Estimating the distribution of welfare effects using quantiles

Authors: Stefan Hoderlein, Anne Vanhems

Persistent link: http://hdl.handle.net/2345/bc-ir:104977

This work is posted on eScholarship@BC,
Boston College University Libraries.

Estimating the Distribution of Welfare Effects Using Quantiles[*]

Stefan Hoderlein[†]   Anne Vanhems[¶]

First Version: July, 1 2008
This Version: September 10, 2013

Abstract

This paper proposes a framework to model empirically welfare effects that are associated with a price change in a population of heterogeneous consumers which is similar to Hausman and Newey (1995), but allows for more general forms of heterogeneity. Individual demands are characterized by a general model which is nonparametric in the regressors, as well as monotonic in unobserved heterogeneity. In this setup, we first provide and discuss conditions under which the heterogeneous welfare effects are identified, and establish constructive identification. We then propose a sample counterpart estimator, and analyze its large sample properties. For both identification and estimation, we distinguish between the cases when regressors are exogenous and when they are endogenous. Finally, we apply all concepts to measuring the heterogeneous effect of a chance of gasoline price using US consumer data and find very substantial differences in individual effects across quantiles.

[†]Department of Economics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA, Tel. +1-617-552-6042. email: stefan_hoderlein@yahoo.com

[¶]University of Toulouse, Toulouse Business School and Toulouse School of Economics. a.vanhems@esc-toulouse.fr.

1   Introduction

Motivation. Quantifying the welfare effects of a price change is one of the fundamental topics both in economic theory as well as in applied policy analysis. While measures of this welfare change like compensating variation or equivalent variation are theoretically well understood, the empirical side of welfare analysis in a heterogeneous population is less well developed. The challenge comes from the fact that in the common cross section data sets we observe every single individual only once, and, in particular, we do not observe the same individual under both the old and new price regime. However, to recover the exact welfare effect, we would have to observe the individual at every price level, clearly an impossible task. Hence, we have to infer the effect by looking at comparable individuals. Needless to mention, any analysis is then faced with the problem of unobserved (preference) heterogeneity, i.e., the fact that even after accounting for all observable variables, individuals remain profoundly different. Thus, adequate means and methods for controlling this complication are called for when evaluating welfare effects.

Moreover, in a heterogeneous population the effects of price change may differ substantially across the population. Any method we advocate thus also has to be able to capture this variation. To allow for unobserved heterogeneity, models that allow the error to enter non-additively have become increasingly popular in the recent econometrics literature. This paper aims at applying such a framework to welfare analysis. The setup is as follows: Following economic theory, we assume that there exists a relationship

$$Y = \varphi(X, A), \tag{1.1}$$

where Y is the quantity of gasoline consumed by a household, a real valued continuously distributed random scalar, X is a real valued vector of continuously distributed observable regressors, and A denotes unobservables, in particular heterogeneous preference parameters. While this allows for individuals to have arbitrarily different utility functions, we will invoke the common assumption that the utility of gasoline is separable from all other goods, implying that demand for gasoline is only a function of own price and a price (index) of all other goods. Moreover, we impose homogeneity of degree zero by normalizing the price for all other goods to be unity. Therefore the vector X contains $(P, S, \tilde{Z}')'$ where P is the relative price of gasoline,

S is real income, $\tilde{Z}$ denotes all observable characteristics of an individual, and $\varphi(\cdot, a)$ is hence the Marshallian demand function for an individual with preferences $A = a$.

To determine the welfare effect, we use the measure of exact consumer surplus known as compensated variation (CV), i.e., the income amount necessary to compensate the utility loss associated with a gasoline price change from p^0 to p, see Willig (1976), Hausman (1981), Vartia (1983). This functional parameter of interest is denoted by $\lambda(p) = \lambda(p, s, \tilde{z}, a)$ (to simplify the notations we suppress all variables other than p). The link between this welfare effect and the heterogeneous Marshallian demand function φ is given by the following differential equation:

$$\begin{cases} \lambda'(p) &= \varphi(p, s + \lambda(p), \tilde{z}, a) \\ \lambda(p^0) &= 0, \end{cases} \tag{1.2}$$

where p^0 is a reference price and $s, \tilde{z}, a$ represent fixed characteristics of the consumer. This system of equations defines the problem; the solution λ determines the welfare (relative to the reference price) of a single individual whose preferences are defined through $(\tilde{z}, a)$. Ideally, we would like to assume that A is infinite dimensional, however, in this case λ is not point identified. In the case of exogeneity of the unobservables A, we therefore assume that A is a scalar and that φ is strictly monotonic in A. As an extension, we consider endogeneity of P, too, in which case we assume that $A = (A_1, A_2) \in \mathbb{R}^2$. We use this specification of equation (1.1) to study the solution to the system (1.2).

In this paper, we moreover establish the asymptotic distribution of the estimated solution to this system of equations when a nonparametric estimator for $\varphi(x, a)$ is plugged in. For instance, in the exogenous case, under the additional assumption that A is uniformly distributed on $[0, 1]$, we use the fact that $\varphi(x, a)$ is identified by the a-quantile of Y given $X = x$. A natural estimator for $\varphi(x, a)$ is hence a nonparametric kernel quantile estimator, and we derive the properties of an estimator $\hat{\lambda}(p)$ which uses such an estimator as building block. Moreover, we extend this type of analysis to allow for endogeneities in price.

We would like to emphasize that the purpose of the present analysis is limited. We do neither propose a new model of demand, nor do we contribute to the methodological understanding of limitations in the data, e.g., nonlinear measurement errors, or limited price variation as in Blundell, Kristensen and Matzkin (2013, BKM henceforth). In parts, one may argue that they are not present in our data, as we use cross section data; in other parts we simply do not focus on these issues. This is, simply put, to the best of our knowledge the first study that focuses on the distribution of welfare effects, extending the framework pioneered by Matzkin (2003). Needless to mention, further research in particular involving panel data, is called for, and this can only be viewed as a first step. However, we do feel that the topic - the distribution of

welfare effects - is of great importance, as any policy maker will undoubtedly worry about this type of heterogeneity when thinking about implementing economic policies.

Related Literature. This paper extends traditional welfare analysis to evaluating the distributions of welfare effects in a heterogeneous population by use of quantile regression methods. The economic foundations are based on Willig (1976), Hausman (1981) and Vartia (1983). Slesnick (1996) provides a lucid discussion of the literature on estimating welfare effects, with particular emphasis on the issue of aggregation. Recent contributions that are closely related to ours are Hausman and Newey (1995), who considers nonparametric mean regressions which allow for great flexibility in the way regressors enter but are more restrictive in the way unobserved heterogeneity enters, Vanhems (2006), who revisits the approach of Hausman and Newey (1995) using tools from functional analysis, and Vanhems (2010) who extends the previous analysis considering price endogeneity. Blundell, Horowitz and Parey (2010a, BHP), propose a nonparametric estimator similar to Hausman and Newey's (1995), but additionally imposing Slutsky negativity conditions coming from economic theory while retaining the mean regression framework. In related work, Blundell, Horowitz and Parey (2010b) use quantile methods to estimate heterogeneous demand functions subject to Slutsky negativity restrictions, but do not focus on welfare effects, or derive the asymptotic distribution of an estimator for the CV measure. Closely related is the recent paper of Hausman and Newey (2011), who consider a very similar setup like ours but allow for high dimensional unobservables, and thus remove one of the main restrictive assumptions in this paper. The drawback is that their less restrictive assumptions only allow bounding average effects, and not being able to make statements about the distribution of welfare effects. Finally, more widely related is the papers by Foster and Hahn (2000), who analyze welfare effects using a linear specification similar to Hausman (1981), but model heterogeneity through random coefficients. While this paper is mainly empirical, it offers a competing, nonnested model for unobserved heterogeneity in welfare effects.

We apply insights from the the recent econometric literature about nonseparable models, see Altonji and Matzkin (2005), Chesher (2003), Imbens and Newey (2009), Matzkin (2003), or Hoderlein and Mammen (2007) (for an overview, see Matzkin (2005)). In particular, we would like to refer to the applications of nonseparable models and mean regressions in Lewbel (2001), and Hoderlein (2010), and the papers of Torgovitsky (2013) and d'Haultfoeuille and Fevrier (2013) on point identification of models like the one analyzed in this paper. Relative to this literature, we do not provide new identification results for the function itself, however, we present, to the best of our knowledge, the first application of such a framework to a question of social decision making.

Heterogeneity of individuals in demand applications has recently been emphasized, see,

e.g., Crawford and Pendakur (2012). We abstract in this paper from the problem that typically the data are provided as household data, and that there is an additional layer of unobserved heterogeneity that stems from the fact that there are several individuals within a household, however, see the paper by Cherchye, De Rock and Vermeulen (2007).

Also widely related is the paper by BKM (2013), who provide bounds on the nonparametric estimation of consumer demand. Their analysis shares similarities and differences with ours. First, the focus is very different; as already discussed above, we are largely focussing on the distribution of welfare effects, while BKM (2013) focus on the derivation of demand bounds and do not discuss welfare analysis. The bounds analysis in BKM (2013) is motivated by the fact that in repeated cross sections as the one employed in their paper, there is only insufficient price variation, which affects our analysis to a lesser degree, as we are largely conducting cross sectional analysis, but is very relevant for commonly used repeated cross section data sets. Despite these differences, there are also important parallels, as both papers assume monotonicity in a scalar unobservable. This assumption may be rightfully critizised, but there is no way of estimating a distribution of marginal effects, without either restricting heterogeneity or functional form. In this present paper we follow the former route, in ongoing research (Hoderlein and Vanhems 2011), we follow the latter (see also Hoderlein and Mammen (2007) on the impossibility of recovering the distribution of unrestricted heterogeneous marginal effects). One heuristic argument for assuming a scalar unobservable in the gasoline demand case may be that the good in question is used for one purpose exactly, which is driving. Moreover, there may be difference in the liking or disliking of driving that explain most of the variation once one conditions on observables, and it is conceivable that individuals gasoline consumption can be pretty well ordered along this one dimensional preference. This single purpose, and the ability to rank people according to (dis)like of driving may make this assumption palatable in the current application. Needless to mention, this does not have to be the case with more complex goods.

Finally, demand for gasoline has been extensively studied in the literature. It is analyzed in the paper by Hausman and Newey (1995), more recent references are Schmalensee and Stoker (1999) and Yatchew and No (2001), and BHP (2010a,b). See the paper of Hausman and Newey (1995) for more information about the economic framework, as well as additional older references to the literature.

Structure of the Paper. We start by discussing identification of $\lambda(p)$ in the second section. In the third section, we analyze the behavior of a sample counterpart estimator. We apply our estimation procedure to US gasoline consumption data in the fourth section, and find results that are roughly in line with the literature, but show a large variety of interesting

distributional effects that justify the focus on heterogeneity advocated in this paper. Finally, we conclude with an outlook.

2   Identification

In this section, we discuss the identification of $\lambda(p)$, and what the required assumptions mean in economic terms. We first start by stating the conditions under which the function φ is identified if the regressors are exogenous. Then we proceed to discuss how this model can be used as building block to identify the distribution of welfare effects. Finally, we extend our approach to the case of endogenous regressors.

2.1   Individual Demand

In the case of purely exogenous regressors, we consider is the following setup:

$$Y = \varphi(X, A)$$

where $Y \in R$ denotes the observed demand for gasoline, $X = (P, Z)$ is a vector of observed variables and A is a scalar disturbance. More precisely, the first component P represents the price of gasoline. Moreover, the vector Z includes income, denoted S along with other exogenous characteristics, $\tilde{Z} \in \mathbb{R}^L$. In this setup $A \in R$ represents unobserved heterogeneity. We assume that A is uniformly distributed on $[0, 1]$. At last, we consider the function $\varphi : \mathcal{X} \times [0, 1] \to R$, continuous in both arguments where $\Theta_X \subset R^{L+2}$ is the support of X. We denote by F the cumulative distribution function (hereafter cdf) of the vector (Y, X). In addition, we make the following assumptions:

[A1] A independent from X

[A2] for all $x \in \Theta_X$, $\varphi(x, .)$ is strictly increasing in a

The following result is standard, see Matzkin (2003).

Proposition 1. *Under Assumptions* $[A1] - [A2]$*, the function* φ *is identified by*

$$\varphi(x, a) = F_{Y|X}^{-1}(a; x)$$

where $F_{Y|X}^{-1}(a; x)$ *is the conditional a-quantile of* Y *given* $X = x$.

