

STOCHASTIC DEMAND AND REVEALED PREFERENCE*

RICHARD BLUNDELL[†]

DENNIS KRISTENSEN[‡]

ROSA MATZKIN[§]

THIS VERSION: 17TH MARCH 2010

Preliminary Draft

Abstract

This paper develops new techniques for the estimation and testing of stochastic consumer demand models. For general non-additive stochastic demand functions, we demonstrate how preference inequality restrictions can be utilized to improve on the nonparametric estimation and testing of demand responses. We show how bounds on demands can be estimated non-parametrically, and derive their asymptotic properties utilizing recent results on estimation and testing of parameters characterized by moment inequalities. We also devise a test for rationality of the consumers. An empirical application using data from the U.K. Family Expenditure Survey illustrates the usefulness of the methods.

JEL: C20, D12

Keywords: microeconometrics, consumer behaviour, revealed preference

*We would like to thank Ian Crawford and Roger Koenker for sharing Matlab code with us and helping with the implementation. We also wish to thank participants at the Latin-American Meetings in Buenos Aires, 2009, and the CAM workshop at University of Copenhagen, 2009 for helpful comments and suggestions.

[†]Department of Economics, UCL. E-mail: r.blundell@ucl.ac.uk.

[‡]Department of Economics, Columbia University and CREATES. E-mail: dk2313@columbia.edu

[§]Department of Economics, UCLA. E-mail: matzkin@econ.ucla.edu.

1 Introduction

This paper develops new nonparametric techniques for the estimation and prediction of consumer demand responses. The objectives are two-fold: First, to impose minimum restrictions on how unobserved heterogeneity enters the demand function of the individual consumer. Second, to utilize restrictions arriving from choice theory to improve demand estimation and prediction.

Our analysis takes place in the commonly occurring empirical setting where only a relatively small number of market prices are observed, but within each of those markets the demand responses of a large number of consumers are reported. In this setting, it is not possible to point identify the predicted demand response to a new, unobserved price. We will instead use restrictions derived from revealed preference theory to establish bounds on the demand responses: If consumers behave according to the axioms of revealed preference their vector of demands at each relative price will satisfy certain well known inequalities (see Afriat, 1973 and Varian, 1982). If, for any individual, these inequalities are violated then that consumer can be deemed to have failed to behave according to the optimisation rules of revealed preference. In this paper, we will employ the results of Blundell, Browning and Crawford (2003, 2008) who extend the analysis of Varian (1982) and obtain ‘expansion path based bounds’ (E-bounds) which are the tightest possible bounds, given the data and the theory. These restrictions will be used to obtain improved demand estimates at observed prices, and to establish bounds on predicted demands at new prices.

We first develop nonparametric sieve estimators of the stochastic demand functions at any given set of observed prices, and show how revealed preferences can be imposed in the estimation. In the estimation, we impose weak assumptions on how the unobserved heterogeneity enters the demand function. In particular, we allow for non-additive heterogeneity which is in contrast to most of the existing literature on the estimation of Engel curves. We show how the estimator can be formulated as a constrained nonparametric quantile estimation problem, and derive its asymptotic properties.

The second part of the paper is concerned with predicting consumer demand responses to a new relative price level that has not been previously observed. Given the limited price variation it is not possible to nonparametrically point identify demand at new, previously unobserved, relative prices. However, the revealed preference restrictions developed in Blundell, Browning and Crawford (2003, 2008) allow us to establish bounds on the demands for the nonseparable heterogeneity case that is the focus here. We show how the demand functions estimated at observed prices, can be used to recover these bounds nonparametrically. The estimation problem is non-standard and falls within the framework of partially identified models (see e.g. Manski, 1993). We employ the techniques developed in, amongst others, Chernozhukov, Hong and Tamer (2003) to establish the properties of the nonparametric demand bounds estimators. As in that part of the literature, the asymptotic distribution is non-standard and confidence bands can in general not be computed directly.

The estimation and prediction strategy outlined above heavily exploits revealed preference inequalities. In the third part of the paper, we develop a nonparametric test for this rationality assumption. This testing problem is again non-standard since the testable restrictions take the form of a set of inequality constraints. As such the testing problem is similar to one of estimation

and testing when the parameter is on the boundary of the parameter space as analyzed in, for example, Andrews (1999, 2001) and Andrews and Guggenberger (2009). By importing the techniques developed there, we derive the asymptotic properties of the test statistic. Again, this is highly non-standard and requires the use of either simulations or resampling techniques.

Our empirical analysis is based on data from the U.K. Family Expenditure Survey where the relative price variation occurs over time, and samples of consumers, each of a particular household type, are observed at specific points in time in particular regional locations. Employing the developed methodology, we obtain demand function and E-bounds estimates for own- and cross-price demand responses using these expansion paths. We find that it is indeed important to relax the additive restriction normally imposed in empirical consumer demand analysis, and that the bounds are informative.

A key ingredient of the analysis we conduct is the Engel curve which describes the expansion path for demand as total expenditure changes. The modelling and estimation of this relationship has a long history. Working (1943) and Leser (1963) suggested standard parametric regression models where budget shares are linear functions of log total budget; the so-called Piglog-specification. This simple linear model has since then been generalised in various ways since empirical studies suggested that higher order logarithmic expenditure terms are required for certain expenditure share equations, see e.g. Hausman, Newey and Powell (1995), Lewbel (1991), Banks, Blundell and Lewbel (1997). An obvious way to detect the presence of such higher order terms is non- and semiparametric methods which have been widely used in the econometric analysis of Engel curves, see for example Blundell, Chen and Kristensen (2007), Blundell and Duncan (1998) and Härdle and Jerison (1988).

In most of these empirical studies it has been standard to simply impose an additive error structure which greatly facilitates the econometric analysis since the demand model collapses to a regression. This type of specification allows relatively straightforward identification and estimation of the structural parameters. Additive heterogeneity on the other hand imposes very strong assumptions on the class of underlying utility functions, see e.g. Lewbel (2001), and as such is in general inconsistent with economic theory; see also Beckert (2007). In response to this problem with existing empirical consumer demand models, we allow for non-additive heterogeneity structures thereby allowing for observed and unobserved characteristics to interact in a much more involved manner compared to standard additive specifications. See Lewbel and Pendakur (2009) for one of the few parametric specification that allows non-additive interaction.

An early treatment of nonparametric identification of non-additive models is Brown (1983) whose results were extended in Roehrig (1988). Brown and Matzkin (1998, 2003) and Matzkin (2003a,b) building on their work propose estimators. A number of other papers have addressed identification and estimation: Chesher (2001, 2002a, 2002b, 2003) consider quantile-driven identification; Ma and Koenker (2003) make use of his results to construct parametric estimators; Imbens and Newey (2009) and Chernozhukov, Imbens and Newey (2003) establish estimators which allow for endogeneity. Our unconstrained demand function estimator is similar to the ones developed in Imbens and Newey (2009).

An important feature of the estimation of Engel curve is the possible presence of endogeneity in the total expenditure variable. In a parametric framework this can be dealt with using standard IV-techniques. In recent years, a range of different methods have been proposed to deal with this problem in a nonparametric setting. The two main approaches proposed in the literature is nonparametric IV (Ai and Chen (2003), Hall and Horowitz (2003), Darolles, Florens and Renault (2002)) and control functions (Newey, Powell and Vella (1998)). Both these methods have been applied in the empirical analysis of Engel curves (Blundell, Chen and Kristensen (2003) and Blundell, Duncan and Pendakur (1998) respectively). We briefly discuss how our estimators and tests can be extended to handle endogeneity of explanatory variables by using recent results in Chen and Pouzo (2009), Chernozhukov, Imbens and Newey (2007) and Imbens and Newey (2009).

Finally, we note that other papers have combined nonparametric techniques and economic theory to estimate and test demand systems; see, for example, Haag, Hoderlein and Pendakur (2009), Hoderlein (2008), Hoderlein and Stoye (2009), Lewbel (1995).

The remainder of the paper is organized as follows: In Section 2, we set up the basic econometric framework. In Section 3-4, we develop estimators of the demand functions at observed prices. The estimation of demand bounds is considered in Section 5, while a test for rationality is developed in Section 6. In Sections 7 and 8, we discuss the implementation of the estimators and how to compute confidence bands. We briefly discuss how to allow for endogenous explanatory variables in Section 9. Section 10 contains an empirical application on British household data. We conclude in Section 11. All proofs have been relegated to the Appendix.

2 The Framework

Suppose we have observed a consumers' market over T periods. Let $\mathbf{p}(t)$ be the set of prices for the goods that all consumers face at time $t = 1, \dots, T$. At each time point t , we draw a new random sample of $n \geq 1$ consumers. For each consumer, we observe his demands and income level (and potentially some other characteristics such as age, education etc.). Let $\mathbf{q}_i(t)$ and $x_i(t)$ be consumer i 's ($i = 1, \dots, n$) vector of demand and income level at time t ($t = 1, \dots, T$). We stress that the data $\{\mathbf{p}(t), \mathbf{q}_i(t), x_i(t)\}$, $t = 1, \dots, T$ and $i = 1, \dots, n$, is not a panel data set since we do not observe the same consumer over time.

We focus on the situation where there are only two goods in the economy such that $\mathbf{q}(t) = (q_1(t), q_2(t))' \in \mathbb{R}_+^2$ and $\mathbf{p}(t) = (p_1(t), p_2(t))' \in \mathbb{R}_+^2$.¹ The demand for good 1 is assumed to arrive from the following demand function,

$$q_1(t) = d_1(x(t), \mathbf{p}(t), \varepsilon(t)),$$

where $\varepsilon(t) \in \mathbb{R}$ is an individual specific error term that may reflect individual heterogeneity in preferences (tastes). To ensure that the budget constraint is met, the demand for good two must

¹By the end of the paper, we discuss extensions to general, multidimensional goods markets.

satisfy:

$$q_2(t) = d_2(x(t), \mathbf{p}(t), \varepsilon(t)) := \frac{x(t) - p_1(t) d_1(x(t), \mathbf{p}(t), \varepsilon(t))}{p_2(t)}.$$

We collect the two demand functions in $\mathbf{d} = (d_1, d_2)$. The demand function \mathbf{d} should be thought of as the solution to an underlying utility maximization problem that the individual consumer solves.

The demand function could potentially depend on other observable characteristics besides income, but to keep the notation at a reasonable level we suppress such dependence in the following. If additionally explanatory variables are present, all the following assumptions, arguments and statements are implicitly made conditionally on those.

We here consider the often occurring situation where the time span T over which we have observed consumers and prices is small (in the empirical application $T = 6$). In this setting, we do not observe enough price variation to identify how these impact demand; thus, we are not able to identify the mapping $\mathbf{p} \mapsto \mathbf{d}(x, \mathbf{p}, \varepsilon)$. To emphasize this, we will in the following write

$$\mathbf{d}(x(t), t, \varepsilon(t)) := \mathbf{d}(x(t), \mathbf{p}(t), \varepsilon(t)).$$

So we have a sequence of T demand functions, $\{\mathbf{d}(x, t, \varepsilon)\}_{t=1}^T$.

