

Nonparametric Estimation of a Nonseparable Demand Function under the Slutsky Inequality Restriction

Richard Blundell*, Joel Horowitz[†] and Matthias Parey[‡]

October 2013, this version February 2016[§]

Abstract

We derive conditions under which a demand function with nonseparable unobserved heterogeneity in tastes can be estimated consistently by nonparametric quantile regression subject to the shape restriction from the Slutsky inequality. We consider nonparametric estimation of the nonseparable demand for gasoline in the U.S. The estimated function detects differences in behavior between heavy and moderate gasoline users, and reveals systematic variation in the responsiveness of demand to plausible changes in prices across the income distribution. We test for exogeneity of prices and develop a new method for estimating quantile instrumental variables to allow for endogeneity of prices. The empirical results illustrate the improvements in finite-sample performance of a nonparametric estimator from imposing shape restrictions based on economic theory.

JEL: C14, C21, D12

Keywords: consumer demand, nonseparable models, quantile regression, nonparametric methods.

*Ricardo Professor of Political Economy, Department of Economics, University College London (UCL), and Institute for Fiscal Studies (IFS). 7 Ridgmount Street, London WC1E 7AE, United Kingdom, email: r.blundell@ucl.ac.uk, tel: 44 (0)20 7679 5863, fax: 44 (0)20 7916 2775 (corresponding author).

[†]Charles E. and Emma H. Morrison Professor of Market Economics, Department of Economics, Northwestern University, and Cemmap. 2001 Sheridan Road, Evanston, Illinois 60208, USA, email: joel-horowitz@northwestern.edu.

[‡]Lecturer (Assistant Professor), Department of Economics, University of Essex, and Institute for Fiscal Studies (IFS). Wivenhoe Park, Colchester CO4 3SQ, United Kingdom, email: m.parey@essex.ac.uk.

[§]We thank Stefen Hoderlein, Whitney Newey, two anonymous referees, the editor and seminar participants at Bristol, Cologne, Essex, Manchester, Southampton, Northwestern, and the Statistical Meeting (DAGStat) in Freiburg for helpful comments. We are grateful to the ESRC Centre for the Microeconomic Analysis of Public Policy at IFS and the ERC grant MICROCONLAB at UCL for financial support.

1 Introduction

Although the microeconomic theory of consumer choice provides shape restrictions on individual demand behavior, it does not provide a finite-dimensional parametric model of demand, see Mas-Colell et al. (1995). This motivates use of nonparametric methods in the study of empirical demand behavior on micro-level data, see Matzkin (2007), and references therein. However, typically nonparametric methods apply to conditional mean regression and will only recover interpretable individual demand when unobserved heterogeneity is additively separable in the regression model. Additive separability occurs under restrictive assumptions about preferences. As Brown and Walker (1989) and Lewbel (2001) have shown, demand functions generated from random utility functions do not typically yield demand functions where the unobserved tastes are additive.

The objective in this paper is to present an analysis of individual demands and welfare, as well the distribution of these objects, for plausible changes in prices. For example, in our application to gasoline demand, the approach detects strong differences in behavior between heavy and moderate gasoline users, and reveals systematic idiosyncratic variation in demand and welfare costs for discrete price changes across the income and taste distribution. The identification and estimation of individual consumer demand models that are consistent with unobserved taste variation require analyzing demand models with nonadditive random terms. Matzkin (2003, 2008) derives general identification results for models that are nonseparable in unobserved heterogeneity in tastes. Under suitable restrictions, quantile estimation allows us to recover demand at a specific point in the distribution of unobservables. This motivates our interest in a quantile estimator and represents a significant development on work estimating average demands, for example Blundell, Horowitz, and Parey (2012).

We utilize a monotonicity assumption on scalar unobserved heterogeneity to recover individual demands. Although we provide a class of preferences which generate individual demands with these properties, the scalar heterogeneity assumption is restrictive but not testable empirically. Dette et al (2016) show that the quantile demand function satisfies the Slutsky restriction even when there is multidimensional heterogeneity. As Hausman and Newey (2015) note, when the quantile satisfies the Slutsky condition and the budget constraint, there is a demand model with scalar (uniform)

heterogeneity that gives the same conditional distribution of quantity given income as the quantile. They refer to this model where the quantile is the demand function as quantile demand. In the absence of scalar heterogeneity, the quantile demands can no longer be interpreted as individual demands and the complete distribution of counterfactual demands is no longer identified.

To diminish the concerns over the scalar heterogeneity assumption, we restrict our empirical analysis of gasoline demand to a group of relatively homogeneous households. This approach is similar in spirit to that of Graham et al. (2015) which also includes only a single dimension of unobserved heterogeneity but reduces the strength of the assumption by conditioning on all leads and lags of the regressors. We further show that the consumer expenditure function obtained under a possibly incorrect assumption of univariate unobserved heterogeneity is a good approximation to the correct expenditure function based on multivariate unobserved heterogeneity if the additional components of unobserved heterogeneity are small.

We point out the benefits of imposing the Slutsky assumption even in the multidimensional heterogeneity case as a way of stabilizing the nonparametric estimator of the quantile demand function. Without adding further structure, nonparametric estimates of the demand function have the drawback of being noisy due to random sampling errors. The estimated function can be wiggly and nonmonotonic. Consequently predictions of individual demand can be erratic and some estimates of individual deadweight losses can have signs that are noninterpretable within the usual consumer choice model, even though true preferences may be well-behaved. One solution is to impose a parametric or semiparametric structure on the demand function. But there is no guarantee that such a structure is correct or approximately correct and demand estimation using a misspecified model can give seriously misleading results.

We impose the Slutsky restriction of consumer theory on an otherwise fully nonparametric estimate of the nonseparable demand function. This yields well-behaved estimates of the demand function but avoids arbitrary and possibly incorrect parametric or semiparametric restrictions. We show that Slutsky constrained nonparametric estimates reveal features of the demand function that are not present in simple parametric models. Where prices take only a few discrete values a related approach is to impose the Afriat revealed preference inequalities, see Blundell et al. (2014). Our method is quite different, directly using the Slutsky condition rather than the sequence of revealed preference

inequalities that obtain in the discrete price case.

We do not carry out inference based on the constrained estimator. Under the assumption that the Slutsky restriction is not binding in the population, the constrained and unconstrained estimators are equal with probability approaching unity as the sample size increases, and the two estimators have the same asymptotic distribution. Therefore, asymptotic inference based on the constrained estimator is the same as asymptotic inference based on the unconstrained estimator. However, in finite samples such as that in this paper, the two estimators are different and have different sampling distributions. Consequently, the relevant distribution for inference based on the constrained estimator is the finite-sample distribution, not the asymptotic distribution. Chernozhukov et al. (2009) have developed methods for carrying out finite-sample inference with unconstrained quantile estimators. Finite-sample methods are not available for constrained estimators such as the one used here.¹ We use the bootstrap based on the unconstrained estimator to obtain confidence bands for the demand function. In terms of statistical precision we expect the additional structure provided by the shape restriction to improve the finite-sample performance of our estimator, analogous to the way sign restrictions in parametric models reduce the Mean Squared Error (MSE). Nonparametric estimation often requires the choice of bandwidth parameters, such as kernel bandwidths or number of knots for a spline. These parameters are optimally chosen in a way which balances bias and variance of the estimates. The use of shape restrictions, reducing the variance of the estimates, modifies this trade-off, and therefore allows potentially for smaller optimal bandwidth choices. Shape restrictions can therefore be thought of as a substitute for bandwidth smoothing, helping to recover the features of interest of the underlying relationship.

We provide a substantive illustration of the methods with an application to the demand for gasoline in the U.S. Given the changes in the price of gasoline that have been observed in recent years, and the role of taxation in the gasoline market, understanding the elasticity of demand is of key policy interest. We pay particular attention to the question of how demand behavior varies across the income distribution, and ask whether the welfare implications of price changes are uniform across the income distribution. Using household-level data from the 2001 National Household Travel Survey

¹See Wolak (1991) on asymptotically valid hypothesis tests involving inequality restrictions in nonlinear models.

(NHTS) complemented by travel diaries and odometer readings, we find constrained estimates are monotonic and reveal features not easily found with parametric models. This is an example where very simple parametric models impose strong restrictions on the behavioral responses allowed for, which may in turn affect resulting policy conclusions.

Our work on the specification of gasoline demand relates to a longstanding literature. Hausman and Newey (1995) develop the nonparametric estimation of conditional mean of gasoline demand. Schmalensee and Stoker (1999) further consider the nonparametric demand curve for gasoline. Yatchew and No (2001) estimate a partially linear model of gasoline demand. Blundell, Horowitz, and Parey (2012) extend this work to the nonparametric estimation of conditional mean demand under Slutsky inequality shape restriction and also consider the possible endogeneity of the price variable. Hoderlein and Vanhems (2011) incorporate endogenous regressors in a control function approach. The approach developed here identifies and estimates the complete distribution of individual demands in the nonseparable case, thereby relaxing the strong assumptions on unobserved heterogeneity necessary to interpret the conditional mean regression. Hausman and Newey (2013) estimate certain features of average behavior in a framework with multidimensional unobserved heterogeneity. The approach we take allows us to study differential effects of price changes and welfare costs across the distribution of unobservables. For example, quantile estimation allows us to compare the price and income responses of heavy users with those of moderate or light users.² We show that there is systematically more responsive price behavior among the middle income consumers. This remains true across consumers with different intensity of use. We also estimate the deadweight loss of a tax by integrating under the demand function to obtain the expenditure function. Some estimates of deadweight losses using unconstrained demand function are negative. This is unsurprising given non-monotonicity of the unconstrained estimated demand function. Our constrained estimates show that the middle income group has the largest loss.

The paper proceeds as follows. The next section develops our nonseparable model of demand behavior and the restrictions required for a structural interpretation. Section 3 presents our estimation method, where we describe the nonparametric estimation method for both the unconstrained esti-

² In the context of alcohol demand, for example, Manning et al. (1995) show that price responsiveness differs at different quantiles.

mates and those obtained under the Slutsky constraint. We also present our procedure for quantile estimation under endogeneity. In Section 4 we discuss the data we use in our investigation and present our empirical findings. We compare the quantile demand estimates to those from a conditional mean regression. The endogeneity of prices is considered in Section 5 where we present the results of an exogeneity test and our quantile instrumental variables procedure. Section 6 concludes.

2 Unobserved Heterogeneity and Demand Functions

The consumer model of interest in this paper is

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

where W is demand (measured as budget share), P is price, Y is income, and U represents (non-separable) unobserved heterogeneity. We impose two types of restrictions on this demand function: The first set of restrictions addresses the way unobserved heterogeneity enters demand, and its relationship to price and income. The second are shape restrictions from consumer choice theory.

In terms of the restrictions on unobserved heterogeneity, we assume that demand g is monotone in the unobserved heterogeneity U . To ensure identification, we for now assume that U is statistically independent of (P, Y) . We relax this assumption in Section 3.4. Given these assumptions, we can also assume without loss of generality that $U \sim U[0; 1]$. This allows recovery of the demand function for specific types of households from the observed conditional quantiles of demand: the α quantile of W , conditional on (P, Y) , is

$$\mathcal{Q}_\alpha(W|P, Y) = g(P, Y, \alpha) \equiv G_\alpha(P, Y). \tag{2}$$

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

In contrast, the conditional mean is

$$\begin{aligned} E(W|P = p, Y = y) &= \int g(p, y, u) f_U(u) du \\ &\equiv m(p, y), \end{aligned}$$

where $f_U(u)$ is the probability density function of U . Given that we are interested in imposing shape restrictions based on consumer theory, estimating the demand function at a specific value of $U = \alpha$ using quantile methods is attractive because economic theory informs us about $g(\cdot)$ rather than $m(\cdot)$. It is possible therefore that $m(\cdot)$ does not satisfy the restrictions even though each individual consumer does (see also Lewbel (2001)).

To illustrate these points we consider a class of preferences that generate nonseparable demands that satisfy monotonicity in unobserved heterogeneity. There are two goods, q_1 and numeraire q_0 . Suppose preferences have the form:

$$\begin{aligned} U(q_1, q_0, u) &= v(q_1, q_0) + w(q_1, u) \\ \text{subject to } &p q_1 + q_0 \leq y \end{aligned} \tag{3}$$

where we have normalised the price of q_0 to unity. Matzkin (2007) shows that provided the functions v and w are twice continuously differentiable, strictly increasing and strictly concave, and that $\partial^2 w(q_1, u) / \partial q_1 \partial u > 0$, then the demand function for q_1 is invertible in u . Hence, the demand function for q_1 will satisfy the restrictions of consumer choice (the Slutsky inequality in this case) for each value u . Similarly, budget shares will be monotonic in u .

