Quantitative Economics 3 (2012), 29–51 1759-7331/20120029

Measuring the price responsiveness of gasoline demand: Economic shape restrictions and nonparametric demand estimation

Richard Blundell
Department of Economics, University College London, and Institute for Fiscal Studies

Joel L. Horowitz
Department of Economics, Northwestern University

Matthias Parey
Department of Economics, University of Essex

This paper develops a new method for estimating a demand function and the welfare consequences of price changes. The method is applied to gasoline demand in the United States and is applicable to other goods. The method uses shape restrictions derived from economic theory to improve the precision of a nonparametric estimate of the demand function. Using data from the U.S. National Household Travel Survey, we show that the restrictions are consistent with the data on gasoline demand and remove the anomalous behavior of a standard nonparametric estimator. Our approach provides new insights about the price responsiveness of gasoline demand and the way responses vary across the income distribution. We find that price responses vary non-monotonically with income. In particular, we find that low- and high-income consumers are less responsive to changes in gasoline prices than are middle-income consumers. We find similar results using comparable data from Canada.

Keywords. Consumer demand, nonparametric estimation, gasoline demand, deadweight loss.

JEL classification. D120, H310, C140.

1. Introduction

This paper develops a new method for estimating a demand function and the welfare consequences of price changes. The method is applied to gasoline demand in the United

Richard Blundell: r.blundell@ucl.ac.uk
Joel L. Horowitz: joel-horowitz@northwestern.edu
Matthias Parey: mparey@essex.ac.uk
We thank José-Víctor Ríos-Rull, two anonymous referees, and seminar participants at IFS, UCL, Oxford, Alicante, and Berlin for helpful comments. This paper is part of the program of research of the ESRC Centre for the Microeconomic Analysis of Public Policy at IFS. Financial support from the ESRC is gratefully acknowledged. The research of Joel L. Horowitz was supported in part by NSF Grants SES-0352675 and SES-0817552.

DOI: 10.3982/QE91

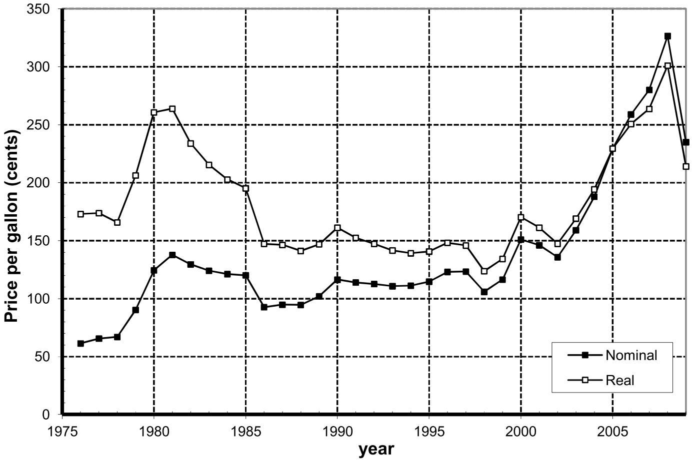


FIGURE 1. Retail motor gasoline price 1976–2009 (unleaded regular). Source: EIA (2010c, Table 5.24). U.S. city average gasoline prices. Real values are in chained (2005) dollars based on gross domestic product implicit price deflators. See source for details.

States and is applicable to other goods. In the United States, as in many other countries, the price of gasoline rose rapidly from 1998 until mid 2008. Figure 1 shows how the average price of gasoline in the United States has varied over the last three decades. Prices began rising steeply in about 1998 following a period of price stability that began in about 1986. Between March 2007 and March 2008, the average gasoline price increased by 25.7 percent in nominal terms.[1] In real terms, gasoline prices reached levels similar to those seen during the second oil crisis of 1979–1981. Although prices have decreased since mid 2008, due at least in part to the global economic downturn, many observers expect prices to rise again in the future as economic activity increases.

The measurement of the welfare consequences of price changes begins with estimating the demand function for the good in question. This is often done by using a linear model in which the dependent variable is the log of demand and the explanatory variables are the logs of price and income. This model is easy to interpret because it gives constant income and price elasticities. However, economic theory provides no guidance on the specific form of the gasoline demand function. This motivates us to use nonparametric estimation methods. We build on Hausman and Newey (1995), who also used nonparametric methods to estimate gasoline demand. We also draw on earlier work on imposing restrictions from consumer theory in a nonparametric setting including Varian (1982, 1983). In a statistical setting, Epstein and Yatchew (1985) and Yatchew and Bos (1997) developed procedures for incorporating and testing additional restrictions, including constraints on derivatives or homotheticity.

Deviations from the constant-elasticity model are not simply a technical concern. It is likely to matter greatly how peoples' responses to prices vary according to the price level and over the income distribution. Therefore, a flexible modeling approach such

[1] Our own calculation based on the Energy Information Administration (EIA (2008, Table 9.4)).

as nonparametric regression seems attractive. However, nonparametric regression can yield implausible and erratic estimates. One way to deal with this problem is to impose a parametric form such as log–log linearity on the demand function. But any parametric form is essentially arbitrary and, as discussed further in Section 4, may be misspecified in ways that produce seriously erroneous results. As a compromise between the desire for flexibility and the need for structure, one may use a semiparametric model, such as a partially linear or single-index model. These impose parametric restrictions on some aspects of the function of interest, but leave other parts unrestricted. In this paper, we take a different approach and impose structure through shape restrictions based on economic theory. Specifically, we impose the Slutsky restriction of consumer theory on an otherwise nonparametric estimate of the demand function. We show that this approach yields well behaved estimates of the demand function and price responsiveness across the income distribution while avoiding the use of arbitrary and possibly misspecified parametric models. We implement our approach by making use of a kernel-type estimator in which observations are weighted in a way that ensures satisfaction of the Slutsky restrictions. This maintains the flexibility of nonparametric regression while using restrictions of economic theory to avoid implausible estimation results. The constrained nonparametric estimates are consistent with observed behavior and provide intuitively plausible, well behaved descriptions of price responsiveness across the income distribution.

One important use of demand function estimates is to compute deadweight loss (DWL) measures of tax policy interventions. We show how the different estimates of the demand function translate into important differences in DWL estimates.

We find that there is substantial variation in price sensitivity across both price and income. In particular, we find that price responses are nonmonotonic in income. Our estimates indicate that households at the median of the income distribution respond more strongly to an increase in prices than do households at the lower or upper income group. We do not speculate on why this is the case, but we show that it implies that our DWL measure is typically higher at the median of the income distribution than in the lower or upper income group.

Section 2 explains our approach to nonparametric estimation of demand functions and DWL subject to the Slutsky shape restrictions. Section 3 describes our data, which are taken from the U.S. National Household Travel Survey (NHTS). Section 4 presents the estimates of the demand function and shows how price responsiveness varies across the income distribution. Section 4 also presents the DWLs associated with price changes and shows how they vary across the income distribution. We also derive comparable results from the Canadian Private Vehicle Use Survey. Section 5 presents results from a nonparametric test for endogeneity in the gasoline price variable. Section 6 concludes.

2. Shape restrictions and the estimation of demand and deadweight loss

We begin this section by describing our approach to estimating the demand function subject to the Slutsky shape restriction. Then we describe how we estimate the DWL of a tax-induced price increase.