This result allows us to characterize the demand behavior of the entire population by identifying the a-quantile of Y given $X = x$ with an individual: We associate the demand behavior of individual i with his quantile position at $X = x_i$, he becomes "type a" if he is at the a quantile of the distribution of Y given his observed vector x_i. One implication of this model is that the individuals never change their relative position; if an individual is "type a" for $X = x_i$ he would also be "type a" for $X = x_j \neq x_i$. The fact that we identify every individuals' demand function enables us to determine the welfare effect for the entire population, even though we only observe every individual once. However, we can infer his demand behavior by looking at comparable individuals with the same a. The philosophy is very much in the spirit of the matching approach to treatment effects; see Hoderlein and Mammen (2007) for more details of the (restrictive) implications of the monotonicity assumption. A more general analysis with unrestricted and high dimensional unobservables remains desirable; in the absence of functional form restrictions we conjecture that this leads at best to partial identification of features of the distribution of (welfare) effects of interest, as an extension to e.g., the average effects analyzed in Hausman and Newey (2011). We leave such an analysis for future research.

2.2 Exact Consumer Surplus

To identify the distribution of welfare effects, consider the inverse problem defined by equation (1.2). To state the conditions under which an unique solution in a neighborhood of the initial condition p^0 exists, we need the following notation: First, fix an income level s as well as specific values for the exogenous variables $\tilde{z}$ and a. Next, let $I = [p^0 - \epsilon_1, p^0 + \epsilon_1]$, for $\epsilon_1 > 0$ denote a closed neighborhood of p^0, let $J = [s - \epsilon_2, s + \epsilon_2]$ with $\epsilon_2 > 0$, and define $D = I \times J$.

With this notation, the regularity conditions required are as follows: For fixed values $(\tilde{z}, a)$ in the support,

- [i] $max_{(p,\tilde{s}) \in D} |\varphi(p, \tilde{s}, \tilde{z}, a)| < \epsilon_2 / \epsilon_1$

- [ii] $|\varphi(p, s_2, \tilde{z}, a) - \varphi(p, s_1, \tilde{z}, a)| \leq k|s_2 - s_1|, \forall (p, s_i) \in D$ such that $c = k\epsilon_1 < 1$

Note that the more substantial condition is the second, a Lipschitz continuity condition which rules out certain rather pathological demand patterns. A sufficient condition on φ to satisfy this assumption is that φ be one time continuously differentiable in s on D. Assumption $[i]$ is a pure regularity condition. In particular, if the function φ is assumed to be continuous, this assumption is easily shown to hold. Under these conditions, the Cauchy-Lipschitz theorem proves existence and uniqueness of a solution defined on I; the proof in Vanhems (2006) extends

to this case with additional arguments. In summary, given identification of φ, the identification of λ follows under these regularity conditions on φ. From now on, we assume tacitly that these conditions hold, and hence obtain:

Proposition 2. *For fixed values $s, \tilde{z}, a$, under assumptions [i] and [ii], there exists a unique solution to (1.2) defined on I.*

2.3 Extensions to Endogenous Regressors

To deal with this situation, we follow Imbens and Newey (2009), and employ a two step control function approach. The first step involves the construction of the control variable; in a second step we obtain the conditional quantile of the demand given the endogenous variable and the control variable (plus some additional exogenous factors). The control function can be thought of as capturing the correlated part of the error; once it is accounted for prices are no longer endogenous.

We give now an economic discussion about the type of endogeneity we can handle. To this end, we first introduce our model formally. It is exactly as in the previous section, i.e.

$$Y = \varphi(X, A)$$

where Y and X are as before, but A is now a two dimensional disturbance vector, i.e., $A = (A_1, A_2) \in \mathbb{R}^2$ represent now the more complex unobserved heterogeneity, we maintain the assumption that one of the unobservables A_2 enters monotonically conditional on all other variables. We assume that P is endogenous and correlated with A_1, however, we will assume that there is a triangular structure involving an exogenous factor/instrument $W \in \mathbb{R}$ that allows us to deal with this problem. In particular, we assume that W enters through a second equation that relates it to the endogenous regressor, i.e.,

$$P = h(Z, W, A_1)$$

We normalize the model by assuming that A_1 and A_2 be uniformly distributed on $[0, 1]$, and we impose the following additional assumptions:

[A'1] $A_1 \perp (Z, W)$,

[A'2] For all (z, w), $h(z, w, .)$ is strictly increasing,

which imply identification of h, see again Matzkin (2003). More precisely,

Proposition 3. *Under Assumptions $[A'1]$ and $[A'2]$, the function h is identified by $h(z, w, a_1) = F_{P|Z,W}^{-1}(a_1; z, w)$ where $F_{P|Z,W}^{-1}(a_1; z, w)$ is the conditional a_1 quantile of P given $(Z, W) = (z, w)$. Moreover, we can also identify the unobserved heterogeneity variable $A_1 = F_{P|Z,W}(P, Z, W)$.*

In order to identify the function φ, we impose the additional assumptions:

[A'3] $A_2 \perp (X, W)|A_1$,

[A'4] For all (x, a_1), $\varphi(x, a_1, .)$ is strictly increasing

[A'5] For all $X \in \mathcal{X}$, the support of A_1 conditional on X equals the support of A_1.

We remark that $[A'1], [A'3]$ are implied by $Z \perp A$ in this system. Note moreover, that the overall model is only partially compatible with the previous, exogenous section in the following sense. Assume that the original model has a monotonic scalar unobservable, denoted $\tilde{A}$, which is correlated with X. Assume moreover, that there is a mapping τ s. th. $\tilde{A} = \tau(A_1, A_2)$, and τ is strictly monotone in a scalar A_2 (in the same direction), with A_2 satisfying assumption $[A3]$, i.e., we decompose a scalar random variable $\tilde{A}$ into two parts, one monotonic and independent, and one the rest. Using this, we obtain:

$$Y = \varphi(X, \tilde{A}) = \varphi(X, \tau(A_1, A_2)) = \phi(X, A_1, A_2),$$

with ϕ strictly monotonic in A_2. In slight abuse of notation, we use φ to denote now both functions. These steps is illustrates that the endogenous model can be related to the exogenous one, but only if one is willing to impose nontrivial structure on the endogeneity. As such, our approach is structural in the sense that it depends on the precise modeling of the endogeneity structure. Since our goal is to recover the entire distribution of welfare effects, this is probably not surprising.

Despite being more structural than would be obvious at first glance, these assumptions are standard, as is the following result that we restate in our notation for completeness purposes, see, in particular, Chesher (2003), and Imbens and Newey (2009)[1]

Proposition 4. *Under Assumptions $[A'1] - [A'5]$, the function φ is identified by:*

$$\varphi(x, a_1, a_2) = F_{Y|X,A_1}^{-1}(a_2; x, a_1)$$

where $F_{Y|X,A_1}^{-1}(a_2; x, a_1)$ is the conditional a_2 quantile of Y given $X = x, A_1 = a_1$.

---

[1]As recalled in Imbens and Newey 2009, assumption $[A'5]$ is stronger than the usual rank condition on the function $F_{P|Z,W}$ and ensures there is a one-to-one mapping between the two variables for any values x, that is required to characterize the change of variable between W and A_1 for any values x.

Given identification of φ, identification of λ goes through with the augmented set of regressors (X, A_1), and an obvious adaptation in the regularity conditions. There are three main scenarios in which this structure can arise, and we believe our application to contain elements of all three of them. Because they are prototypical, we list them in the following:

The first is *simultaneity*, i.e., we assume that prices and incomes are determined by a two equation demand and supply system, where quantities Y are a function of prices P, other determinants Z, and unobservables A_2, while prices would be determined by a quantities, an exogenous cost shifter W, in our case the distance from the (refineries at the) Gulf of Mexico also used in BHP (2010a) which determines transportation costs, as well as other unobservables. We would like to rearrange this system of equation to a triangular above, which is monotonic in A_1, given Z, W. This is not possible in general, however, Blundell and Matzkin (2010) provide conditions under which this holds, in particular, a full rank and a control function separability condition. While the former is less controversial, the latter places nontrivial structure on the unobserved structural equations.

The second one is that the *true structural model is triangular* from the outset: In this interpretation, to fix ideas, think of A_1 as a part of preferences that reflects an attitude towards public goods, in particular, the higher A_1 the more individuals care about the environment. Prices are ceteris paribus higher in areas where the taxes are high, which reflects a population with a higher willingness to sacrifice money for a clean environment. This causes correlation as the driving behavior and the price may have joint determinants. To complete the description of variables, A_2 may reflect a desire for driving, in parts determined by factors like distance to school and workplace that we only partially control for. We assume that these are independent of A_1 and X, and enter monotonically.

Controlling for the distance to the Gulf as well as compositional effects of the population (e.g., how many people live in rural areas), the differences in prices may well be attributed to different attitudes towards public goods like the environment and towards taxes: Ceteris paribus prices are high were individuals are less concerned by paying a higher tax to support public (environmental) issues. Therefore we can use this second equation to isolate the control functions A_1, which captures the feature in the individuals' preference ordering - in our application the willingness to accept higher taxes - that is correlated with price. Once we control for this factor, the remaining unobserved heterogeneity (in our application, the desire to drive) is orthogonal to prices and can be dealt with in the same fashion as before. This scenario is of course not directly compatible with the first, as the structural models are different.

Finally, there may be *measurement error*. Prices in our application are averages across counties; individual specific prices may differ from that and the deviation is hence contained in

the error. Observe that the averages of these differences may vary from county to county. If we think of Z in the h relationship to be independent of the measurement error on individual level, then the same is also true for the county level. Moreover, for any given $Z = z$, the average price in a county varies with the average measurement error in the county, the larger and positive the error, the larger P, and the larger and negative the average error is, the smaller P. Hence both monotonicity and independence in this equation may be warranted.

To argue the conditional monotonicity in the demand equation is harder: First, for the true price P^*, we invoke the standard assumption that $P = P^* + \eta$, with $\eta \perp P^*, Z$, as argued above. Finally, we assume that $A_2 = \eta + \tilde{A}_2$ has the same interpretation as in the first example. We strengthen the marginal independence assumptions to $\left(A_1, \tilde{A}_2, \eta\right) \perp Z$, which implies our independence assumptions. What is more debatable in this scenario is the monotonicity in the index ξ, where $\xi = \eta + \tilde{A}_2$; at this stage we simply remark that this strictly generalizes the classical approach to measurement errors in the linear regression model.

3   Estimation and Asymptotic Properties

The data consists of i.i.d. observations $\{(Y_i, X_i, W_i) : i = 1, ..., n\}$ where $X_i = (P_i, S_i, \tilde{Z}_i)$. In what follows, we use nonparametric kernel method to estimate the demand function as well as the consumer surplus.

3.1   Exogenous Regressors

Estimation. In the case of exogenous regressors, the nonparametric counterpart of the demand function φ is derived from Matzkin (2003) as $\hat{\varphi}(x, a) = \hat{F}^{-1}_{Y|X}(a; x)$ where $\hat{F}^{-1}_{Y|X}(a; x)$ represents the kernel estimator of the a quantile of Y given $X = x$.

The function $\widehat{\lambda}(p)$ is then defined as solution of the estimated differential equation system:

$$\begin{aligned} \widehat{\lambda}'(p) &= \widehat{\varphi}(p, s + \widehat{\lambda}(p), \tilde{z}, a) &\qquad (3.1) \\ \widehat{\lambda}(p^0) &= 0, \end{aligned}$$

The solution can be approximated using numerical methods. Various classical algorithms can be used to calculate a solution, like the Euler-Cauchy algorithm, Heun's method, the Runge Kutta method, or the Buerlisch-Stoer algorithm (as in Hausman and Newey (1995)). Let us briefly outline the general methodology. Consider a grid of equidistant points $p_1, ..., p_n$ where $p_{i+1} = p_i + h$ and $p_0 = p^0$. The differential equation is transformed into a discretized version

where $\widehat{\varphi}_{\tilde{h}}$ is an approximation of $\widehat{\varphi}$.:

$$
\begin{cases}
\widehat{\lambda}_{(i+1)} & = \widehat{\lambda}_i + \tilde{h}\widehat{\varphi}_{\tilde{h}}(p_i, s + \widehat{\lambda}_i, \tilde{z}, a) \\
\widehat{\lambda}_0 & = 0.
\end{cases}
\tag{3.2}
$$

In the particular case of the Euler algorithm, $\widehat{\varphi}_{\tilde{h}} = \widehat{\varphi}$. By similar arguments as discussed in Vanhems (2006) for the mean regression case, the numerical approximation of $\widehat{\lambda}$ does not impact the theoretical properties of the estimator since the steps involving numerical approximation can be chosen to have a faster rate of convergence than the nonparametric estimation methods employed.