The unobserved heterogeneity $\varepsilon(t)$ is assumed (or normalized) to follow a uniform distribution, $\varepsilon(t) \sim U[0, 1]$ and to be independent of $x(t)$.² This combined with the assumption that d_1 is invertible in $\varepsilon(t)$ implies that $d_1(x, t, \tau)$ is identified as the τ th quantile of $q_1(t) | x(t) = x$ (Matzkin, 2003; Newey and Imbens, 2009):

$$d_1(x, t, \tau) = F_{q_1(t)|x(t)=x}^{-1}(\tau), \quad \tau \in [0, 1]. \quad (1)$$

3 Unrestricted Sieve Estimator

We here develop sieve estimators of the sequence of demand functions $\mathbf{d}(x, t, \varepsilon) = (d_1(x, t, \varepsilon), d_2(x, t, \varepsilon))$, $t = 1, \dots, T$.

As a starting point, we assume that for all $t = 1, \dots, T$ and all $\tau \in [0, 1]$, the function $x \mapsto d_1(x, t, \tau)$ is situated in some known function space \mathcal{D}_1 which is equipped with some (pseudo-)norm $\|\cdot\|$.³ We specify the precise form of \mathcal{D}_1 and $\|\cdot\|$ below. Given the function space \mathcal{D}_1 , we choose sieve spaces $\mathcal{D}_{1,n}$ that are finite-dimensional subsets of \mathcal{D} . In particular, we will assume that for any function $d_1 \in \mathcal{D}_1$, there exists a sequence $\pi_n d_1 \in \mathcal{D}_{1,n}$ such that $\|\pi_n d_1 - d_1\| \rightarrow 0$ as $n \rightarrow \infty$. Most standard choices of the function space \mathcal{D}_1 can be written on the form

$$\mathcal{D}_1 = \left\{ d_1 : d_1(x, t, \tau) = \sum_{k \in \mathcal{K}} \pi_k(t, \tau) B_k(x), \quad \pi(t, \tau) \in \mathbb{R}^{|\mathcal{K}|} \right\},$$

²The independence assumption can be relaxed as discussed in Section 9.

³The function space could without problems be allowed to depend on time, t , and the errors, τ . For notational simplicity, we maintain that the function space is the same across different values of (t, τ) .

for known (basis) functions $\{B_k\}_{k \in \mathcal{K}}$, and some (infinite-dimensional) index set \mathcal{K} ; see Chen (2007, Section 2.3) for some standard specifications. A natural choice for sieve is then

$$\mathcal{D}_{1,n} = \left\{ d_{n,1} : d_{n,1}(x, t, \tau) = \sum_{k \in \mathcal{K}_n} \pi_k(t, \tau) B_k(x), \quad \pi(t, \tau) \in \mathbb{R}^{|\mathcal{K}_n|} \right\}, \quad (2)$$

for some sequence of (finite-dimensional) sets $\mathcal{K}_n \subseteq \mathcal{K}$. Finally, we define the space of vector functions,

$$\mathcal{D} = \left\{ \mathbf{d} = (d_1, d_2) : d_1(x, t, \tau) \in \mathcal{D}_{1,n}, \quad d_2(t, x, \tau) := \frac{x - p_1(t) d_1(x, t, \tau)}{p_2(t)} \right\},$$

and with the associated sieve space \mathcal{D}_n defined similarly to $\mathcal{D}_{1,n}$.

Given the function space \mathcal{D} and its associated sieve, we can construct a sieve estimator of the function $\mathbf{d}(\cdot, t, \tau)$. Given that $d_1(x, t, \tau)$ is identified as a conditional quantile for any given value of x , c.f. eq. (1), we may employ standard quantile regression techniques to obtain the estimator: Let

$$\rho_\tau(z) = (\mathbb{I}\{z < 0\} - \tau)z, \quad \tau \in [0, 1],$$

be the standard check function used in quantile estimation (see Koenker and Bassett, 1978). We then propose to estimate $\mathbf{d}(x, t, \tau)$ by

$$\hat{\mathbf{d}}(\cdot, t, \tau) = \arg \min_{\mathbf{d}_n \in \mathcal{D}_n} \frac{1}{n} \sum_{i=1}^n \rho_\tau(q_{1,i}(t) - d_{n,1}(x_i(t), t, \tau)), \quad (3)$$

for any $t = 1, \dots, T$ and $\tau \in [0, 1]$.

The above estimator can be computed using standard numerical methods for linear quantile regressions when the sieve space is on the form in Eq. (2): Define $W_i(t) = \{B_k(x_i(t)) : k \in \mathcal{K}_n\} \in \mathbb{R}^{|\mathcal{K}_n|}$ as the collection of basis functions spanning the sieve $\mathcal{D}_{1,n}$. Then the sieve estimator is given by $\hat{d}_1(x, t, \tau) = \sum_{k \in \mathcal{K}_n} \hat{\pi}_k(t, \tau) B_k(x)$, where

$$\hat{\pi}(t, \tau) = \arg \min_{\pi \in \mathbb{R}^{|\mathcal{K}_n|}} \frac{1}{n} \sum_{i=1}^n \rho_\tau(q_{1,i}(t) - \pi' W_i(t)), \quad \tau \in [0, 1]. \quad (4)$$

That is, the estimator $\hat{\pi}(t, \tau)$ is simply the solution to a standard linear quantile regression problem. Finally, the estimator of the demand function for the "residual" good is given by

$$\hat{d}_2(x, t, \tau) = \frac{x - p_1(t) \hat{d}_1(x, t, \tau)}{p_2(t)}. \quad (5)$$

To derive the asymptotics of $\hat{\mathbf{d}}(\cdot, t, \tau) = (\hat{d}_1(\cdot, t, \tau), \hat{d}_2(\cdot, t, \tau))$, we introduce the following L_2 -function norm which will be used to state our convergence rate results:

$$\|\mathbf{d}\|_2 = \sqrt{E[\|\mathbf{d}(x, t, \tau)\|]}.$$

The following assumptions are imposed on the model:

A.1 The variable $x(t)$ has bounded support, $x(t) \in \mathcal{X} = [a, b]$ for $-\infty < a < b < +\infty$, and is independent of $\varepsilon \sim U[0, 1]$.

A.2 The demand function $d_1(x, t, \varepsilon)$ is invertible in ε and is continuously differentiable in (x, ε) .

These are fairly standard assumptions in the literature on nonparametric quantile estimation. It would be desirable to weaken the restriction of bounded support, but the cost would be more complicated conditions and proof so we maintain (A.1) (see e.g. Chen, Blundell and Kristensen, 2007 for results with unbounded support). The independence assumption rules out endogenous income; in Section 9, we argue how this can be allowed for by adopting nonparametric IV or control function approaches. We refer to Beckert (2007) and Beckert, and Blundell (2008) for more primitive conditions in terms of the underlying utility-maximization problem for (A.2) to hold.

Finally, we need to impose some restrictions on the function space \mathcal{D}_1 . We consider two different choices of function spaces, and assume either of the two following assumptions, (A.3.1) or (A.3.2):

A.3 The function $d_1(\cdot, t, \tau) \in \mathcal{D}_1$, where \mathcal{D}_1 satisfies either of the following two conditions:

- i $\mathcal{D}_1 = \mathcal{W}_2^m([a, b])$ where $\mathcal{W}_p^m([a, b])$ is the Sobolev space of all functions on $[a, b]$ with L_p -integrable derivative up to order $m \geq 0$.
- ii $\mathcal{D}_1 = \mathcal{B}_{p,\infty}^m(c)$, $m > 0$, $p \geq 2$, where $\mathcal{B}_{p,\infty}^m(c)$ is the Besov ball of L_p -functions with $\|d_1\|_{\mathcal{B}_{p,\infty}^m} < c$.

Under (A.3.i), we may use tensor-product splines as a sieve space to approximate d_1 , while under (A.3.ii), tensor-product wavelets may be used; for the definition of those, see Chen (2007, Section 2.3). Given the chosen sieve space, define the following class of functions indexed by $\mathbf{d}(\cdot, \tau) \in \mathcal{D}_n$,

$$\mathcal{F}_n(\delta) = \{Q_\tau(\mathbf{d}(\cdot, \tau)) - Q_\tau(\mathbf{d}_0(\cdot, \tau)) : \|\mathbf{d}(\cdot, \tau) - \mathbf{d}_0(\cdot, \tau)\|_2 \leq \delta_n, \quad \mathbf{d}(\cdot, \tau) \in \mathcal{D}_n\},$$

where

$$Q_\tau(\mathbf{d}_0(\cdot, \tau)) = E[\rho_\tau(q_1 - d_1(x, \tau))].$$

Let $H_\square(w, \mathcal{F}_n(\delta), \|\cdot\|_2) = \log(N_\square(w, \mathcal{F}_n(\delta), \|\cdot\|_2))$, where $N_\square(w, \mathcal{F}_n(\delta), \|\cdot\|_2)$ is the so-called L_2 -covering numbers with bracketing of the function class $\mathcal{F}_n(\delta)$, see Van der Vaart and Wellner (1996) and van de Geer (2000) for the precise definitions. We now have the following result:

Theorem 1 *Assume that (A.1)-(A.2) hold. Then for any $t = 1, \dots, T$ and $\tau \in [0, 1]$:*

$$\|\hat{\mathbf{d}}(\cdot, t, \tau) - \mathbf{d}_0(\cdot, t, \tau)\|_2 = O_P(\max\{\delta_n, \|\pi_n d_{1,0}(\cdot, t, \tau) - d_1(\cdot, t, \tau)\|\})$$

where

$$\delta_n = \inf_{\delta \in (0, 1)} \left\{ \frac{1}{\sqrt{n}\delta^2} \int_{b\delta^2}^\delta \sqrt{H_\square(w, \mathcal{F}_n, \|\cdot\|)} dw \leq \text{const.} \right\},$$

$H_\square(w, \mathcal{F}_n, \|\cdot\|)$ is the metric entropy with bracketing and $\pi_n d_{1,0}$ is an element in $\mathcal{D}_{1,n}$.

In particular, under either (A.3.i) [using splines] or (A.3.ii) [using wavelets],

$$\|\hat{\mathbf{d}}(\cdot, t, \tau) - \mathbf{d}_0(\cdot, t, \tau)\|_2 = O_P\left(n^{-m/(2m+1)}\right),$$

for suitable choice of approximation parameters.

Next, under additional assumptions, we obtain by following Kim (2006) the following result regarding the asymptotic distribution of the sieve estimator:

Theorem 2 [INCOMPLETE!] Assume that [????]. Then for any $x(t) \in \mathbb{R}_+$, $t = 1, \dots, T$, and $\tau \in [0, 1]$,

$$\sqrt{r_n} \begin{pmatrix} \hat{\mathbf{d}}(x(1), 1, \tau) - \mathbf{d}_0(x(1), 1, \tau) \\ \vdots \\ \hat{\mathbf{d}}(x(T), T, \tau) - \mathbf{d}_0(x(T), T, \tau) \end{pmatrix} \rightarrow^d N(0, \Sigma(x, \tau)),$$

where $r_n = \dots$ and $\Sigma(x, \tau) = \dots$

A consistent estimator of $\Sigma(x, \tau)$ is

$$\Sigma(x, \tau) = \dots$$

4 GARP-Restricted Sieve Estimator

We now wish to impose revealed preferences (GARP) restrictions on the function \mathbf{d} in the estimation. First, we briefly introduce the concept of GARP; see Blundell, Browning and Crawford (2003, 2008) for a more detailed introduction to GARP and its empirical implications.