The Slutsky condition is an inequality constraint on the demand function. Our method for estimating the demand function nonparametrically subject to this constraint is adapted from Hall and Huang (2001), who presented a nonparametric kernel estimator of a conditional mean function subject to a monotonicity constraint. We replace their monotonicity constraint with the Slutsky condition. To describe our estimator, let Q, P, and Y, respectively, denote the quantity of gasoline demanded by an individual, the price paid, and the individual's income. We assume that these variables are related by

$$Q = g(P, Y) + U, \tag{1}$$

where g is a function that satisfies smoothness conditions and the Slutsky restriction, but is otherwise unknown, and U is an unobserved random variable satisfying $E(U|P = p, Y = y) = 0$ for all p and y. Our aim is to estimate $g(p, y)$ nonparametrically subject to the Slutsky constraint

$$\frac{\partial g(p, y)}{\partial p} + g(p, y)\frac{\partial g(p, y)}{\partial y} \leq 0. \tag{2}$$

The data are observations $\{Q_i, P_i, Y_i : i = 1, \ldots, n\}$ for n randomly sampled individuals. A fully nonparametric estimate of g that does not impose the Slutsky restriction can be obtained by using the Nadaraya–Watson kernel estimator (Nadaraya (1964), Watson (1964)). The properties of this estimator are summarized in Härdle (1990). We call it the unconstrained nonparametric estimator, denoted by $\hat{g}_U$, because it is not constrained by (2). The estimator is

$$\hat{g}_U(p, y) = \frac{1}{nh_p h_y \hat{f}(p, y)} \sum_{i=1}^{n} Q_i K\left(\frac{p - P_i}{h_p}\right) K\left(\frac{y - Y_i}{h_y}\right), \tag{3}$$

where

$$\hat{f}(p, y) = \frac{1}{nh_p h_y} \sum_{i=1}^{n} K\left(\frac{p - P_i}{h_p}\right) K\left(\frac{y - Y_i}{h_y}\right),$$

K is a bounded, differentiable probability density function that is supported on $[-1, 1]$ and is symmetrical about 0, and h_p and h_y are bandwidth parameters.

Owing to the effects of random sampling errors, $\hat{g}_U$ does not necessarily satisfy (2) even if g does satisfy this condition. Following Hall and Huang (2001), we solve this problem by replacing $\hat{g}_U$ with the weighted estimator

$$\hat{g}_C(p, y) = \frac{1}{h_p h_y \hat{f}(p, y)} \sum_{i=1}^{n} w_i Q_i K\left(\frac{p - P_i}{h_p}\right) K\left(\frac{y - Y_i}{h_y}\right), \tag{4}$$

where $\{w_i : i = 1, \ldots, n\}$ are nonnegative weights satisfying $\sum_{i=1}^{n} w_i = 1$ and the subscript C indicates that the estimator is constrained by the Slutsky condition. The weights are obtained by solving the optimization problem

$$\underset{w_1, \ldots, w_n}{\text{minimize}} : D(w_1, \ldots, w_n) \tag{5}$$

subject to

$$\frac{\partial \hat{g}_C(p_j, y_j)}{\partial p} + \hat{g}_C(p, y)\frac{\partial \hat{g}_C(p_j, y_j)}{\partial y} \leq 0, \quad j = 1, \ldots, J,$$

$$\sum_{i=1}^{n} w_i = 1,$$

and

$$w_i \geq 0, \quad i = 1, \ldots, n,$$

where $\{p_j, y_j : j = 1, \ldots, J\}$ is a grid of points in the (p, y) plane. The objective function is the following measure of the "distance" of the weights from the values $w_i = 1/n$ that correspond to the Nadaraya–Watson estimator:

$$D(w_1, \ldots, w_n) = n - \sum_{i=1}^{n} (nw_i)^{1/2}.$$

When $w_i = 1/n$ for all $i = 1, \ldots, n$, $\hat{g}_C(p_j, y_j) = \hat{g}_U(p_j, y_j)$ for all $j = 1, \ldots, J$. Thus, the weights minimize the distance of the constrained estimator from the unconstrained one. The constraint is not binding at points (p_j, y_j) that satisfy (2). In the empirical application described in Section 4, we solve (5) by using the nonlinear programming algorithm E04UC from the Numerical Algorithms Group library. The bandwidths are selected using a method that is described in Section 4. In some applications, it may be desirable to impose the restriction that the good in question is normal. This can be done by adding the constraints $\partial \hat{g}_C(p_j, y_j)/\partial y \geq 0$ to (5), but we do not take this step here.

The literature on transport demand has documented the importance of accounting for household characteristics in estimating gasoline demand, including urbanization, population density, and transit availability, as well as demographic characteristics such as household size. Since the curse of dimensionality prevents us from estimating a fully nonparametric model in all of these dimensions, we account for these covariates in a partially linear framework. For this purpose, we estimate the effects of the covariates from a double-residual regression (Robinson (1988)), and then estimate the nonparametric demand function of interest after removing the effect of the covariates.

We now describe our method for estimating the DWL of a tax. Let $E(p)$ denote the expenditure function at price p and some reference utility level. The DWL of a tax that changes the price from p^0 to p^1 is

$$L(p^0, p^1) = E(p^1) - E(p^0) - (p^1 - p^0)g[p^1, E(p^1)]. \tag{6}$$

We estimate this by

$$\hat{L}(p^0, p^1) = \hat{E}(p^1) - \hat{E}(p^0) - (p^1 - p^0)\hat{g}[p^1, \hat{E}(p^1)], \tag{7}$$

where $\hat{E}$ is an estimator of the expenditure function and $\hat{g}$ may be either $\hat{g}_U$ or $\hat{g}_C$. We obtain $\hat{E}$ by solving the differential equation

$$\frac{d\hat{E}(t)}{dt} = \hat{g}[p(t), \hat{E}(t)]\frac{dp(t)}{dt}, \tag{8}$$

where $[p(t), \hat{E}(t)]$ $(0 \leq t \leq 1)$ is a price–(estimated) expenditure path. We solve this equation along a grid of points by using Euler's method (Ascher and Petzold (1998)). We have found this method to be quite accurate in numerical experiments.

Inference with the constrained estimator $\hat{g}_C$ is difficult because the estimator's asymptotic distribution is very complicated in regions where (2) is a binding constraint (strict equality). However, if we assume that (2) is a strict inequality in the population, then violation of the Slutsky condition by $\hat{g}_U$ is a finite-sample phenomenon, and we can use $\hat{g}_U$ to carry out asymptotically valid inference. We use the bootstrap to obtain asymptotic joint confidence intervals for $g(p, y)$ on a grid of (p, y) points and to obtain confidence intervals for L. The bootstrap procedure is as follows.

Step 1. Generate a bootstrap sample $\{Q_i^*, P_i^*, Y_i^* : i = 1, \ldots, n\}$ by sampling the data randomly with replacement.

Step 2. Use this sample to estimate $g(p, y)$ on a grid of (p, y) points without imposing the Slutsky constraint. Also, estimate L. Denote the bootstrap estimates by $\hat{g}_U^*$ and L^*.