Asymptotic properties. Consistency and asymptotic normality of the estimator $\hat{\varphi}$ mainly follows from Matzkin (2003). We present the distribution theory in two theorems, depending on whether the regressors are exogenous or not.

In order to derive rates of convergence for $\widehat{\lambda}(p)$, we need to make the link between the solution λ and the function φ explicit. The main issue of this differential inverse problem is its nonlinearity. The methodology used to transform the nonlinear equation into a linear problem is closely related to the functional delta method. Under the assumptions of existence, uniqueness and stability of $\widehat{\lambda}$ and λ, it can be established that:

$$
\forall p \in I, \widehat{\lambda}(p) - \lambda(p) = I(p) + R(p)
\tag{3.3}
$$

where the first term $I(p)$ is linear in $\widehat{F} - F$ and $R_n = o_P\left(\left\|\widehat{F} - F\right\|'\right)$ where F is the cdf of (Y, X) and the norm $\|.\|'$ is a Sobolev norm defined in Appendix I.

Introducing this expansion enables us to transform the nonlinear problem into a linear one, up to a residual term that converges faster. Obviously, under the condition that both terms converge, our estimator is consistent. More precisely, we can analyze the behavior of each term:

- the linear part $I(p)$. The rate of convergence of the estimated solution of the differential equation is expected to be faster than the rate of convergence of the estimator of the function φ since there is a gain in regularity obtained by integration.

- the residual term $R(p)$, which is the counterpart of the remainder in the Taylor expansion.

In the exogenous regressors case, we need the following assumptions (to simplify the notations, we consider a one dimension kernel function K with a generic bandwidth parameter h):

[B1] The random tuples (Y_i, P_i, Z_i), $i = 1, ..., n$, are $i.i.d.$

12

[B2] The density $f(y, p, z)$ of (Y, P, Z) has compact support $\Theta \subset \mathbb{R}^{3+L}$ and is continuously differentiable up to the order $s' \geq 2$.

[B3] The kernel function K vanishes outside a compact set, integrates to 1, is continuously differentiable of order s' with Lipschitz derivatives up to s', and is of order s'.

[B4] As $n- > \infty$, $h- > 0$, $\frac{ln(n)}{nh^{L+4}}- > 0$, $h^{s'}\sqrt{nh^{2(L+2)}}- > 0$, $\sqrt{nh^{L+1}}- > \infty$ where h is the bandwidth parameter associated with kernel estimation

[B5] $0 < f(p, z) < \infty$ for $(p, z) \in \Theta_X$

Then, following Matzkin (2003), for $s' = 2$, it can be shown that the nonparametric quantile estimator is consistent and converges asymptotically pointwise to a normal distribution at rate $\sqrt{nh^{L+2}}$.

The next theorem proves consistency and asymptotic normality for the estimated surplus.

Theorem 1. *Suppose that assumptions $[A1] - [A2]$, $[B1] - [B5]$ are satisfied with $s' = 2$ and consider fixed values $s, \tilde{z}, a$. Moreover, assume that the assumptions required for identification hold. Then, the estimated solution $\widehat{\lambda}$ is unique in a neighborhood I of p_0. Moreover, we get, for all $p \in I$:*

$$\sqrt{nh^{L+1}}(\hat{\lambda}(p) - \lambda(p)) \xrightarrow[n \to \infty]{d} \mathcal{N}(0, V) \text{ in distribution}$$

where

$$V = \frac{1}{nh^{L+1}}\|K\|_2^2 \int_{p^0}^{p} \gamma^2(p, t, s, \tilde{z}, a)var\left[1(Y \leq \varphi(t, s + \lambda(t), \tilde{z}, a)|P = t, S = s + \lambda(t), \tilde{Z} = \tilde{z}\right] dt$$

and $\gamma(p, t, s, \tilde{z}, a) = \frac{exp\left[\int_t^p \frac{\partial \varphi}{\partial e_2}(u, s+\lambda(u), \tilde{z}, a)du\right]}{f_{Y|X}(\varphi(t, s+\lambda(t), \tilde{z}, a), t, s+\lambda(t), \tilde{z})}$

Corollary 5. *Under the assumptions of the previous theorem, with the assumption that $s' = 2$, we derive the asymptotic mean square error for the linear term $\forall p \in I$, $\mathbb{E}[I(p)^2] = (V + B^2).(1 + o(1))$ where V is the asymptotic variance and the asymptotic squared bias B^2 is equal to:*

$$\begin{aligned}
B^2 &= \frac{h^4}{4}\left(\int u^2 K(u)du\right)^2 \times [\int_{p^0}^{p} \frac{\gamma(p, t, \tilde{z}, a)}{f(t, s+\lambda(t), \tilde{z})}\int(a - 1(y \leq \varphi(t, s+\lambda(t), \tilde{z}, a))) \\
&\times \left(\sum_{e_k} \frac{\partial^2 f}{\partial e_k^2}(y, t, s+\lambda(t), \tilde{z})\right)dydt]^2
\end{aligned}$$

where $\frac{\partial^2 f}{\partial e_k^2}$ denotes the second order derivative of f with respect to the argument e_k. Under the additional assumption that the kernel function K is of order 3 and the density function f is continuously differentiable of order 3 with respect to z, we obtain that $B^2 = O(h^6)$.

Note that the rate of convergence obtained for the estimated surplus is faster than for the estimated demand function. This gain is due to the smoothing effect involved by solving the differential equation. Moreover, the kernel of order 3 assumption allows to reduce the bias term further. In either case, we have to undersmooth our surplus estimator compared to the optimal choice of bandwidth parameter for the estimation of the demand function.

3.2 Endogenous regressors

In the case of endogenous regressors, we first need to estimate the regressor A_1. We define the observed heterogeneity $A_{1i} = F_{P|Z,W}(P_i, Z_i, W_i) : i = 1, ..., n$ where $F_{P|Z,W}(p, z, w)$ represents the conditional cdf of (P, Z, W) and denote by $\hat{A}_{1i} = \hat{F}_{P|Z,W}(P_i, Z_i, W_i)$ the associated kernel estimator. To simplify the formula, we consider two kernel functions $K_1 : \mathbb{R}- > \mathbb{R}$ and $K_2 : \mathbb{R}^{2+L}- > \mathbb{R}$ and we denote by h the generic bandwidth parameter. The estimated heterogeneity variable $\hat{A}_1$ is defined as follows:

$$\hat{A}_{1i} = \frac{\sum_{j=1,j\neq i}^{n} \tilde{K}_1(\frac{P_i-P_j}{h})K_2(\frac{Z_i-Z_j}{h}, \frac{W_i-W_j}{h})}{\sum_{j=1,j\neq i}^{n} K_2(\frac{Z_i-Z_j}{h}, \frac{W_i-W_j}{h})}$$

where $\tilde{K}_1(u) = \int_{-\infty}^{u} K_1(s)ds$. A nonparametric estimator for φ is then given by

$$\hat{\varphi}(x, a) = \hat{F}_{Y|X,\hat{A}_1}^{-1}(a_2; x, a_1) \tag{3.4}$$

The numerical computation of the associated estimated surplus follows the same steps as in the exogenous case (i.e., using the numerical algorithm presented in (3.2)).

In order to derive asymptotic properties for $\hat{\lambda}$, we follow the same methodology as in the exogenous case and make use of the following assumptions:

[B'1] The random tuples (Y_i, P_i, Z_i, W_i), $i = 1, ..., n$, are $i.i.d.$

[B'2] the density $f(y, p, z, w)$ has compact support $\Theta \subset \mathbb{R}^{4+L}$ and is continuously differentiable up to the order $s' \geq 2$.

[B'3] The kernel function K vanishes outside a compact set, integrates to 1, is continuously differentiable of order s' with Lipschitz derivatives up to s', and is of order s'.

[B'4] As $n- > \infty$, $h- > 0$, $\frac{ln(n)}{nh^{L+5}}- > 0$ and $h^{s'}\sqrt{nh^{2(L+3)}}- > 0$, $\sqrt{nh^{L+2}}- > \infty$ where h is the bandwidth parameter associated with kernel estimation

[B'5] $0 < f(p, z, w) < \infty$ for $(p, z, w) \in \Theta_{X,W}$ where $\Theta_{X,W}$ is the compact support of X, W.

Under these assumptions, the following theorem establishes consistency and asymptotic normality for the associated estimated heterogeneous surplus:

Theorem 2. *Suppose that assumptions $[A'1] - [A'5]$, $[B'1] - [B'5]$ with $s' = 2$ are satisfied and consider fixed values $s, \tilde{z}, a$. Then, there exists a unique consistent estimated solution $\widehat{\lambda}$ which is defined on a common neighborhood I of p^0 with the true solution λ. Moreover, we get, for all $p \in I$:*

$$\sqrt{nh^{L+2}}(\hat{\lambda}(p) - \lambda(p)) \to \mathcal{N}(0, V) \ \text{in distribution}$$

where

$$V = \frac{1}{nh^{L+2}} \|K\|_2^2 \int_{p^0}^{p} \gamma^2(p, t, s, \tilde{z}, a) var\left[1(Y \leq \varphi(t, s + \lambda(t), \tilde{z}, a)|P = t, S = s + \lambda(t), \tilde{Z} = \tilde{z}, A_1 = a_1\right] dt$$

and $\gamma(p, t, s, \tilde{z}, a) = \dfrac{exp\left[\int_t^p \frac{\partial \varphi}{\partial e_2}(u, s+\lambda(u), \tilde{z}, a) du\right]}{f_{Y|X, A_1}(\varphi(t, s+\lambda(t), \tilde{z}, a), t, s+\lambda(t), \tilde{z}, a_1)}$

As we can see from the previous section, plugging a nonparametric estimator of the control variable on the surplus function gives similar results as in the exogenous case with one supplementary regressor.

Corollary 6. *Under the assumptions of the previous theorem, with the assumption that $s' = 2$, we derive the asymptotic mean square error for the linear term $\forall p \in I$, $\mathbb{E}[I(p)^2] = (V + B^2)(1 + o(1))$ where V is the asymptotic variance and the asymptotic squared bias B^2 is equal to:*

$$
\begin{aligned}
B^2 \ =& \ \frac{h^4}{4}\left(\int u^2 K(u) du\right)^2 \times [\int_{p^0}^{p} \frac{\gamma(p, t, \tilde{z}, a)}{f(t, s+\lambda(t), \tilde{z}, a_1)} \int (a_2 - 1(y \leq \varphi(t, s+\lambda(t), \tilde{z}, a))) \\
& \times \left(\sum_{e_k} \frac{\partial^2 f}{\partial e_k^2}(y, t, s+\lambda(t), \tilde{z}, a_1)\right) dy dt]^2
\end{aligned}
$$

Under the additional assumption that the kernel function K is of order 3 and under the assumption that the density function f is continuously differentiable of order 3 with respect to z and a_1, we obtain that $B^2 = O(h^6)$.

4   Application

This section discusses the details of the empirical implementation. We start our discussion by presenting the data employed, which are similar to the data used by Blundell, Horowitz and Parey (2010a). Then we present the details of the kernel based estimation procedure. Finally, we show the results of our (first) experiment, where we consider an increase in the price from p_0 to p, for various choices of p, and a (arbitrary) fixed value p_0.

4.1 Data Description

The data we use come from the 2001 National Household Travel Survey (NHTS), which was conducted between March 19th, 2001 and May 9th, 2002 under the sponsorship of the Bureau of Transportation Statistics (BTS), the Federal Highway Administration (FHWA) and also the National Highway Traffic Safety Administration (NHTSA). The data are essentially identical to the ones used by Blundell, Horowitz and Parey (2010a, henceforth BHP). As discussed, we extend their analysis by focusing on welfare effects.

The NHTS is a survey of the civilian, non-institutionalized population of the U.S. that collects a) information on household characteristics such as income, education, size and further demographics b) data on each household vehicle, including year, model, make and estimates of annual miles traveled and c) precise information on trips made in designated periods of time, which is of minor importance for our purposes. Household and most vehicle information were gathered via telephone interviews and complemented by written travel diaries and odometer readings. The households are sampled from a random-dialing list of telephone numbers[2] that covers all geographic areas of the U.S. Eventually, interviews were conducted in all 50 states and the District of Columbia.

The key variables used in our analysis are gasoline consumption, price per gallon of gasoline and household income. Gasoline consumption is derived from odometer readings and estimates of the vehicle fuel economy (miles per gallon), and is aggregated over different vehicles owned by the household [3].