Under these assumptions quantile demands will recover individual demands and will satisfy Slutsky inequality restrictions. However, apart from very special cases, neither demands nor budget shares, will be additive in u . Consequently, average demands will not recover individual demands. For the nonseparable demand case, where there are *high dimensional unobservables*, Dette et al. (2016) show that the Slutsky inequality holds for quantiles if individual consumers satisfy Slutsky restrictions. This is a key result as it provides a more general motivation for Slutsky constrained estimation of the kind developed in this paper. However, in their framework, quantile demands do not identify individual demand behavior. A central objective of our study is to estimate the impact of plausible

changes in prices on individual demands and on individual welfare, as well the distribution of these objects.

Hausman and Newey (2013) consider the case of multi-dimensional unobserved heterogeneity; they show that in this case neither the demand function nor the dimension of heterogeneity is identified. They estimate quantile demands and use bounds on the income effect to derive bounds for average surplus. In Appendix A.3 we argue that the individual consumer expenditure function obtained under a possibly incorrect assumption of univariate unobserved heterogeneity is a good approximation to the correct expenditure function based on multivariate unobserved heterogeneity if the additional components of unobserved heterogeneity are small. To diminish the concerns over our scalar heterogeneity assumption, in our empirical analysis we restrict the analysis to a group of relatively homogeneous households.

In the context of scalar heterogeneity, Hoderlein and Vanhems (2011) consider identification of welfare effects, and allow for endogenous regressors in a control function approach. Hoderlein (2011) studies the testable implications of negative semidefiniteness as well as symmetry of the Slutsky matrix in a heterogeneous population. Hoderlein and Stoye (2014) investigate how violations of the Weak Axiom of Revealed Preference (WARP) can be detected in a heterogeneous population based on repeated cross-sectional data. Using copula methods they relax the monotonicity restriction and bound the fraction of the population violating WARP.

We impose the Slutsky constraint by restricting the price and income responses of the demand function g . Preference maximization implies that the Slutsky substitution matrix is symmetric negative semidefinite (Mas-Colell et al., 1995). Ensuring that our estimates satisfy this restriction is however not only desirable because of the increase in precision from additional structure, it is also a necessary restriction in order to be able to perform welfare analysis. Welfare analysis requires knowledge of the underlying preferences. The question under which conditions we can recover the utility function from the observed Marshallian demand function, referred to as the integrability problem, has therefore been of long-standing interest in the analysis of consumer behavior (Hurwicz and Uzawa, 1971). A demand function which satisfies adding up, homogeneity of degree zero, and a symmetric negative semidefinite Slutsky matrix allows recovery of preferences (Deaton and Muellbauer, 1980). As Deaton and Muellbauer (1980) emphasize, these characteristics also represent

the *only* structure that is implied by utility maximization. Slutsky negative semidefiniteness is therefore critical for policy analysis of changes in the prices consumers face. In the context of the two good model considered here, these integrability conditions are represented through the negative compensated price elasticity of gasoline demand.³

In previous work household demographics or other household characteristics have been found to be relevant determinants of transport demand. One possibility of accounting for these characteristics would be to incorporate them in a semiparametric specification. However, in order to maintain the fully nonparametric nature of the model, we instead condition on a set of key demographics in our analysis.⁴ Thus we address the dimension-reduction problem by conditioning on a particular set of covariates. This exploits the fact that the relevant household characteristics are all discrete in our application. We then estimate our nonparametric specification on this sample which is quite homogeneous in terms of household demographics.

Note finally that prices could be endogenous in the demand function (2). We will later relax the assumption of independence between U and the price P , test for endogeneity following the cost-shifter approach in Blundell et al. (2012) and present instrumental variables estimates. Imbens and Newey (2009) define the quantile structural function (QSF) as the α -quantile of demand $g(p, y, U)$, for fixed p and y ; under endogeneity of prices, the QSF will be different from the α -quantile of $g(P, Y, U)$, conditional on $P = p$ and $Y = y$.

³See Lewbel (1995) and Haag et al. (2009) on testing and imposing Slutsky symmetry. The quantile approach we adopt for identifying individual demand and welfare counterfactuals applies only to the scalar heterogeneity two-good case examined in this paper. The treatment of multiple goods is a topic for future research.

⁴ These characteristics include household composition and life-cycle stage of the household, race of the survey respondent, and as well as the urban-rural location of the household. We describe these selection criteria in detail in Section 4.1 below.

3 Nonparametric Estimation

3.1 Unconstrained Nonparametric Estimation

From equation (2), we can write

$$W = G_\alpha(P, Y) + V_\alpha; \quad P(V_\alpha \leq 0 \mid P, Y) = \alpha, \quad (4)$$

where V_α is a random variable whose α quantile conditional on (P, Y) is zero. We estimate G_α using a truncated B-spline approximation with truncation points M_1 and M_2 chosen by cross-validation.⁵

Thus

$$G_\alpha(P, Y) = \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} c_{m_1, m_2; \alpha} B_{m_1}^p(P) B_{m_2}^y(Y),$$

where B^p and B^y (with indices m_1 and m_2) are spline functions following Powell (1981)⁶ and $c_{m_1, m_2; \alpha}$ is the finite-dimensional matrix of coefficients.

We denote the data by $\{W_i, P_i, Y_i : i = 1, \dots, n\}$. The estimator is defined in the following optimization problem:

$$\min_{\{c_{m_1, m_2; \alpha}\}} \sum_{i=1}^n \rho_\alpha(W_i - G_\alpha(P_i, Y_i)), \quad (5)$$

where $\rho_\alpha(V) = (\alpha - 1 [V < 0]) V$ is the check function.

3.2 Estimation Subject to the Slutsky Inequality

One contribution of this paper is to provide estimates of the quantile demand function subject to the Slutsky inequality restriction. As Dette et al. (2016) have shown, even in the presence of multidimensional unobserved heterogeneity, the Slutsky condition will hold at each quantile provided it holds for every individual in the sample. Our estimation results show that the Slutsky restriction considerably improves the properties of the estimated quantile demand function, removing the wiggly behavior of the nonparametric estimator. Under the assumption of scalar heterogeneity, the Slutsky constrained quantile demand function further identifies the individual demand function allowing us to recover the impact of changes in prices on the distribution of individual demands and the

⁵See Section 4.1 for details on the cross-validation procedure.

⁶See Powell (1981), equation (19.25).

distribution of individual welfare measures.

The Slutsky condition is imposed on the nonparametric estimate of the conditional quantile function. Writing this condition in terms of shares, and taking price and income to be measured in logs⁷, gives

$$\frac{\partial \hat{G}_\alpha^C(P, Y)}{\partial p} + \hat{G}_\alpha^C(P, Y) \frac{\partial \hat{G}_\alpha^C(P, Y)}{\partial y} \leq \hat{G}_\alpha^C(P, Y) \left(1 - \hat{G}_\alpha^C(P, Y)\right), \quad (6)$$

where the superscript C indicates that the estimator is constrained by the Slutsky condition.

The Slutsky constrained estimator is obtained by solving the problem (5), subject to (6), for all (P, Y) . This problem has uncountably many constraints. We replace the continuum of constraints by a discrete set, thereby solving:

$$\min_{\{c_{m_1, m_2; \alpha}\}} \sum_{i=1}^n \rho_\alpha \left(W_i - \hat{G}_\alpha^C(P_i, Y_i) \right)$$

subject to

$$\frac{\partial \hat{G}_\alpha^C(p_j, y_j)}{\partial p} + \hat{G}_\alpha^C(p_j, y_j) \frac{\partial \hat{G}_\alpha^C(p_j, y_j)}{\partial y} \leq \hat{G}_\alpha^C(p_j, y_j) \left(1 - \hat{G}_\alpha^C(p_j, y_j)\right), \quad j = 1, \dots, J,$$

where $\{p_j, y_j : j = 1, \dots, J\}$ is a grid of points. To implement this, we use a standard optimization routine from the NAG library (E04UC). In the objective function we use a check function which is locally smoothed in a small neighborhood around zero (Chen, 2007). We show that the resulting demand figures are not sensitive to a range of alternative values of the corresponding smoothing parameter. For imposing the constraints, we choose a fine grid of points along the price dimension, at each of the 15 income category midpoints.

No method currently exists for carrying out inference based on the Slutsky restricted estimator. Therefore, we use the bootstrap based on the unconstrained estimator to obtain confidence bands for the demand function. Asymptotically, these bands satisfy the Slutsky restriction if it does not

⁷Denoting $Q = H_\alpha^C(P, Y)$ as the Marshallian demand function for the good in question, Slutsky negativity requires

$$\frac{\partial H_\alpha^C(P, Y)}{\partial P} + \frac{\partial H_\alpha^C(P, Y)}{\partial Y} G_\alpha^C(P, Y) \leq 0.$$

Using the definition of the share and substituting leads to equation (6). See also Deaton and Muellbauer (1980) for a similar share specification in the context of the Almost Ideal Demand System.

bind in the population. If the Slutsky restriction is binding in the population or in the sample, then the bands based on the unconstrained estimator are at least as wide as bands based on the restricted estimator would be if methods for obtaining such bands were available. This is because the Slutsky restriction reduces the size of the feasible region for estimation. Regions within the unconstrained confidence band for which the restriction is violated are excluded from a confidence band based on the restricted estimator.

To investigate the gain in precision from imposing the shape restriction, we have conducted a simulation study. In this exercise, we draw simulated data in a setting similar to our empirical application. For each simulation draw, we estimate both the constrained and the unconstrained demand function. To compare the variability of these two sets of estimates, we compute confidence intervals across the simulation runs. We summarize the findings from this exercise below.

3.3 Individual Welfare Measures

The estimates of the Slutsky constrained demand function can then be used to recover the distribution of individual welfare measures, including deadweight loss (DWL). For this purpose, we consider a hypothetical *discrete* tax change which moves the price from p^0 to p^1 . Let $e(p)$ denote the expenditure function at price p and some reference utility level. The DWL of this price change is given by

$$L(p^0, p^1) = e(p^1) - e(p^0) - (p^1 - p^0) H_\alpha [p^1, e(p^1)],$$

where $H_\alpha(p, y)$ is the Marshallian demand function. $L(p^0, p^1)$ is computed by replacing e and H with consistent estimates. The estimator of e , \hat{e} , is obtained by numerical solution of the differential equation

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

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

Notice that our focus here is on discrete changes in prices. If one is interested in the impact of marginal changes in prices on average welfare measures, see Chetty (2009) for example, then local derivatives of average behavior and weaker conditions on unobserved heterogeneity may be sufficient. Although, as Hausman and Newey (2015) show, nonlinearities in income lead to the average surplus

for quantile demand being different than average surplus for the true demand.

3.4 Quantile Instrumental Variable Estimation

To recognize potential endogeneity of prices, we introduce a cost-shifter instrument Z for prices. In the application this is a distance measure to gulf supply refinery to reflect transport costs. Consider again equation (4) from above, where now we impose the quantile restriction conditional on the distance instrument (and household income):

$$W = G_\alpha(P, Y) + V_\alpha; \quad P(V_\alpha \leq 0 \mid Z, Y) = \alpha.$$

The identifying relation can be written as

$$P(W - G_\alpha(P, Y) \leq 0 \mid Z, Y) = \alpha.$$

Let $f_{Z,Y}$ be the probability density function of (Z, Y) . Then we have

$$\int_{Z \leq z, Y \leq y} P(W - G_\alpha(P, Y) \leq 0 \mid Z, Y) f_{Z,Y}(Z, Y) dZ dY = \alpha P(Z \leq z, Y \leq y)$$

for all (z, y) . An empirical analog is

$$n^{-1} \sum_{i=1}^n 1 [W_i - G_\alpha(P_i, Y_i) \leq 0] 1 [Z_i \leq z, Y_i \leq y] = \frac{\alpha}{n} \sum_{i=1}^n 1 [Z_i \leq z, Y_i \leq y].$$

Define

$$Q_n(G_\alpha, z, y) = n^{-1} \sum_{i=1}^n \{1 [W_i - G_\alpha(P_i, Y_i) \leq 0] - \alpha\} 1 [Z_i \leq z, Y_i \leq y]. \quad (7)$$

Estimate G_α by solving

$$\min_{G_\alpha \in \mathcal{H}_n} \int Q_n(G_\alpha, z, y)^2 dz dy,$$

where \mathcal{H}_n is the finite-dimensional space consisting of truncated series approximations and includes the shape restriction when we impose it.

3.5 A Test of Exogeneity

Building on the work for the conditional mean case in Blundell and Horowitz (2007), we follow Fu et al. (2015) and develop a nonparametric exogeneity test in a quantile setting. As Blundell and Horowitz (2007), this approach does not require an instrumental variables estimate, and instead tests the exogeneity hypothesis directly. By avoiding the ill-posed inverse problem, it is likely to have substantially better power properties than alternative tests.

We require a test of the hypothesis that an explanatory variable P in a quantile regression model is exogenous against the alternative that P is not exogenous.