Step 3. Form percentile confidence intervals for L by repeating Steps 1 and 2 many times. Also, use the bootstrap samples to form joint percentile-t confidence intervals for g on the grid of points $\{p_j, y_j : j = 1, \ldots, J\}$. The joint confidence intervals at a level of at least $1 - \alpha$ are

$$\begin{aligned} \hat{g}_U(p_j, y_j) - z_\alpha(p_j, y_j)\hat{\sigma}(p_j, y_j) &\leq g(p_j, y_j) \\ &\leq \hat{g}_U(p_j, y_j) + z_\alpha(p_j, y_j)\hat{\sigma}(p_j, y_j), \end{aligned} \tag{9}$$

where

$$\hat{\sigma}^2(p, y) = \frac{B_K}{[nh_p h_y \hat{f}(p, y)]^2} \sum_{i=1}^{n} \hat{U}_i^2 K\left(\frac{p - P_i}{h_p}\right) K\left(\frac{y - Y_i}{h_y}\right), \tag{10}$$

with $B_K = \int K(v)^2\, dv$ and $\hat{U}_i = Q_i - \hat{g}_U(P_i, Y_i)$, is a consistent estimate of $\mathrm{Var}[\hat{g}_U(p, y)]$. The critical value $z_\alpha(p_j, y_j)$ is chosen following the approach in Härdle and Marron (1991) for computing joint confidence intervals. For this purpose, we partition the grid into intervals of $2h_p$. Within each of these M neighborhoods, $z_\alpha(p_j, y_j)$ is the solution to

$$P^*\left[\frac{|\hat{g}_U^*(p_j, y_j) - \hat{g}_U(p_j, y_j)|}{\hat{\sigma}^*(p_j, y_j)} \leq z_\alpha(p_j, y_j)\right] = 1 - \beta,$$

where P^* is the probability measure induced by bootstrap sampling, and $\hat{\sigma}^*(p, y)$ is the version of $\hat{\sigma}(p, y)$ that is obtained by replacing $\hat{U}_i$, P_i, and Y_i in (10) by their bootstrap analogs, and β is a parameter. We then choose β such that the simultaneous size in each neighborhood equals $1 - \frac{\alpha}{M}$. As Härdle and Marron (1991) showed using the Bonferroni inequality, the resulting intervals over the full grid form simultaneous confidence intervals at a level of at least $1 - \alpha$. Hall (1992) showed that the bootstrap consistently estimates the asymptotic distribution of the Studentized form of $\hat{g}_U$. It is necessary to undersmooth $\hat{g}_U$ and $\hat{g}_U^*$ (that is, use smaller than asymptotically optimal bandwidths) in (9) and Step 2 of the bootstrap procedure to obtain a confidence interval that is centered at g. We discuss bandwidth selection in Section 4.

3. Data

Our analysis is based on the 2001 National Household Travel Survey. The NHTS was sponsored by the Bureau of Transportation Statistics and the Federal Highway Administration. The data were collected through a telephone survey of the civilian, noninstitutionalized population of the United States. The survey was conducted between March 2001 and May 2002 (ORNL (2004, Chap. 3)). The telephone interviews were complemented with written travel diaries and odometer readings.

The key variables used in our study are annual gasoline consumption, the gasoline price, and household income. Gasoline consumption is derived from odometer readings and estimates of the fuel efficiencies of vehicles. Details of the computations are described in an Oak Ridge National Laboratory report (ORNL (2004, Appendices J and K)). The gasoline price for a given household is the average price in dollars per gallon, including taxes, in the county where the household is located. This price variable is a county average, rather than the price actually paid by a household. It precludes an intracounty analysis (see Schmalensee and Stoker (1999)), but does capture variation in prices consumers face in different regions. Price differences across local markets reflect proximity of supply, short-run shocks to supply, competition in the local market, and local differences in taxes and environmental programs (EIA (2010a)). We return to this in Section 5, where we investigate the role of proximity of supply as a cost shifter and test for endogeneity of prices.

Household income in dollars is available in 18 groups. In our analysis, we assign each household an income equal to the midpoint of its group. The highest group, consisting of incomes above \$100,000, is assigned an income of \$120,000.[2] To investigate how price responsiveness of gasoline demand varies across the income distribution, we focus on three income levels of interest: a middle-income group at \$57,500, which corresponds to median income in our sample; a low-income group (\$42,500), which corresponds to the first quartile; and a high-income group (\$72,500).[3] To obtain gasoline demand at the

[2]Assuming log-normality of income, we estimated the corresponding mean and variance by using a simple Tobit model, right-censored at \$100,000. Excluding households with very high incomes above \$150,000, the median income in the upper group corresponds to about \$120,000.

[3]The income point \$72,500 occupies the 59.6–63.3th percentile. This point was chosen to avoid the problems created by the interval nature of the income variable which becomes especially important in the upper quartile of the income distribution: income brackets are relatively narrow (with widths of \$5000) up to \$80,000, but substantially wider for higher incomes. However, estimates using higher quantiles yielded similar results and did not change our conclusions on price responsiveness across the income distribution.

household level, we aggregate gasoline consumption in gallons over multicar households. We do not investigate the errors-in-variables issues raised by the use of county-average prices or the interval censoring issues raised by the grouping of household incomes in the data. These potentially important issues are left for future research.

Previous research on determinants of gasoline demand has shown the importance of accounting for demographic characteristics of the household. In our analysis, we include the age of the household respondent, household size, and the number of drivers in the household (all measured in logs). We also include the number of employed household members.

We measure population density in eight categories. Urbanity is measured in five categories (rural, small town, suburban, second city, urban), and public transit availability is an indicator for whether the household is located in a Metropolitan Statistical Area (MSA) or a Consolidated Metropolitan Area (CMSA) of one million or more with rail. In one specification, we also include region fixed effects, corresponding to the nine U.S. census divisions.

We exclude from our analysis households where the number of drivers is zero or whose variables of interest are not reported, and we require gasoline consumption of at least one gallon. Due to its special geographic circumstances, we also exclude households that are located in Hawaii. In addition, we restrict our sample to households with a white respondent, two or more adults, and at least one child under 16 years of age. We take vehicle ownership as given and do not investigate how changes in prices affect vehicle purchases or how vehicle ownership varies across the income distribution (Poterba (1991), West (2004), Bento, Goulder, Henry, Jacobsen, and von Haefen (2005), Bento, Goulder, Jacobsen, and von Haefen (2009)). The results of Bento et al. (2005) indicate that over 95 percent of the reduction in gasoline demand in response to price changes is due to changes in miles traveled rather than fleet composition. We limit attention to vehicles that use gasoline as fuel, rather than diesel, natural gas, or electricity. The resulting sample consists of 5254 observations (4812 observations when we condition on regions as well). Table 1 shows summary statistics.[4]

4. Estimates of demand responses

A. The constant elasticity model

We begin by using ordinary least squares to estimate the log–log linear demand model

$$\log Q = \beta_0 + \beta_1 \log P + \beta_2 \log Y + U; \qquad E(U|P = p, Y = y) = 0. \tag{11}$$

This constant elasticity model is one of the most frequently estimated (e.g., Dahl (1979), Hughes, Knittel, and Sperling (2008)). It has been criticized on many grounds (e.g., Deaton and Muellbauer (1980)) but its simplicity and frequent use make it a useful parametric reference model. Later in this section, we compare the estimates obtained from model (11) with those obtained from the nonparametric analysis.

[4]At the bottom of Table 1, we also report a variance decomposition of log price by income groups, indicating that most of the observed price variation is within income groups.

Table 1. Sample descriptives.