Gasoline prices represent a weighted average of monthly prices, including taxes, provided by the U.S. Energy Information Administration (EIA) at the state level. The NHTS made use of monthly fuel economy estimates per vehicle (these take individual driving circumstances such as temperature, wind and traffic into account) and the distribution of traveled miles over the course of the year to estimate the level of fuel consumption by month. Gasoline prices are then derived by dividing the households fuel expenditures by the level of his fuel consumption. Households report their annual income, before taxes, in 18 different ranges[4]. We set the households income equal to the midpoint of the respective interval and assigned an income of $120,000 if households reported to earn more than $100,000 annually[5].

---

[2]This excludes telephones in motels, hotels, group quarters, such as nursing homes, prisons, barracks, convents and monasteries and any living quarters with 10 or more unrelated roommates.

[3]See Appendix J and K of ORNL for a detailed description, http://nhts.ornl.gov/2001/usersguide/UsersGuide.pdf.

[4]See Appendix E for the various sources of income.

[5]This benchmark is taken from Blundell, Horowitz and Parey (2009), who estimate the first two moments of a log-normal income distribution. Dropping very high incomes, above $150,000, suggests an average income

We devote our attention to households in the national sample that provide information on all of the three key variables. We exclude those households that are located in Hawaii and those who do not report any drivers. Finally, we drop vehicles that use diesel, electricity or natural gas as fuel and end up with a sample size of 22,204 observations. Table 1 gives an overview on both key variables and further household plus regional characteristics.

Table 1: Summary Table

	Mean	10%	Median	90%	Stdv
Gasoline Demand in 100 Gallons	12.03	2.63	9.75	23.66	10.12
Gasoline Price in $ per Gallon	1.33	1.24	1.34	1.44	0.08
Annual HH Income in 1000 $	53.77	17.50	47.50	120.00	33.85
Distance of State From Gulf in 1000 km	1.73	0.88	1.59	2.86	0.72
# of Drivers per HH	1.92	1.00	2.00	3.00	0.74
HH Size	2.64	1.00	2.00	4.00	1.36
Mean Age of Drivers	48.15	29.50	45.33	72.00	15.91
Some College Education (Highest HH)	0.67	0.00	1.00	1.00	0.47
Rail in Metropolitan Statistical Area	0.23	0.00	0.00	1.00	0.42
Pop. Dens. [100 Pers./Block]	38.61	0.50	15.00	70.00	53.85
Rural Area	0.23	0.00	0.00	1.00	0.42
Small Town	0.24	0.00	0.00	1.00	0.43
Suburban Area	0.24	0.00	0.00	1.00	0.43
Second City	0.18	0.00	0.00	1.00	0.38
Urban Area	0.10	0.00	0.00	1.00	0.31

In the appendix we also display tables that report the results of standard demand analysis. Specifically, in table A.1 we report the results of a log log regression of log gasoline demand on log own price, log income, and several dummies that indicate geographical regions, varying degrees of urbanity, varying population density, the availability of public transportation, and the mean age of the driver. The result display very much the expected signs and magnitudes. The own price elasticities is with -0.45 somewhat at the lower end of the usually reported results, which range between -0.6 and -0.9, see Hausman and Newey (1995), Yatchew and No (2001) and Schmalensee and Stoker (1999), but fits exactly with the results in BHP, which may

---

of $120,000 in this upper income bracket.

in parts be due that more (low elasticity) medium grade gasoline is consumed in our data. The income elasticity is 0.31, which is in line with the entire literature. Finally, the more urban the areas, the less gasoline individuals consume, on average 20-30%, depending on the specification, which is in line with the literature. Population density has an effect in excess of urbanity, which again matches BHP. The age of the driver has only a limited effect, and the same is true of the availability of public transport. The significance of both variables depends on the specification. Other than this, changes in the specification, e.g., removing the marginally significant variables, do not have a fundamental effect on the results. In particular, the elasticity of own price is marginally higher in absolute value, but remains around - 0.5, see also BHP for various similar specifications. Like Yatchew and No (2001) we find that correcting for household demographics reduces the price elasticity somewhat, compared to results by Hausman and Newey (1995).

The result of correcting for endogeneity using the control function version of 2SLS are displayed in table 2. The relevant own price elasticity is with -0.77, while the coefficients on the other variables remain materially unchanged. Changes in the specification as above yield generally to own price coefficients that are, if anything, marginally higher, so that the effect is around -0.65. The coefficient on the control functions is significant, indicating that correcting for endogeneity is important, and results in a more price elastic demand, with coefficients that are about 20-40% larger in absolute value. With this in the back of our mind, one would expect the welfare effects to be larger. While all of the three above reasons for the validity of the distance from the Gulf of Mexico as an instrument, we personally view the measurement error component as the largest contributor to the difference in results, because the fact that the exogenous parametric coefficient is much smaller (see also BHR, (2010a)) compared to the literature (see Yatchew and No (2001) for an overview), seems to be indicative of attenuation associated with measurement error. As in any given application, the reality is probably more complex and has not only features of exactly one explanation. The overriding issue for our analysis, however, is that the levels are almost not affected by the correction for endogeneity, as we shall see below.

Since the focus of this paper is on heterogeneity using quantiles, we have also implemented linear quantile regression models. The price coefficient does not change materially when we perform median regression. In the comparable specification to table A.1, the price elasticity is - 0.43, and the income elasticity is 0.28. Materially, the same other variables remain significant. The price elasticities seem to decrease in absolute values for lower deciles (to -0.28 for the tenth, to be specific), and stay approximately constant for higher quantiles. The overriding feature is the change in intercept across quantiles; in every other respect they look like parallel lines. If we add control function residuals, the price elasticities increase again in absolute value to

about -0.75, depending on the specification, confirming that correcting for endogeneity has a material impact on the outcomes.

4.2 Details of the Econometric Implementation

When using the above data, we are mainly concerned with the relationship between demand, income and prices. Consequently, $\tilde{Z}$ are not of primary importance and act only as controls, and we thus reduce them to two approximately continuously distributed principal components, which capture the bulk of the variation. While this is arguably ad hoc, we have experimented with varying the number of principal components, as well as selecting different ones, without materially affecting the results (which are available from the authors upon request). Hence, we feel that our approach is justified as it allows the use of nonparametric quantile methods.

We implement two different estimators for the quantile regressions: first, we implement our estimator as described in the text, where all estimates of conditional distributions are obtained by nonparametric kernel estimators. The inversion we have to perform is computationally quite expensive, in particular since we also obtain the standard errors via bootstrap. Hence we also apply a quantile estimator as the optimizer of a local linear quantile regression problem, which is asymptotically equivalent to the estimator we have analyzed in the theoretical part. In either case, we make use of a standard second order Epanechnikov kernel.

Also in either case, our large sample theory suggests that the integration step involved in computing the welfare effect acts reversely to estimating derivatives - it increases the rate of convergence. Hence we chose the bandwidth by first performing cross validation for the conditional mean, and then choosing a smaller bandwidth. There is no theoretical guidance on how much smaller the bandwidth should be chosen; however, our results were not sensitive to changes in the bandwidth. The integration was performed by ordering the prices from p_0 to p as $p_0, p_1, ..., p_{T-1}, p_T$, $p_T = p$, and computing $\hat{\lambda}(p)$ recursively through

$$\widehat{\lambda}(p_i) = \widehat{\lambda}(p_{i-1}) + (p_i - p_{i-1})\,\hat{k}^{\alpha}_{Y|X}(p_{i-1}, q + \widehat{\lambda}(p_{i-1}), z), \quad i = 1, ..., T. \tag{4.1}$$

for a consumer with $S = q, Z = z$, and $A = a$. In the endogenous case, we first estimate A_1 by $\widehat{A}_1 = \widehat{F}_{P|S,Z,W}(P; S, Z, W)$, where $\widehat{F}_{P|S,Z,W}(p; q, z, w)$ is a local linear mean regression estimator of the conditional cdf, using $1\{P \leq p\}$ as dependent variable, and S, Z, W as regressors. The bandwidth in this regression in chosen by cross validation to focus on eliminating the bias, since the variance of this estimation error averages out. Then, in (4.1), we simply replace $\hat{k}^{\alpha}_{Y|X}(p_{i-1}, q - \widehat{\lambda}(p_{i-1}), z)$ by $\hat{k}^{\alpha}_{Y|XA_1}(p_{i-1}, q + \widehat{\lambda}(p_{i-1}), \hat{a}_1, z)$.

Standard errors are obtained using the bootstrap, with an undersmoothed bandwidth, as

is common in the nonparametric literature to account for potential biases. In particular, we draw from the data with replacement a similar sized sample, and apply the same procedure as above.

4.3 Results

The policy experiment that we are conducting is increasing the price of gasoline from the (median) level of USD 1.31 to USD 1.40, equaling a 7% increase in gas price. We already start with the most general setup where the price is assumed to be endogenous. We focus on compensating variation, as computed in equation (4.1) using the same mechanism as outlined in the previous subsection, for a continuous increase from 1.31 to 1.40 for different quantiles. The results at mean characteristics and income is shown in Fig. 1 and Fig.2. Specifically, we look at the 10, 30, and 50th percentile (i.e., the median) of the conditional distribution of the demand for gasoline in Fig.1, and at the 50, 70 and 90th of the conditional distribution of the demand for gasoline in Fig.2.

—— Fig. 1 approx here —-

The median effect of the price change on welfare is slightly below 123 USD (the point estimate is 122.81), which is plausible given the summary statistic. Indeed, back of the envelope calculations reveal that this is approximately the order of magnitude we expect[6]. The standard errors around this quantity are rather tight, indication of the fact that the integration really stabilizes estimation. Since demand only shrinks by about 4% across the price range of our experiment, the income effect is rather small, and the income elasticity of demand is not very large, the linearity of effects is to be expected - the quantiles of demand simply do not vary a lot, so the essential part of the welfare effect is that of the price change. As is shown below, welfare also does not vary too much according to demographics. We conclude that the median welfare effect is largely as we would have expected.

What is not revealed, however, when looking at the median is an astonishing variation by quantiles. While the effects are less than half as large as the median for the first decile (55.26), they are almost twice as large for the 9th decile (239.01). Put another way, the effect is close to five fold as strong on the 9th decile than on the first, compare Fig.1 and Fig.2:

---

[6]Given parametric elasticities of demand of around -0.5 and a price change of 7%, we expect to see a reduction in gas demand of 3-4%, i.e., demand does not vary too much. If it were constant, at conditional median gasoline demand of 1450 gallons, an increase of 9 cents per gallon means a value of 130 USD. The difference is readily explained by our detailed analysis that takes substitution effects into account, as well as the sampling error.

—— Fig. 2 approx here —-

These results do not change significantly, if we do not control for endogeneity. The order of magnitude of the decrease in CV is around 1% for the median, see fig. 3. It is somewhat higher at the upper end of the distribution.

—— Fig. 3 approx here —-

While we consider the specification with control function residuals to produce the more plausible results in general, the welfare effects do not differ significantly. This is in line with the nonparametric mean regression results in BHP (2010a), who have performed a nonparametric test for endogeneity in the (nonparametric) mean regression case, and concluded that the regressors are not endogenous. The results are even smaller than in the parametric specification employed above. One reason why the difference in welfare effects are smaller may be due to the fact that the linear model is misspecified, and thus overemphasizes the effect of endogeneity.

As was to be expected given the asymptotic results and the tight standard errors, the difference in estimation methods is neglectable. At the median, the estimator which is based on the inversion of the cdf produces within 1% of the same result, and the difference is not statistically significant, see Fig.4. The same is true at other quantiles.

—— Fig. 4 approx here —-

In the following we slice the population by demographics, however, we do retain the correction for endogeneity. In particular, from the log-log specification we conclude that the question of residence in an urban environment plays are large role. Therefore we compare urban versus rural households by stratifying the population, and performing all of our analysis on two separate samples, including the conditioning on covariates. We find, as was to be expected, larger welfare effects for the rural households which have to commute more, see Fig. 5:

—— Fig. 5 approx here —-

As in the parametric regression example, rural households have a 30% higher welfare effect. In other analysis we found that the spread in results is approximately comparable between both urban and rural households. Similar results where obtain when the population is sliced according to population density. This is compatible with a theory where driving is determined by your needs (i.e., go to work, dive kids to school, go shopping etc.), and the larger distances in rural areas account for larger welfare effects. However, both in rural and in urban areas,

there are enormous differences between households within each subpopulation that are larger in magnitude than the differences between populations.

In summary, we find pronounced variations in welfare effects between households, using our method that dwarf effects of household covariates. This seems to underscore the need for accounting for unobserved heterogeneity. The downside is that any method that point identifies the distribution of welfare effects has to impose some structure, in our case a scalar monotonic error term. In a companion paper (Hoderlein and Vanhems (2011)), we develop an approach that allows for random coefficients. Both approaches are non-nested, and give potentially diverging results, a fact that one has to bear in mind.