The object of interest is the unknown function G_α that is identified by

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

and

$$P(V_\alpha \leq 0 \mid Y = y, Z = z) = \alpha$$

for almost every $(y, z) \in \text{supp}(Y, Z)$, where W , P , Y , and Z are observable, continuously distributed random variables; Z is an instrument for P ; V_α is an unobservable continuously distributed random variable; and α is a constant satisfying $0 < \alpha < 1$. Equivalently, G_α is the solution to

$$P[W - G_\alpha(P, Y) \leq 0 \mid Y = y, Z = z] = \alpha$$

for almost every $(y, z) \in \text{supp}(Y, Z)$. Now consider the unknown function K_α that is identified by

$$W = K_\alpha(P, Y) + V_\alpha$$

and

$$P(V_\alpha \leq 0 \mid P = p, Y = y) = \alpha.$$

The null hypothesis to be tested is⁸

$$H_0 : K(p, y) = G(p, y)$$

for almost every $(p, y) \in \text{supp}(P, Y)$. The alternative hypothesis is

$$H_1 : P[K(P, Y) \neq G(P, Y)] > 0.$$

K can be estimated consistently by nonparametric quantile regression, and G can be estimated consistently by nonparametric instrumental variables quantile regression. Denote the estimators of K and G by \hat{K} and \hat{G} , respectively. H_0 can be tested by determining whether the difference between \hat{K} and \hat{G} in some metric is larger than can be explained by random sampling error. H_0 is rejected if the difference is too large. However, this approach to testing H_0 is unattractive because estimation of G is an ill-posed inverse problem. The rate of convergence of \hat{G} to G is unavoidably slow, and the resulting test has low power.

However, as in Blundell and Horowitz (2007), estimation of G and the ill-posed inverse problem can be avoided by observing that under H_0 ,

$$P[W - K(P, Y) \leq 0 \mid Y = y, Z = z] = \alpha. \tag{8}$$

Equation (8) can then be used to obtain a test statistic for H_0 . More details on the derivation, properties and computation of the test statistic are given in Appendix A.1.

4 Estimation Results

4.1 Data

The data are from the 2001 National Household Travel Survey (NHTS). The NHTS surveys the civilian non-institutionalized population in the United States. This is a household-level survey

⁸ Note that to simplify the notation, we drop the α -subscript from G_α and K_α in the remainder of this section.

conducted by telephone and complemented by travel diaries and odometer readings.⁹ We select the sample to minimize heterogeneity as follows: we restrict the analysis to households with a white respondent, two or more adults, at least one child under age 16, and at least one driver. We drop households in the most rural areas, given the relevance of farming activities in these areas.¹⁰ We also restrict attention to those localities where the state of residence is known, and omit households in Hawaii due to its different geographic situation compared to continental U.S. states. Households where key variables are not reported are excluded and we restrict attention to gasoline-based vehicles (rather than diesel, natural gas, or electricity), requiring gasoline demand of at least one gallon; we also drop one observation where the reported gasoline share is larger than 1. We take vehicle ownership as given and do not investigate how changes in gasoline prices affect vehicle purchases or ownership. The results by Bento et al. (2009) indicate that price changes operate mainly through vehicle miles traveled rather than through fleet composition: they find that more than 95% of the reduction in gasoline consumption in response to an increase in gasoline tax is due to a reduction in vehicle miles traveled.

The resulting sample contains 3,640 observations. The key variables of interest are gasoline demand, price of gasoline, and household income. Corresponding sample descriptives are reported in Table 1; further detail on these variables can be found in Blundell et al. (2012).¹¹

[TABLE 1 ABOUT HERE]

The nonparametric estimates are shown below for the three income groups whose midpoints in 2001 dollars are \$42,500, \$57,500 and \$72,500. These income levels are chosen to compare the behavior of lower, middle and upper income households.¹²

We use cubic B-splines for our nonparametric analysis.¹³ For each quantile of interest, the number

⁹ See ORNL (2004) for further detail on the survey.

¹⁰ These are households in rural localities according to the Claritas urbanization index, indicating a locality in the lowest quintile in terms of population density (ORNL, 2004, Appendix Q).

¹¹ In the nonparametric analysis below, we impose two additional restrictions to avoid low-density areas in the data. For this purpose, we restrict attention to households with (2001) household income of at least \$15,000, facing a price of at least \$1.20.

¹² These three income points occupy the 19.1-22.8th, 34.2-42.3th, and 51.7-55.9th percentiles of the income distribution in our data (see Table 1).

¹³ In the income dimension, we place the knots at equally-spaced percentiles of a normal distribution,

of knots is obtained by cross-validation, separately for each quantile.¹⁴ The resulting number of (interior) knots is shown in Panel (1) of Table 2. In particular, at the median, the procedure indicates 4 interior knots in the price dimension and 3 knots in the income dimension. Across the quartiles, we obtain the same number of knots in the income dimension, while in the price dimension the cross-validation procedure indicates a more restrictive B-spline for the first quartile ($\alpha = 0.25$). In the subsequent analysis we follow these knot choices for both the unconstrained and the constrained quantile estimates under exogeneity. We have also investigated whether this cross-validation outcome is sensitive to outliers in the share variable. For this purpose, we have repeated the cross-validation procedure, leaving out the 10 highest and the 10 lowest gasoline budget share observations. The results are reported in Panel (2) of Table 2, suggesting that overall the number of knots is not very sensitive to this exercise.

[TABLE 2 ABOUT HERE]

4.2 Implications for the Pattern of Demand

Parametric benchmark specifications using linear quantile estimates can be found in Table 3, where we regress log quantity on log price and log income:

$$\log Q = \beta_0 + \beta_1 \log P + \beta_2 \log Y + U; \quad \mathcal{Q}_\alpha(U|P, Y) = 0.$$

For comparison we also report estimates obtained using an OLS estimator (see column (4)). These indicate a price elasticity of -0.83 and an income elasticity of 0.34. These are similar to those reported by others (see Hausman and Newey (1995); Schmalensee and Stoker (1999); West (2004); Yatchew and No (2001)).

The quantile regression estimates are reported in columns (1)-(3), revealing plausible and interesting

where we have estimated the corresponding mean and variance in our data. In the (log) price dimension we space the knots linearly.

¹⁴ This allows for different number of knots by quantile. Following equation (1) we use the budget share as dependent variable in the cross-validation. Given that our analysis focuses on the demand behavior for the three income levels of interest, we evaluate the cross-validation function only for observations which are not too far from these income points, and use 0.5 (in the log income dimension) as cutoff. The objective function in our cross-validation reflects the corresponding sum of the check function evaluated at the residual from the leave-one-out quantile regression.

patterns in the elasticities across quantiles. At lower quantiles, the estimated price elasticity is much higher (in absolute values) than at higher quantiles.¹⁵ Similarly, the estimated income elasticity declines strongly as we move from the first quartile to the median, and from the median to the third quartile. Thus, low-intensity users appear to be substantially more sensitive in their demand responses to price and income variation than high-intensity users.

A natural question is whether this benchmark specification is appropriately specified. To investigate this, we perform the specification test for the linear quantile regression model developed in Horowitz and Spokoiny (2002).¹⁶ The results are reported in Table 4. We clearly reject our baseline specification at a 5% level. This holds whether we measure our dependent variable as log quantity or as gasoline budget share.

[TABLES 3–4 ABOUT HERE]

We have also augmented the specification reported in Table 3 with squares and cubes of price and income and found these to be significant. This suggests that the parametric benchmark model may be misspecified. We therefore now proceed to the nonparametric analysis.

Figure 1 shows the nonparametric estimates, where we show the log of demand (measured in gallons per year) implied by our estimates of equation (5). Each panel corresponds to a particular point in the income distribution. The line shown with open markers represents the unconstrained estimates, together with the corresponding bootstrapped confidence intervals (solid lines). As discussed earlier, we also use these intervals for the constrained estimator. In the next paragraph, we use the confidence bands to support our claim that non-monotonicity of the unconstrained estimates is caused by random sampling errors whose effects are reduced by use of the constrained estimator. As can be seen in panel (b) for the middle income level, for example, the unconstrained estimates show overall a downward-sloping trend, but there are several instances where the estimated demand is upward sloping. A similar pattern is also found in Hausman and Newey (1995). Although here we plot the Marshallian demand estimate, these instances of upward sloping demand also point to violations of the Slutsky negativity when we compensate the household for an increase in prices. The line shown

¹⁵ A similar pattern is reported in Frondel et al. (2012) using travel diary data for Germany.

¹⁶ See Zheng (1998) and Escanciano and Goh (2014) for alternative nonparametric tests of a parametric quantile regression model.

as filled markers represents the estimate constrained by the Slutsky shape restriction.¹⁷ By design, the constrained estimates are consistent with economic theory.

The constrained and unconstrained estimates are both well contained in a 90% confidence band based on the unconstrained estimates. This pattern is consistent with our interpretation of the non-monotonicity of the unconstrained estimates as the consequence of random sampling errors whose effects are diminished by imposition of the Slutsky restriction. At the same time, the constrained estimates show that imposing the shape constraint can also be thought of as providing additional smoothing. Focusing on the constrained estimates, we compare the price sensitivity across the three income groups. The middle income group appears to be more price sensitive than either the upper or the lower income group.¹⁸

[FIGURE 1 ABOUT HERE]

In Appendix A.2 we present the results from a simulation exercise, in which we repeatedly estimate constrained and unconstrained estimates in simulated data modelled on our empirical application. As true data generating function we take the constrained estimate reported in our main results. We then compute (joint) confidence intervals across simulation draws for the unconstrained and the constrained estimates, respectively. We find that imposing the constraint substantially narrows the confidence bands. For example, for the middle income group, we find that the width of the constrained intervals is about 63% of the unconstrained intervals (averaging over the price range). For the higher and the lower income group, this ratio is 47% and 55%, respectively. (This is consistent with the finding that the middle income group is more price sensitive, and the constraint therefore having less effect for this income group.) Overall, this simulation evidence suggests that in applications similar to ours the gain in precision from imposing the shape constraint can be substantial. The results also support our argument that the unconstrained confidence intervals are conservative relative to those taking the constraint into account.

¹⁷ In Appendix Figure A.1, we show that the resulting demand figures are not sensitive to a range of alternative values of the smoothing parameter discussed in Section 3.2.

¹⁸This is a pattern also noted in Blundell et al. (2012).

4.3 Comparison Across Quantiles and the Conditional Mean Estimates

Figure 2 compares the quantile estimates across the three quartiles, holding income constant at the middle income group. In the unconstrained estimates, the differences in flexibility (corresponding to the cross-validated number of knots in the price dimension) are clearly visible. The constrained estimates, however, are quite similar in shape, suggesting that they may approximately be parallel shifts of each other. This would be consistent with a location-scale model together with conditional homoskedasticity (Koenker, 2005). Under this model, conditional mean estimates would show the same shape as seen in the conditional quartile results, and we turn to this comparison in the following.

[FIGURE 2 ABOUT HERE]

As noted in the introduction, we have previously investigated gasoline demand, focussing on the conditional mean (Blundell et al. (2012)). That analysis used a kernel regression method, in which the shape restriction is imposed by reweighting the data in an approach building on Hall and Huang (2001). As in the quantile demand results here we found strong evidence of differential price responsiveness across the income distribution, suggesting a stronger price responsiveness in the middle income group. Figure 3 shows the conditional mean regression estimates, where we use the same B-spline basis functions as in the quantile results presented above (see Figure 1). The shape of these two sets of estimates is remarkably similar, especially for the constrained estimates; in terms of levels, the mean estimates are somewhat higher than the median estimates (by around 0.1 on the log scale).

[FIGURE 3 ABOUT HERE]

4.4 The Measurement of Individual Welfare Distribution

The Slutsky constrained demand function estimates can in turn be used for welfare analysis of changes in prices. For this purpose we consider a change in price from the 5th to the 95th percentile in our sample for the nonparametric analysis, and we report Deadweight Loss measures corresponding to this price change. Table 5 shows the DWL estimates for the three quartiles of unobserved

heterogeneity and three income groups. In the constrained estimates, we find that the middle income group has the highest DWL at all quartiles. This is consistent with the graphical evidence presented in Figure 1 above. The table also shows the DWL estimates implied by the parametric estimates corresponding to a linear specification. The uniform patterns in the corresponding DWL figures (within each quantile) reflect the strong assumptions underlying these functional forms, which have direct consequences for the way DWL measures vary across these subgroups in the population.

There are two instances (both for the lower-income group) where the unconstrained DWL shows the wrong sign. This underscores that DWL analysis is only meaningful if the underlying estimates satisfy the required properties of consumer demand behavior.

One feature of the estimates in Table 5 is the variation in DWL seen across different quantiles. More generally, we can ask how DWL is distributed over the entire population of types. Such an analysis is presented in Figure A.2. In this figure we show for each income group the density of DWL across the range of quantiles (from $\alpha = 0.05$ to $\alpha = 0.95$), comparing unconstrained and constrained estimates.