A. Means and Standard Deviations		
Log gasoline demand	7.170	[0.670]
Log price	0.287	[0.057]
Log income	10.955	[0.613]
Log age of household respondent	3.628	[0.240]
Log household size	1.385	[0.234]
Log number of drivers	0.781	[0.240]
Number of workers in household	1.868	[0.745]
Public transit indicator	0.216	[0.411]
Rural	0.252	[0.434]
Small town	0.285	[0.452]
Suburban	0.256	[0.436]
Second city	0.144	[0.352]
Urban	0.062	[0.241]
Population density	8 categories	
B. Variance Decomposition of log Price by Income Groups		
Overall variance	0.003297	
Within-group variance	0.003274	
Between-group variance	0.000023	
Observations	5254	

The estimates of the coefficients of (11) are shown in Table 2. The estimates in column 1, where we include no further covariates beyond price and income, imply a price elasticity of demand of −0.92 and an income elasticity of 0.29. These estimates are similar to those reported by others. Hausman and Newey (1995) reported estimates of −0.81 and 0.37, respectively, for price and income elasticities based on U.S. data collected between 1979 and 1988. Schmalensee and Stoker (1999) reported price elasticities between −0.72 and −1.13 and income elasticities between 0.12 and 0.33, depending on the survey year and control variables, in their specifications without regional fixed effects. Yatchew and No (2001) estimated a partially linear model using Canadian data for 1994–1996 and found an income elasticity of 0.28 and an average price elasticity of −0.89.[5] West (2004) reported a mean price elasticity of −0.89 using 1997 data. In columns 2–5, we add further covariates. Although the number of drivers and the number of workers are highly significant, the effect on the estimated price elasticity is relatively limited. Adding public transport availability (column 3) has only a small effect on the estimated elasticities. In column 4, we add indicators for urbanity and population density. While the income elasticity changes little, the price elasticity goes down to −0.50. In the last column, we also add regional fixed effects. The main effect of including regional fixed effects is that the standard error of the price elasticity increases sharply, and we see a modest further

[5]The dependent variable is log of distance traveled. See Yatchew and No (2001, Figure 2) for details.

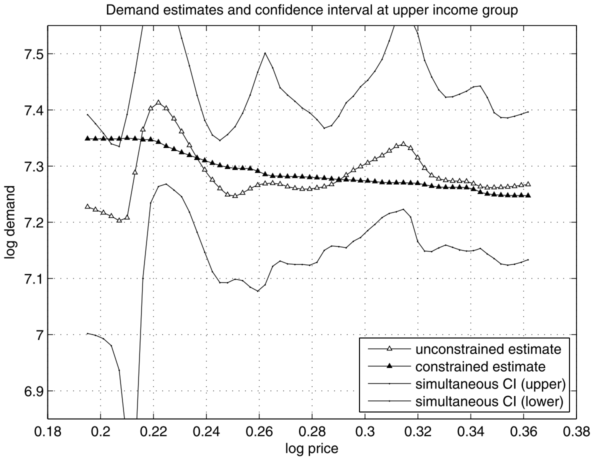
Table 2. OLS regression.[a]

	(1)	(2)	(3)	(4)	(5)
Log price	−0.925	−0.879	−0.830	−0.495	−0.358
	[0.155]**	[0.149]**	[0.148]**	[0.147]**	[0.272]
Log income	0.289	0.246	0.269	0.298	0.297
	[0.0145]**	[0.0143]**	[0.0146]**	[0.0147]**	[0.0153]**
Log age of household respondent		−0.0520	−0.0343	−0.0265	−0.0182
		[0.0366]	[0.0365]	[0.0356]	[0.0372]
Log household size		0.0586	0.0662	0.0539	0.0634
		[0.0395]	[0.0393]	[0.0383]	[0.0399]
Log number of drivers		0.601	0.582	0.542	0.510
		[0.0454]**	[0.0453]**	[0.0442]**	[0.0463]**
Number of workers in household		0.0877	0.0857	0.0893	0.0928
		[0.0137]**	[0.0136]**	[0.0133]**	[0.0139]**
Public transit indicator			−0.152	−0.0458	−0.0286
			[0.0212]**	[0.0219]*	[0.0249]
Small town				−0.0464	−0.0359
				[0.0296]	[0.0313]
Suburban				−0.165	−0.146
				[0.0368]**	[0.0386]**
Second city				−0.184	−0.164
				[0.0382]**	[0.0404]**
Urban				−0.178	−0.149
				[0.0523]**	[0.0541]**
Constant	4.264	4.200	3.914	3.722	3.642
	[0.163]**	[0.194]**	[0.198]**	[0.196]**	[0.223]**
Population density (8 categories)	No	No	No	Yes	Yes
Regions (9 categories)	No	No	No	No	Yes
Test of equality of coefficients on price and income (compared to previous specification)					
χ^2 test statistic		51.20	44.68	90.72	0.35
p-value		0.000	0.000	0.000	0.841
Observations	5254	5254	5254	5254	4812
R-squared	0.0741	0.154	0.163	0.207	0.209

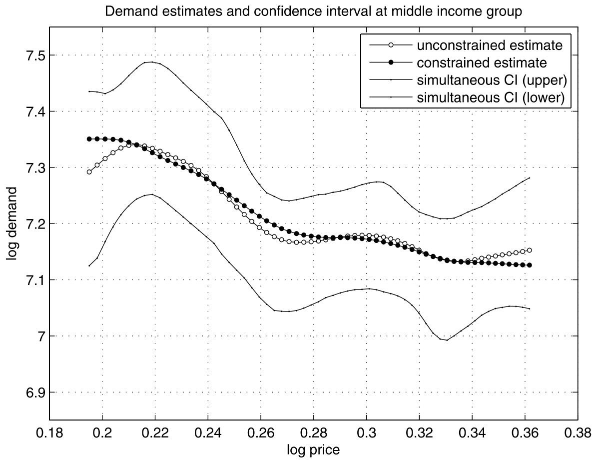
[a]The dependent variable is log of annual household gasoline demand in gallons. * indicates significance at 5%; ** indicates significance at 1% level. See text for details.

reduction in the price elasticity. As reported in the bottom panel of the table, we cannot reject that the price and income elasticities are the same between specifications 4 and 5. In the following analysis, we include the set of covariates that correspond to column 4.

Although the estimates we obtain from model (11) are similar to those reported by others, it is possible that (11) is misspecified. For example, West (2004) found evidence for dependence of the price elasticity on income. One possibility would be to add the interaction term $(\log P)(\log Y)$ to model (11). However, if the structure imposed by such an augmented linear model remains misspecified, this may lead to inconsistent estimators whose properties are unknown. Nonparametric estimators, by contrast, are consistent.



(a) upper income group

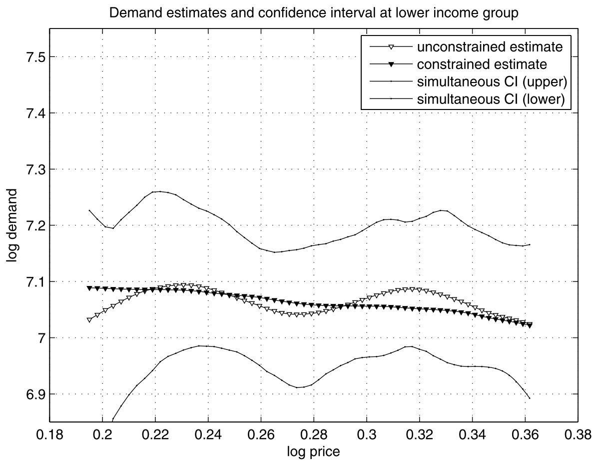


(b) middle income group

FIGURE 2. Demand estimates and simultaneous confidence intervals at different points in the income distribution. Income groups correspond to \$72,500, \$57,500, and \$42,500. Confidence intervals shown refer to bootstrapped symmetrical, Studentized simultaneous confidence intervals with a confidence level of 90%, based on 5000 replications. See text for details.