5   Summary and Outlook

This paper proposes a framework to model empirically direct welfare effects in a population of heterogeneous consumers, as are associated with, e.g., the introduction of a tax on gasoline. We aim in particular at modeling the heterogeneity in effects. Using nonseparable models combined with monotonicity assumptions, we identify the variation in consumer data with preference heterogeneity, which may be restrictive. Under this assumption, however, we can precisely characterize the distribution of welfare effects. For every consumer (characterized by an either one or two dimensional unobserved parameter) it is given by the solution to a partial differential equation. The parameters vary from individual to individual, but are point identified from the distribution of the data. Given estimators of the cumulative distribution function (cdf), we then propose an estimator for the distribution of welfare effects. Moreover, using nonparametric estimators of the cdf as building blocks, we can characterize the large sample behavior of our estimator.

When implementing our estimator with US data from the early 2000s, we find a large spread of welfare effects across the population. Indeed, a gasoline price change of nine cents, from USD 1.31 per gallon to USD 1.40 per gallon has a welfare effect on the median person of 123 USD; however, the effect at the 90% is more than 110 USD higher, while the 10 th percentile is, with around USD 55, hardly affected (and this does not even include the subpopulation who does not drive at all). While these estimates may be slightly inflated due to the fact that we identify all the observed variation with preference heterogeneity, the fact that we observe few outliers as well as implausible values lead us to believe that the order of magnitude of the variation is essentially correct. However, a more detailed analysis using repeated measurements and a corresponding econometric framework is definitely required to underscore these findings.

Finally, we also believe that different specifications for the unobserved heterogeneity should be analyzed. If one insists on point identification of individual effects, a natural alternative are random coefficient models which allow for several sources of unobserved heterogeneity at the expense of constraining the functional form of individual demands. Indeed, in ongoing research, we analyze the same question - heterogeneity in welfare effects - in such a framework, and we plan a comparison of the findings between the two approaches. Given the findings of this paper, we believe the issue of modeling heterogeneity to be of great importance for the evaluation of welfare effects of economic policies in the future.

References

[1] Altonji, J., and R. Matzkin, 2005. Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors, *Econometrica*, 73, 1053 - 1103.

[2] Blundell, R., Horowitz J., and M. Parey, 2010a. Measuring the Price Responsiveness of Gasoline Demand, Economic Shape Restrictions and Nonparametric Demand Estimation CeMMAP working papers CWP11/09, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

[3] Blundell R., Horowitz J., and M. Parey, 2010b. Semi-nonparametric Estimation of a Nonseparable Demand Function under Shape Restrictions, slides, IFS web page.

[4] Blundell, R., D. Kristensen and R. Matzkin 2013, Bounding Quantile Demand Functions using Revealed Preference Inequalities, *Journal of Econometrics*, forthcoming.

[5] Blundell, R., and R. Matzkin 2010, Conditions for the Existence of Control Functions in Nonseparable Simultaneous Equations Models, CeMMAP working papers CWP28/10.

[6] Cherchye, L., B. De Rock, and F. Vermeulen, 2007. The Collective Model of Household Consumption: A Nonparametric Characterization, *Econometrica*, 75, 553-574.

[7] Crawford, I., and K. Pendakur, 2012, How Many Types Are There?, *Economic Journal*, forthcoming.

[8] Deaton, A and J.Muellbauer, 1980, Economics and Consumer Behaviour,Cambridge University Press.

[9] D'Haultfoeuille, X., and P. Février, Identification of Nonseparable Models with Endogeneity and Discrete Instruments, Working Paper, CREST.

[10] Hall, P. and J.L. Horowitz, 2005. Nonparametric methods for inference in the presence of instrumental variables, *Annals of Statistics* **33**, 2904-2929.

[11] Hausman, J., 1981. Exact Consumers Surplus and Deadweight Loss, *American Economic Review* **71**, 662-676.

[12] Hausman, J. and W. Newey, 1995. Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss, *Econometrica* **63** , 1445-1476.

[13] Hausman, J. and W. Newey, 2011,Individual Heterogeneity and Average Welfare, *Working Paper*, MIT.

[14] Hoderlein, S., 2010, How Many Consumers are Rational?, *Working Paper*, Boston College.

[15] Hoderlein, S. and E. Mammen, 2009, Identification and Estimation of Marginal Effects in Nonseparable, Nonmonotonic Models, *Econometrics Journal*, **12**, 1-25.

[16] Hoderlein, S. and J. Klemelä and E. Mammen, 2010. Reconsidering the Random Coefficient Model, *Econometric Theory, forthcoming.*.

[17] Imbens, G., and W. Newey (2009) "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity Corresponding" *Econometrica*, Vol. 77, No. 5, pp 1481-1512

[18] Jorgensen, D, L. Lau and T. Stoker, 1982. The Transcendental Logarithmic Model of Aggregate Consumer Behaviour, *Advances in Econometrics* **1**, 97-238.

[19] Kirman, A., 1992. Whom or What Does the Representative Individual Represent, *Journal of Economic Perspectives* **6:2**, 117-136.

[20] Lewbel, A. (2001); Demand Systems With and Without Errors, *American Economic Review*, 611-18.

[21] Lewbel, A. and K. Pendakur, (2009), Tricks with Hicks: The EASI Implicit Marshallian Demand System for Unobserved Heterogeneity and Flexible Engel Curves, *American Economic Review*, 99(3): 827-63.

[22] Matzkin, R., 2003. Nonparametric estimation of nonadditive random functions, *Econometrica* **71:5**, 1339-1375.

24

[23] Matzkin, R., 2005. Heterogeneous Choice, for *Advances in Economics and Econometrics*, edited by Richard Blundell, Whitney Newey, and Torsten Persson, Cambridge University Press; presented at the Invited Symposium on Modeling Heterogeneity, World Congress of the Econometric Society, London, U.K.

[24] Schmalensee, R. and Stoker, T.M. 1999. Household Gasoline Demand in the United States, *Econometrica*, **67**, 645-662

[25] Slesnick, D., 1996. Empirical Approaches to the Measurement of Welfare, *Journal of Economic Literature* **4**, 2108-2165.

[26] Torgovitsky, A. (2011) "Identification and Estimation of Nonparametric Quantile Regressions with Endogeneity," Working Paper, Northwestern.

[27] Van der Vaart, A.W. and J.A. Wellner 1996. Weak Convergence and Empirical Processes with Applications to Statistics, *Springer Series in Statistics*.

[28] Vanhems, A. 2006 nonparametric study of solutions of differential equations, *Econometric Theory*, **22**, 127-157.

[29] Vanhems 2010 nonparametric estimation of exact consumer surplus with endogeneity in price, *Econometrics Journal*, **13**, 80-98.

[30] Vartia, Y., 1983. Efficient Methods of Measuring Welfare Change and Compensated Income in Terms of Ordinary Demand Functions, *Econometrica* **51:1**, 79-98

[31] Willig, R., 1976. Consumer's Surplus Without Apology, *American Economic Review* **66**, 589-597.

[32] Yatchew, A., and No 2001. Household Gasoline Demand in Canada, *Econometrica*, **69**, 1697-1709

6 Appendix I - Proofs

Proof of Theorem 1 This proof uses arguments from both Matzkin (2003) and Vanhems (2006). For any fixed values $(s, \tilde{z}, a)$, existence and uniqueness of the estimated surplus $\hat{\lambda}$ in I follows from the Cauchy-Lipschitz theorem. In order to establish consistency of the estimated solution $\widehat{\lambda}$, in addition to the regularity conditions discussed in section 2.2 we need

an additional assumption about the convergence of the Lipschitz factor k_n. This assumption will guarantee the stability of the estimated solution $\hat{\lambda}$ and its consistency and can be expressed using the derivatives of the function φ as follows (see Vanhems (2006) for more details): $\sup_{x,a} |\frac{\partial}{\partial e_2}\widehat{\varphi(x,a)} - \frac{\partial}{\partial e_2}\varphi(x,a)|$ converges to 0 a.s. where $\frac{\partial}{\partial e_2}$ denotes the derivative with respect to the second argument[7]. This stability condition is fulfilled thanks to Assumption [$B4$] and conditions on the rate of decay of the bandwidth parameter (see Vanhems (2006), Hoderlein and Mammen (2009)).

Under this last condition, both solutions $\hat{\lambda}$ and λ can be defined on a common subset I, and the inverse problem defined by the differential equation is stable and well-posed. Both solutions $\hat{\lambda}$ and λ can be characterized with the same operator Φ:

$$\begin{aligned}
\lambda(p) &= \Phi(F)(p) \\
\hat{\lambda}(p) &= \Phi(\hat{F})(p)
\end{aligned}$$

In order to derive the asymptotic normality result, we need to linearize the differential equation defined in (1.2). Let us first introduce some notation (cf. Matzkin (2003)). In what follows, F denotes the joint cdf of (Y, X), f denotes its probability density function (pdf) and $F_{Y|X}$ denotes the conditional cdf of Y given X. The function $F_{Y|X}^{-1}(a; x)$ applied to (x, a) denotes the inverse (in y) of the conditional cdf, i.e., the quantile function due to continuity of Y, evaluated at (x, a). To simplify the notations, $f(x)$ denotes the marginal pdf of X in x. For any continuously differentiable function $G : \mathbb{R}^{L+3} \to R$ we define the function $g(y, x) = \partial^{L+3} G(y, x)/\partial y \partial x$, $g(x) = \int g(y, x)dy$ and $G_{Y|X}(y, x) = \int_{-\infty}^{y} g(u, x)du/g(x)$. Let C denote a compact set in $\mathbb{R}^{L+3}$ that strictly includes Θ. Let $\mathcal{E}$ denote the set of all continuously differentiable functions $G : \mathbb{R}^{L+3} \to R$ such that $g(y, x)$ vanishes outside C.

Consider first the following operator Ψ defined by:

$$\begin{aligned}
\Psi : \mathcal{E} &\to C^1(\Theta_X \times [0, 1]) \\
G &\mapsto G_{Y|X}^{-1}
\end{aligned}$$

The space $C^1(\Theta_X \times [0, 1])$ is the space of continuously differentiable functions defined on $\Theta_X \times [0, 1]$ and Θ_X is the compact support of X. So, for all $(x, a) \in \Theta_X \times [0, 1]$,

$$\begin{aligned}
\Psi(F)(x, a) &= F_{Y|X}^{-1}(a; x) \\
&= \varphi(x, a)
\end{aligned}$$

---

[7]Note that since $x \in R^{L+2}$, $\frac{\partial}{\partial e_2}$ means the derivative with respect to the second argument of x

We also introduce the operator A defined by:

$$A : \mathcal{E} \times C^1_{\epsilon_1, \epsilon_2}(I) \quad \to \quad C(I)$$
$$(G, \lambda) \quad \mapsto \quad \lambda'(.) - \Psi(G)(., s + \lambda(.), \tilde{z}, a)$$

where $C(I)$ is the space of continuous functions defined on I and $C^1_{\epsilon_1, \epsilon_2}(I)$ is the space of continuously differentiable functions on I, satisfying both assumptions (i) and (ii) in section 2.2. Note that both spaces endowed with the L^2 norm $\|.\|$ are Banach spaces. Consider now the following norm on $C^1_{\epsilon_1, \epsilon_2}(I)$: $\forall v \in C^1_{\epsilon_1, \epsilon_2}(I), \|v\|' = max(\|v\|, \|v'\|)$. Then $\left(C^1_{\epsilon_1, \epsilon_2}(I), \|.\|' \right)$ and $(\mathcal{E}, \|.\|')$ are also Banach spaces. Following Matzkin (2003) and Vanhems (2006, 2010), it can be shown that both operators are continuous and continuously differentiable on the Banach spaces previously defined.