[TABLE 5 AND FIGURE A.2 ABOUT HERE]

5 Price Endogeneity

So far we have maintained the assumption of exogeneity on prices. There are many reasons why prices vary at the local market level. These include cost differences on the supply side, short-run supply shocks, local competition, as well as taxes and government regulation (EIA, 2010). However, one may be concerned that prices may also reflect preferences of the consumers in the locality, so that prices faced by consumers may potentially be correlated with unobserved determinants of gasoline demand.

To address this concern, we follow Blundell et al. (2012) and use a cost-shifter approach to identify the demand function. An important determinant of prices is the cost of transporting the fuel from the supply source. The U.S. Gulf Coast Region accounts for the majority of total U.S. refinery net production of finished motor gasoline, and for almost two thirds of U.S. crude oil imports. It is also the starting point for most major gasoline pipelines. We therefore expect that transportation cost

increases with distance to the Gulf of Mexico, and implement this with the distance between one of the major oil platforms in the Gulf of Mexico and the state capital (see Blundell et al. (2012) for further details and references). Figure 5 shows the systematic and positive relationship between log price and distance (in 1,000 km) at state level.

[FIGURE 5 ABOUT HERE]

In the following, we first present evidence from a nonparametric exogeneity test. We then estimate a nonparametric quantile IV specification, incorporating the shape restriction.

5.1 Exogeneity Test

We use the nonparametric exogeneity test for the quantile setting discussed earlier. To simplify the computation we focus on the univariate version of the test here. For this purpose, we split the overall sample according to household income, and then run the test for each household income group separately.¹⁹ We select income groups to broadly correspond to our three reference income levels in the quantile estimation; we select a low income group of households (household income between \$35,000 and \$50,000), a middle income group of households (household income between \$50,000 and \$65,000), and an upper income group of households (household income between \$65,000 and \$80,000). Given that we perform the test three times (for these three income groups) we can adjust the size for a joint 0.05-level test. Given the independence of the three income samples, the adjusted p -value for a joint 0.05-level test of exogeneity, at each of the three income groups, is $1 - (0.95)^{(1/3)} = 0.01695$.

Table 6 shows the test results, where column (1) presents our baseline estimates, and columns (2) and (3) show a sensitivity with respect to the bandwidth parameter choice required for the kernel density estimation. For the median case, the p -values are above 0.1 throughout and thus there is no evidence of a violation of exogeneity at the median. The evidence for the first quartile is similar. The only instance of a borderline p -value is for the lower income group for the upper quartile, with

¹⁹ The test makes use of the vector of residuals from the quantile model under the null hypothesis. Even though we implement the test separately for three income groups, we use the residuals from the bivariate model using all observations, so that these residuals correspond to the main (unconstrained) specification of interest (see e.g. Figure 1).

a baseline p -value of 0.041, which is still above the adjusted cutoff value for a test 0.05-level test. Overall, we interpret this evidence as suggesting that we do not find strong evidence of endogeneity in this application. This finding is also consistent with our earlier analysis focusing on the conditional mean (see Blundell et al. (2012)). In order to allow a comparison, we nonetheless present quantile IV estimates in the following.

[TABLE 6 ABOUT HERE]

5.2 Quantile Instrumental Variable Estimates

Figure 6 presents our quantile IV estimates of demand under the shape restriction. These estimates are shown as filled markers, and compared with our earlier shape-constrained estimates assuming exogeneity of prices (see Figure 1), shown as open markers.²⁰ Overall, the shape of the IV estimates is quite similar to those obtained under the assumption of exogeneity. This is consistent with the evidence from the exogeneity test presented above. As before the comparison across income groups suggests that the middle income group is more elastic than the two other income groups, in particular over the lower part of the price range.

[FIGURE 6 ABOUT HERE]

6 Conclusions

The paper has made a number of contributions. We have presented a quantile estimator which incorporates shape restrictions. We have developed a new estimator for the case of quantile estimation under endogeneity. We have applied these methods in the context of individual gasoline demand with nonseparable unobserved heterogeneity. The nonparametric estimate of the demand function

²⁰ To simplify the computation of the IV estimates we set the number of interior knots for the cubic splines to 2 in both the income and the price dimension here, and impose the Slutsky constraint at five points in the income dimension (\$37,500, \$42,500, \$57,500, \$72,500, and \$77,500). We use the NAG routine E04US together with a multi-start procedure to solve the global minimization problem. The resulting demand function estimates do not appear sensitive to specific starting values. In the implementation of the objective function (see equation (7)), we smooth the indicator function corresponding to the term $1[W_i - G_\alpha(P_i, Y_i) \leq 0]$ in the neighborhood of 0 using a Gaussian kernel.

was found to be noisy due to random sampling errors. The estimated function is non-monotonic, and there are instances where the estimate, taken at face value, is inconsistent with economic theory. When we imposed the Slutsky restriction of consumer theory on the demand function, our approach yielded well-behaved estimates of the demand function and welfare costs across the income and taste distribution. Comparing across income groups and quantiles, our work allowed us to document differences in demand behavior across both observables and unobservables.

The starting point for our analysis was the following two observations: First, when there is heterogeneity in terms of usage intensity, the patterns of demand may potentially be quite different at different points in the distribution of the unobservable heterogeneity. Under suitable exogeneity assumptions and a monotonicity restriction, quantile methods allow us to recover the demand function at different points in the distribution of unobservables. This allows us to estimate demand functions for specific types of individuals, rather than averaging across different types of consumers.

Second, we want to be able to allow a flexible effect of price and income on household demand, and in particular allow price responses to differ by income level. Nonparametric estimates eliminate the risk of specification error but can be poorly behaved due to random sampling errors. Fully nonparametric demand estimates can be non-monotonic and may violate consumer theory. In contrast, a researcher choosing a tightly specified model is able to precisely estimate the parameter vector; however simple parametric models of demand functions can be misspecified and, consequently, yield misleading estimates of price sensitivity and DWL. We argue that in the context of demand estimation, this apparent trade-off can be overcome by constraining nonparametric estimates to satisfy the Slutsky condition of economic theory. We have illustrated this approach by estimating a gasoline demand function. The constrained estimates are well-behaved and reveal features not found with typical parametric model specifications. We present estimates across income groups and at different points in the distribution of the unobservables.

These estimates are obtained initially under the assumption of exogenous prices, and the reader may therefore be concerned about potential endogeneity of prices. We investigate this in two ways. First, we implement an exogeneity test to provide direct evidence on this. As instrument, we use a cost shifter variable measuring transportation cost. The results suggest that endogeneity is unlikely to be of first order relevance. Nonetheless, we investigate the shape of the demand function

without imposing exogeneity of prices. For this purpose, we develop a novel estimation approach to nonparametric quantile estimation with endogeneity. We estimate IV quantile models under shape restrictions. The results are broadly similar to the estimates under exogeneity.

The analysis showcases the value of imposing shape restrictions in nonparametric quantile regressions. These restrictions provide a way of imposing structure and thus informing the estimates without the need for arbitrary functional form assumptions which have no basis in economic theory.

References

- Bento, Antonio M., Lawrence H. Goulder, Mark R. Jacobsen, and Roger H. von Haefen, “Distributional and Efficiency Impacts of Increased US Gasoline Taxes,” *American Economic Review* 99 (June 2009), pp. 667–699.
- Blundell, Richard and Joel L. Horowitz, “A Non-Parametric Test of Exogeneity,” *Review of Economic Studies* 74 (October 2007), 1035–1058.
- Blundell, Richard, Joel L. Horowitz, and Matthias Parey, “Measuring the price responsiveness of gasoline demand: Economic shape restrictions and nonparametric demand estimation,” *Quantitative Economics* 3 (March 2012), 29–51.
- Blundell, Richard, Dennis Kristensen, and Rosa L. Matzkin, “Bounding Quantile Demand Functions using Revealed Preference Inequalities,” *Journal of Econometrics* 179 (April 2014), 112–127.
- Brown, Bryan W. and Mary Beth Walker, “The Random Utility Hypothesis and Inference in Demand Systems,” *Econometrica* 57 (July 1989), 815–829.
- Chen, Colin, “A Finite Smoothing Algorithm for Quantile Regression,” *Journal of Computational and Graphical Statistics* 16 (March 2007), 136–164.
- Chernozhukov, Victor, Christian Hansen, and Michael Jansson, “Finite Sample Inference in Econometric Models via Quantile Restrictions,” *Journal of Econometrics* 152 (October 2009), 93–103.
- Chetty, Raj, “Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods,” *Annual Review of Economics* 1 (2009), 451–488.
- Deaton, Angus and John Muellbauer, *Economics and consumer behavior*, Cambridge University Press (1980). Reprinted 1999.
- Dette, Holger, Stefan Hoderlein, and Natalie Neumeyer, “Testing multivariate economic restrictions using quantiles: the example of Slutsky negative semidefiniteness,” *Journal of Econometrics* 191 (2016), 129–144.
- EIA, “Regional gasoline price differences,” (2010). Energy Information Administration, U.S. Department of Energy. Available at http://www.eia.gov/energyexplained/index.cfm?page=gasoline_regional.
- Escanciano, Juan Carlos and Sze-Chuan Goh, “Specification Analysis of Linear Quantile Models,” *Journal of Econometrics* 178 (January 2014), 495–507.

- Frondel, Manuel, Nolan Ritter, and Colin Vance, “Heterogeneity in the rebound effect: Further evidence for Germany,” *Energy Economics* 34 (March 2012), 461–467.
- Fu, Jia-Young Michael, Joel Horowitz, and Matthias Parey, “Testing Exogeneity in Nonparametric Instrumental Variables Identified by Conditional Quantile Restrictions,” *cemmap working paper CWP68/15* (2015).
- Graham, Bryan S, Jinyong Hahn, Alexandre Poirier, and James L Powell, “Quantile Regression with Panel Data,” *NBER Working Paper 21034* (2015).
- Haag, Berthold, Stefan Hoderlein, and Krishna Pendakur, “Testing and Imposing Slutsky Symmetry in Nonparametric Demand Systems,” *Journal of Econometrics* 153 (November 2009), 33–50.
- Hall, Peter and Li-Shan Huang, “Nonparametric Kernel Regression Subject to Monotonicity Constraints,” *The Annals of Statistics* 29 (June 2001), 624–647.
- Hausman, Jerry A. and Whitney K. Newey, “Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss,” *Econometrica* 63 (November 1995), 1445–1476.
- , “Individual Heterogeneity and Average Welfare,” *cemmap working paper CWP34/13* (July 2013).
- , “Nonparametric Welfare Analysis,” (2015). Forthcoming, Annual Reviews.
- Hoderlein, Stefan, “How Many Consumers are Rational?” *Journal of Econometrics* 164 (October 2011), 294–309.
- Hoderlein, Stefan and Jörg Stoye, “Revealed Preferences in a Heterogeneous Population,” *Review of Economics and Statistics* 96 (May 2014), 197–213.
- Hoderlein, Stefan and Anne Vanhems, “Welfare Analysis using Nonseparable Models,” *cemmap working paper CWP01/11* (January 2011).
- Horowitz, Joel L. and Vladimir G. Spokoiny, “An Adaptive, Rate-Optimal Test of Linearity for Median Regression Models,” *Journal of the American Statistical Association* 97 (September 2002), 822–835.
- Hurwicz, Leonid and Hirofumi Uzawa, “On the Problem of Integrability of Demand Functions,” in Chipman, J.S., L. Hurwicz, M. K. Richter, and H. F. Sonnenschein (eds.), “Preferences, Utility, and Demand. A Minnesota Symposium,” Harcourt Brace Jovanovich, New York (1971), pp. 114–148.
- Imbens, Guido W. and Whitney K. Newey, “Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity,” *Econometrica* 77 (September 2009), 1481–1512.
- Jones, M. C., “Simple boundary correction for density estimation,” *Statistics and Computing* 3 (September 1993), 135–146.
- Koenker, Roger, *Quantile Regression*, Econometric Society Monographs, Cambridge University Press (2005).
- Lewbel, Arthur, “Consistent Nonparametric Hypothesis Tests with an Application to Slutsky Symmetry,” *Journal of Econometrics* 67 (June 1995), 379–401.
- , “Demand Systems with and without Errors,” *The American Economic Review* 91 (June 2001), 611–618.
- Manning, Willard G., Linda Blumberg, and Lawrence H. Moulton, “The demand for alcohol: The differential response to price,” *Journal of Health Economics* 14 (June 1995), 123–148.

- Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green, *Microeconomic Theory*, Oxford University Press (1995).
- Matzkin, Rosa L., “Nonparametric Estimation of Nonadditive Random Functions,” *Econometrica* 71 (September 2003), 1339–1375.
- , “Heterogeneous Choice,” in Blundell, Richard, Whitney Newey, and Torsten Persson (eds.), “Advances in Economics and Econometrics, Theory and Applications,” Ninth World Congress of the Econometric Society, chap. 4, Cambridge University Press (2007), pp. 75–110.
- , “Identification in Nonparametric Simultaneous Equations Models,” *Econometrica* 76 (September 2008), 945–978.
- ORNL, “2001 National Household Travel Survey. User Guide,” (2004). Oak Ridge National Laboratory. Available at <http://nhts.ornl.gov/2001/>.
- Powell, M. J. D., *Approximation Theory and Methods*, Cambridge University Press (1981).
- Schmalensee, Richard and Thomas M. Stoker, “Household Gasoline Demand in the United States,” *Econometrica* 67 (May 1999), 645–662.
- West, Sarah E., “Distributional effects of alternative vehicle pollution control policies,” *Journal of Public Economics* 88 (March 2004), 735–757.
- Wolak, Frank A., “The Local Nature of Hypothesis Tests Involving Inequality Constraints in Non-linear Models,” *Econometrica* 59 (July 1991), 981–995.
- Yatchew, Adonis and Joungyeo Angela No, “Household Gasoline Demand in Canada,” *Econometrica* 69 (November 2001), 1697–1709.
- Zheng, John Xu, “A Consistent Nonparametric Test of Parametric Regression Models under Conditional Quantile Restrictions,” *Econometric Theory* 14 (February 1998), 123–138.

Tables and Figures

Table 1: Sample descriptives

	<i>Mean</i>	<i>St. dev.</i>
Log gasoline demand	7.127	0.646
Log price	0.286	0.057
Log income	11.054	0.580
Observations	3640	

Note: See text for details.

Table 2: Cross-validation results

quantile (α)	number interior knots	
	price	income
<i>(1) Base case</i>		
0.25	1	3
0.50	4	3
0.75	3	3
<i>(2) Leaving out largest 10 and lowest 10 share observations</i>		
0.25	1	3
0.50	4	4
0.75	1	3

Note: Table shows cross-validation results by quantile.

Table 3: Log-log model estimates

	$\alpha = 0.25$	$\alpha = 0.50$	$\alpha = 0.75$	OLS
	(1)	(2)	(3)	(4)
$\log(p)$	-1.00 [0.23]	-0.72 [0.19]	-0.60 [0.22]	-0.83 [0.18]
$\log(y)$	0.41 [0.02]	0.33 [0.02]	0.23 [0.02]	0.34 [0.02]
Constant	2.58 [0.27]	3.74 [0.21]	5.15 [0.26]	3.62 [0.20]
N	3640	3640	3640	3640

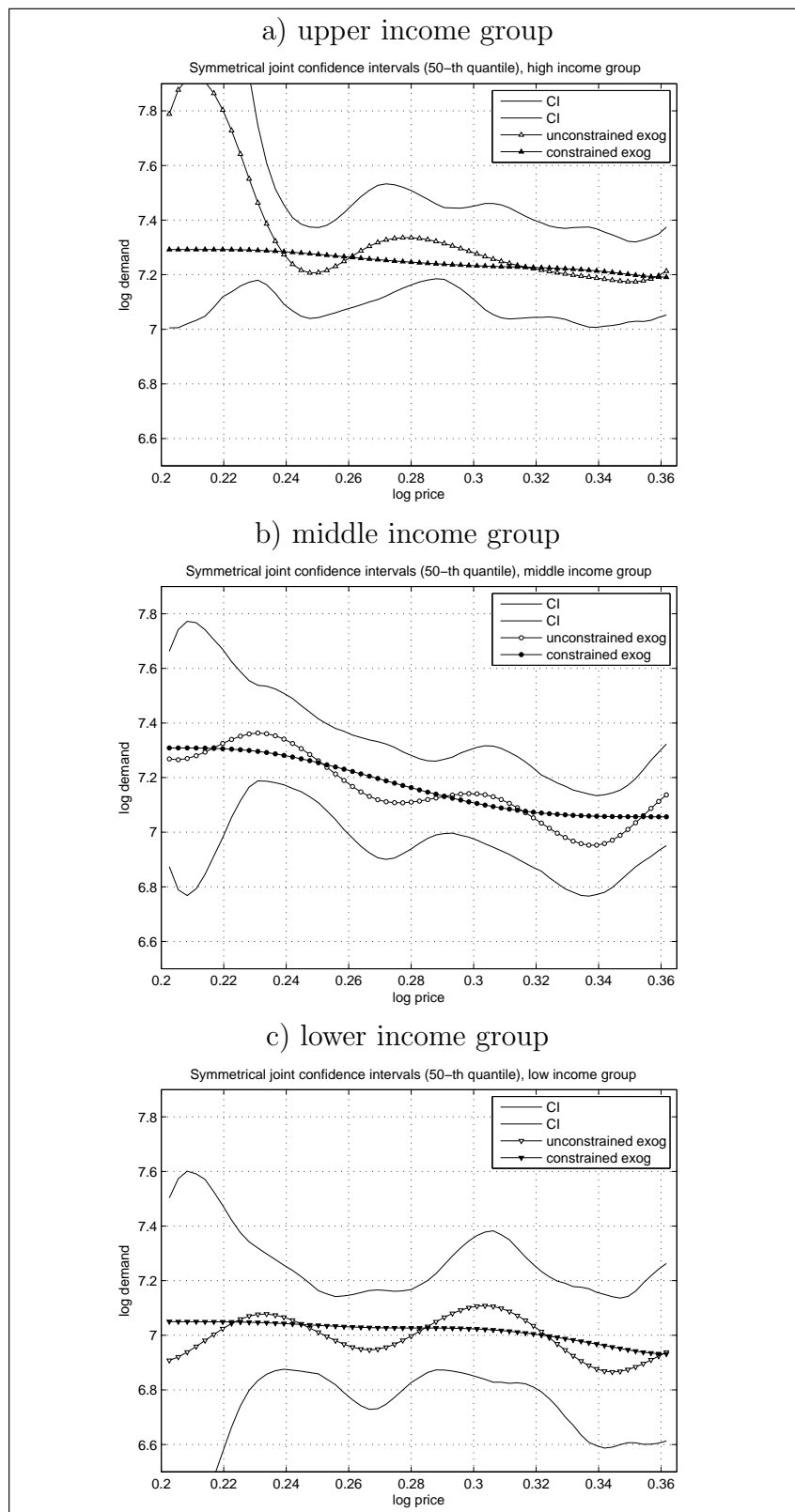
Note: Dependent variable is log gasoline demand. See text for details.

Table 4: Specification test

Dependent var.	test statistic	critical value		p -value	reject?
		0.05 level	0.01 level		
gasoline share	2.52	1.88	2.69	0.0120	yes
log quantity	2.71	1.82	2.43	0.0020	yes

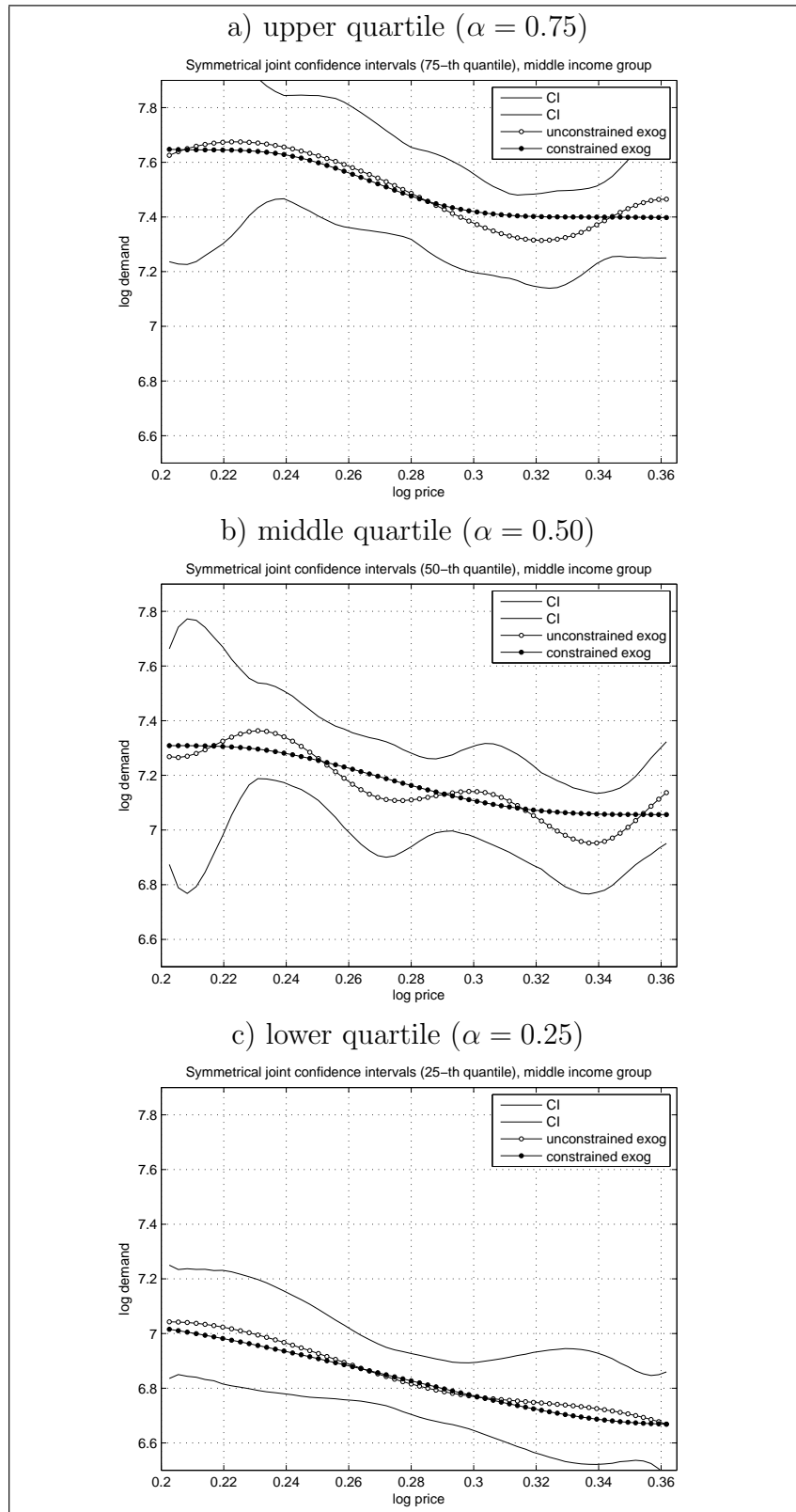
Note: Test implements Horowitz and Spokoiny (2002) for the median case. The first row reports the test results for gasoline demand measured as budget share, the second row for log quantity. Under the null hypothesis, the model is linear in log price and log income. See text for details.

Figure 1: Quantile regression estimates: constrained versus unconstrained estimates



