by equivariance of quantiles to monotonic transformations the conditional quantile of q_i given $x_i = x$ will be $q(x, Q_\eta(\tau))$, where $Q_\eta(\tau)$ is the τ^{th} quantile of the distribution of η_i, i.e.

$$q(x, Q_\eta(\tau)) = Q(\tau|x).$$

Thus the demand function at the τ^{th} quantile of η is equal to the τ^{th} conditional quantile of quantity given x in the data.

If the scalar heterogeneity specification is really correct then one can estimate demand functions by estimating conditional quantile functions. A nonparametric estimator $\hat{Q}(\tau|x)$ of the conditional quantile of quantity conditional on price and income x (and possibly covariates) will estimate the demand function $q(x, \eta)$ at $\eta = Q_\eta(\tau)$. The corresponding surplus can then then be estimated as the numerical solution $\hat{S}(\tau) = \hat{s}(0, \tau)$ to the ordinary differential equation

$$\frac{d\hat{s}(t,\tau)}{dt} = -\hat{Q}(\tau|p^0 + t\Delta p, y - \hat{s}(t,\tau))\Delta p, \quad \hat{s}(1,\tau) = 0.$$

Standard errors can be formed by the bootstrap as discussed earlier or by analytical methods. Average surplus and the distribution of surplus can be estimated by integrating over τ. One can generalize this model to allow for endogeneity by using the control function approach sketched above. Alternatively one could assume that an instrument is independent of η and estimate $q(x, Q_\eta(\tau))$ using quantile nonparametric instrumental variables estimation as in Chernozhukov and Hansen (2005) and Chernozhukov, Imbens, and Newey (2007). Blundell Horowitz, and Parey (2016) use this approach.

Blundell, Horowitz, and Parey (2016), Hoderlein and Vanhems (2013), and Blundell, Kristensen, and Matzkin (2014) consider two goods and scalar heterogeneity. Hoderlein and Vanhems (2013) propose unrestricted conditional quantile estimation, using a control function to account for endogeneity. Their application is to gasoline demand with the distance from the Gulf of Mexico as an instrument, as suggested by Blundell, Horowitz, and Parey (2012). Blundell, Horowitz and Parey (2016) propose quantile IV estimation assuming that the instrument is independent of η. They find in their gasoline demand application that imposing the Slutzky condition smooths out the demand estimates substantially.

The results of Blundell, Horowitz, and Parey (2016) and Hoderlein and Vanhems (2013) depend on the assumption of scalar heterogeneity. As discussed in Section 6. the question of whether heterogeneity is scalar cannot be answered from cross-section data. If heterogeneity is not scalar then it is not clear how we should interpret surplus estimated at various quantiles. We do know that when income effects are bounded the average of the quantile surplus across quantiles will be between the bounds on average surplus in the Hoderlein and Vanhems (2013) setting, as discussed in Section 7. The width of the bounds on average surplus thus provide a partial sensitivity check on how the assumption of scalar heterogeneity affects the estimate of average surplus. A corresponding check for quantile estimates is not available. Indeed we do not know how to interpret such estimates when there is general heterogeneity. In the instrumental variables setting of Blundell, Horowitz, and Parey (2016) we do not know of any sensitivity check even for the average surplus. We know very little about what is identified or about bounds for welfare analysis in a model with an instrument that is independent of general heterogeneity.

Lewbel and Pendakur (2016) estimate surplus using a demand model that is possibly nonlinear and nonparametric in random coefficients. They give a straightforward approach to identification and estimation when the coefficients are independent of one another. Random coefficient specifications are an important approach to heterogeneity and indeed is the way we approached the problem in the general two good model of Section 8.