In the same vein as in Vanhems (2006), we apply the implicit function theorem to the operator A and define $\mathcal{F} \subset \mathcal{E}$ to be an open subset around the true cdf F, and $\mathcal{L}$ to be an open subset around λ such that: $\forall G \in \mathcal{F}, A(G, u) = 0$ has a unique solution in V. We denote by $u = \Phi(G)$ this unique solution, and by construction, Φ is continuously differentiable on $\mathcal{F}$. We can now differentiate the relation $A(G, u) = 0$ and apply it to (F, λ). For all $H = (H_1, H_2) \in \mathcal{F} \times \mathcal{L}$ and $\forall p \in I$, we obtain:

$$
\begin{aligned}
dA(F, \lambda)(H)(p) &= d_1 A(F, \lambda) dF(H)(p) + d_2 A(F, \lambda) d\lambda(H)(p) \\
&= d_1 A(F, \lambda) H_1(p, s + \lambda(p), \tilde{z}, a) + d_2 A(F, \lambda) H_2(p) \\
&= -d\Psi(F) H_1(p, s + \lambda(p), \tilde{z}, a) + H_2'(p) - \frac{\partial}{\partial e_2} \Psi(F)(p, s + \lambda(p), \tilde{z}, a) H_2(p) \\
&= 0
\end{aligned}
$$

So, the differential of A leads to a linear differential equation in H_2 that can be solved for H_2:

$$H_2(p) = \int_{p^0}^{p} d\Psi(F) H_1(t, s + \lambda(t), \tilde{z}, a) exp \left(\int_t^p \frac{\partial \Psi(F)}{\partial e_2}(u, s + \lambda(u), \tilde{z}, a) du \right) dt \qquad (6.1)$$

Next, compute the differential function $d\Psi(F) H_1(t, s + \lambda(t), \tilde{z}, a)$. Define first the two following operators: for any $G \in \mathcal{F}$, let

$$\Psi_1 : G \quad \mapsto \quad G_{Y|X}$$
$$\Psi_2 : G_{Y|X} \quad \mapsto \quad G_{Y|X}^{-1}$$

such that $\Psi(G) = \Psi_2 \circ \Psi_1(G)$. For all $H_1 \in \mathcal{F}$ we have:

$$
\begin{aligned}
\Psi(F + H_1) - \Psi(F) &= (F + H_1)_{Y|X}^{-1} - F_{Y|X}^{-1} \\
&= d\Psi(F)(H_1) + o(\|H_1\|)
\end{aligned}
$$

27

Then, following Matzkin (2003), for all $p \in I$, we obtain:

$$(F + H_1)_{Y|X} \left((F + H_1)_{Y|X}^{-1}(a; x), x \right) - (F + H_1)_{Y|X} \left(F_{Y|X}^{-1}(a; x), x \right)$$

$$= a - (F + H_1)_{Y|X} \left(F_{Y|X}^{-1}(a; x), x \right)$$

$$= \frac{\partial (F + H_1)_{Y|X}}{\partial e_1} \left(F_{Y|X}^{-1}(a; x), x \right) \cdot \left((F + H_1)_{Y|X}^{-1}(a; x) - F_{Y|X}^{-1}(a; x) \right) + o(\|H_1\|)$$

Therefore, we get:

$$(F + H_1)_{Y|X}^{-1}(a; x) - F_{Y|X}^{-1}(a; x) \;=\; \frac{a - (F + H_1)_{Y|X} \left(F_{Y|X}^{-1}(a; x), x \right)}{\frac{\partial (F+H_1)_{Y|X}}{\partial e_1} \left(F_{Y|X}^{-1}(a; x), x \right)} + o(\|H_1\|)$$

$$= \frac{-d\Psi_1(F)(H_1)(\left(F_{Y|X}^{-1}(a; x), x \right)}{\frac{\partial (F+H_1)_{Y|X}}{\partial e_1} \left(F_{Y|X}^{-1}(a; x), x \right)} + o(\|H_1\|)$$

$$= \frac{-d\Psi_1(F)(H_1)(\left(F_{Y|X}^{-1}(a; x), x \right)}{f_{Y|X} \left(F_{Y|X}^{-1}(a; x), x \right)} + o(\|H_1\|)$$

Again, following Theorem 1 of Matzkin (2003), we obtain:

$$d\Psi_1(F)(H_1)(\left(F_{Y|X}^{-1}(a; x), x \right) = \frac{ah(x) - \int_{-\infty}^{\varphi(x,a)} h(y, x)dy}{f(x)} + o(\|H_1\|)$$

Plugging these results into equation (6.1) for $x = (t, s + \lambda(t), \tilde{z})$ leads to:

$$H_2(p) = \int_{p^0}^{p} \frac{ah_1(t, s + \lambda(t), \tilde{z}) - \int_{-\infty}^{\varphi(t, s + \lambda(t), \tilde{z}, a)} h(y, t, s + \lambda(t), \tilde{z})dy}{f(t, s + \lambda(t), \tilde{z})} . \gamma(p, t, \tilde{z}, a) dt \qquad (6.2)$$

where

$$\gamma(p, t, \tilde{z}, a) = \frac{exp \left(\int_t^p \frac{\partial \varphi}{\partial e_2}(u, s + \lambda(u), \tilde{z}, a)du \right)}{f_{Y|X} \left(F_{Y|X}^{-1}(a; t, s + \lambda(t), \tilde{z}), t, s + \lambda(t), \tilde{z} \right)}$$

Finally, note that both solutions $\hat{\lambda}$ and λ can be characterized with the same operator Φ:

$$\lambda(p) = \Phi(F)(p)$$
$$\hat{\lambda}(p) = \Phi(\hat{F})(p)$$

The definition of differentiability of the operator Φ gives:

$$(\hat{\lambda} - \lambda)(p) = (\Phi(\hat{F}) - \Phi(\hat{F}))(p)$$
$$= d\Phi(F)(\hat{F} - F)(p) + o_P(\|\hat{F} - F\|')$$

28

Apply then equation (6.2) in $H_1 = \hat{F} - F$ and $H_2 = d\Phi(F)(\hat{F} - F)$ in order to get:

$$(\hat{\lambda} - \lambda)(p) = I(p) + R(p)$$

where

$$
\begin{aligned}
I(p) &= \int_{p^0}^{p} \frac{a(\hat{f} - f)(t, s + \lambda(t), \tilde{z}) - \int_{-\infty}^{\varphi(t, s + \lambda(t), \tilde{z}, a)} (\hat{f} - f)(y, t, s + \lambda(t), \tilde{z}) dy}{f(t, s + \lambda(t), \tilde{z})} . \gamma(p, t, \tilde{z}, a) dt \\
&= \int \int (\hat{f} - f)(y, t, s + \lambda(t), \tilde{z}) \left(a - 1(y \leq \varphi(t, s + \lambda(t), \tilde{z}, a)) \right) . 1(p^0 \leq t \leq p) . \frac{\gamma(p, t, \tilde{z}, a)}{f(t, s + \lambda(t), \tilde{z})} dy dt
\end{aligned}
$$

and

$$R(p) = o_P(\|\hat{F} - F\|')$$

The asymptotic normality result follows from Theorem 3.9.4 in Van der Vaart and Wellner (1996) and Assumption $[B4]$. The computation of the asymptotic variance is derived as follows (for simplicity, we consider two kernel functions $K_1 : \mathbb{R}- > \mathbb{R}$ and $K_2 : \mathbb{R}^L - > \mathbb{R}$ and we denote by h the generic bandwidth parameter):

$$
\begin{aligned}
\mathbf{var}\,(I(p)) = \frac{1}{nh^{2(L+3)}} \mathbf{var}[\quad &\int \int K_1(\frac{y - Y}{h}) K_1(\frac{t - P}{h}) K_1(\frac{\lambda(t) - S}{h}) K_2(\frac{\tilde{z} - \tilde{Z}}{h}) \\
. \quad &(a - 1(y \leq \varphi(t, s + \lambda(t), \tilde{z}, a))) . 1(p^0 \leq t \leq p) . \frac{\gamma(p, t, \tilde{z}, a)}{f(t, s + \lambda(t), \tilde{z})} dy dt]
\end{aligned}
$$

After changes of variables, we obtain:

$$
\begin{aligned}
\mathbf{var}\,(I(p)) = \frac{1}{nh^{2(L+1)}} \mathbf{var}[\quad &K_1(\frac{\lambda(P) - S}{h}) K_2(\frac{\tilde{z} - \tilde{Z}}{h}) (a - 1(Y \leq \varphi(t, s + \lambda(t), \tilde{z}, a))) \\
. \quad &1(p^0 \leq P \leq p) . \frac{\gamma(p, P, \tilde{z}, a)}{f(P, s + \lambda(P), \tilde{z})}] (1 + o(1))
\end{aligned}
$$

It then follows, by standard calculations, that:

$$
\begin{aligned}
\mathbf{var}\,(I(p)) &= \frac{1}{nh^{L+1}} \|K\|_2^2 \int_{p^0}^{p} \gamma^2(p, t, s, \tilde{z}, a) \\
&\quad . \quad \mathbf{var}\left[1(Y \leq \varphi(t, s + \lambda(t), \tilde{z}, a) | P = t, S = s + \lambda(t), \tilde{Z} = \tilde{z} \right] dt\, (1 + o(1))
\end{aligned}
$$

where K is a generic notation including K_1 and K_2. That concludes for the formula of the asymptotic variance.

Proof of Corollary 5 The result follows from standard calculus on the bias of $I(p)$:

$$E\,(I(p)) = \int_{p^0}^{p} \mathbb{E}\left[\int (\hat{f} - f)(y, t, s + \lambda(t), \tilde{z}) (a - 1(y \leq \varphi(t, s + \lambda(t), \tilde{z}, a))) \, dy \right] . \frac{\gamma(p, t, \tilde{z}, a)}{f(t, s + \lambda(t), \tilde{z})} dt$$

Therefore, we get:

$$
\begin{aligned}
E\left(I(p)\right)^2 &= \frac{h^4}{4}\left(\int u^2 K(u)du\right)^2 \times [\int_{p^0}^{p} \frac{\gamma(p,t,\tilde{z},a)}{f(t,s+\lambda(t),\tilde{z})} \int (a-1(y \le \varphi(t,s+\lambda(t),\tilde{z},a))) \\
&\quad \times \left(\sum_{e_k} \frac{\partial^2 f}{\partial e_k^2}(y,t,s+\lambda(t),\tilde{z})\right) dydt]^2(1+o(1))
\end{aligned}
$$

The faster rate of convergence for the bias term follows from Vanhems (2006), Theorem 4.2. It comes from the assumption of third order of the kernel function and the fact that the surplus is obtained by integrating the demand function over price. This means that we can go further in the Taylor expansion with respect to the price argument to derive a bias term with a faster rate.

Proof of Theorem 2 The way to proceed is very similar to the previous proof except that F represents now the cdf of (Y, X, W). Again, for any fixed values $(s, \tilde{z}, a)$, existence, uniqueness and consistency of the estimated surplus $\hat{\lambda}$ in I follows from Cauchy-Lipschitz theorem and the assumption of stability: $\sup_{x,a} |\frac{\partial}{\partial e_2}\widehat{\varphi(x,a)} - \frac{\partial}{\partial e_2}\varphi(x,a)|$ fulfilled thanks to Assumption $[B'4]$.

In order to prove the asymptotic normality result, we also need to linearize the differential equation defined in (1.2). The notations used are similar to the previous proof except that we consider now any continuously differentiable function $G : \mathbb{R}^{L+4} \to R$ and define the function $g(y,x,w) = \partial^{L+4}G(y,x,w)/\partial y \partial x \partial w$. Let C denote a compact set in $\mathbb{R}^{L+4}$ that strictly includes Θ. Let $\mathcal{E}$ denote the set of all continuously differentiable functions $G : \mathbb{R}^{L+4} \to R$ such that $g(y,x,w)$ vanishes outside C.