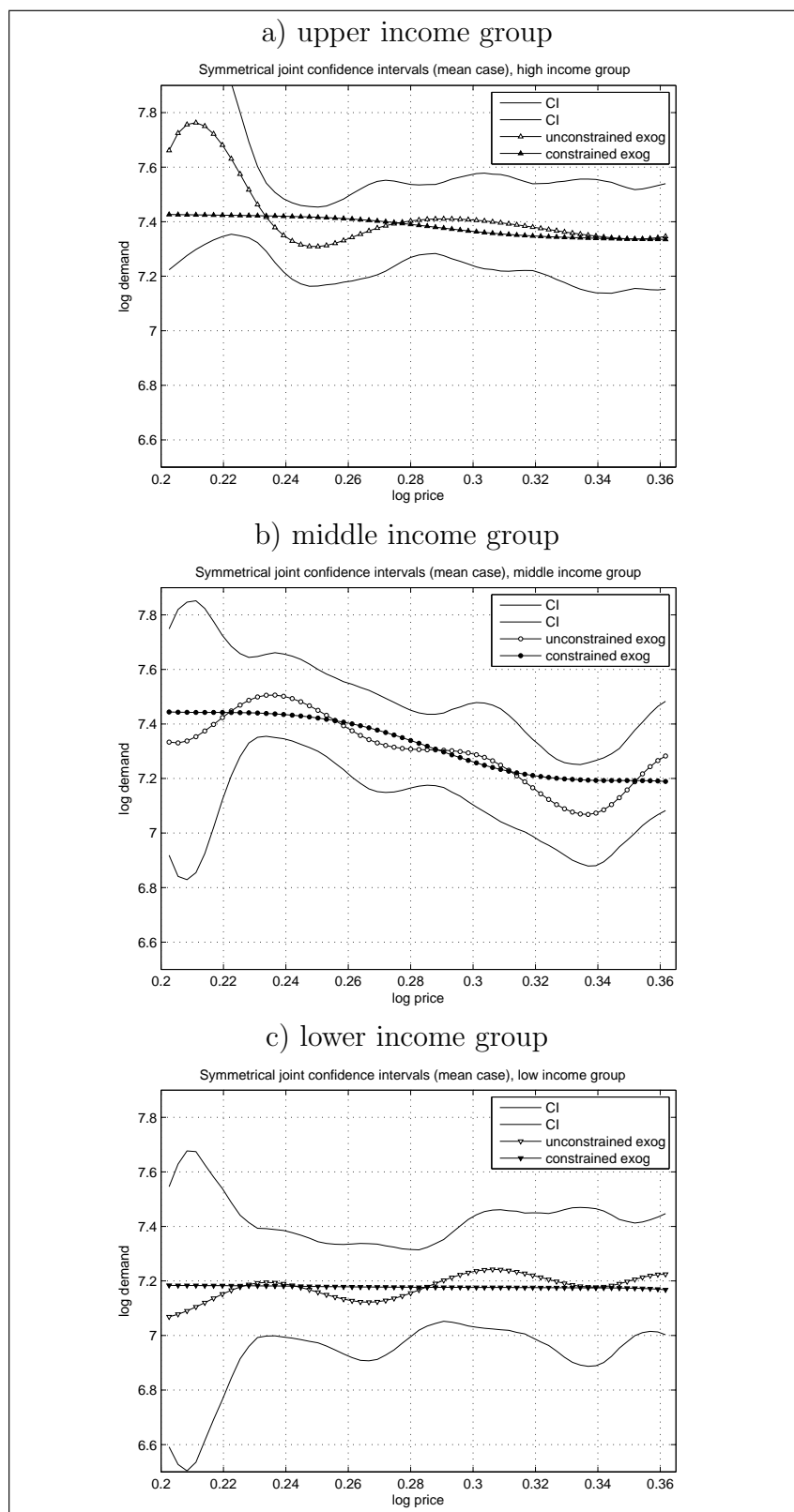
Note: Figure shows unconstrained nonparametric quantile demand estimates (open markers) and constrained nonparametric demand estimates (filled markers) at different points in the income distribution for the median ($\alpha = 0.5$), together with simultaneous confidence intervals. Income groups correspond to \$72,500, \$57,500, and \$42,500. Confidence intervals shown refer to bootstrapped symmetrical, simultaneous confidence intervals with a confidence level of 90%, based on 4,999 replications. See text for details.

Figure 2: Quantile regression estimates: constrained versus unconstrained estimates (middle income group)



Note: Figure shows unconstrained nonparametric quantile demand estimates (filled markers) and constrained nonparametric demand estimates (filled markers) at the quartiles for the middle income group (\$57,500), together with simultaneous confidence intervals. Confidence intervals shown refer to bootstrapped symmetrical, simultaneous confidence intervals with a confidence level of 90%, based on 4,999 replications. See text for details.

Figure 3: Mean regression estimates: constrained versus unconstrained estimates



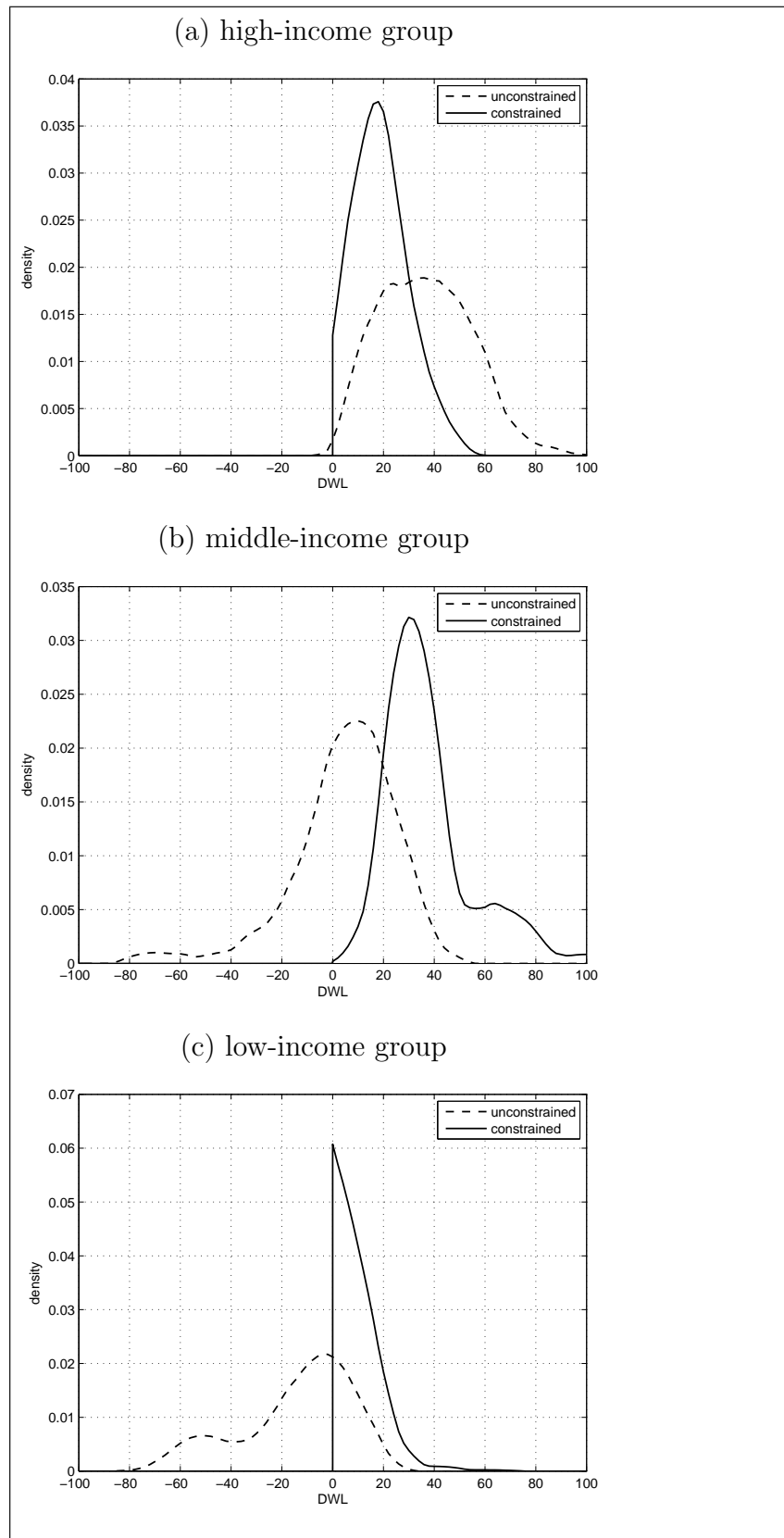
Note: Figure shows unconstrained nonparametric mean regression demand estimates (filled markers) and constrained nonparametric demand estimates (filled markers) at different points in the income distribution, together with simultaneous confidence intervals. Income groups correspond to \$72,500, \$57,500, and \$42,500. Confidence intervals shown refer to bootstrapped symmetrical, simultaneous confidence intervals with a confidence level of 90%, based on 4,999 replications. See text for details.

Table 5: DWL estimates

income	unconstrained			constrained			linear quantile estimates		
	DWL	DWL/tax	DWL/inc*	DWL	DWL/tax	DWL/inc*	DWL	DWL/tax	DWL/inc*
lower quartile ($\alpha = 0.25$)									
72500	11.76	5.72%	1.62	12.74	6.21%	1.76	13.89	7.12%	1.92
57500	33.24	20.01%	5.78	29.18	17.54%	5.08	12.88	7.24%	2.24
42500	-15.40	-8.91%	-3.62	0.85	0.54%	0.20	11.30	7.35%	2.66
median ($\alpha = 0.50$)									
72500	49.64	17.30%	6.85	16.32	5.81%	2.25	20.33	7.26%	2.80
57500	5.86	2.20%	1.02	30.20	12.30%	5.25	19.06	7.36%	3.32
42500	12.81	5.87%	3.01	18.57	8.56%	4.37	16.90	7.45%	3.98
upper quartile ($\alpha = 0.75$)									
72500	23.07	5.71%	3.18	20.64	5.07%	2.85	19.29	4.76%	2.66
57500	15.98	4.35%	2.78	39.40	11.42%	6.85	19.77	5.22%	3.44
42500	-43.60	-11.25%	-10.26	1.17	0.35%	0.28	18.86	5.63%	4.44

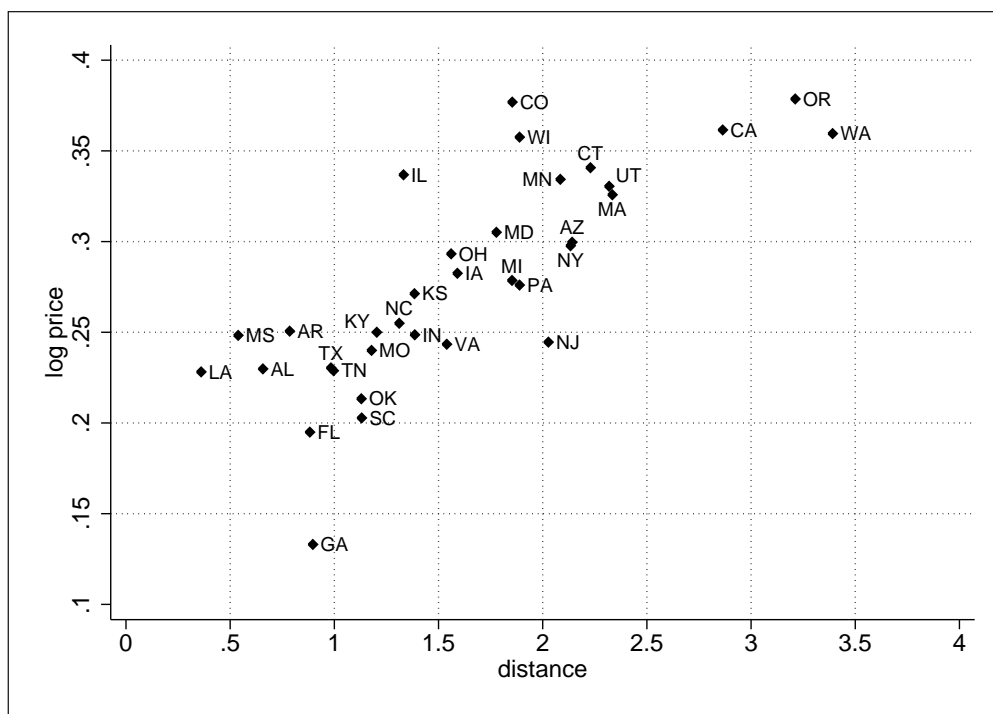
Note: Table shows DWL estimates, corresponding to a change in prices from the 5th to the 95th percentile, that is from \$1.225 to \$1.436. For comparability all three sets of estimates are based on the sample for the nonparametric analysis, and use budget share as dependent variable. * DWL per income figures are rescaled by factor 10^4 for better readability.

Figure 4: Distribution of DWL, constrained versus unconstrained



Note: Graphs show density estimates for the distribution of DWL estimates. Based on estimates for the 5th to the 95th percentile ($\alpha = 0.05$ to 0.95 in steps of 0.005). Density estimates computed using an Epanechnikov kernel. Since DWL is nonnegative in the constrained case, density is renormalized in the boundary area (Jones (1993)). Estimates computed using the same knot choice throughout as crossvalidated for the median.