Blundell, Kristensen, and Matzkin (2014) use the model with scalar heterogeneity for a different purpose, to use revealed preference bounds to predict demand outside the range of the data. This work is robust to the presence of general heterogeneity. As discussed in Section 5, the Slutzky condition is satisfied for each quantile if and only if there is a demand model generating the data. For two goods revealed preference conditions are equivalent to the Slutzky condition, so imposing the revealed preference bounds on the quantiles is equivalent to imposing the Slutzky condition. Thus what is done in this approach can be viewed as imposing restrictions on the quantile sufficient for it to be a demand function. The goal can then be viewed as predicting quantile demand outside the range of the data. Since quantile demand can also be interpreted as the quantile of the true demand process, this work can be thought of as predicting the quantiles of a general demand process outside the range of the data, while imposing all the restrictions implied by utility maximization. In an application to British expenditure data they find quite tight bounds on demand over a range of income and prices.

Blomquist, Kumar, Liang, and Newey (2015) make choice predictions subject to all the restrictions imposed by utility maximization with general heterogeneity. Their goal is to predict the effect of tax changes on the expected value of taxable income when taxes are nonlinear. They give a way of estimating the conditional mean of taxable income conditional on nonlinear budget constraints that imposes utility maximization.

Their approach allows for measurement error in taxable income, unlike quantile based methods. In an application to data from Sweden they obtain accurate predictions of the effect of tax reforms.

11 Estimation and Welfare Analysis of Gasoline Demand

In this section we investigate how estimates of average consumer surplus and deadweight loss for gasoline taxes in the US vary with income. We use data from the 2001 U.S. National Household Transportation Survey (NHTS) from Hausman and Newey (2016). 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. The data we use consists of 8,908 observations.

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

$$\widehat{\bar{q}(x,w)} = \sum_{j,k,\ell=1}^{3} \hat{\beta}_{j,k,\ell}(\ln p)^j(\ln y)^k(v(w,\hat{\delta}))^\ell \tag{11.14}$$

We estimate equation (11.14) allowing 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 Parey (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 (11.14), constructing

$$E[\widehat{q_i|x,w},\xi] = \sum_{j,k,\ell,m=1}^{3} \tilde{\beta}_{j,k,\ell}(\ln p)^j(\ln y)^k(w'\delta)^\ell(\hat{\xi})^m, \tag{11.15}$$

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 by averaging over the estimated residuals $\hat{\xi}_i$ holding p, y, and w fixed. We used a 3rd degree polynomial after finding high standard errors with a 4th degree polynomial.

To set bounds on income effects we assume that gasoline is a normal good and so choose the lower bound $\underline{B}$ to be 0.0. To set the upper bound we estimate a local linear quantile regression of log of gasoline demand on log price and log income and evaluate the income derivative of the gasoline quantile at median price and income. We find that this income effect is increasing in the quantile τ. We take the upper bound on the income effect to be .0197, which is 20 times the quantile derivative at $\tau = .9$. This income effect is very large, corresponding to more than two cents of every additional dollar of income being spent on gasoline. We are confident that no one would have such a large income effect for gasoline, as further discussed in Hausman and Newey (2105).

We estimate bounds on average equivalent variation and deadweight loss at each of the deciles of the income distribution in our data. These bounds are based on income effect bounds given in the previous paragraph. We form 95 % confidence intervals for the identified set via Beresteanu and Molinari (2008) method, using an estimator of the joint asymptotic variance of the upper and lower bounds obtained via bootstrapping the estimates of the bounds, including all steps used in estimation. In Figures 1-4 we plot the bounds for surplus and deadweight loss as a function of income, where we evaluate at the .1, ..., .9 deciles of the income distribution in our sample. In Figures 1 and 2 we graph the bounds on the deadweight loss and the associated confidence intervals. Figure 1 gives the results for a price change from 1.2 to 1.3 while Figure 2 gives the corresponding graphs for a price change from 1.2 to 1.4. We find that the deadweight loss is quite flat as a function of income though there is some evidence that deadweight loss is largest at smaller income values and tends to decline with income. Figures 3 and 4 plot equivalent variation for a price change from 1.2 to 1.3 and 1.2 to 1.4 respectively, along with confidence intervals. We find that the equivalent variation tends to increase with income. When combined with the deadweight loss results this result implies that tax receipts will tend to increase with income.