Consider a given income level $x(T)$ at time T , and compute recursively for $t = T-1, T-2, \dots, 1$, an income expansion path $\{x(t)\}$ by

$$x(t) = \mathbf{p}(t)' \mathbf{d}(x(t+1), t+1, \tau).$$

We then require that the demand of a given consumer characterised by $\tau \in [0, 1]$ satisfies:

$$\mathbf{p}(t)' \mathbf{d}(x(t), t, \tau) = x(t) \leq \mathbf{p}(t)' \mathbf{d}(x(s), s, \tau), \quad s < t, \quad t = 2, \dots, T. \quad (6)$$

If the demand functions $\mathbf{d}(x, t, \tau)$, $t = 1, \dots, T$, satisfy these inequalities for any given income level $x(T)$ we say that " \mathbf{d} satisfies GARP".

We expect that if indeed the consumers in our sample satisfy GARP, more precise estimates of $\mathbf{d}(\cdot, \cdot, \tau)$ can be obtained by imposing this restriction. Furthermore, once a GARP restricted estimator has been obtained, it can be used to test for GARP by comparing it with the unrestricted estimator developed in the previous section. A GARP-restricted sieve estimator is easily obtained in principle: First observe that the unrestricted estimator of $\{\mathbf{d}(\cdot, t, \tau)\}_{t=1}^T$ of the previous section can be expressed as the solution to the following joint estimation problem.

$$\{\hat{\mathbf{d}}(\cdot, t, \tau)\}_{t=1}^T = \arg \min_{\{\mathbf{d}_n(\cdot, t, \tau)\}_{t=1}^T \in \mathcal{D}_n^T} \frac{1}{n} \sum_{t=1}^T \sum_{i=1}^n \rho_\tau(q_{1,i}(t) - d_{1,n}(t, x_i(t))), \quad \tau \in [0, 1],$$

where $\mathcal{D}_n^T = \otimes_{t=1}^T \mathcal{D}_n$ and \mathcal{D}_n is defined in the previous section. Since there are no restrictions across goods and t , the above definition of $\{\hat{\mathbf{d}}(\cdot, t, \tau)\}_{t=1}^T$ is equivalent to the unrestricted estimators in eqs. (3) and (5). In order to impose the GARP restrictions, we simply define the constrained function set as

$$\mathcal{D}_C^T := \mathcal{D}^T \cap \{\mathbf{d}(\cdot, \cdot, \tau) \text{ satisfies GARP}\}, \quad (7)$$

and similarly the constrained sieve as

$$\mathcal{D}_{C,n}^T := \mathcal{D}_n^T \cap \{\mathbf{d}_n(\cdot, \cdot, \tau) \text{ satisfies GARP}\}.$$

We define the constrained estimator by:

$$\{\hat{\mathbf{d}}_C(\cdot, t, \tau)\}_{t=1}^T = \arg \min_{\{\mathbf{d}_n(\cdot, t, \tau)\}_{t=1}^T \in \mathcal{D}_{C,n}^T} \frac{1}{n} \sum_{t=1}^T \sum_{i=1}^n \rho_\tau(q_{1,i}(t) - d_{1,n}(t, x_i(t))), \quad \tau \in [0, 1]. \quad (8)$$

Since GARP imposes restrictions across both goods ($l = 1, 2$) and time ($t = 1, \dots, T$), the above estimation problem can no longer be split up into individual subproblems as in the unconstrained case.

The proposed estimator shares some similarities with the ones considered in, for example, Beresteanu (2004), Gallant and Golub (1984), Mammen and Thomas-Agnan (1999) and Yatchew and Bos (1997) who also consider constrained sieve estimators. However, they focus on least-squares regression while ours is a least-absolute deviation estimator, and they furthermore restrict themselves to linear constraints. There are some results for estimation of monotone quantiles and other linear constraints, see Chernozhukov et al (2006), Koenker and Ng (2005) and Wright (1984), but again their constraints are simpler to analyze and implement. These two issues, a non-smooth criterion function and non-linear constraints, complicate the analysis and implementation of our estimator, and we cannot readily import results from the existing literature.

In order to derive the convergence rate of the constrained sieve estimator, we employ the same proof strategy as found elsewhere in the literature on nonparametric estimation under shape constraints, see e.g. Birke and Dette (2007), Mammen (1991), Mukerjee (1988): We first demonstrate that as $n \rightarrow \infty$, the unrestricted estimator, $\hat{\mathbf{d}}$, satisfies GARP almost surely. This implies that $\{\hat{\mathbf{d}}(\cdot, t, \tau)\}_{t=1}^T \in \mathcal{D}_{C,n}^T$ with probability approaching one (w.p.a.1) which in turn means that $\hat{\mathbf{d}} = \hat{\mathbf{d}}_C$ w.p.a.1, since $\hat{\mathbf{d}}_C$ solves a constrained version of the minimization problem that $\hat{\mathbf{d}}$ is a solution to. We are now able to conclude that $\hat{\mathbf{d}}_C$ is asymptotically equivalent $\hat{\mathbf{d}}$, and all the asymptotic properties of $\hat{\mathbf{d}}$ are inherited by $\hat{\mathbf{d}}_C$.

For the above argument to go through, we need to slightly change the definition of the constrained estimator though. We introduce the following generalized version of GARP: We say that " \mathbf{d} satisfies GARP(ϵ)" for some ("small") $\epsilon \geq 0$ if for any income expansion path,

$$x(t) \leq \mathbf{p}(t)' \mathbf{d}(x(s), s, \tau) + \epsilon, \quad s < t, \quad t = 2, \dots, T.$$

The definition of GARP(ϵ) is akin to Afriat (1973) who suggests a similar modification of GARP to allow for waste ("partial efficiency"). We then define the constrained function space and its

associated sieve as:

$$\begin{aligned}\mathcal{D}_C^T(\epsilon) &= \mathcal{D}^T \cap \{\mathbf{d}(\cdot, \cdot, \tau) \text{ satisfies GARP}(\epsilon)\}, \\ \mathcal{D}_{C,n}^T(\epsilon) &= \mathcal{D}_n^T \cap \{\mathbf{d}_n(\cdot, \cdot, \tau) \text{ satisfies GARP}(\epsilon)\}.\end{aligned}$$

We note that the constrained function space \mathcal{D}_C^T as defined in eq. (7) satisfies $\mathcal{D}_C^T = \mathcal{D}_C^T(0)$. Moreover, it should be clear that $\mathcal{D}_C^T(0) \subset \mathcal{D}_C^T(\epsilon)$ and $\mathcal{D}_{C,n}^T(0) \subset \mathcal{D}_{C,n}^T(\epsilon)$ since $\text{GARP}(\epsilon)$, $\epsilon > 0$, imposes weaker restrictions on the demand functions.

We now re-define our GARP constrained estimators to solve the same optimization problem as before, but now the optimization takes place over $\mathcal{D}_{C,n}(\epsilon)$. We let $\hat{\mathbf{d}}_C^\epsilon$ denote this estimator, and note that $\hat{\mathbf{d}}_C^0 = \hat{\mathbf{d}}_C$, where $\hat{\mathbf{d}}_C$ is given in Eq. (8). The assumption that $\{\mathbf{d}_0(\cdot, t, \tau)\}_{t=1}^T \in \mathcal{D}_C^T(0)$ implies that $\{\hat{\mathbf{d}}(\cdot, t, \tau)\}_{t=1}^T \in \mathcal{D}_{C,n}^T(\epsilon)$ w.p.a.1. Since $\hat{\mathbf{d}}_C^\epsilon$ is a constrained version of $\hat{\mathbf{d}}$, this implies that $\hat{\mathbf{d}}_C^\epsilon = \hat{\mathbf{d}}$ w.p.a.1. Similar assumptions and proof strategies have been employed in Birke and Dette (2007) [Mammen (1991)]: They assume that the function being estimated is *strictly* convex [monotone], such that the unconstrained estimator is convex [monotone] w.p.a.1. Since $\mathcal{D}_{C,n}^T(0) \subset \mathcal{D}_{C,n}^T(\epsilon)$, our new estimator will in general be less precise than the one defined as the optimizer over $\mathcal{D}_{C,n}^T(0)$, but for small values of $\epsilon > 0$ the difference should be negligible.

Theorem 3 *Assume that (A.1)-(A.3) hold, and that $\mathbf{d}_0 \in \mathcal{D}_C^T(0)$. Then for any $\epsilon > 0$: $\|\hat{\mathbf{d}}_C^\epsilon(\cdot, t, \tau) - \mathbf{d}_0(\cdot, t, \tau)\|_2 = O_P(n^{-m/(2m+1)})$, $t = 1, \dots, T$, and the estimator has the same asymptotic distribution as the unrestricted estimator given in Theorem 2.*

It should be noted that in terms of convergence rate our unconstrained and constrained estimators are asymptotically equivalent. In terms of asymptotic convergence rate, we are not able to show that our additional constraints lead to an improvement of the estimator. This is similar to other results in the literature on constrained nonparametric estimation. Kiefer (1982) and Beresteanu (2004) establish optimal nonparametric rates in the case of constrained densities and regression functions respectively when the constraints are not binding. In both cases, the optimal rate is the same as for the unconstrained one. However, as demonstrated both analytically and through simulations in Mammen (1991) for monotone restrictions (see also Beresteanu, 2004 for simulation results for other restrictions), there may be significant finite-sample gains.

We conjecture that the above result will not in general hold for the estimator $\hat{\mathbf{d}}_C$ defined as the minimizer over $\mathcal{D}_{C,n}^T(0)$. In this case the GARP constraints would be binding, and we can no longer ensure that the unconstrained estimator is situated in the interior of the constrained function space. This in turn means that the unconstrained and constrained estimator most likely are not asymptotically equivalent and very different techniques have to be used to analyze the constrained estimator. In particular, the rate of convergence and/or asymptotic distribution of it would most likely be non-standard. This is, for example, demonstrated in Andrews (1999,2001), Anevski and Hössjer (2006) and Wright (1981) who give results for inequality-constrained parametric and nonparametric problems respectively.

Finally, we note that the above theorem is not specific to our particular quantile sieve estimator. One can by inspection easily see that the arguments employed in our proof can be carried over

without any modifications to show that for any unconstrained demand function estimator, the corresponding RP-constrained estimator will be asymptotically equivalent.

5 Estimation of Demand Bounds

Once an estimator of the demand function has been obtained, either unrestricted or restricted, we can proceed to estimate the associated demand bounds. We will here utilize the machinery developed in Chernozhukov, Hong and Tamer (2007) and use their results to develop the asymptotic theory of the proposed demand bound estimators.