We consider the following operator Ψ defined by:

$$
\begin{aligned}
\Psi : \mathcal{E} &\to C^1(\Theta_X \times [0,1]^2) \\
G &\mapsto G_{Y|X,A_1}^{-1}
\end{aligned}
$$

The space $C^1(\Theta_X \times [0,1]^2)$ is the space of continuously differentiable functions defined on $\Theta_X \times [0,1]^2$. Compared to the decomposition given in the proof of Theorem 1, the operator Ψ is now defined using three functionals $\Psi = \Psi_2 \circ \Psi_1 \circ \Psi_0$ and each functional is defined by:

$$
\begin{aligned}
\Psi_0 : G &\mapsto G_{Y,X,A_1} \\
\Psi_1 : G_{Y,X,A_1} &\mapsto G_{Y|X,A_1} \\
\Psi_2 : G_{Y|X,A_1} &\mapsto G_{Y|X,A_1}^{-1}
\end{aligned}
$$

where $A_1 = G_{P|Z,W}(P;Z,W) = g_1(G)(P,Z,W)$. So, for all $(x,a) \in \Theta_X \times [0,1]^2$,

$$
\begin{aligned}
\Psi(F)(x,a) &= F_{Y|X,A_1}^{-1}(a_2; x, a_1) \\
&= \varphi(x,a)
\end{aligned}
$$

30

The differentiability of the operator Psi is, as in the previous proof, proved using continuity and differentiability on Banach spaces endowed with Sobolev norms. More precisely, under Assumption $[A'5]$, we get:

$$
\begin{aligned}
f_{YXA_1}(y, x, a_1) &= \frac{f_{YXW}(y, x, g_1(F_{YXW})^{-1}(a_1; x))}{\frac{\partial g_1(F_{YXW})}{\partial e_{L+2}}(x, g_1(F_{YXW})^{-1}(a_1; x))} \\
F_{YXA_1}(y, x, a_1) &= F_{YXW}(y, x, g_1(F_{YXW})^{-1}(a_1; x)) \\
\Psi(F_{YXW})(x, a) &= F_{Y|XW}^{-1}(a_2; x, g_1(F_{YXW})^{-1}(a_1; x))
\end{aligned}
$$

Using the previous definition for the operator A, we can derive the same equation as in (6.1) from the differential of A in the two arguments $(H_1, H_2) \in \mathcal{F} \times \mathcal{L}$ and $p \in I$:

$$
H_2(p) = \int_{p^0}^{p} d\Psi(F) H_1(t, s + \lambda(t), \tilde{z}, a) exp\left(\int_t^p \frac{\partial \Psi(F)}{\partial e_2}(u, s + \lambda(u), \tilde{z}, a) du \right) dt \tag{6.3}
$$

The difference then lies in the computation of $d\Psi(F) H_1(t, s + \lambda(t), \tilde{z}, a)$. Indeed, $\forall H_1 \in \mathcal{F}$ and $\forall (x, a) \in \Theta_X \times [0, 1]^2$, we have:

$$
\begin{aligned}
\Psi(F + H_1)(x, a) - \Psi(F)(x, a) &= (F + H_1)_{Y|XW}^{-1}(a_2; p, z, g_1(F + H_1)^{-1}(a_1; p, z)) \\
&\quad - (F)_{Y|PZW}^{-1}(a_2; p, z, g_1(F)^{-1}(a_1; p, z)) \\
&= (I) + (II) \tag{6.4}
\end{aligned}
$$

where

$$
\begin{aligned}
(I) &= (F + H_1)_{Y|XW}^{-1}(a_2; x, g_1(F + H_1)^{-1}(a_1; x)) \\
&\quad - (F + H_1)_{Y|XW}^{-1}(a_2; x, g_1(F)^{-1}(a_1; x)) \\
(II) &= (F + H_1)_{Y|XW}^{-1}(a_2; x, g_1(F)^{-1}(a_1; x)) \\
&\quad - (F)_{Y|XW}^{-1}(a_2; x, g_1(F)^{-1}(a_1; x))
\end{aligned}
$$

We analyze each term separately. The analysis of the second term is equivalent to the exogenous case with one supplementary regressor. Indeed, we have:

$$
\begin{aligned}
(II) &= (F + H_1)_{Y|XA_1}^{-1}(a_2; x, a_1) - (F)_{Y|XA_1}^{-1}(a_2; x, a_1) \\
&= \frac{a_2 h_1(x, a_1) - \int_{-\infty}^{\varphi(x,a)} h_1(y, x, a_1) dy}{f(x, a_1) f_{Y|XA_1}(\varphi(x, a), x, a_1)} + o(\|H_1\|')
\end{aligned}
$$

Then, the first term can be developed the following way:

$$
\begin{aligned}
(I) \;=\;& \frac{\partial (F+H_1)^{-1}_{Y|XW}}{\partial e_{L+4}}(a_2;x,g_1(F)^{-1}(a_1;x))\left[g_1(F+H_1)^{-1}(a_1;x)-g_1(F)^{-1}(a_1;x)\right]+o(\|H_1\|') \\[2mm]
\;=\;& \frac{\partial F^{-1}_{Y|XW}}{\partial e_{L+4}}(a_2;x,g_1(F)^{-1}(a_1;x))\left[g_1(F+H_1)^{-1}(a_1;x)-g_1(F)^{-1}(a_1;x)\right]+o(\|H_1\|') \\[2mm]
\;=\;& \frac{\partial F^{-1}_{Y|XA_1}}{\partial e_{L+4}}(a_2;x,a_1)\left[\frac{a_1 h_1(z,a_1)-\int_{-\infty}^{p}h_1(u,z,F^{-1}_{P|ZW}(a_1;x)dy}{f(z,F^{-1}_{P|ZW}(a_1;x))\frac{\partial F_{P|ZW}}{\partial e_{L+3}}(p,z,F^{-1}_{P|ZW}(a_1;x))}\right]+o(\|H_1\|') \\[2mm]
\;=\;& \frac{\partial F^{-1}_{Y|XA_1}}{\partial e_{L+4}}(a_2;x,a_1)\left[\frac{a_1 h_1(z,a_1)-\int_{-\infty}^{p}h_1(u,z,a_1)dy}{f(z,a_1)\frac{\partial F_{P|ZA_1}}{\partial e_{L+3}}(p,z,a_1)}\right]+o(\|H_1\|')
\end{aligned}
$$

At last, we use the definition of both solutions λ and $\hat{\lambda}$ by the same operator Φ to obtain the following characterization: $(\hat{\lambda}-\lambda)(p)=d\Phi(F)(\hat{F}-F)(p)+o_P(\|\hat{F}-F\|')$. Apply then equation (6.4) in $H_1=\hat{F}-F$ and $H_2=d\Phi(F)(\hat{F}-F)$ in order to get:

$$(\hat{\lambda}-\lambda)(p)=I(p)+R(p)$$

where $I(p)=(I')+(II')$ and $R(p)=o_P(\|\hat{F}-F\|')$. The linear term is then decomposed into two parts:

$$
\begin{aligned}
(I') \;=\;& \int_{p^0}^{p}\frac{a_1\left(\hat{f}-f\right)(s+\lambda(t),\tilde{z},a_1)-\int_{-\infty}^{t}\left(\hat{f}-f\right)(u,s+\lambda(t),\tilde{z},a_1)dy}{f(s+\lambda(t),\tilde{z},a_1)}.\delta(p,t,\tilde{z},a)dt \\[3mm]
(II') \;=\;& \int_{p^0}^{p}\frac{a_2\left(\hat{f}-f\right)(t,s+\lambda(t),\tilde{z},a_1)-\int_{-\infty}^{\varphi(t,s+\lambda(t),\tilde{z},a)}\left(\hat{f}-f\right)(y,t,s+\lambda(t),\tilde{z},a_1)dy}{f(t,s+\lambda(t),\tilde{z},a_1)}\gamma(p,t,\tilde{z},a)dt
\end{aligned}
$$

where

$$
\begin{aligned}
\delta(p,t,\tilde{z},a) \;=\;& exp\left(\int_{t}^{p}\frac{\partial\varphi}{\partial e_2}(u,s+\lambda(u),\tilde{z},a)du\right).\frac{\frac{\partial\varphi}{\partial e_{L+4}}(t,s+\lambda(t),\tilde{z},a)}{\frac{\partial F_{P|ZA_1}}{\partial e_{L+3}}(t,s+\lambda(t),\tilde{z},a_1)} \\[3mm]
\gamma(p,t,\tilde{z},a) \;=\;& \frac{exp\left(\int_{t}^{p}\frac{\partial\varphi}{\partial e_2}(u,s+\lambda(u),\tilde{z},a)du\right)}{f_{Y|X,A_1}\left(\varphi(t,s+\lambda(t),\tilde{z},a),t,s+\lambda(t),\tilde{z},a_1\right)}
\end{aligned}
$$

The asymptotic normality follows again, as in Matzkin (2003), from Theorem 3.9.4 in van der Vaart and Wellner (1996). Since the first term (I') converges faster than the second one (II'), the asymptotic variance computation is driven by the second term and the calculus is similar to the exogenous case with one supplementary regressor.

Proof of Corollary 6 The proof follows the same argument as in Corollary 5. Now we compute the bias term for:

$$(II') = \int_{p^0}^{p} \left[\int (\hat{f} - f)(y, t, s + \lambda(t), \tilde{z}, a_1) \left(a_2 - 1(y \le \varphi(t, s + \lambda(t), \tilde{z}, a)) \right) dy \right] \frac{\gamma(p, t, \tilde{z}, a)}{f(t, s + \lambda(t), \tilde{z}, a_1)} dt$$

The result follows from classical calculus on the bias decomposition and from Theorem 4.2 in Vanhems 2006.

7 Appendix II - Tables and Figures

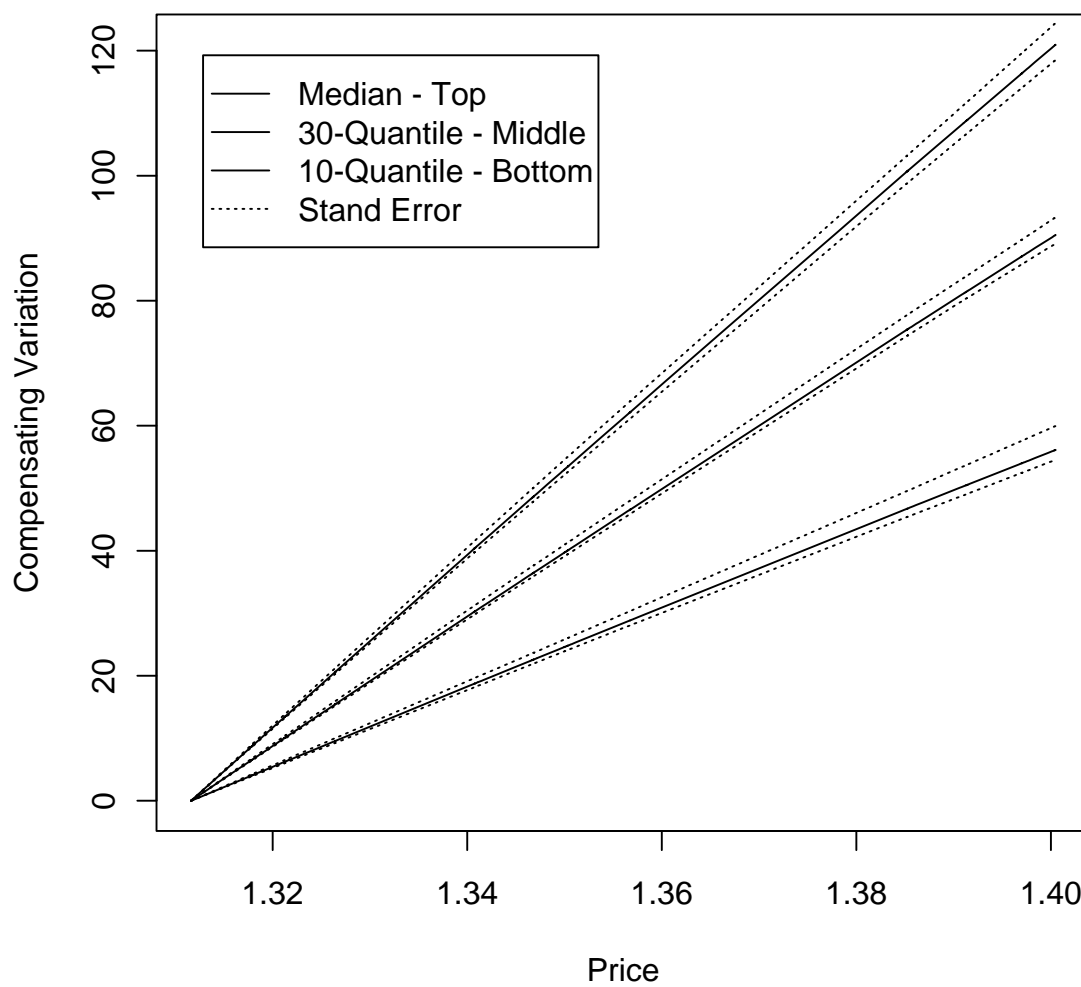Table A.1: Standard log-log Regression, OLS estimates

	Coefficient	Standard Error	t-Value	p-Value
intercept	2.965784	0.203957	14.541	2e-16
log own price	-0.449457	0.159608	-2.816	0.004882
log income	0.310499	0.015488	20.048	2e-16
log_drvrcnt	0.672372	0.037596	17.884	2e-16
region	0.006715	0.003796	1.769	0.076965
cl5_smtown_d	-0.046965	0.031962	-1.469	0.141791
cl5_suburban_d	-0.155659	0.039523	-3.938	8.32e-05
cl5_secondcity_d	-0.165686	0.041350	-4.007	6.25e-05
cl5_urban_d	-0.155687	0.055374	-2.812	0.004950
popdensity_d1	0.673887	0.121166	5.562	2.82e-08
popdensity_d2	0.602190	0.118517	5.081	3.90e-07
popdensity_d3	0.517959	0.119333	4.340	1.45e-05
popdensity_d4	0.519924	0.117398	4.429 9	69e-06
popdensity_d5	0.494301	0.116170	4.255	2.13e-05
popdensity_d6	0.432112	0.114282	3.781	0.000158
popdensity_d7	0.339010	0.119141	2.845	0.004454
public transport	-0.042310	0.023339	-1.813	0.069919
mean age driver	-0.001697	0.001195	-1.420	0.155761