We have used our bounds approach to estimate household gasoline demand functions allowing for unrestricted heterogeneity. While the welfare measures are not point identified, we find that the lower and upper bound estimates are close to each other and provide precise information about exact surplus with general heterogeneity.

12 Conclusion

Nonparametric welfare analysis with general heterogeneity is now straightforward. Bounds on income effects lead to simple bounds on welfare. Bounds can also be constructed using only the Slutzky condition. Average demand or share estimates can be used to construct measures of welfare that average over general heterogeneity but vary with income and covariates. Variation of the bounds with income and covariates allow us to assess how average welfare effects vary across different groups, as important for evaluating policies. All of this analysis can be accomplished while allowing for general heterogeneity.

An important open question is how the independence of budget sets and preferences can be relaxed. As we have seen independence can be dropped where there is an estimable control function where budget sets and preferences are independent conditional on the control function. It would also be good to explore the power for identifying welfare effects of instruments that are independent of multi dimensional heterogeneity. Other open questions include how to do welfare analysis for many goods while only imposing the Slutzky condition on demand, how to estimate bounds for the distribution of welfare effects, and how to extend existing results to other settings such as those with nonlinear budget sets. Ways to do welfare analysis with general heterogeneity are now available but there remain many topics to be explored.

13 References

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

Bhattacharya, D. (2015): "Nonparametric Welfare Analysis for Discrete Choice," *Econometrica* 83, 617–649.

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

Blomquist, S., A Kumar, Che-Yuan Liang, and W.K. Newey (2015): "Individual Heterogeneity, Nonlinear Budget Sets, and Taxable Income," CEMMAP working paper CWP21/15.

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 Econonometrics: Theory and Applications, Eighth World Congress, Vol. II*, Cambridge: Cambridge University Press.

Blundell, R., D. Kristensen, and R. Matzkin (2014): "Bounding Quantile Demand Functions Using Revealed Preference Inequalities," *Journal of Econometrics* 179, 112–127.

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.

Blundell, R., J. Horowitz, and M. Parey (2016): "Nonparametric Estimation of a Nonseparable Demand Function under the Slutsky Inequality Restriction," *Review of Economics and Statistics*, forthcoming.

Blundell, R. and R. Matzkin (2014): "Control Functions in Nonseparable Simultaneous Equations Models," *Quantitative Economics* 5, 271–295.

Brown, B.W. and M.B. Walker (1989): "The Random Utility Hypothesis and Inference in Demand Systems," *Econometrica* 57, 815-829.

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.

Chernozhukov, V., I. Fernandez-Val, S. Hoderlein, H. Holzman, and W. Newey, (2015): "Nonparametric Identification in Panels Using Quantiles," *Journal of Econo-*

metrics 188, 378–392.

Chernozhukov, V., G. Imbens, and W. Newey (2007) "Instrumental Variable Identification and Estimation of Nonseparable Models," *Journal of Econometrics* 139, 4-14.

Chernozhukov, V., and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica* 73, 245–261.

Cosaert, S. and T. Demuynck (2014): "Nonparametric Welfare and Demand Analysis with Unobserved Individual Heterogeneity," working paper RM/14/10, Maastricht University.

Dette, H., S. Hoderlein, and N. Neumeyer (2016): "Testing Multivariate Economic Restrictions Using Quantiles: The Example of Slutsky Negative Semidefiniteness," *Journal of Econometrics* 191, 129-144.

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

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.

Hausman, J.A. and W. Newey (2016): "Individual Heterogeneity and Average Welfare," *Econometrica* 84, 1225-1248.

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

Hoderlein, S. and J. Stoye (2014): "Revealed Preferences in a Heterogeneous Population," *Review of Economics and Statistics* 96, 197-213

Hoderlein, S. and A. Vanhems (2013): "Estimating the Distribution of Welfare Effects Using Quantiles," working paper, Boston College.

Hotelling, H. (1938): "The General Welfare in Relation to Problems of Taxation and of Railway 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.