B. Unconstrained semi-nonparametric estimates

Our unconstrained semi-nonparametric estimates of the demand function, $\hat{g}_U$, are displayed in Figure 2 (shown as open dots). They were obtained by using the Nadaraya–Watson kernel estimator with a biweight kernel (Silverman (1986)). In principle, the bandwidths h_p and h_y can be chosen by applying least-squares cross-validation (Härdle (1990)) to the entire data set, but this yields bandwidths that are strongly influenced by low-density regions. To avoid this problem, we used the following method to choose h_p and h_y. We are interested in $g(p, y)$ for y values corresponding to our three income groups and price levels between the 5th and 95th percentiles of the observed prices.



(c) lower income group

FIGURE 2. *Continued.*

We defined three price–income rectangles consisting of prices between the 5th and 95th percentiles and incomes within 0.5 of each income level of interest (measured in logs). We then applied least-squares cross-validation to each price–income rectangle separately to obtain bandwidth estimates appropriate to each rectangle. This procedure yielded $(h_p, h_y) = (0.0431, 0.2143)$ for the lower-income group, $(0.0431, 0.2061)$ for the middle-income group, and $(0.0210, 0.2878)$ for the upper-income group. The estimation results are not sensitive to modest variations in the dimensions of the price–income rectangles. As was discussed in Section 2, $\hat{g}_U$ and $\hat{g}_U^*$ must be undersmoothed to obtain properly centered confidence intervals. To this end we multiplied each of the foregoing bandwidths by 0.8 when computing confidence intervals.

Figure 2 shows the unconstrained semi-nonparametric estimates of gasoline demand as a function of price at three points across the income distribution (open dots in the figure). The figure gives some overall indication of downward sloping demand curves with slopes that differ across the income distribution, but there are parts of the estimated demand curves that are upward sloping and, therefore, implausible. We interpret the implausible shapes of the curves in Figure 2 as indicating that fully nonparametric methods are too imprecise to provide useful estimates of gasoline demand functions with our data. Figure 2 shows several instances in which the semi-nonparametric estimate of the (Marshallian) demand function is upward sloping. This anomaly is also present in the results of Hausman and Newey (1995). The theory of the consumer requires the compensated demand function to be downward sloping. Combined with a positive income derivative, an upward-sloping Marshallian demand function implies an upward-sloping compensated demand function and, therefore, is inconsistent with the theory of the consumer. At the median income, our semi-nonparametric estimate of $\partial g/\partial y$ is positive over the range of prices of interest. Therefore, the semi-nonparametric estimates are inconsistent with consumer theory. As is discussed in more detail in Section 4D, we believe this result to be an artifact of random sampling errors and the consequent imprecision of the unconstrained semi-nonparametric estimates. This motivates

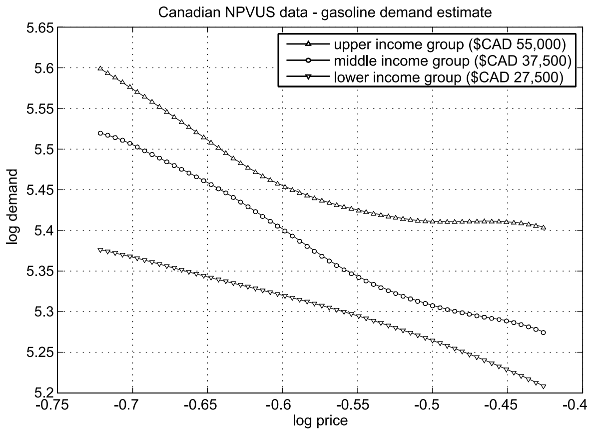


FIGURE 3. Canadian NPVUS data—gasoline demand estimate. Based on the Canadian NPVUS data as in Yatchew and No (2001). The dependent variable is log of total monthly gasoline consumption. The sample size in this analysis is 5001, where we have restricted age to be greater than or equal to 20, grade of gasoline to be regular, and the price of gasoline (measured in Canadian dollars per liter) to be at least 0.4. Taking midpoints of the income brackets (measured in Canadian dollars), the quartiles of the income variable in the sample are \$27,500, \$37,500, and \$55,000. We follow the same procedure for bandwidth choice as for the NHTS.

the use of the constrained estimation procedure, which increases estimation precision by imposing the Slutsky condition.

C. Comparison to the Canadian National Private Vehicle Use Survey

One of the advantages of the Canadian gasoline demand data used in the analysis of Yatchew and No (2001) is that price information is based on fuel purchase diaries rather than local averages. Here we briefly provide comparison estimates obtained from the Canadian National Private Vehicle Use Survey (NPVUS). These data were collected between 1994 and 1996. The dependent variable is log of total monthly gasoline consumption. Apart from price and income effects, we control for household size, number of drivers, and age (all measured in logs), an indicator for whether the age variable is censored at 65, an urbanity indicator, and month and year effects.[6] With regard to the grade of gasoline, we restrict the analysis here to regular gas.[7] In a parametric reference model, we obtain a price elasticity of -0.99 and an income elasticity of 0.19. Figure 3 shows the semi-nonparametric estimates at the quartiles of the income distribution. The figure suggests that the Canadian data yield smoother demand functions than the U.S. data do, but exhibit evidence of differences in price elasticity across the income groups. The

[6]This set of covariates is similar to the one used in Yatchew and No (2001). Reflecting the different focus of their study, one difference is that their specification allows for more general age effects than we do here.

[7]Since the NPVUS collects gasoline consumption for a representative vehicle in the household (rather than for all vehicles), we multiply the consumption corresponding to the representative vehicle by the number of vehicles. The resulting sample size is 5001, where we have restricted age to be greater or equal to 20, and the price of gasoline (measured in Canadian dollars per liter) to be at least 0.4.

estimated differences across the three income groups also matter for the resulting DWL estimates, which we return to below. For the purposes of the analysis in this paper, a limitation of the Canadian data is that income is reported in only 9 brackets, compared to 18 in the NHTS, and the main focus of this paper is therefore on the NHTS data.

D. Semi-nonparametric estimates under the Slutsky condition

Figure 2 also shows the constrained semi-nonparametric estimates of the demand function, $\hat{g}_C$, at each of the three income levels of interest (solid dots). These estimates are constrained to satisfy the Slutsky condition and were obtained using the methods described in Section 2. The solid lines in Figure 2 connect the endpoints of joint 90% confidence intervals for $g(p, y)$. These were obtained using the bootstrap procedure described in Section 2. Table A1 in the Appendix reports the estimates from the partially linear component.

In contrast to the unconstrained estimates, the constrained estimates are downward sloping everywhere and similar in appearance to those obtained with the Canadian data. The constrained estimates are also less wiggly than the unconstrained ones. In contrast to ad hoc "ironing procedures" for producing monotonic estimates, $\hat{g}_C$ is consistent with the theory of the consumer and everywhere differentiable. This is important for estimation of DWL. Except for one instance for the upper income group, the 90% confidence bands shown in Figure 2 contain both the constrained and unconstrained estimates. This is consistent with our view that the anomalous behavior of the unconstrained estimates is due to imprecision of the unconstrained estimator. It also indicates that the Slutsky constraint is consistent with the data.

The results in Figure 2 indicate that the middle-income group is more sensitive to price changes than are the other two groups. In particular, the slope of the constrained estimate of g is noticeably larger for the middle group than for the other groups.

A possible way to summarize the nonparametric evidence in a parsimonious parametric specification, an approach suggested in Schmalensee and Stoker (1999), would be to interact the price and income effects of the log–log specification described in (11) with indicators for three income groups. The resulting estimates corresponding to such a specification are presented in Table 3.

The differential responsiveness to price changes across the income distribution described in the semi-nonparametric estimates suggests that the DWL of a tax increase is larger for the middle-income group than for the others. We investigate this further in Section 4F.