Figure 5: The Instrument Variable for Price: Distance to the Gulf of Mexico



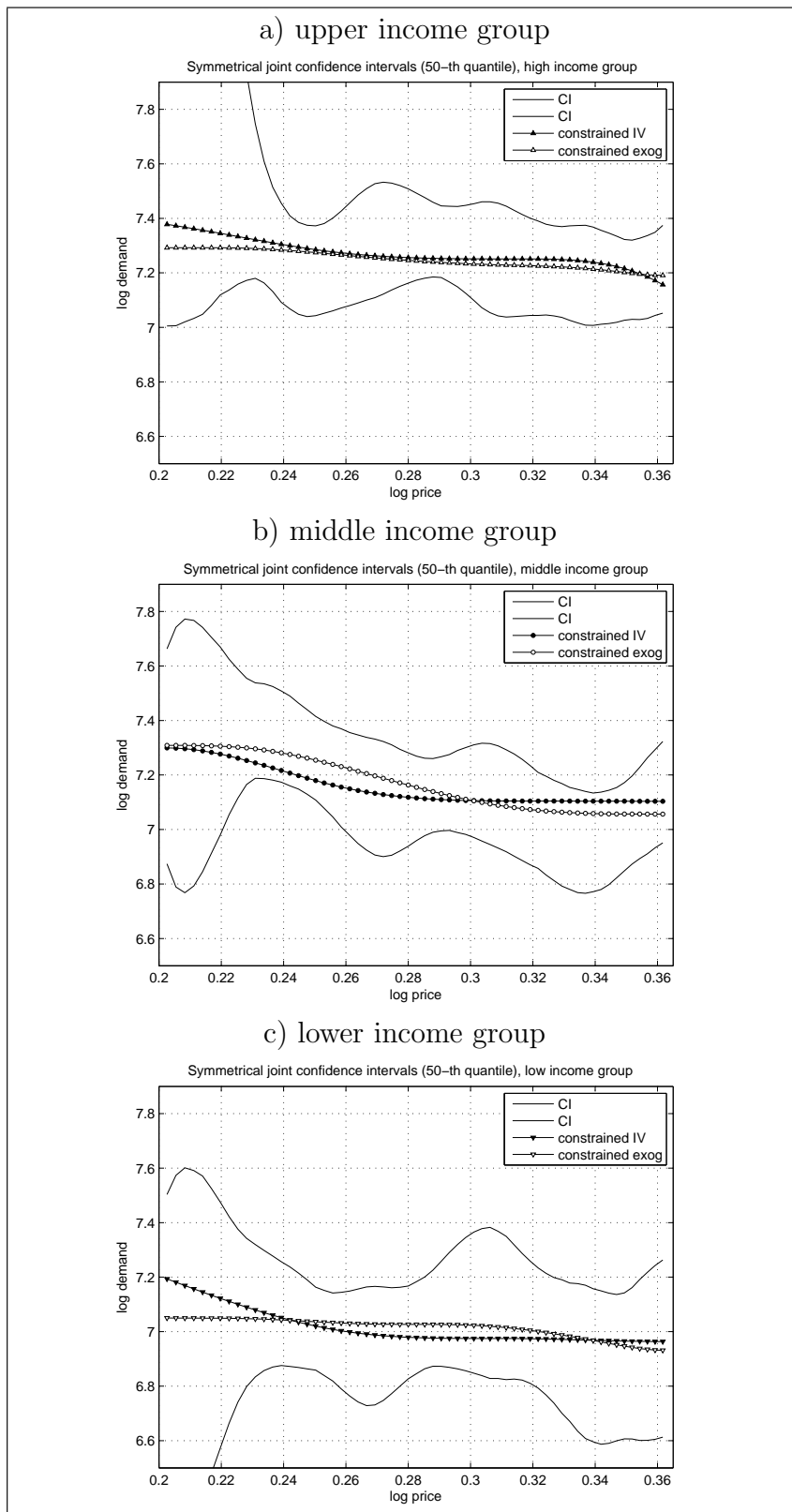
Source: BHP (2012, Figure 5).

Table 6: Exogeneity test (p -values)

	income	base	bandwidth sensitivity	
	range	case	factor 0.8	factor 1.25
		(1)	(2)	(3)
first quartile ($\alpha = 0.25$)	low	0.343	0.284	0.452
	middle	0.209	0.197	0.192
	high	0.313	0.256	0.372
median ($\alpha = 0.50$)	low	0.261	0.179	0.341
	middle	0.137	0.170	0.118
	high	0.754	0.709	0.814
third quartile ($\alpha = 0.75$)	low	0.041	0.055	0.029
	middle	0.624	0.748	0.503
	high	0.402	0.467	0.377

Note: Table shows p -values for the exogeneity test from Fu et al. (2015). Endogenous variable is price, instrumented with distance. We run separate tests for three income groups; for this test, these groups are defined as follows: ‘low’: income between \$35,000 and \$50,000, ‘middle’: \$50,000 – \$65,000, ‘high’: \$65,000 – \$80,000. The specification we test is the unconstrained nonparametric quantile estimate as shown e.g. in Figure 1 for the median. In implementing this test, required bandwidth choices for the kernel density estimates use Silverman’s rule of thumb. Columns (2) and (3) vary all bandwidth inputs by the indicated factor.

Figure 6: Quantile regression estimates under the shape restriction: IV estimates versus estimates assuming exogeneity



Note: Figure shows constrained nonparametric IV quantile demand estimates (filled markers) and constrained quantile demand estimates under exogeneity (open markers) at different points in the income distribution for the median ($\alpha = 0.5$), together with simultaneous confidence intervals. Income groups correspond to \$72,500, \$57,500, and \$42,500. Confidence intervals shown correspond to the unconstrained quantile estimates under exogeneity as in Figure 1. See text for details.

A Online Appendix

A.1 Exogeneity test

Using the notation of Section 3.5, the null hypothesis is

$$H_0 : K_\alpha(p, y) = G_\alpha(p, y)$$

for almost every $(p, y) \in \text{supp}(P, Y)$. The alternative hypothesis is

$$H_1 : P [K_\alpha(P, Y) \neq G_\alpha(P, Y)] > 0.$$

H_0 is equivalent to

$$P [W - K_\alpha(P, Y) \leq 0 | Y = y, Z = z] = \alpha. \quad (9)$$

The test of H_0 is based on a sample analog of $P [W - K_\alpha(P, Y) \leq 0 | Y = y, Z = z] - \alpha$. H_0 is rejected if the sample analog differs from 0 by more than can be explained by random sampling errors.

To obtain a test statistic, let F_{WPYZ} denote the cumulative distribution function of (W, P, Y, Z) and f_{YZ} denote the probability density function of (Y, Z) . Assume without loss of generality that

$$\text{supp}(P, Y, Z) \subset [0, 1]^3.$$

This condition can always be satisfied by, if necessary, carrying out monotone increasing transformations of P , Y , and Z . Therefore (9) is equivalent to

$$\int_0^1 F_{WPYZ} [K_\alpha(p, y), p, y, z] dp = \alpha f_{YZ}(y, z).$$

Define

$$\tilde{S} = \int_{[0,1]^2} \left\{ \int_0^1 F_{WPYZ} [K_\alpha(p, y), p, y, z] dp - \alpha f_{YZ}(y, z) \right\}^2 dydz \quad (10)$$

H_0 is equivalent to $H_0 : \tilde{S} = 0$ and can be tested by determining whether an empirical analog of

\tilde{S} exceeds 0 by more than can be explained by random sampling errors. \tilde{S} does not depend on G_α , so a test based on \tilde{S} does not require estimation of G_α and does not have the low power caused by the ill-posed inverse problem. However, a test based on \tilde{S} requires nonparametric estimation of K_α and f_{YZ} . The rates of convergence of the nonparametric estimators and the resulting rate of testing are slower than $n^{-1/2}$, where n is the sample size. That is, the smallest difference between K_α and G_α that the test can detect is larger than $O(n^{-1/2})$.

This problem can be overcome by smoothing the quantity in braces $\{\}$ on the right-hand side of (10). To this end, let $l(y, z, \zeta, \eta)$ denote the kernel of a one-to-one integral operator, L , on $L_2[0, 1]^2$. That is, L is defined by

$$(L\psi)(\zeta, \eta) = \int_{[0,1]^2} l(y, z, \zeta, \eta)\psi(y, z)dydz$$

and is one-to-one, where ψ is any function in $L_2[0, 1]^2$. Define

$$S = \int_{[0,1]^2} \left[\int_{[0,1]^2} \left\{ \int_0^1 F_{WPYZ} [K_\alpha(p, y), p, y, z] dp - \alpha f_{YZ}(y, z) \right\} l(y, z, \zeta, \eta) dydz \right] d\zeta d\eta.$$

H_0 is equivalent to

$$H_0 : S = 0. \tag{11}$$

H_1 is equivalent to the statement that (11) does not hold. The test statistic used in this paper is based on a sample analog of S .

To define the statistic, let $\{W_i, P_i, Y_i, Z_i : i = 1, \dots, n\}$ denote an independent random sample from the distribution of (W, P, Y, Z) . Let $\hat{K}_\alpha^{(-i)}(p, y)$ denote the leave-observation- i -out nonparametric quantile regression estimator

$$\hat{K}_\alpha^{(-i)}(p, y) = \arg \inf_a \sum_{\substack{j=1 \\ j \neq i}}^n \rho_\alpha(W_j - a) M\left(\frac{p - P_j}{h}\right) M\left(\frac{y - Y_j}{h}\right),$$

where $\rho_\alpha(\nu) = \nu[\alpha - I(\nu \leq 0)]$ is the check function, M is a kernel function, and h is a bandwidth. The asymptotic properties of the test statistic do not change if a nonparametric series estimator of $K_\alpha(p, y)$ is used instead of the kernel estimator. The sample analog of S is obtained by replacing K_α with $\hat{K}_\alpha^{(-i)}$, F_{WPYZ} with the empirical distribution function of $Y_i - \hat{K}_\alpha^{(-i)}(P_i, Y_i)$ ($i = 1, \dots, n$), and $\int f_{YZ} q$ for any function $q(y, z)$ with the sample average of $q(Y_i, Z_i)$. The resulting test statistic is

$$\tau_n = \int_{[0,1]^2} \left\{ n^{-1/2} \sum_{i=1}^n \left\{ I \left[W_i \leq \hat{K}_\alpha^{(-i)}(P_i, Y_i) \right] - \alpha \right\} l(Y_i, Z_i, \zeta, \eta) \right\}^2 d\zeta d\eta. \quad (12)$$

The asymptotic properties of τ_n are stated below without regularity conditions or proofs. Regularity conditions and proofs are given by Fu et al. (2015).

1. Under H_0 ,

$$\tau_n \xrightarrow{d} \sum_{j=1}^{\infty} \omega_j \chi_{1j}^2,$$

where the ω_j 's are non-negative weights and the χ_{1j}^2 's are independent random variables that are distributed as chi-square with one degree of freedom.

2. Under H_1 and as $n \rightarrow \infty$, $P(\tau_n > z) \rightarrow 1$ for any finite z .

3. Under the sequence of local alternative hypotheses $k(p, y) = K_\alpha(p, y) + n^{-1/2}\Delta(p, y)$,

$$\tau_n \xrightarrow{d} \sum_{j=1}^{\infty} \omega_j \chi_{1j}^2(\mu_j^2/\omega_j),$$

where the ω_j 's are the weights in item 1 above, the μ_j 's are constants, and the $\chi_{1j}^2(\mu_j^2/\omega_j)$'s are independent random variables that are distributed as non-central chi-square with one degree of freedom and non-central parameter μ_j^2/ω_j .

The statistic τ_n is not asymptotically pivotal, so its asymptotic distribution cannot be tabulated. Fu et al. (2015) gives a simulation method for computing critical values of τ_n that is the quantile analog of the method of Blundell and Horowitz (2007) for computing critical values of a test of exogeneity in a nonparametric mean regression model.

As is discussed in Section 5.1, the test carried out in this paper conditions on income group. For each income group g , we test for endogeneity of P in the model

$$W = G_{\alpha g}(P) + V_\alpha; \quad P(V_\alpha \leq 0|Z) = \alpha.$$

The null hypothesis is

$$H_0 : K_{\alpha g}(p) = G_{\alpha g}(p)$$

for almost every $p \in \text{supp}(P)$, where $K_{\alpha g}$ satisfies

$$P[W - K_{\alpha g}(P)|P] = \alpha.$$

The alternative hypothesis is

$$H_1 : P [K_{\alpha g}(P) \neq G_{\alpha g}(P)] > 0.$$

The test statistic is the univariate version of τ_n in (12). This is

$$\tau_{n1} = \int_0^1 \left\{ n^{-1/2} \sum_{i=1}^n \left\{ I [W_i \leq \hat{K}_{\alpha g}^{(-i)}(P_i)] - \alpha \right\} l(Z_i, \zeta) \right\}^2 d\zeta,$$

where $\hat{K}_{\alpha g}^{(-i)}$ is the leave-observation- i -out nonparametric quantile regression estimator of $K_{\alpha g}$ for income group g , and l is the kernel of the one-to-one integral operator