Consider a particular consumer characterized by some $\tau \in [0, 1]$ with associated sequence of demand functions $\mathbf{d}(x, t, \tau)$, $t = 1, \dots, T$. Since we keep τ fixed, we suppress the dependence of this variable in the following. Suppose that the consumer faces a given new price \mathbf{p}_0 at an income level x_0 . Define the consumer's budget set as

$$\mathcal{B}_{\mathbf{p}_0, x_0} = \{\mathbf{q} \in \mathbb{R}_+^2 \mid \mathbf{p}_0' \mathbf{q} = x_0\},$$

which is compact and convex. The closure of consumer's so-called demand support set can then be represented as follows:

$$\mathcal{S}_{\mathbf{p}_0, x_0} = \{\mathbf{q} \in \mathcal{B}_{\mathbf{p}_0, x_0} \mid \mathbf{p}(t)' \mathbf{q} \geq \mathbf{p}(t)' \mathbf{d}(x(t), t, \tau), \quad t = 1, \dots, T\},$$

where $\{x(t)\}$ solves

$$\mathbf{p}_0' \mathbf{d}(x(t), t, \tau) = x_0, \quad t = 1, \dots, T.$$

A natural estimator is then to simply substitute the estimated demand functions for the unknown ones. Define $\{\hat{x}(t)\}$ as the solution to

$$\mathbf{p}_0' \mathbf{d}_C(\hat{x}(t), t, \tau) = x_0, \quad t = 1, \dots, T.$$

We then define the estimator of the support set as

$$\hat{\mathcal{S}}_{\mathbf{p}_0, x_0}(c) = \{\mathbf{q} \in \mathcal{B}_{\mathbf{p}_0, x_0} \mid \mathbf{p}(t)' \mathbf{q} \geq \mathbf{p}(t)' \mathbf{d}_C(\hat{x}(t), t, \tau) - c/n, \quad t = 1, \dots, T\},$$

for some positive sequence $c \rightarrow \infty$ with $c/n \rightarrow 0$. In order to do inference, we wish to choose a (possibly data dependent) sequence \hat{c} such that

$$P(\mathcal{S}_{\mathbf{p}_0, x_0} \subseteq \hat{\mathcal{S}}_{\mathbf{p}_0, x_0}(\hat{c})) \rightarrow 1 - \alpha,$$

for some given confidence level $1 - \alpha$. Here, we need to choose $\hat{c} > 0$ to obtain the correct coverage, even if $\hat{\mathcal{S}}(0)$ is consistent.

To derive the asymptotic properties of $\hat{\mathcal{S}}_{\mathbf{p}_0, x_0}(\hat{c})$, we employ the general results of Chernozhukov, Hong and Tamer (2007). To see how our estimator fits into their general framework, define the following set of "moments" $m(\mathbf{q})$,

$$m(\mathbf{q}) = -P\mathbf{q} + b, \quad \hat{m}(\mathbf{q}) = -P\mathbf{q} + \hat{b},$$

where $\mathbf{q} \in \mathbb{R}^2$,

$$\mathbf{P} = [\mathbf{p}(1), \dots, \mathbf{p}(T)]' \in \mathbb{R}_+^{T \times 2}, \quad (9)$$

and

$$b = [\mathbf{p}(1)' \mathbf{d}(x(1), 1, \tau), \dots, \mathbf{p}(T)' \mathbf{d}(x(T), T, \tau)]', \quad \hat{b} = [\mathbf{p}(1)' \hat{\mathbf{d}}_C(x(1), 1, \tau), \dots, \mathbf{p}(T)' \hat{\mathbf{d}}_C(x(T), T, \tau)]'$$

We can then express $\mathcal{S}_{\mathbf{p}_0, x_0}$ and $\hat{\mathcal{S}}_{\mathbf{p}_0, x_0}(\hat{c})$ in terms of a set of moment inequalities:

$$\mathcal{S}_{\mathbf{p}_0, x_0} = \{\mathbf{q} \in \mathcal{B}_{\mathbf{p}_0, x_0} | m(\mathbf{q}) \leq 0\}, \quad \hat{\mathcal{S}}_{\mathbf{p}_0, x_0}(c) = \{\mathbf{q} \in \mathcal{B}_{\mathbf{p}_0, x_0} | \hat{m}(\mathbf{q}) \leq c/n\}.$$

The asymptotic results for $\hat{\mathcal{S}}_{\mathbf{p}_0, x_0}(c)$ will be stated in terms of the Hausmann norm given by:

$$d_H(\mathcal{A}_1, \mathcal{A}_2) = \max \left\{ \sup_{y \in \mathcal{A}_1} \rho(y, \mathcal{A}_2), \sup_{y \in \mathcal{A}_2} \rho(y, \mathcal{A}_1) \right\}, \quad \rho(y, \mathcal{A}) = \inf_{x \in \mathcal{A}} \|x - y\|,$$

for any two sets $\mathcal{A}_1, \mathcal{A}_2 \in \mathbb{R}^2$. We impose the following additional conditions are imposed on the demand functions \mathbf{d} and the observed prices \mathbf{p} :

A.4 $x(t) \mapsto \mathbf{d}(x(t), t, \tau)$ is monotonically increasing, $t = 1, \dots, T$.

A.5 The matrix $\mathbf{P} \in \mathbb{R}_+^{T \times 2}$ defined in Eq. (9) has rank 2.

Theorem 4 Assume that (A.1)-(A.5) hold. Then for any sequence $c \propto \log(n)$,

$$d_H(\hat{\mathcal{S}}_{\mathbf{p}_0, x_0}(c), \mathcal{S}_{\mathbf{p}_0, x_0}) = O_P(\sqrt{\log(n)/r_n}).$$

Also,

$$P(\mathcal{S}_{\mathbf{p}_0, x_0} \subseteq \hat{\mathcal{S}}_{\mathbf{p}_0, x_0}(\hat{c})) \rightarrow 1 - \alpha,$$

where $\hat{c} = \hat{c}_{1-\alpha} + O_P(\log(n))$ with $\hat{c}_{1-\alpha}$ being an estimator of $(1 - \alpha)$ th quantile of $\mathcal{Q}_{\mathbf{p}_0, x_0}$ given by

$$\mathcal{Q}_{\mathbf{p}_0, x_0} := \sup_{\mathbf{q} \in \mathcal{S}_{\mathbf{p}_0, x_0}} \|Z(y) + \xi(y)\|_+^2.$$

Here, $\mathbf{q} \mapsto Z(\mathbf{q})$ is a zero-mean Gaussian process with covariance matrix

$$\Sigma(\mathbf{q}_1, \mathbf{q}_2) = \begin{pmatrix} \mathbf{q}_1' \otimes I_T & O_{T \times T} \\ O_{T \times 2T} & -I_T \end{pmatrix} \begin{pmatrix} O & O \\ O & \Omega \end{pmatrix} \begin{pmatrix} \mathbf{q}_2 \otimes I_T & O_{2T \times T} \\ O_{T \times T} & -I_T \end{pmatrix},$$

with

$$\Omega = P(I_{2T}, I_{2T}) \Psi \Sigma(x, \tau) \Psi' (I_{2T}, I_{2T})' P',$$

where Ψ is given in Eq. (25), while $\xi(\mathbf{q}) = (\xi_1(\mathbf{q}), \dots, \xi_T(\mathbf{q}))'$ is given by

$$\xi_t(\mathbf{q}) = \begin{cases} -\infty, & \mathbf{p}(t)' \mathbf{q} > \mathbf{p}(t)' \mathbf{d}(x(t), t, \tau) \\ 0, & \mathbf{p}(t)' \mathbf{q} = \mathbf{p}(t)' \mathbf{d}(x(t), t, \tau) \end{cases}, \quad t = 1, \dots, T.$$

In order to employ the above result, one either have to simulate the asymptotic distribution by Monte Carlo methods, or use resampling methods. In the latter case, one can either use the modified bootstrap procedures developed in Bugni (2009,2010) and Andrews and Soares (2010) or the subsampling procedure of Chernozhukov, Hong and Tamer (2007); we discuss in more detail in Section 8 how the bootstrap procedure of Bugni (2009,2010) can be implemented in our setting.

6 Testing for Rationality

In the previous two sections, we have developed estimators of the demand responses under revealed preferences constraints. It is of interest to test whether the consumers in the data set indeed do satisfy these restrictions: First, from an economic point of view it is highly relevant to test the axioms underlying standard choice theory. Second, from an econometric point of view, we wish to test whether the imposed constraints are actually satisfied in data.

We here develop a test for whether the consumers satisfy the revealed preference axiom; that is, are they rational? A natural way of testing this hypothesis would be to compare the unrestricted and restricted demand function estimates, and rejecting if they are "too different" from each other. Unfortunately, since we have only been able to develop the asymptotic properties of the constrained estimator under the hypothesis that none of the inequalities are binding, the unrestricted and restricted estimators are asymptotically equivalent under the null. Thus, any reasonable test comparing the two estimates would have a degenerate distribution under the null.

Instead, we take the same approach as in Blundell et al (2008) and develop a minimum-distance statistic. For given price and income levels \mathbf{p}_0 and x_0 , define the associated intersection demands,

$$\hat{\mathbf{q}}(\mathbf{p}_0, x_0, t, \tau) = \hat{\mathbf{d}}(\hat{x}(t), t, \tau) \in \mathbb{R}^2, \quad (10)$$

where $\hat{\mathbf{d}}$ is the unrestricted demand function estimator, and $\{\hat{x}(t)\}$ solves

$$\mathbf{p}'_0 \hat{\mathbf{d}}(\hat{x}(t), t, \tau) = x_0, \quad t = 1, \dots, T. \quad (11)$$

We then introduce the following statistic to measure discrepancies between a given alternative set of demands, $\underline{\mathbf{q}} = (\mathbf{q}(1), \dots, \mathbf{q}(T)) \in \mathbb{R}^{2T}$, and $\hat{\mathbf{q}}$:

$$MD_n(\underline{\mathbf{q}}|\mathbf{p}_0, x_0, \tau) = \sum_{t=1}^T (\mathbf{q}(t) - \hat{\mathbf{q}}(\mathbf{p}_0, x_0, t, \tau))' W_t (\mathbf{q}(t) - \hat{\mathbf{q}}(\mathbf{p}_0, x_0, t, \tau)),$$

where $W_t \in \mathbb{R}^{2 \times 2}$ is some weighting matrix; it could for example be chosen as an estimator of the inverse of the asymptotic covariance matrix of $\hat{\mathbf{q}}(\mathbf{p}_0, x_0, t, \tau)$ as given in Theorem 2. We then define the projection of the unrestricted demand prediction as:

$$\underline{\mathbf{q}}^* = \arg \min_{\underline{\mathbf{q}} \in \mathbb{S}_{\mathbf{p}_0, x_0}} MD_n(\underline{\mathbf{q}}|\mathbf{p}_0, x_0, \tau),$$

where $\mathbb{S}_{\mathbf{p}_0, x_0}$ is the set of intersection demands that satisfy GARP. As demonstrated in Blundell et al (2008), this can be written as:

$$\mathbb{S}_{\mathbf{p}_0, x_0} = \{\underline{\mathbf{q}} \in \mathcal{B}_{\mathbf{p}_0, x_0}^T \mid \exists V > 0, \lambda \geq 1 : V_t - V_s \geq \lambda_t \mathbf{p}(t)' (\mathbf{q}(s) - \mathbf{q}(t)), \quad t = 1, \dots, T\}, \quad (12)$$

where $\mathcal{B}_{\mathbf{p}_0, x_0}$ was defined in the previous section. The associated Wald-type statistic is given by

$$MD_n^*(\mathbf{p}_0, x_0, \tau) = MD_n(\underline{\mathbf{q}}^*|\mathbf{p}_0, x_0, \tau),$$

which will be used to test for rationality. The idea is that if indeed the consumer is rational, then $\hat{\mathbf{d}}(\hat{x}(t), t, \tau)$, $t = 1, \dots, T$, in the limit will be situated in $\mathbb{S}_{\mathbf{p}_0, x_0}$ and as such $MD_n^*(\mathbf{p}_0, x_0, \tau) \xrightarrow{P} 0$. Conversely, if the consumer is irrational, then $\lim_P MD_n^*(\mathbf{p}_0, x_0, \tau) \neq 0$. In other words, we are testing the hypothesis that

$$\underline{\mathbf{q}}_0 := \mathbf{d}(\mathbf{p}_0, x_0, t, \tau) = (\mathbf{d}(x(1), 1, \tau), \dots, \mathbf{d}(x(T), T, \tau)), \quad (13)$$

satisfies $\underline{\mathbf{q}}_0 \in \mathbb{S}_{\mathbf{p}_0, x_0}$, where $\{x(t)\}$ is given as

$$\mathbf{p}'_0 \mathbf{d}(x(t), t, \tau) = x_0, \quad t = 1, \dots, T. \quad (14)$$