E. Comparison using an alternative price variable

In this section, we briefly explore the robustness of our results to using a different gasoline price measure. For this purpose, we draw on price data collected for the ACCRA Cost of Living Index by the Council for Community and Economic Research. These data report the price of a gallon of gasoline (regular unleaded, national brand, including all taxes) for a sample of about 300 cities across the United States. Similar data have been used, for example, in Li, Timmins, and von Haefen (2009).

TABLE 3. Log–log model interacted with income group.[a]

log price × upper-income group	−0.225
	[0.240]
	($p = 0.348$)
log price × middle-income group	−1.316
	[0.423]**
	($p = 0.002$)
log price × lower-income group	−0.441
	[0.283]
	($p = 0.119$)
log income × upper-income group	0.233
	[0.0345]**
	($p = 0.000$)
log income × middle-income group	0.260
	[0.0376]**
	($p = 0.000$)
log income × lower-income group	0.229
	[0.0378]**
	($p = 0.000$)
Test on equality of price effects: upper- vs. middle-income group	
F-statistic	5.09
p-value	0.0241
Test on equality of price effects: middle- vs. lower-income group	
F-statistic	2.98
p-value	0.0842
Set of covariates	Yes
Observations	4902

[a]This table shows estimates of a log–log specification interacted with income group. For the purpose of this regression, three income groups are defined as below \$50,000 (lower-income group), \$50,000–65,000 (middle-income group), and above \$65,000 (upper-income group). Households with incomes below \$15,000 are excluded in this exercise, and log prices are restricted to the range of 0.18–0.38. The set of covariates is the same as in Table 2, column 4, that is, age of household respondent, household size, number of drivers (all in logs), number of workers in the household, public transit availability, urbanity indicators, and population density indicators. Numbers in square brackets are standard errors; numbers in round brackets are corresponding p-values. * indicates significance at 5%; ** indicates significance at the 1% level.

In the NHTS, large MSAs (of one million population or more) are separately identified. We aggregate the ACCRA gasoline price observations, on a quarterly basis, to the level of these MSAs, as well as to state level (excluding these large MSAs), using 2001 population estimates as weights.[8] We then average the resulting prices over the four quarters 2001/Q2–2002/Q1, a period over which most of the NHTS data collection took place. For households located in large MSAs in the NHTS, we assign the corresponding MSA-level price, and for households outside of these MSAs, we assign the corresponding state-level price. This results in a sample of 4847 households.

[8]Population estimates are 2001 county-level Census estimates (U.S. Census Bureau (2010)); links between different geographic identifiers are based on U.S. Census Bureau (2011).

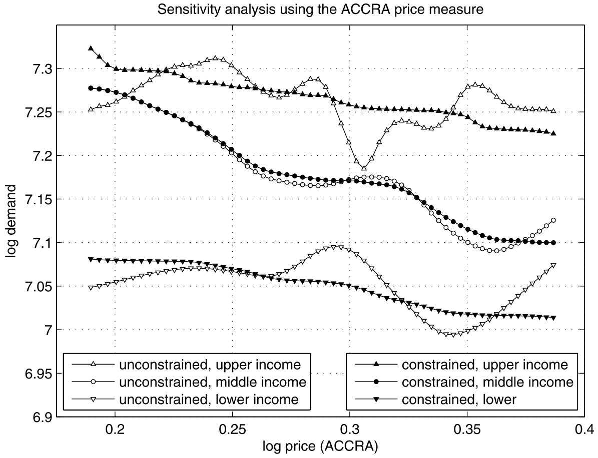


FIGURE 4. Sensitivity analysis using the ACCRA price measure. Income groups correspond to \$72,500, \$57,500, and \$42,500. Estimates are shown over the range of the 5th to the 95th percentile of the ACCRA-based gasoline price. See text for details.

We then repeat our nonparametric analysis using this ACCRA-based gasoline price. We use the same specification and bandwidth choices as before, but we add an indicator for location in a large MSA to the vector of partially linear covariates. Figure 4 shows the resulting nonparametric unconstrained and constrained estimates. These results are very similar to our main results reported above, in particular with regard to the differences in price sensitivity across the three income groups.

F. Estimates of deadweight loss

We now investigate the DWLs associated with an increase in gasoline taxes. The increases considered in the literature typically are quite large and often out of the support of the data. We take an intervention that moves prices from the 5th to the 95th percentile of the price distribution in our sample (from \$1.215 to \$1.436). We compute DWL as follows. Over the range of the intervention, we evaluate the Marshallian demand estimates presented in the previous section for the three estimators (parametric, unconstrained semi-nonparametric, and constrained semi-nonparametric) on a grid of 61 points.[9] We then use this demand estimate and the corresponding derivatives to compute the expenditure function and the DWL by following the methods described in Section 2.

We study DWL relative to tax paid, which we interpret as a "price" for raising tax revenue. We refer to this measure as relative DWL. Results are shown in panel A of Table 4.[10] The differences in the demand estimates between the different estimation methods translate into differences in relative DWLs. Comparing across income levels, the

[9]For consistency, we use the same grid for the computation of the DWL measures as we do when we impose the Slutsky constraint. Using a finer grid for computing DWL would lead to slightly different deadweight loss estimates, but not affect the pattern we find or our conclusions.

[10]Confidence intervals for the unconstrained and the parametric model are reported in Table A2.

TABLE 4. Deadweight loss estimates.[a]

Income	Semi-Nonparametric: Unconstrained (1)	Semi-Nonparametric: Constrained (2)	Parametric: Log–Log (3)
		A. DWL (as % of tax paid)	
\$72,500	1.71%	4.27%	4.13%
\$57,500	6.06%	9.19%	4.12%
\$42,500	3.86%	3.91%	4.10%
		B. DWL (relative to income) $\times 10^4$	
\$72,500	0.75	1.83	1.69
\$57,500	2.98	4.39	1.98
\$42,500	2.26	2.28	2.44

[a]This table shows deadweight loss estimates corresponding to moving prices from the 5th to the 95th percentile in the data (\$1.215 to \$1.436). Deadweight loss is shown as a percentage of tax paid after the (compensated) intervention (panel A) and relative to baseline income (panel B). See text for details.

log–log linear model estimates relative DWL to be almost identical for the three income groups and indicates that the cost of taxation is about 4.1% of revenue raised, irrespective of income level. In contrast, the constrained semi-nonparametric estimates indicate that the cost of taxation is higher for the middle-income group than for the other two groups. This result is consistent with our earlier finding that the middle-income group is more responsive to price changes than are the other groups. We note that the Canadian NPVUS data yield a similar pattern.[11] These results also illustrate how the functional form assumptions of the parametric model affect estimates of consumer behavior and the effects of taxation.

Although not the case for the intervention we study here, the DWL obtained from the unconstrained semi-nonparametric estimate of demand may be negative for specific interventions. This anomalous result can occur because, due to random sampling errors, the unconstrained estimate of the demand function does not decrease monotonically and does not satisfy the integrability conditions of consumer theory. The constrained semi-nonparametric model yields DWL estimates that are positive and, for the middle-income group, more than double those obtained from the parametric model.

One can also study DWL relative to income so as to reflect the household's utility loss relative to available resources. The results for this analysis are shown in panel B of Table 4. The estimates from the parametric model and constrained semi-nonparametric model give different indications of the effects of the tax increase across income groups. The parametric estimates indicate that the relative utility loss increases as income decreases. However, the constrained semi-nonparametric estimates indicate that the relative utility loss is greater for the middle-income group than for the other groups.