$$(L\psi)(\zeta) = \int_0^1 l(z, \zeta)\psi(z)dz.$$

In the tests reported in Section 5.1, we set

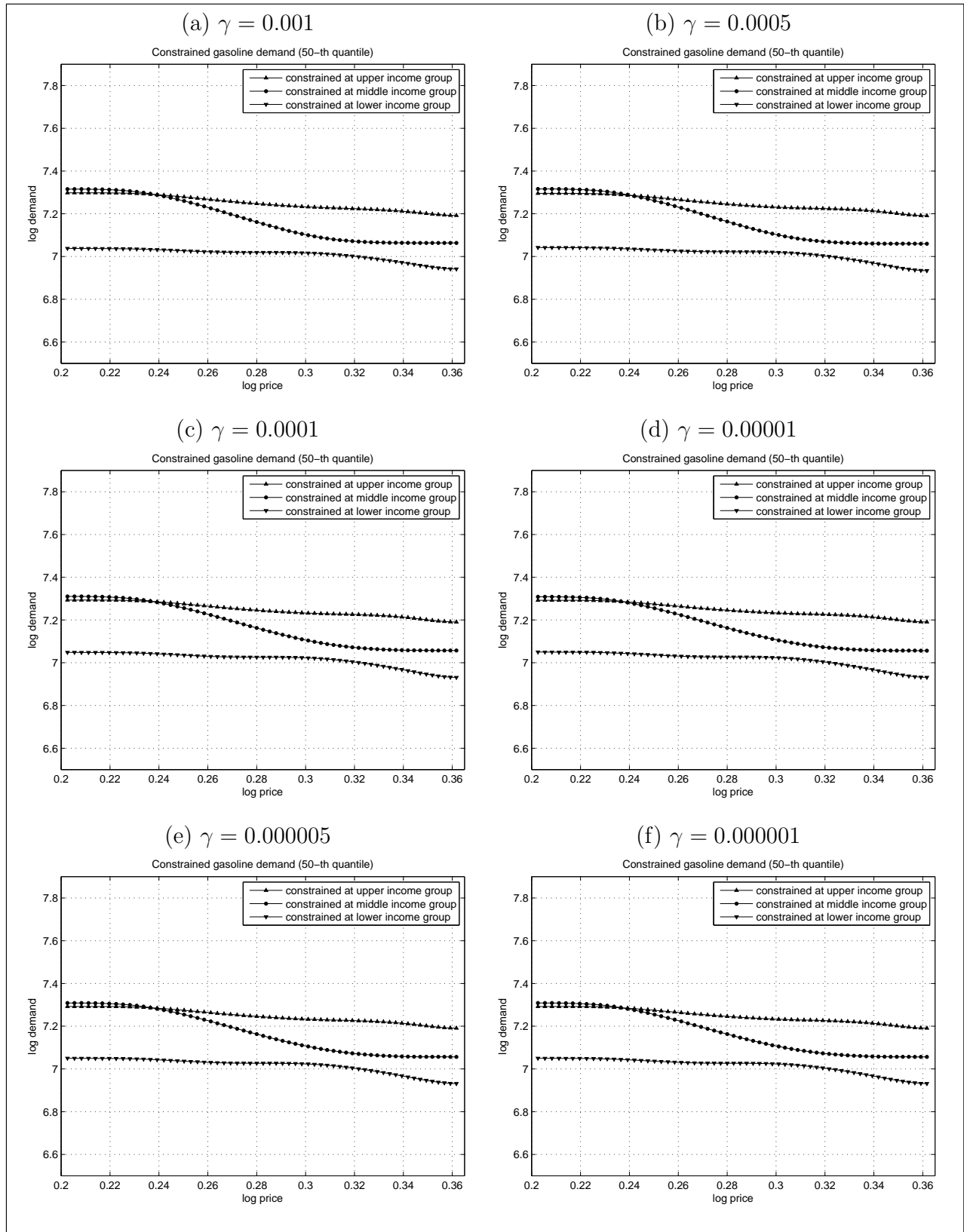
$$l(z, \zeta) = K \left(\frac{z - \zeta}{h_l} \right),$$

where

$$K(u) = \frac{15}{16} (1 - u^2)^2 I(|u| \leq 1)$$

and h_l is a bandwidth parameter. Table 6 shows the p -values obtained with several different values of h_l .

Figure A.1: Sensitivity of constrained quantile regression estimates to smoothing parameter



Note: In the computation of the constrained quantile estimates (see Section 3.2), the check function is smoothed in a small neighborhood around 0, using a quadratic approximation over the range $[-(1 - \alpha)\gamma; \alpha\gamma]$ (see Chen (2007)), where γ is a bandwidth parameter. This figure shows the constrained quantile regression estimates for the median ($\alpha = 0.5$), resulting from alternative choices of γ . The figures presented in the main text correspond to panel (f) of Figure A.1.

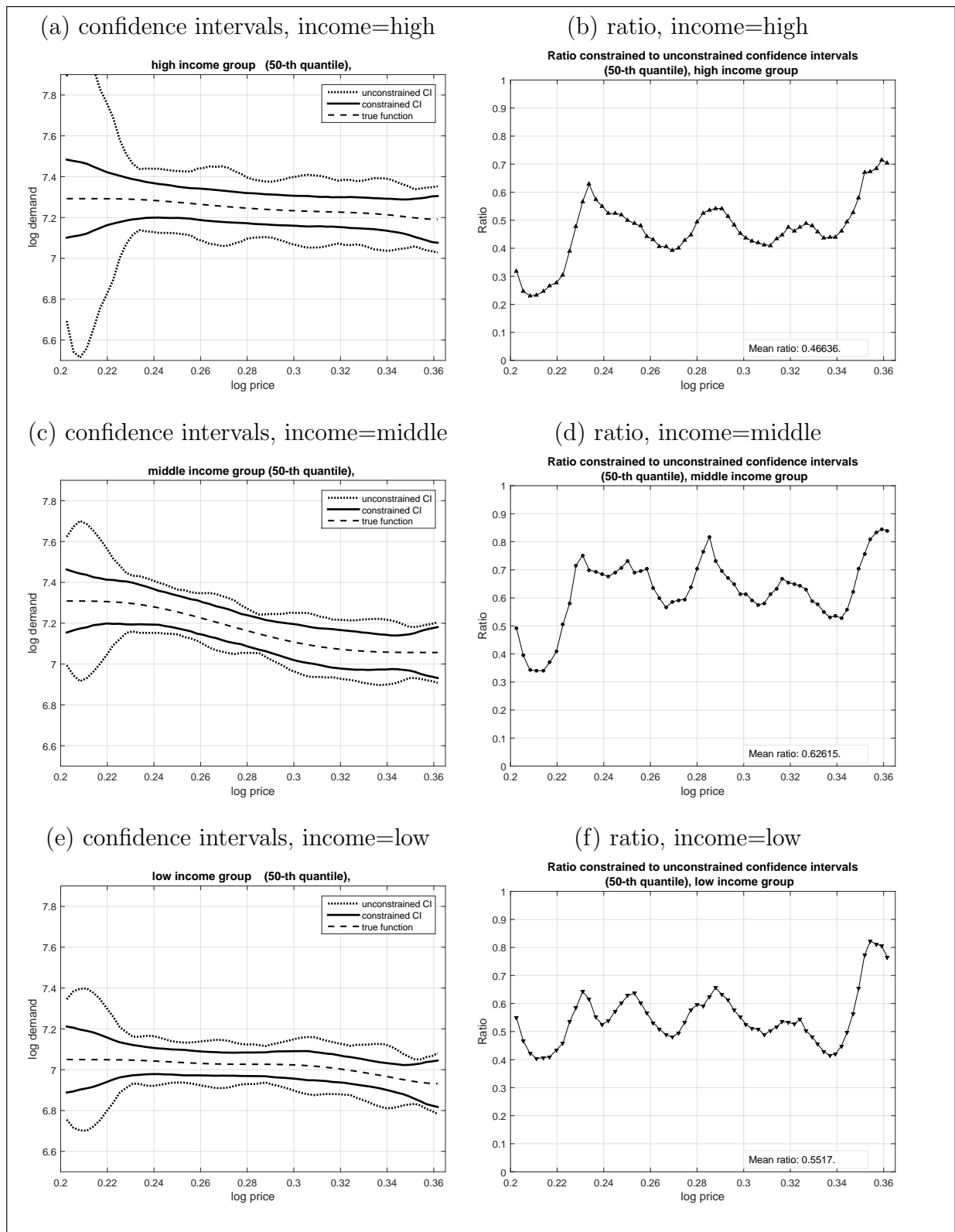
A.2 Simulation-based comparison between constrained and unconstrained estimates

In this subsection we present results from a simulation exercise, in order to provide a sense of the potential gain in precision from imposing the shape restriction. We set up the simulation exercise to be similar to our empirical application. Starting from the observed sample of covariates (price and income), we take our constrained estimates (see Figure 1) as the true function, and add an error term that is drawn randomly from the distribution of estimated residuals.²¹ We then estimate both the constrained and the unconstrained model for these simulated data. We repeat this many times (999 replications). Based on the sample of simulations, we then construct joint symmetrical confidence intervals for the constrained and the unconstrained estimates, respectively, around the true data generating function.

The left panel of Figure A2 compares the resulting confidence intervals between the constrained and the unconstrained median estimates. The results show a narrowing in the width of the bands from imposing the constraint. The gain in precision is higher for the low and the high income group, consistent with the finding that middle income households are more price elastic, and so the Slutsky constraint has less bite for this group. The right-hand figure shows the ratio of the width of the confidence intervals at each price grid point. Averaging across the grid points, we find that the width of the constrained confidence intervals is 47% of the unconstrained ones for high income, 63% for middle income, and 55% for low income. This simulation evidence suggests that in applications similar to ours the gain in precision from imposing the shape constraint can be substantial.

²¹Our model is estimated in shares. When the drawn error term is such that the simulated share would be negative for a particular household, we set the gasoline share for this particular household to zero.

Figure A.2: Simulation results: constrained versus unconstrained confidence interval ($\tau = 0.50$)



Note: Confidence intervals shown are joint symmetrical confidence intervals, based on a simulation with 999 replications, constructed around the true function. The true underlying function is based on the shape-constrained demand estimate, see Figure 1. See text for details.

A.3 Approximating a Heterogeneous Expenditure Function

The argument in this section is informal to minimize its length and complexity. It can be formalized at the cost of a lengthier and more complex presentation. Assume that the demand model is

$$Q = G(p, y, u), \quad (13)$$

where Q is quantity, y is income, p is price, and u is a scalar unobserved heterogeneity variable that is independent of (p, y) . We also assume that G is monotone increasing in its third argument. Without further loss of generality, we can assume that $u \sim U[0, 1]$, so the α quantile of demand is

$$G_\alpha(p, y) = G(p, y, \alpha).$$

We obtain the expenditure function $e(p)$ from G_α by solving the differential equation

$$\frac{de}{dp} = G_\alpha(p, e). \quad (14)$$

Now suppose that unobserved heterogeneity is multivariate. Denote the additional unobserved heterogeneity variables by the vector ε . Then the correct demand function is

$$Q = G(p, y, u, \varepsilon).$$

The demand function (13) and the expenditure function obtained from (14) are correct only if $\varepsilon = 0$. The α quantile of demand conditional on ε is

$$G_\alpha(p, y, \varepsilon) = G(p, y, \alpha, \varepsilon),$$

and the expenditure function $e(p, \varepsilon)$ is the solution to

$$\frac{de}{dp} = G_\alpha(p, e, \varepsilon). \quad (15)$$

Assume that G_α is twice differentiable with respect to its third argument and that the second derivative is bounded. Define

$$g_\varepsilon(p, y, 0) = \|\varepsilon\|^{-1} \varepsilon' \nabla_\varepsilon G_\alpha(p, y, 0).$$

Then a Taylor series expansion gives

$$\begin{aligned} G_\alpha(p, y, \varepsilon) &= G_\alpha(p, y, 0) + \varepsilon' \nabla_\varepsilon G_\alpha(p, y, 0) + O(\|\varepsilon\|^2) \\ &= G_\alpha(p, y) + \|\varepsilon\| g_\varepsilon(p, y, 0) + O(\|\varepsilon\|^2). \end{aligned}$$

Then (15) is equivalent to

$$\frac{de}{dp} = G_\alpha(p, e) + \|\varepsilon\| g_\varepsilon(p, e, 0) + O(\|\varepsilon\|^2). \quad (16)$$

Make the following Taylor series approximation to the solution to (16):

$$e(p, \varepsilon) = e(p, 0) + \varepsilon' \nabla_e e(p, 0) + O(\|\varepsilon\|^2). \quad (17)$$

Define

$$e_{1\varepsilon}(p) = \|\varepsilon\|^{-1} \varepsilon' \nabla_e e(p, 0).$$

Then

$$\begin{aligned} e(p, \varepsilon) &= e(p, 0) + \|\varepsilon\| e_{1\varepsilon}(p) + O(\|\varepsilon\|^2) \\ &= e(p) + \|\varepsilon\| e_{1\varepsilon}(p) + O(\|\varepsilon\|^2), \end{aligned}$$

and (16) becomes

$$\frac{de}{dp} + \|\varepsilon\| \frac{de_{1\varepsilon}}{dp} = G_\alpha(p, e) + \|\varepsilon\| g_\varepsilon(p, e, 0) + O(\|\varepsilon\|^2). \quad (18)$$

Now equate terms that have the same power of $\|\varepsilon\|$ to obtain

$$\frac{de}{dp} = G_\alpha(p, e)$$

and

$$\frac{de_{1\varepsilon}}{dp} = g_\varepsilon(p, e, 0).$$

It follows that the leading term of the expansion (17) is $e(p)$, the expenditure function obtained from the incorrect demand function (13), and that $e(p)$ differs from the correct expenditure function by $O(\|\varepsilon\|)$. Therefore, the expenditure function obtained under a possibly incorrect assumption of univariate unobserved heterogeneity is a good approximation to the correct expenditure function based on multivariate unobserved heterogeneity if the additional components of unobserved heterogeneity are small.