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

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 (2016): "Unobserved Preference Heterogeneity in Demand Using Generalized Random Coefficients," *Journal of Political Economy*, forthcoming.

Manski, C.F. and E. Tamer (2002): "Inference On Regressions With Interval Data On a Regressor or Outcome," *Econometrica* 70, 519–47.

McFadden, D.L. (2004): "Welfare Economics at the Extensive Margin: Giving Gorman Polar Consumers Some Latitude," working paper, UC Berkeley.

McFadden, D.L. (2005): "Revealed Stochastic Preference: A Synthesis," *Economic Theory* 26, 245-264.

McFadden, D.L. 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.

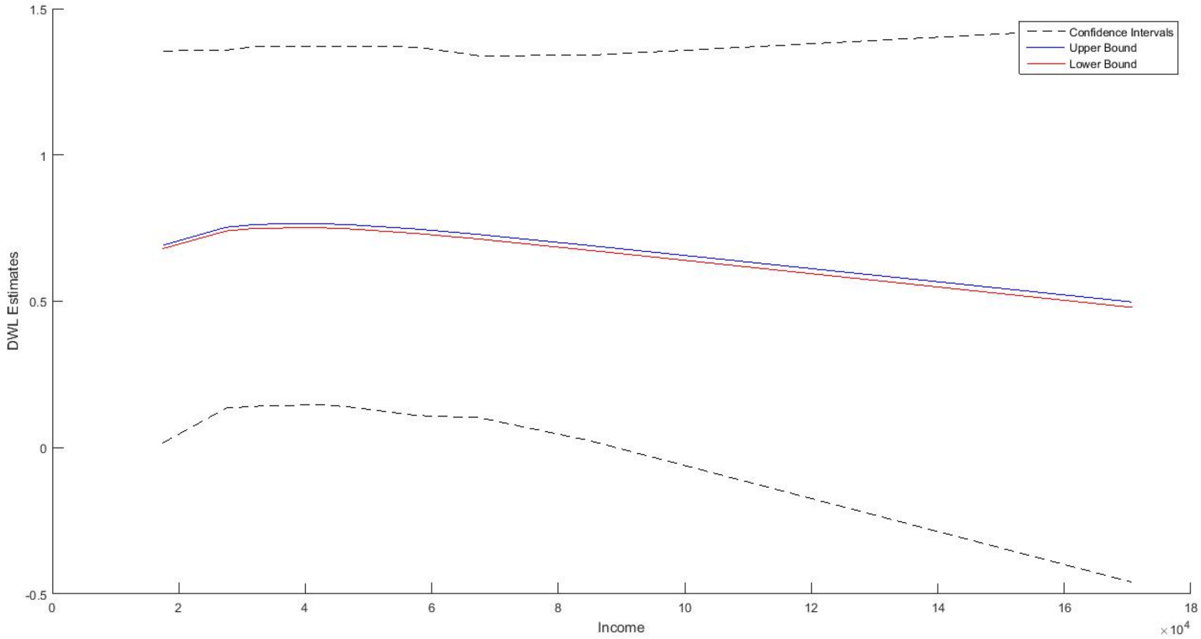
Newey, W.K. (1994): "Kernel Estimation of Partial Means and a General Variance Estimator," *Econometric Theory* 10, 233-253.