The asymptotics of $MD_n^*(\mathbf{p}_0, x_0, \tau)$ under the null hypothesis of rationality are non-standard due to the fact that under the null, $\underline{\mathbf{q}}_0$ is situated on the boundary of $\mathbb{S}_{\mathbf{p}_0, x_0}$. Thus, the problem falls within the framework of Andrews (1999, 2001) who consider estimation and testing of a parameter on the boundary of the (restricted) parameter space. We employ his general results to derive the asymptotic distribution of $MD_n^*(\mathbf{p}_0, x_0, \tau)$. In the Appendix, we demonstrate that there exists mappings $V(\underline{\mathbf{q}})$ and $\lambda(\underline{\mathbf{q}})$ that takes a given demand and maps them into the corresponding utility levels and marginal utilities. Defining $B(\underline{\mathbf{q}}) = \{B_{s,t}^{(1)}(\underline{\mathbf{q}}), B_t^{(2)}(\underline{\mathbf{q}})\}_{1 \leq s, t \leq T}$ where

$$\begin{aligned} B_{s,t}^{(1)}(\underline{\mathbf{q}}) &: = V_s(\underline{\mathbf{q}}) - V_t(\underline{\mathbf{q}}) + \lambda_t(\underline{\mathbf{q}}) \mathbf{p}(t)' (\mathbf{q}(s) - \mathbf{q}(t)), \\ B_t^{(2)}(\underline{\mathbf{q}}) &= -y(t), \end{aligned}$$

for $s, t = 1, \dots, T$, we can then write the constrained set as $\mathbb{S}_{\mathbf{p}_0, x_0} = \{\underline{\mathbf{q}} | B(\underline{\mathbf{q}}) \leq 0\}$, and define the cone Λ by

$$\Lambda = \left\{ v \in \mathbb{R}^{2T} : \frac{\partial B(\underline{\mathbf{q}}_0)}{\partial \underline{\mathbf{q}}} v \leq 0 \right\}, \quad (15)$$

where $\partial B(\underline{\mathbf{q}})/\partial \underline{\mathbf{q}} = \{\partial B_{s,t}^{(1)}(\underline{\mathbf{q}})/\partial \underline{\mathbf{q}}, \partial B_t^{(2)}(\underline{\mathbf{q}})/\partial \underline{\mathbf{q}}\}_{1 \leq s, t \leq T}$ with

$$\begin{aligned} \frac{\partial B_{s,t}^{(1)}(\underline{\mathbf{q}})}{\partial \underline{\mathbf{q}}} &: = \frac{\partial V_s(\underline{\mathbf{q}})}{\partial \underline{\mathbf{q}}} - \frac{\partial V_t(\underline{\mathbf{q}})}{\partial \underline{\mathbf{q}}} + \frac{\partial \lambda_t(\underline{\mathbf{q}})}{\partial \underline{\mathbf{q}}} \mathbf{p}(t)' \Delta_{s,t}, \\ \frac{\partial B_t^{(2)}(\underline{\mathbf{q}})}{\partial \underline{\mathbf{q}}} &= -1 \text{ if } \tau = t \text{ and } = 0 \text{ for } \tau \neq t, \end{aligned}$$

and $\Delta_{s,t} = 1$ if $\tau = s$, $\Delta_{s,t} = -1$ if $\tau = t$, and $\Delta_{s,t} = 0$ otherwise.

Theorem 5 Assume that (A.1)-(A.5) hold and $W_n \xrightarrow{P} W > 0$. Then

$$r_n MD_n^*(\mathbf{p}_0, x_0, \tau) \xrightarrow{d} \inf_{\lambda \in \Lambda} (\lambda - Z(\mathbf{p}_0, x_0, \tau))' W (\lambda - Z(\mathbf{p}_0, x_0, \tau)),$$

where Λ is the convex cone given in Eq. (15) and $Z(\mathbf{p}_0, x_0, \tau) \sim N(0, \Sigma(\mathbf{p}_0, x_0, \tau))$. Here, $\Sigma(\mathbf{p}_0, x_0, \tau)$ is the asymptotic variance matrix given in Theorem 2 with $\underline{x} = (x(1), \dots, x(T))$ defined by Eq. (14).

Sometimes, it may also be of interest to draw inference regarding the constrained set of demands, \hat{y}^* . Applying Andrews (1999, Theorem 3) we obtain under the null that

$$\sqrt{r_n}(\hat{y}^* - y_0) \rightarrow^d \lambda^*(\mathbf{p}_0, x_0, \tau), \quad \lambda^*(\mathbf{p}_0, x_0, \tau) = \arg \inf_{\lambda \in \Lambda} (\lambda - Z(\mathbf{p}_0, x_0, \tau))' W(\lambda - Z(\mathbf{p}_0, x_0, \tau)).$$

The distributions of the estimator and the test statistic are non-standard. Andrews, (1999, Theorem 5) shows that the asymptotic distribution λ^* can be written as a linear projection of $Z(\mathbf{p}_0, x_0)$. Alternatively, it may be simulated.

The proposed test only examines rationality for a particular income level, x_0 , and set of prices, \mathbf{p}_0 . A stronger test should examine rationality uniformly over incomes and prices would be $\sup_{\mathbf{p}_0, x_0} MD_n^*(\mathbf{p}_0, x_0)$ where the sup is taken over some compact set. If $\hat{\mathbf{d}}(\mathbf{p}_0, x_0)$ is stochastically equicontinuous and $V(\mathbf{p}_0, x_0)$ is continuous as functions of (\mathbf{p}_0, x_0) , the asymptotic distribution of this test statistic would be

$$\sup_{\mathbf{p}_0, x_0} MD_n^*(\mathbf{p}_0, x_0, \tau) \rightarrow^d \sup_{\mathbf{p}_0, x_0} \inf_{\lambda \in \Lambda} (\lambda - Z(\mathbf{p}_0, x_0, \tau))' W(\lambda - Z(\mathbf{p}_0, x_0, \tau)),$$

c.f. Van der Vaart and Wellner (1996). Unfortunately, we have not been able to establish stochastic equicontinuity of our sieve estimator of $\mathbf{d}(\mathbf{p}_0, x_0, \tau)$.

7 Practical Implementation

In this section, we discuss in further detail how the unconstrained and constrained estimators can be implemented in the leading class of sieves on the form of Eq. (2). In the following, we again suppress the dependence on τ since this is kept fixed throughout. Let $\mathbf{d}_n \in \mathcal{D}_n^T$ be some given function in the sieve space. This function can be represented by its corresponding set of basis function coefficients, $\pi = [\pi(1)', \dots, \pi(T)']' \in \mathbb{R}^{|\mathcal{K}_n|T}$ be a given set of parameter values.

Also, choose (a large number of) M income "termination" values, $x_m^*(T)$, $m = 1, \dots, M$. The latter will be used to generate income paths. The idea is that as $M \rightarrow \infty$, we cover all possible income paths in the limit.

We then first compute M SMP paths $\{x_m^*(t)\}$, $m = 1, \dots, M$:

$$x_m^*(t) = \mathbf{p}(t)' \mathbf{d}_n(x_m^*(t+1), t+1), \quad (16)$$

where

$$d_{1,n}(x, t) = \pi(t)' B(x), \quad d_{2,n}(x, t) = \frac{x - p_1(t) d_{1,n}(x, t)}{p_2(t)}.$$

Note that $x^*(t)$ implicitly depends on π . For any of these paths, say, $\{x^*(t)\}$, we can rewrite the restriction in Eq. (16) as:

$$a(s, t, \pi) \pi(s) \leq b(s, t, \pi), \quad s < t,$$

where

$$\begin{aligned} a(s, t, \pi) &= \left\{ \frac{p_2(t)}{p_2(s)} p_1(s) - p_1(t) \right\} B(x^*(s))' \in \mathbb{R}^{|\mathcal{K}_n|}, \\ b(s, t, \pi) &= \frac{p_2(t)}{p_2(s)} x^*(s) - x^*(t) \in \mathbb{R}, \end{aligned} \quad (17)$$

for $s < t$. For the given set of M income paths, we collect these inequalities and write them on matrix form,

$$A(\pi)\pi \leq b(\pi),$$

where

$$A(\pi) = [O_{1 \times (s-1)|\mathcal{K}_n|}, a_m(s, t, \pi), O_{1 \times (T-s)|\mathcal{K}_n|}]_{m=1, \dots, M, s < t}, \quad b = [b_m(s, t, \pi)]_{m=1, \dots, M, s < t}.$$

Here, $a_m(s, t, \pi)$ and $b_m(s, t, \pi)$ denote the coefficients in Eq. (17) associated with the income path $\{x_m^*(t)\}$, $m = 1, \dots, M$. For example, with $T = 3$ and $M = 1$, we have

$$\begin{bmatrix} a(1, 2, \pi) & O_{1 \times |\mathcal{K}_n|} & O_{1 \times |\mathcal{K}_n|} \\ a(1, 3, \pi) & O_{1 \times |\mathcal{K}_n|} & O_{1 \times |\mathcal{K}_n|} \\ O_{1 \times |\mathcal{K}_n|} & a(2, 3, \pi) & O_{1 \times |\mathcal{K}_n|} \end{bmatrix} \begin{bmatrix} \pi(1) \\ \pi(2) \\ \pi(3) \end{bmatrix} \leq \begin{bmatrix} b(1, 2, \pi) \\ b(1, 3, \pi) \\ b(2, 3, \pi) \end{bmatrix}.$$

We now see that the original estimation problem is an inequality constrained quantile estimation problem:

$$\hat{\pi}_C = \arg \min_{\pi} \frac{1}{n} \sum_{t=1}^T \sum_{i=1}^n \rho_{\tau}(q_{1,i}(t) - \pi(t)' W_i(t)) \quad \text{s.t. } A(\pi)\pi \leq b(\pi). \quad (18)$$

Unfortunately, $A(\pi)$ and $b(\pi)$ both depend on π ; otherwise, the estimator would be a simple linearly constrained quantile estimator as discussed in Koenker and Ng (2005).

In some cases, it may be simpler to compute an least-squares projection estimator instead:

$$\hat{\pi}_C = \min_{\pi} \|\hat{\pi} - \pi\| \quad \text{s.t. } A(\pi)\pi \leq b(\pi), \quad (19)$$

where $\hat{\pi}$ is the unconstrained estimator of the coefficients given in Eq. (4).