[11]For the NPVUS data, the relative DWL from the estimates shown in Figure 3 follow the same pattern across income groups as in the NHTS, but at overall higher levels: DWL relative to tax paid amounts to 5.8% for the high-income group, 11.1% for the middle-income group, and 9.4% for the low-income group. These estimates correspond to moving the price in the NPVUS sample from the 5th to the 95th percentile, that is, from CAD\$0.486 to CAD\$0.653 per liter.

5. Testing for endogeneity of prices

A longstanding concern in demand estimation is the potential endogeneity of prices (Working (1927)). This aspect has also been emphasized in the literature on discrete choice with differentiated products in the market for automobiles (Berry, Levinsohn, and Pakes (1995)). Throughout the analysis so far, we have maintained the mean independence assumption on the error term. A natural way to proceed is to test for endogeneity of gasoline price. One possible approach would be to estimate the demand function using nonparametric instrumental variables (IV) methods (see Hall and Horowitz (2005), and Blundell, Chen, and Kristensen (2007)) and then to test by comparing the IV estimate with the estimate under the exogeneity assumption. Such a test is likely to have low power, though, because of the low rate of convergence associated with the nonparametric IV regression estimates. We therefore take a different approach to testing for endogeneity, and apply the nonparametric test developed in Blundell and Horowitz (2007). An important benefit of this test is that it is likely to have better power properties because it avoids the difficulties associated with the ill-posed inverse problem.

To identify the demand function, we use the following cost shifter as the instrumental variable: Due to transportation cost, an important determinant of interregional differences in gasoline prices faced by consumers is the distance from the source of supply. The U.S. Gulf Coast Region (PADD 3) accounts for 56% of total U.S. refinery net production of finished motor gasoline;[12] it accounts for about 56% of U.S. field production of crude oil and about 64% of U.S. imports of crude oil entered the United States through this region in the year of our survey.[13] This region is also the starting point for most major gasoline pipelines. Thus, we expect prices to increase with distance from the U.S. Gulf Coast. We construct a distance measure (in 1000 km) from the source of supply in the Gulf of Mexico to the capital of the state in which the household is located. To implement this, we take as starting point a major oil platform located in the Green Canyon area, an area of the Gulf of Mexico where many of the major oil fields are located. We compute distance to the state capitals using the Haversine formula.

Figure 5 documents the relationship between log price and distance in our data.[14] The correlation coefficient between log price and our distance measure is 0.78 and highly significant.[15] In the following, we assume that our cost shifter variable satisfies the required independence assumption relating to the error term U. To account for the role of covariates, we take the same approach as in the nonparametric estimation above and remove the estimated partially linear component in a first step. Table 5 shows the results from this exogeneity test. The test statistic (see panel A of Table 5) is substantially below the critical value, so we fail to reject the null hypothesis of price exogeneity in this application. We have experimented with varying the bandwidth parameters in this

[12]Source: EIA (2010b), data for 2005 (earlier data not available).

[13]Source: EIA (2010b), data for 2001.

[14]This analysis is based on the 34 biggest states in terms of population; smaller states are not separately identified in the data for confidentiality reasons.

[15]We have also studied the effect of including our full set of controls (as in Table 2) in a regression of log price on our distance measure. The coefficient on distance remains stable and highly significant ($p < 0.01$). With the covariates we use in our analysis (Table 2, column 4), the corresponding t-statistic is 84.7.

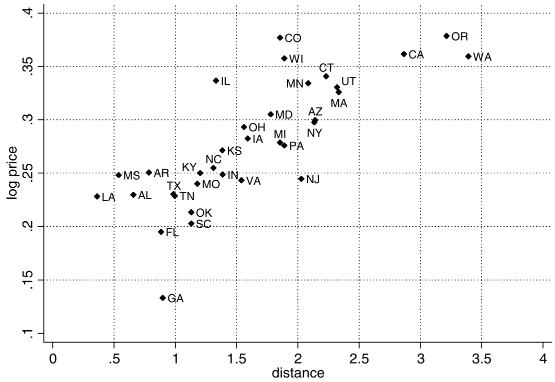


FIGURE 5. Price of gasoline and distance to the Gulf of Mexico. Distance to the respective state capital is measured in 1000 km. See text for details.

test. Panel B shows that modifications to the bandwidth parameters do not affect the conclusions from this test.

6. CONCLUSIONS

Simple parametric models of demand functions can yield misleading estimates of price sensitivity and welfare measures such as DWL, owing to misspecification. Fully nonparametric or semi-nonparametric estimation of demand reduces the risk of misspecification, but, because of the effects of random sampling errors, can yield imprecise estimates with anomalous properties such as non-monotonicity. This paper has shown that these problems can be overcome by constraining semi-nonparametric estimates to satisfy the Slutsky condition of economic theory. This stabilizes the semi-nonparametric estimates without the need for fully parametric or other restrictions that have no basis in economic theory.

TABLE 5. Exogeneity test.[a]

	Test Stat. (1)	Crit. Value (5%) (2)	p-Value (3)	Rejection (4)
A. Main estimate	0.066	0.174	0.692	no
B. Sensitivity to bandwidth choice: All bandwidths multiplied by				
factor 0.80	0.084	0.197	0.621	no
factor 1.25	0.050	0.155	0.751	no
factor 1.50	0.042	0.149	0.781	no

[a]This table shows results from the exogeneity test from Blundell and Horowitz (2007). In a first step, we remove the partially linear component as before, using the bandwidth choice corresponding to the middle-income group. In the second step, we implement the exogeneity test. For this, we restrict the sample to incomes above \$15,000 and log prices to the range between 0.18 and 0.38 (resulting in 4520 observations). We rescale price, income, and distance into the [0; 1] range and adjust bandwidths accordingly. For the distance dimension, we set the bandwidth to 0.15 (panel A, after transforming this variable into the unit interval).

We have implemented this approach by using a modified kernel estimator that weights the observations so as to satisfy the Slutsky restriction. To illustrate the method, we have estimated a gasoline demand function for a class of households in the United States. We find that a semi-nonparametric estimate of the demand function is non-monotonic. The estimate that is constrained to satisfy the Slutsky condition is well behaved. Moreover, the constrained semi-nonparametric estimates show patterns of price sensitivity that are very different from those of the simple parametric model. We find price responses vary non-monotonically with income. In particular, we find that low- and high-income consumers are less responsive to changes in gasoline prices than are middle-income consumers. Similar results are found for comparable Canadian data.

We have also computed the DWL of an increase in the price of gasoline. The constrained semi-nonparametric estimates of DWL are quite different from those obtained with the parametric model. Mirroring the results on price responsiveness, the DWL estimates are highest for middle-income groups. These results illustrate the usefulness of nonparametrically estimating demand functions subject to the Slutsky condition.

Appendix

Table A1. Estimates of the partially linear component.[a]

	\$42,500 (1)	\$57,500 (2)	\$72,500 (3)
Log age of household respondent	−0.024	−0.024	−0.015
	[−0.103; 0.057]	[−0.101; 0.054]	[−0.089; 0.062]
Log household size	0.055	0.055	0.070
	[−0.022; 0.133]	[−0.022; 0.133]	[−0.006; 0.148]
Log number of drivers	0.522	0.522	0.500
	[0.417; 0.617]	[0.418; 0.618]	[0.396; 0.595]
Number of workers in household	0.091	0.091	0.096
	[0.065; 0.12]	[0.066; 0.119]	[0.071; 0.125]
Public transit indicator	−0.042	−0.042	−0.037
	[−0.082; 0.002]	[−0.083; 0.003]	[−0.078; 0.011]
Small town	−0.045	−0.045	−0.049
	[−0.106; 0.016]	[−0.108; 0.017]	[−0.108; 0.014]
Suburban	−0.165	−0.165	−0.168
	[−0.242; −0.09]	[−0.242; −0.09]	[−0.242; −0.089]
Second city	−0.175	−0.175	−0.175
	[−0.257; −0.093]	[−0.257; −0.092]	[−0.252; −0.091]
Urban	−0.169	−0.169	−0.162
	[−0.277; −0.059]	[−0.277; −0.058]	[−0.265; −0.052]
Population density (8)	Yes	Yes	Yes
Observations	5254	5254	5254

[a]Bootstrapped standard errors based on 5000 replications.