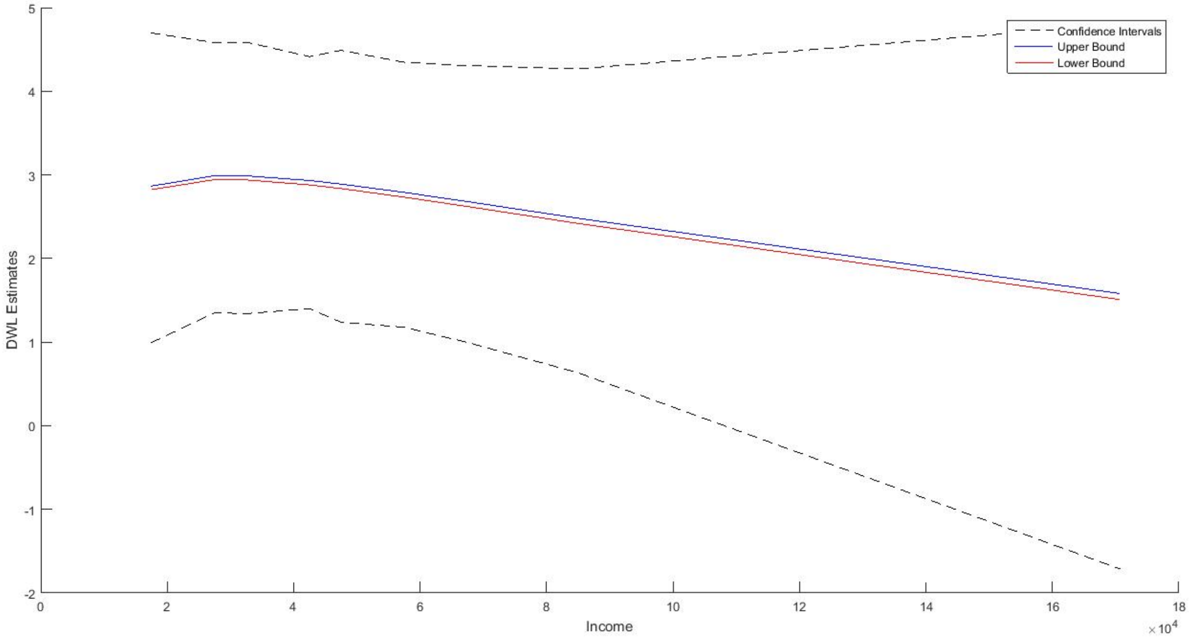
Newey, W.K. (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics* 79, 147-168.

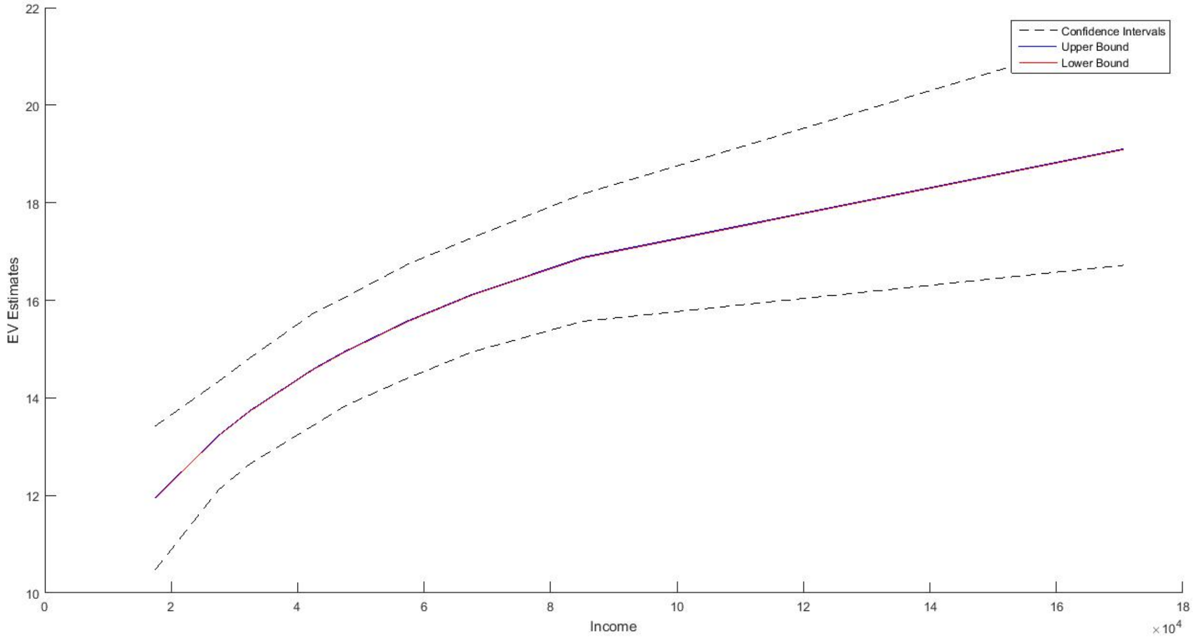
Vanhems, A. (2006): "Nonparametric Study of Solutions of Differential Equations," *Econometric Theory* 22, 127-157.