8 Bootstrap Inference

We here propose to employ bootstrap procedures to compute confidence regions and critical values of the estimators and tests developed in the previous sections. Since the estimators and test statistics are non-standard in the sense that they suffer from boundary problems (binding constraints) and/or the population parameters are not point identified, standard bootstrap procedures will not be valid. Instead, we base our proposed bootstrap procedures on the ideas developed in Bugni (2009,2010) in the context of moment inequalities. Both estimation problems in Sections 5 and 6 can be expressed as a set of moment inequalities and as such fit into the framework of Bugni (2009,2010).

As an alternative to the proposed bootstrap procedure, one can employ "plug-in" methods, where nuisance parameters appearing in the relevant asymptotic distributions are estimated from the sample such that one can directly evaluate quantiles from the "estimated" asymptotic distribution (using simulations). These can then be used to obtain confidence regions for the population parameters of interest.

Below, we demonstrate that each of our estimators and test statistics fit into the framework of Bugni (2009,2010), and describe how his bootstrap procedure can be used in our setting. First, we briefly summarize his procedure: Suppose we are given a set of "moments" , $m(\theta) \in \mathbb{R}^q$, that defines the set of parameters through equality constraints,

$$\Theta_I := \{\theta \in \Theta : m(\theta) \leq 0\}.$$

This set may be a singleton such that we have point identification. We have at our disposal a sample estimator of $m(\theta)$, say $\hat{m}(\theta) \in \mathbb{R}^q$, such that $\{\sqrt{r_n}(\hat{m}(\theta) - m(\theta)) : \theta \in \Theta\}$ has a well-defined, tight weak limit. This is then used to define the estimator of Θ_I by

$$\hat{\Theta}_I(c_n) := \{\theta \in \Theta : G(\hat{m}(\theta)) \leq c_n/\sqrt{r_n}\},$$

where c_n is a slack variable satisfying $c_n/\sqrt{r_n} \rightarrow 0$ and $\sqrt{\log \log(r_n)}/c_n \rightarrow 0$, and $G(z)$ is given by either $G(z) = \sum_{i=1}^q w_i z_i$ or $G(z) = \max_{i=1, \dots, q} w_i z_i$, for some weights $w_i > 0$. Bugni (20010) then proposes the following bootstrap procedure given our estimator $\hat{\Theta}_I(c_n)$:

1. For $b = 1, \dots, B$: Draw a bootstrap sample with replacement from the data and compute the moment estimator based on the bootstrap sample, $m_b^*(\theta)$.
2. Compute

$$e_{b,i}^*(\theta) := \sqrt{r_n}(\hat{m}_{b,i}^*(\theta) - \hat{m}_i(\theta)) \times \mathbb{I}\{|\hat{m}_i(\theta)| \leq c_n/\sqrt{r_n}\}, \quad i = 1, \dots, q,$$

and

$$\Gamma_b^* := \sup_{\theta \in \hat{\Theta}_I(c_n)} G(e_b^*(\theta)).$$

The empirical $(1 - \alpha)$ quantile of $\{\Gamma_b^* : b = 1, \dots, B\}$, $\hat{c}_{1-\alpha}$, is then used to estimate the $(1 - \alpha)$ quantile of the statistic $\sup_{\theta \in \Theta_I} G(\hat{m}(\theta))$. Moreover, the $(1 - \alpha)$ confidence set of Θ_I is estimated by

$$\hat{\Theta}_I(1 - \alpha) = \{\theta \in \Theta : G(\hat{m}(\theta)) \leq \hat{c}_{1-\alpha}/\sqrt{r_n}\}.$$

Bootstrapping RP-Restricted Demand Estimates: We here wish to bootstrap the constrained version of the demand function, \mathbf{d} . We focus on the restricted least-squares estimator given in Eq. (19). Let $\pi(\mathbf{d})$ be the set of coefficients corresponding to a given demand function \mathbf{d} situated in the function space \mathcal{D} . Define

$$\hat{m}_1(\mathbf{d}) = \mathbf{d} - \hat{\mathbf{d}}, \quad \hat{m}_2(\mathbf{d}) = \hat{\mathbf{d}} - \mathbf{d}, \quad \hat{m}_3(\mathbf{d}) = A(\pi(\mathbf{d}))\pi(\mathbf{d}) - b(\pi(\mathbf{d})).$$

We now have that $\mathcal{D}_I = \{\mathbf{d}_0\}$ and $\hat{\mathcal{D}}_I(\epsilon) = \{\hat{\mathbf{d}}_\epsilon\}$, where $\hat{\mathcal{D}}_I(c_n) = \{\pi : \hat{m}(\mathbf{d}) \leq c_n/\sqrt{r_n}\}$. We then propose to use the following bootstrap procedure to obtain a confidence region for the RP-constrained demand functions:

1. For $b = 1, \dots, B$: Draw a bootstrap sample with replacement from the data,

$$\{Z_{t,i}^*(b) : i = 1, \dots, n, t = 1, \dots, T\}.$$

Compute the unrestricted demand function estimator based on the bootstrap sample, $\hat{\mathbf{d}}^*$,

2. With $\hat{m}_1^*(\mathbf{d}) = \mathbf{d} - \hat{\mathbf{d}}^*$ and $\hat{m}_2^*(\mathbf{d}) = A(\pi(\mathbf{d}))\pi(\mathbf{d}) - b(\pi(\mathbf{d}))$, compute

$$e_{b,i}^*(d) := \sqrt{r_n}(\hat{m}_i^*(\mathbf{d}) - \hat{m}_i(\mathbf{d})) \times \mathbb{I}\{|\hat{m}_i(\mathbf{d})| \leq c_n/\sqrt{r_n}\}, \quad , t = 1, \dots, T,$$

and

$$\Gamma_b^* := \sup_{d \in \hat{\mathcal{D}}_I(c_n)} G(e_b^*(d)).$$

We then compute the confidence region of the demand function by

$$\hat{\mathcal{D}}(1 - \alpha) = \{d \in \mathcal{D}_n : G(\hat{m}(\mathbf{d})) \leq \hat{c}_{1-\alpha}/\sqrt{r_n}\}.$$

Bootstrapping Demand Bounds: Define

$$m(\mathbf{q}) = \mathbf{P}\mathbf{q} - a \in \mathbb{R}^T, \quad \hat{m}(y) = \mathbf{P}\mathbf{q} - \hat{a} \in \mathbb{R}^T,$$

where \mathbf{P} is given in Assumption C.2, while a and \hat{a} are given as

$$a(t) := \mathbf{p}(t)' \mathbf{d}(x(t), t), \quad \hat{a}(t) := \mathbf{p}(t)' \hat{\mathbf{d}}(\hat{x}(t), t), \quad t = 1, \dots, T.$$

Finally, note that \mathcal{B} is the "parameter space." We now have that $\mathcal{S} = \Theta_I$ and $\hat{\mathcal{S}}(c) = \hat{\Theta}_I$, and we use the following bootstrap procedure to obtain a confidence region for the demand bounds:

1. For $b = 1, \dots, B$: Draw a bootstrap sample with replacement from the data,

$$\{Z_{t,i}^*(b) : i = 1, \dots, n, t = 1, \dots, T\}.$$

Compute the demand function estimator based on the bootstrap sample, $\hat{\mathbf{d}}_b^*(x(t), t)$, and the income path associated with (\mathbf{p}_0, x_0) , $\{\hat{x}_b^*(t)\}$,

$$\mathbf{p}_0' \hat{\mathbf{d}}_b^*(\hat{x}_b^*(t), t) = x_0, \quad t = 1, \dots, T.$$

2. With $\hat{m}_{b,t}^*(\mathbf{q}) = \mathbf{p}(t)' \mathbf{q} - \mathbf{p}(t)' \hat{\mathbf{d}}_b^*(\hat{x}_b^*(t), t)$, compute

$$e_{b,t}^*(\mathbf{q}) := \sqrt{r_n}(\hat{m}_{b,t}^*(\mathbf{q}) - \hat{m}_t(\mathbf{q})) \times \mathbb{I}\{|\hat{m}_t(\mathbf{q})| \leq c_n/\sqrt{r_n}\}, \quad , t = 1, \dots, T,$$

and

$$\Gamma_b^* := \sup_{\mathbf{q} \in \hat{\mathcal{S}}(c_n)} G(e_b^*(\mathbf{q})).$$

We then compute the confidence region of the bounds by

$$\hat{\mathcal{S}}(1 - \alpha) = \{\mathbf{q} \in \mathcal{B} : G(\hat{m}(\mathbf{q})) \leq \hat{c}_{1-\alpha}/\sqrt{r_n}\}.$$

Bootstrapping RP Test: To translate the "moment" equalities used to test for rationality into moment inequalities, we define $\hat{m}(\mathbf{q}) = (\hat{m}_1(\mathbf{q}), \hat{m}_2(\mathbf{q}), \hat{m}_3(\mathbf{q})) \in \mathbb{R}^{4T}$, where $\hat{m}_1(\mathbf{q}) = \mathbf{q} - \hat{\mathbf{d}}(\mathbf{p}_0, x_0) \in \mathbb{R}^{2T}$, $\hat{m}_2(\mathbf{q}) = \hat{\mathbf{d}}(\mathbf{p}_0, x_0) - \mathbf{q} \in \mathbb{R}^{2T}$, and $\hat{m}_3(\mathbf{q}) = B(\mathbf{q})$, where $B(\mathbf{q})$ is defined in Section 7. Furthermore, we restrict the weighting matrix in the test statistic, W_n , to be diagonal, $W_n = \text{diag}\{w_t\}$. We then introduce the following set of RP-restricted demands:

$$\hat{\mathbb{S}}_{\text{RP}}(c_n) := \{\mathbf{q} \in \mathbb{S} : G(\hat{m}(\mathbf{q})) \leq c_n/\sqrt{r_n}\},$$

and note that $\hat{\mathbb{S}}_{\text{RP}}(0) = \{\hat{\mathbf{q}}^*\}$ while $MD_n(\mathbf{q}|\mathbf{p}_0, x_0) := \sup_{y \in \hat{\mathbb{S}}_{\text{RP}}(0)} G(\hat{m}(\mathbf{q}))$. Our bootstrap procedure now proceeds as follows:

1. For $b = 1, \dots, B$: Draw a bootstrap sample with replacement from the data,

$$\{Z_{t,i}^*(b) : i = 1, \dots, n, t = 1, \dots, T\}.$$

Compute the demand function estimator based on the bootstrap sample,

$$\hat{\mathbf{d}}_b^*(\mathbf{p}_0, x_0) = (\hat{\mathbf{d}}_b^*(\hat{x}_b^*(1), 1), \dots, \hat{\mathbf{d}}_b^*(\hat{x}_b^*(T), T)) \in \mathbb{R}^{2T}$$

where $\hat{\mathbf{d}}_b^*$ is the unrestricted demand function estimator, and $\{\hat{x}_b^*(t)\}$ solves

$$\mathbf{p}'_0 \hat{\mathbf{d}}_b^*(\hat{x}_b^*(t), t) = x_0, \quad t = 1, \dots, T.$$

2. With $\hat{m}_b^*(\mathbf{q}) = (\hat{m}_{b,1}^*(\mathbf{q}), \hat{m}_{b,2}^*(\mathbf{q}))$, where $\hat{m}_{b,1}^*(\mathbf{q}) = \mathbf{q} - \hat{\mathbf{d}}_b^*(\mathbf{p}_0, x_0)$, $\hat{m}_{b,2}^*(\mathbf{q}) = \hat{\mathbf{d}}_b^*(\mathbf{p}_0, x_0) - \mathbf{q}$ and $\hat{m}_{b,3}^*(\mathbf{q}) = B(\mathbf{q})$, compute

$$e_{b,t}^*(\theta) := \sqrt{r_n}(\hat{m}_{b,t}^*(\theta) - \hat{m}_t(\theta)) \times \mathbb{I}\{|\hat{m}_t(\theta)| \leq c_n/\sqrt{r_n}\}, \quad t = 1, \dots, 2T,$$

and

$$\Gamma_b^* := \sup_{\mathbf{q} \in \hat{\mathbb{S}}_{\text{RP}}(c_n)} G(e_b^*(\mathbf{q})).$$

The $(1 - \alpha)$ critical value of $MD_n(\mathbf{q}|\mathbf{p}_0, x_0)$ is now estimated by the $(1 - \alpha)$ quantile of $\{\Gamma_b^* : b = 1, \dots, B\}$, $\hat{c}_{1-\alpha}$.