R^2	0.1922			
adjusted R^2	0.1893			
$d.o.f.$	4796			

Table A.2: log-log Regression, IV estimates

	Coefficient	Standard Error	t-Value	p-Value
intercept	3.054597	0.207398	14.728	2e-16
log own price	-0.771068	0.210990	-3.655	0.000260
log income	0.310156	0.015482	20.034	2e-16
log_drvrcnt	0.671981	0.037579	17.882	2e-16
region	0.008345	0.003858	2.163	0.030609
cl5_smtown_d	-0.044035	0.031972	-1.377	0.168484
cl5_suburban_d	-0.152186	0.039533	-3.850	0.000120
cl5_secondcity_d	-0.162746	0.041350	-3.963	8.41e-05
cl5_urban_d	-0.154281	0.055352	-2.787	0.005336
popdensity_d1	0.664454	0.121178	5.483	4.39e-08
popdensity_d2	0.593308	0.118523	5.006	5.76e-07
popdensity_d3	0.507848	0.11935	4.255	2.13e-05
popdensity_d4	0.510722	0.117411	4.350	1.39e-05
popdensity_d5	0.485296	0.116181	4.177	3.01e-05
popdensity_d6	0.423496	0.114289	3.705	0.000213
popdensity_d7	0.330401	0.119144	2.773	0.005573
public transport	-0.030002	0.023919	-1.254	0.209795
mean age driver	-0.001632	0.001195	-1.366	0.171927
V	0.767499	0.329511	2.329	0.019889
R^2	0.1922			
adjusted R^2	0.1893			
$d.o.f.$	4793			

Compensating Variation for Various Quantiles

Legend:
- Median - Top
- 30-Quantile - Middle
- 10-Quantile - Bottom
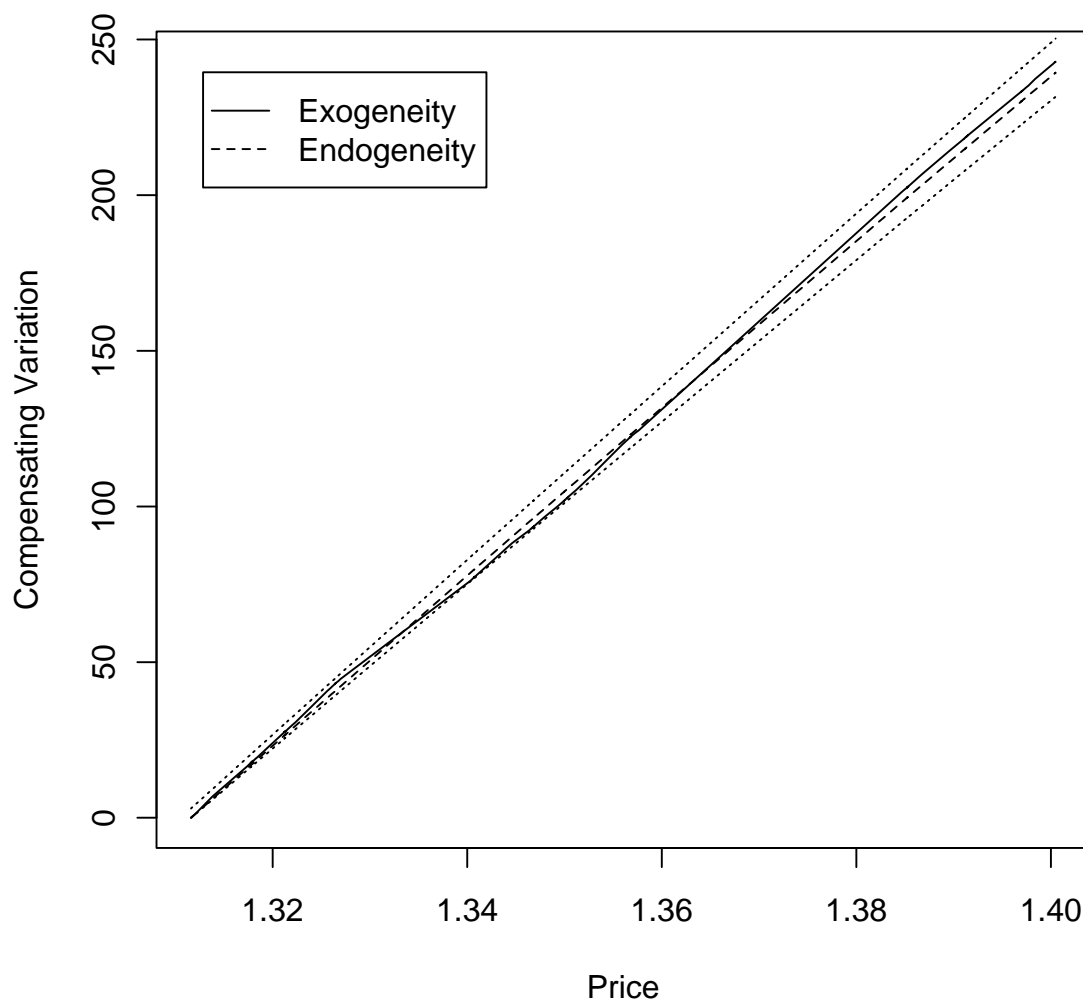- Stand Error

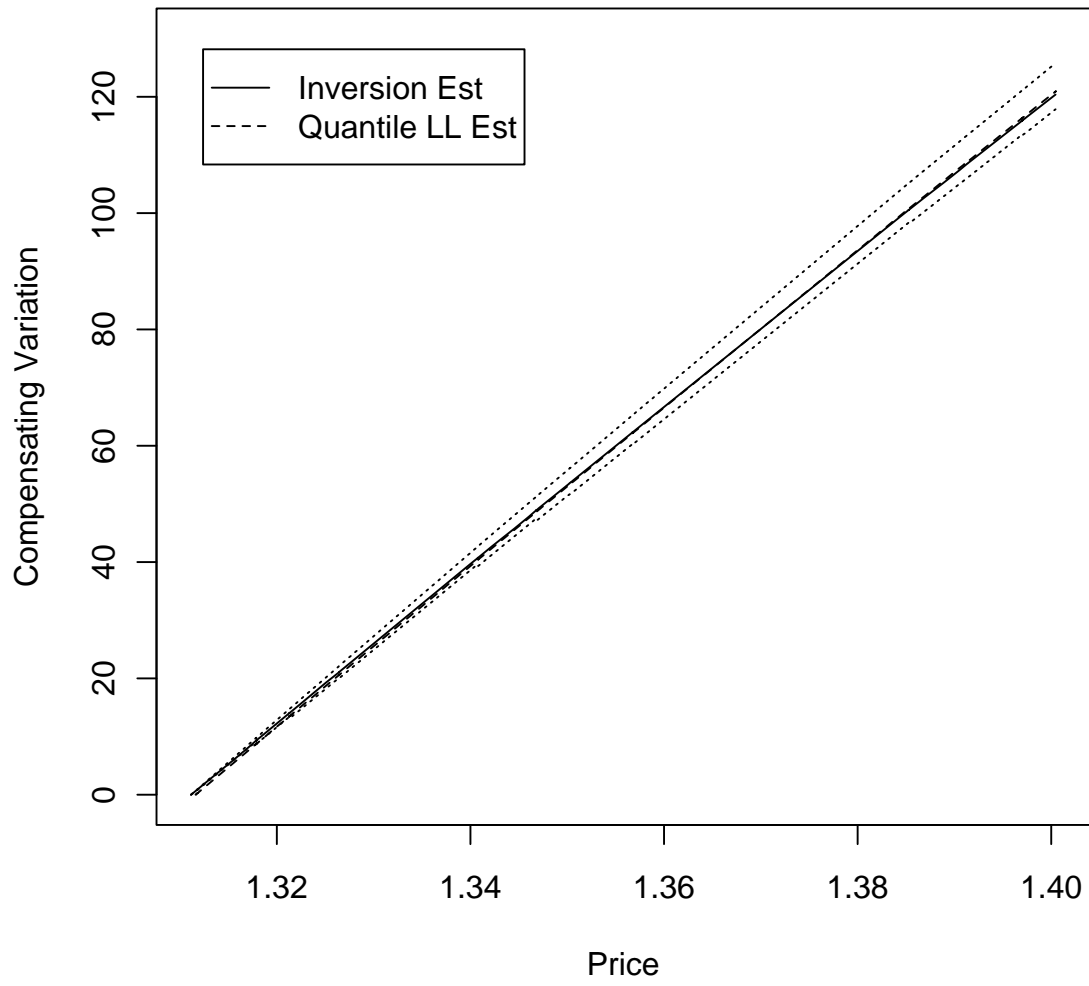Y-axis: Compensating Variation

X-axis: Price

Compensating Variation for Various Quantiles
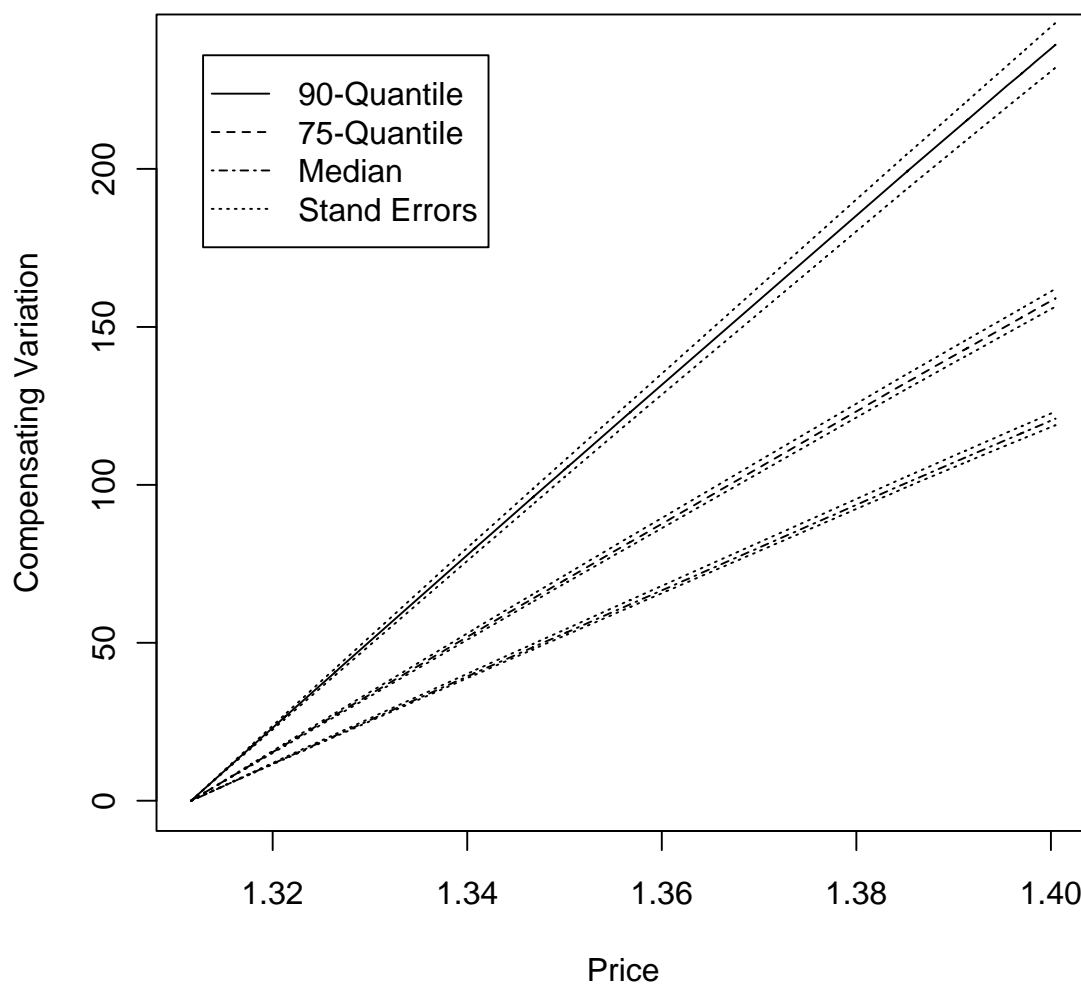
CV Comparison Exogeneity vs Endogeneity

Comparing IE and QLLE

CV for median - Rural vs Urban

Compensating Variation

Price

Rural
Urban
Stand Errors