Table A2. Confidence intervals for DWL measures.[a]

	Semi-Nonparametric		Parametric (Log–Log)	
Income	Lower (1)	Upper (2)	Lower (3)	Upper (4)
		A. DWL (as % of tax paid)		
\$72,500	−7.52%	10.63%	1.60%	6.62%
\$57,500	−4.97%	13.00%	1.77%	6.49%
\$42,500	−7.53%	12.96%	1.65%	6.48%
		B. DWL (relative to income) $\times 10^4$		
\$72,500	−2.90	4.87	0.72	2.69
\$57,500	−1.94	6.61	0.91	3.11
\$42,500	−3.63	7.93	1.08	3.83

[a]This table shows confidence intervals corresponding to estimates reported in Table 4. Confidence intervals are computed with an undersmoothed bandwidth, based on 5000 replications. See text for details.

References

Ascher, U. M. and L. R. Petzold (1998), *Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations*. SIAM, Philadelphia. [34]

Bento, A. M., L. H. Goulder, E. Henry, M. R. Jacobsen, and R. H. von Haefen (2005), "Distributional and efficiency impacts of gasoline taxes: An econometrically-based multimarket study." *American Economic Review*, 95 (2), 282–287. [36]

Bento, A. M., L. H. Goulder, M. R. Jacobsen, and R. H. von Haefen (2009), "Distributional and efficiency impacts of increased U.S. gasoline taxes." *American Economic Review*, 99 (3), 667–699. [36]

Berry, S., J. Levinsohn, and A. Pakes (1995), "Automobile prices in market equilibrium." *Econometrica*, 63 (4), 841–890. [46]

Blundell, R., X. Chen, and D. Kristensen (2007), "Semi-nonparametric IV estimation of shape-invariant Engel curves." *Econometrica*, 75 (6), 1613–1669. [46]

Blundell, R. and J. L. Horowitz (2007), "A non-parametric test of exogeneity." *Review of Economic Studies*, 74 (4), 1035–1058. [46, 47]

Dahl, C. A. (1979), "Consumer adjustment to a gasoline tax." *The Review of Economics and Statistics*, 61 (3), 427–432. [36]

Deaton, A. and J. Muellbauer (1980), *Economics and Consumer Behavior*. Cambridge University Press, Cambridge. Reprinted 1999. [36]

EIA (2008), "Monthly energy review." April 2008, Energy Information Administration, U.S. Department of Energy. [30]

EIA (2010a), "Regional gasoline price differences." Energy Information Administration, U.S. Department of Energy. Available at http://www.eia.gov/energyexplained/index.cfm?page=gasoline_regional. [35]

EIA (2010b), "Petroleum navigator." Energy Information Administration, U.S. Department of Energy. Available at http://www.eia.gov/. [46]

EIA (2010c), "Annual energy review 2009." Report DOE/EIA-0384(2009), Energy Information Administration, U.S. Department of Energy. [30]

Epstein, L. G. and A. J. Yatchew (1985), "Non-parametric hypothesis testing procedures and applications to demand analysis." *Journal of Econometrics*, 30 (1–2), 149–169. [30]

Hall, P. (1992), *The Bootstrap and Edgeworth Expansion*. Springer, New York. [35]

Hall, P. and J. L. Horowitz (2005), "Nonparametric methods for inference in the presence of instrumental variables." *The Annals of Statistics*, 33 (6), 2904–2929. [46]

Hall, P. and L.-S. Huang (2001), "Nonparametric kernel regression subject to monotonicity constraints." *The Annals of Statistics*, 29 (3), 624–647. [32]

Härdle, W. (1990), *Applied Nonparametric Regression*. Econometric Society Monograph Series 19. Cambridge University Press, Cambridge. [32, 39]

Härdle, W. and J. S. Marron (1991), "Bootstrap simultaneous error bars for nonparametric regression." *The Annals of Statistics*, 19 (2), 778–796. [34, 35]

Hausman, J. A. and W. K. Newey (1995), "Nonparametric estimation of exact consumers surplus and deadweight loss." *Econometrica*, 63 (6), 1445–1476. [30, 37, 40]

Hughes, J. E., C. R. Knittel, and D. Sperling (2008), "Evidence of a shift in the short-run price elasticity of gasoline demand." *The Energy Journal*, 29 (1), 93–114. [36]

Li, S., C. Timmins, and R. H. von Haefen (2009), "How do gasoline prices affect fleet fuel economy?" *American Economic Journal: Economic Policy*, 1 (2), 113–137. [42]

Nadaraya, E. A. (1964), "On estimating regression." *Theory of Probability and Its Applications*, 9 (1), 141–142. [32]

ORNL (2004), "2001 National Household Travel Survey." User's Guide, Oak Ridge National Laboratory. Available at http://nhts.ornl.gov/2001/usersguide/UsersGuide.pdf. [35]

Poterba, J. N. (1991), "Is the gasoline tax regressive?" Working Paper 3578, NBER. [36]

Robinson, P. M. (1988), "Root-N-consistent semiparametric regression." *Econometrica*, 56 (4), 931–954. [33]

Schmalensee, R. and T. M. Stoker (1999), "Household gasoline demand in the United States." *Econometrica*, 67 (3), 645–662. [35, 37, 42]

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London. [39]

U.S. Census Bureau (2010), "Annual resident population estimates, estimated components of resident population change, and rates of the components of resident population change for states and counties: April 1, 2000 to July 1, 2009." Population Division, Release Date: March 2010. Available at http://www.census.gov/popest/data/counties/totals/2009/files/CO-EST2009-ALLDATA.csv. [43]

U.S. Census Bureau (2011), "Geographic relationship file: 1999 MA to 2003 CBSA." Available at http://www.census.gov/population/metro/files/CBSA03_MSA99.xls. [43]

Varian, H. R. (1982), "The nonparametric approach to demand analysis." *Econometrica*, 50 (4), 945–973. [30]

Varian, H. R. (1983), "Non-parametric tests of consumer behaviour." *Review of Economic Studies*, 50 (1), 99–110. [30]

Watson, G. S. (1964), "Smooth regression analysis." *Sankhya: The Indian Journal of Statistics, Series A*, 26 (4), 359–372. [32]

West, S. E. (2004), "Distributional effects of alternative vehicle pollution control policies." *Journal of Public Economics*, 88, 735–757. [36, 37, 38]

Working, E. J. (1927), "What do statistical 'demand curves' show?" *The Quarterly Journal of Economics*, 41 (2), 212–235. [46]

Yatchew, A. and L. Bos (1997), "Nonparametric least squares regression and testing in economic models." *Journal of Quantitative Economics*, 13, 81–131. [30]

Yatchew, A. and J. A. No (2001), "Household gasoline demand in Canada." *Econometrica*, 69 (6), 1697–1709. [37, 41]

Submitted June, 2010. Final version accepted October, 2011.