We unfortunately have no theoretical justification for the above two bootstrap procedure when the demand function estimators are nonparametric. The theoretical results in Bugni (20010) are restricted to parametric estimation problems; in particular, the moments are based on i.i.d. averages that converge with \sqrt{n} -rate and are unbiased. In contrast, our "moments" are based on nonparametric estimators that converge with slower rates and are biased (in finite samples). As

such, we can unfortunately not directly appeal to his theoretical results showing the validity of the above bootstrap procedures. It is outside the scope of this paper to demonstrate that the bootstrap procedures proposed here are in fact valid. It should be noted though that if we treat the sieve estimator as a parametric estimator (keeping $|\mathcal{K}_n|$ fixed), then Bugni’s results should apply.

The finite-sample biases of nonparametric estimators can lead to inferior performance of bootstrap procedures if the same smoothing parameter is used for the estimator based on the actual sample, and those based on the bootstrap samples, see Hall (1992, Section 4.5). As in Blundell et al (2007), we handle this issue by computing the demand function estimators in Step 2 of the bootstrap procedures with $|\mathcal{K}_n|$ (and/or the penalization term) chosen slightly larger (smaller) than for the estimator based on the actual sample.

9 Endogenous Income

Suppose that income, $x(t)$, is endogenous such that the independence assumption made in (A.2) fails. The proposed sieve quantile estimator will in this case be inconsistent. One can instead then employ the IV quantile estimators developed in Chen and Pouzo (2009) and Chernozhukov, Imbens and Newey (2007). Alternatively, the control function approach taken in Imbens and Newey (2009) can be used.

With the assumptions and results of either of these three papers replacing our assumptions (A.1)-(A.3) and our Theorem 1, the remaining results of ours as stated in Theorems 2-5 remain valid since these follow from the properties of the unconstrained estimator. Thus, all the results stated in Theorems 2-5 go through except that the convergence rates and asymptotic distributions have to be modified to adjust for the use of another unrestricted estimator.

10 Empirical Application (to be completed)

In our application we apply the methodology for constructing demand bounds under revealed preference restrictions to data from the U.K. Family Expenditure Survey. The data set contains expenditure data and prices from UK households. We use the same sample selection as in Blundell et al (2008) and we refer to that paper for a more detailed description. The distribution of relative prices over the central period of the data is give in Figure 1. It shows periods of quite dense relative prices, in the 1980s for example, and periods of sparse relative prices as in the 1970s.

We choose food as our primary good, and then group the other goods together in this application. The basic distribution of the Engel curve data are described in Figures 2 and 3.

Figure 1: Relative prices in the UK FES: 1975 to 1999

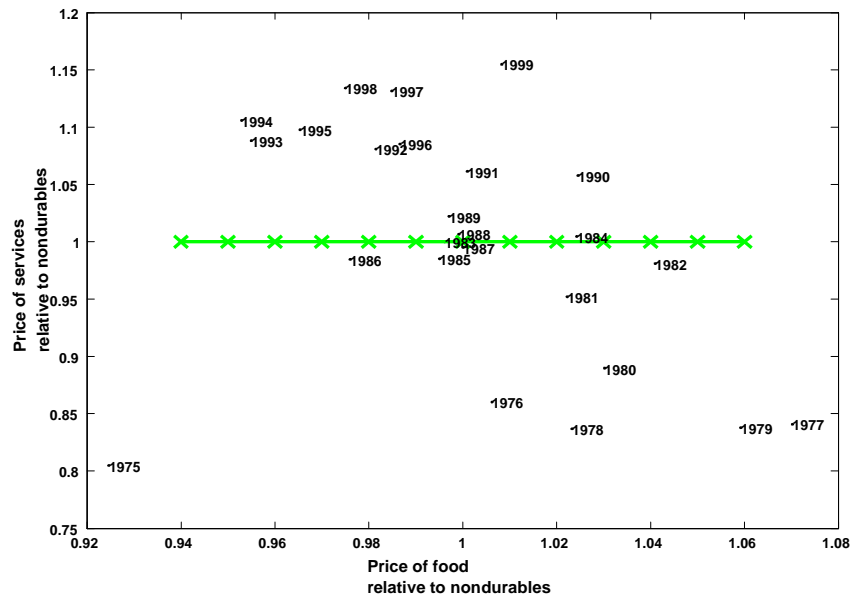


Figure 2: The Engel Curve Distribution

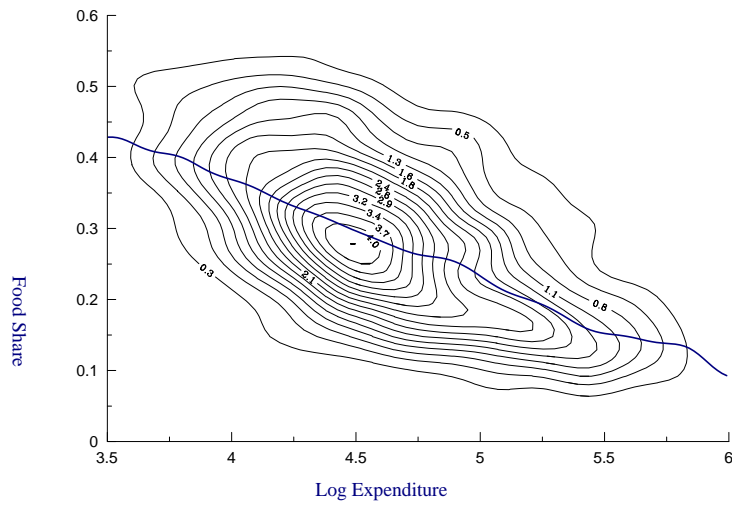
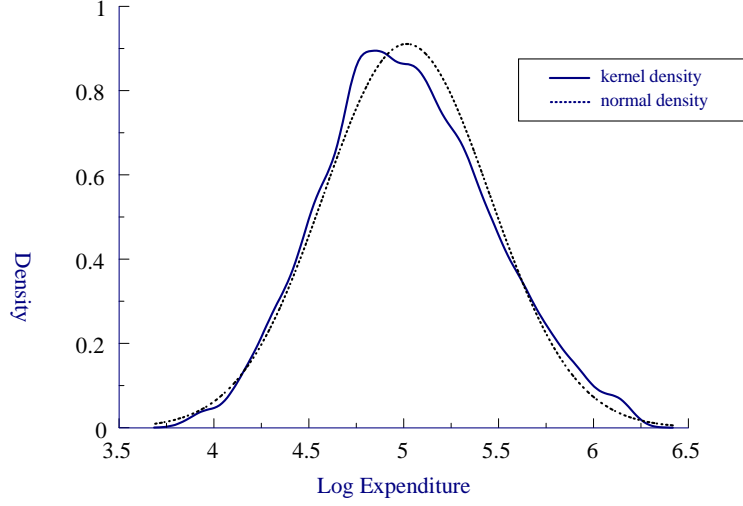


Figure 3: The Density of Log Consumption: UK FES



10.1 The Quantile Sieve Estimates of Expansion Paths

In the estimation, we use polynomial splines

$$d_{1,n}(x, t, \tau) = \pi(t, \tau)' B_{K_n}(y)' = \sum_{j=0}^{q_n} \pi_j(t, \tau) x^j + \sum_{k=1}^{r_n} \pi_{q_n+k}(t, \tau) (x - \nu_k(t))_+^{q_n}, \quad K_n = q_n + r_n + 1, \quad (20)$$

where $q_n \geq 1$ is the order of the polynomial and ν_k , $k = 1, \dots, r_n$, are the knots. For a given choice of r_n , we place the knots according to the sample quantiles of $x_i(t)$, $i = 1, \dots, n$, i.e., $\nu_k(t)$ was chosen as the estimated $k/(r_n + 1)$ -th quantile of $x(t)$.

In the implementation of the quantile sieve estimator, a small penalization term was added to the objective function to robustify the estimators (see Blundell, Chen and Kristensen, 2007 for a similar approach). That is,

$$\hat{\pi}(t, \tau) = \arg \min_{\pi \in \mathbb{R}^{|K_n|}} \frac{1}{n} \sum_{i=1}^n \rho_{\tau}(q_{1,i}(t) - \pi' W_i(t)) + \lambda Q(\pi), \quad \tau \in [0, 1], \quad (21)$$

where

$$W_i(t) = (1, x_i(t), \dots, x_i(t)^q, (x - \nu_1(t))_+^{q_n}, \dots, (x - \nu_{r_n}(t))_+^{q_n})',$$

and $\lambda Q(\pi)$ is an L_1 -penalty term. Here, $Q(\pi)$ is the total variation of $\partial d_{n,1}(x)/(\partial x)$,

$$Q(\pi) = \int_a^b \left| \pi' \frac{\partial^2 B(x)}{\partial x^2} \right| dx \in \mathbb{R}_+,$$

while $\lambda > 0$ is the penalization weight that controls the smoothness of the resulting estimator.

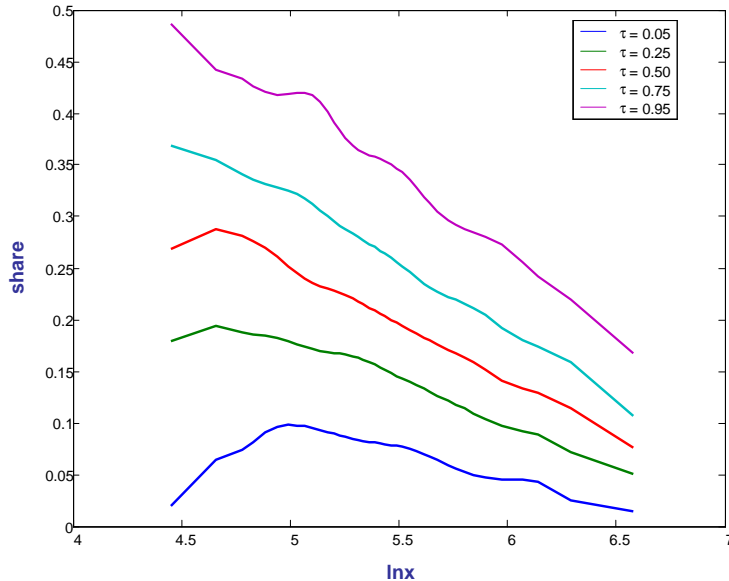
By following the arguments of Koenker, Ng and Portnoy (1994), the above estimation problem can be formulated as a linear programming problem. The computation of the unrestricted estimator was done using Matlab code kindly provided by Roger Koenker.