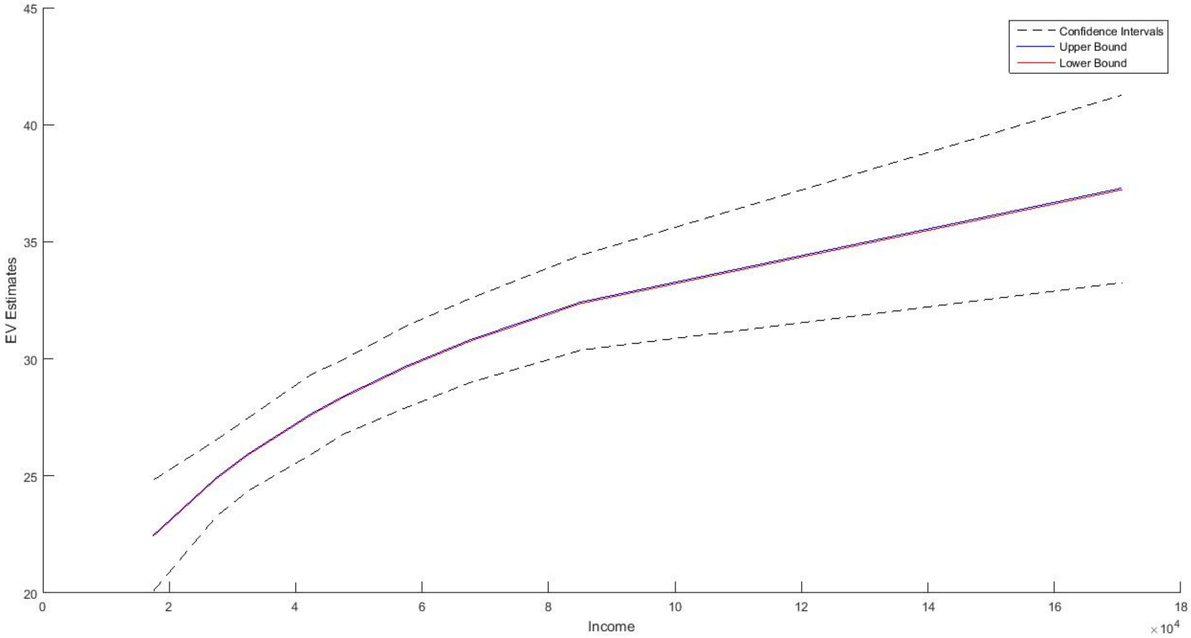
Vartia, Y. (1984): "Efficient Methods of Measuring Welfare Change and Compensated Income in Terms of Ordinary Demand Functions," *Econometrica* 51, 79-98.

Willig, R.D. (1976): ”Consumer Surplus Without Apology,” *American Economic Review* 66, 589-597.


1.5
1
0.5
0
-0.5
DWL Estimates
0
2
4
6
8
10
12
14
16
18
Income
× 10^4
Confidence Intervals
Upper Bound
Lower Bound


DWL Estimates
Income
$\times 10^4$
5
4
3
2
1
0
-1
-2
0
2
4
6
8
10
12
14
16
18
Confidence Intervals
Upper Bound
Lower Bound


EV Estimates
Income
$\times 10^4$
22
20
18
16
14
12
10
0
2
4
6
8
10
12
14
16
18
Confidence Intervals
Upper Bound
Lower Bound


45
40
35
30
25
20
EV Estimates
0
2
4
6
8
10
12
14
16
18
Income
× 10^4
Confidence Intervals
Upper Bound
Lower Bound