The restricted estimator is computed by solving the least-squares problem in Eq. (19). We opted for this estimator instead of the quantile estimator proposed in Eq. (18) since numerically we found it easier to solve the constrained least-squares problem.

In our application, we focus on the six year period 1983-1988. As in Blundell et al (2008) we use a group of demographically homogeneous households and estimate conditional quantile spline expansion paths using a 3rd order polynomial ($q_n = 3$) with $r_n = 5$ knots. The RP restrictions are imposed at 100 x -points over the empirical support x (log total expenditure on non-durables and services).

Each household is defined by a point in the distribution of log income and unobserved heterogeneity (x, ε) . For households at the median of the income (total expenditure) distribution the unrestricted τ -quantile expansion paths for food share are given in Figure 4.

Figure 4: τ -Quantile Expansion Paths for Food Shares



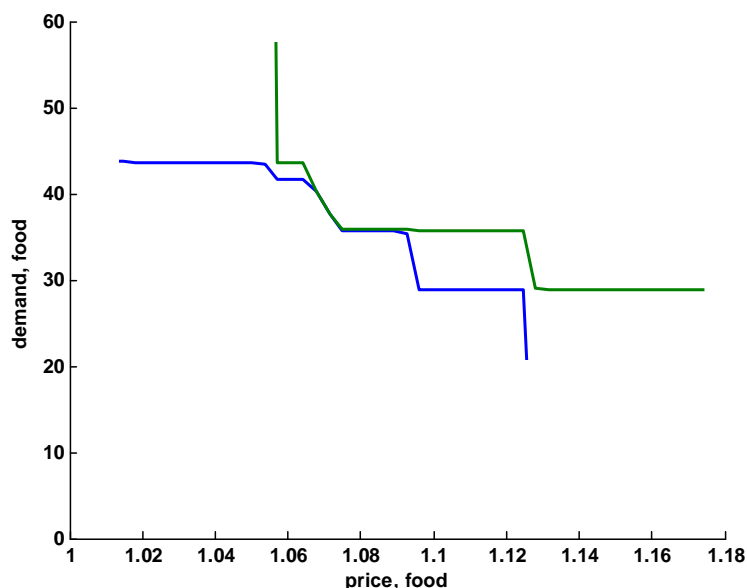
10.2 Estimated Demand Bounds

The key parameter of interest in this study is the consumer response at some new relative price \mathbf{p}_0 and income x or at some sequence of relative prices. The latter defines the demand curve for (x, ε) . The estimated (e-)bounds (support sets) using the revealed preference inequalities and our FES data are given in Figures 5-8.

The Figures display the bounds on demand responses across the two dimensions of individual

heterogeneity - income and unobserved heterogeneity. For a given income we can look at demand bounds for consumers with stronger or weaker preferences for food. Figure 5 shows the bounds on demands at the median income for the 50th percentile ($\tau = .5$) of the unobserved taste distribution. Notice that where the relative prices are quite dense the bounds are correspondingly narrow. Figure 6 contrasts this for a consumer at the 25% percentile of the heterogeneity distribution - a consumer with much weaker taste for food. At all points demands are much lower and the price response is somewhat less steep. Figure 7 considers a consumer with a strong taste for food - at the 75th percentile of the taste distribution. Demand shifts up at all points. The bounds remain quite narrow where the relative prices are dense. Finally, to illustrate the power of this approach, Figure 8 considers a higher income consumer but with median taste for food.

Figure 5: Quantile (RP-Rest) e-Bounds on Demand (Median InC) $\tau = .50$



11 Conclusion

This paper has developed a nonparametric estimator of revealed-preference restricted demand functions and their associated demand bounds for the case of nonseparable heterogeneity. The asymptotic properties of the estimators were derived and the implementation discussed. A test for rationality was proposed and its asymptotic properties analyzed.

Figure 6: Quantile (RP-Rest) e-Bounds on Demand (Median Inc) $\tau = .25$

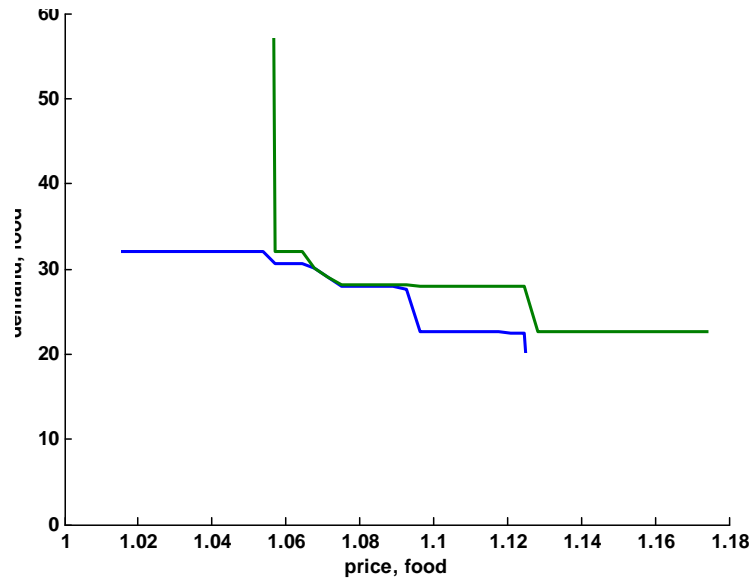


Figure 7: Quantile (RP-Rest) e-Bounds on Demand (Median Inc) $\tau = .75$

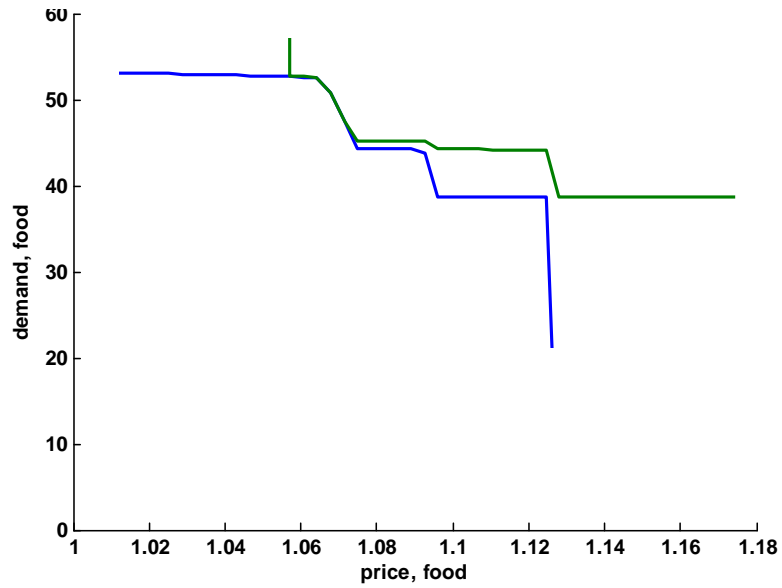
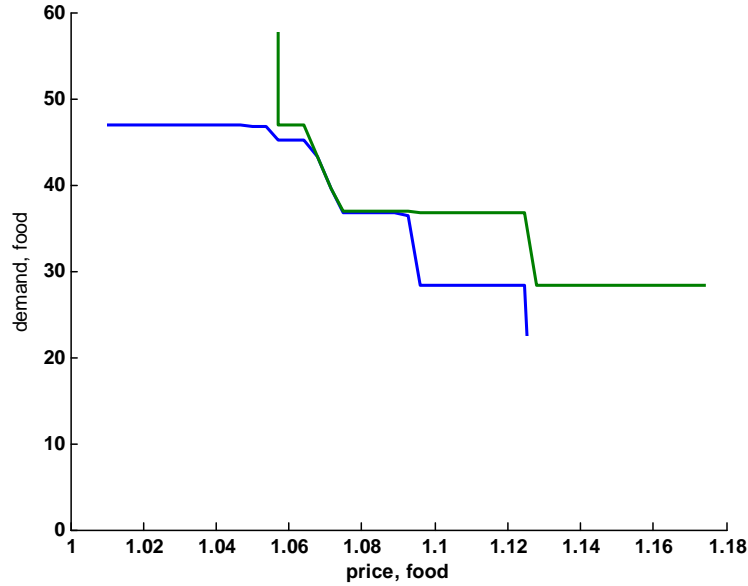


Figure 8: Quantile (RP-Rest) e-Bounds on Demand (75% Inc)



References

- Afriat, S.N. (1973): "On a System of Inequalities in Demand Analysis: An Extension of the Classical Method," *International Economic Review* 14, 460-472.
- Anevski, D. and O. Hössjer (2006): "A General Asymptotic Scheme for Inference Under Order Restrictions," *Annals of Statistics* 34, 1874-1930.
- Andrews, D.W.K. (1997): "Estimation When a Parameter Is on a Boundary: Theory and Applications," Cowles Foundation Discussion Paper No. 1153.
- Andrews, D.W.K. (1999): "Estimation When a Parameter Is on a Boundary," *Econometrica* 67, 1341-1383.
- Andrews, D.W.K. (2001): "Testing When a Parameter Is on the Boundary of the Maintained Hypothesis," *Econometrica* 69, 683-734.
- Andrews, D.W.K. and P. Guggenberger (2009): "Validity of Subsampling and 'Plug-in Asymptotic' Inference for Parameters Defined by Moment Inequalities," *Econometric Theory* 25, 669-709.
- Andrews, D.W.K. and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," forthcoming in *Econometrica*.
- W Beckert (2007): "Specification and Identification of Stochastic Demand Models," *Econometric Reviews* 26, 669-683.

- Beckert, W. and R. Blundell (2008): "Heterogeneity and the nonparametric analysis of consumer choice: conditions for invertibility," *Review of Economic Studies* 75, 1069-1080.
- Beresteanu, A. (2006): "Nonparametric Estimation of Regression Functions under Restrictions on Partial Derivatives," working paper, Duke University.
- Birke, M. and H. Dette (2007): "Estimating a Convex Function in Nonparametric Regression," *Scandinavian Journal of Statistics* 34, 384-404.
- Blundell, R., M. Browning and I. Crawford (2003): "Nonparametric Engel Curves and Revealed Preference," *Econometrica* 71, 205-240.
- Blundell, R., M. Browning and I. Crawford (2008): "Best Nonparametric Bounds on Demand Responses," *Econometrica* 76, 1227-1262.
- Blundell, R., X. Chen and D. Kristensen (2007): "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica* 75, 1613-1669.
- Brown, B.W. (1983): "The Identification Problem in Systems Nonlinear in the Variables," *Econometrica*, 51, 175-196.
- Bugni, F. (2009): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Points of the Identified Set," manuscript, Department of Economics, Duke University.
- Bugni, F. (2010): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," forthcoming in *Econometrica*.
- Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, Vol. 6B (eds. J. Heckman and E. Leamer), 5549-5632. Amsterdam: North-Holland.
- Chen, X. and D. Pouzo (2007): "Estimation of Nonparametric Conditional Moment Models with Possibly Nonsmooth Moments," Cowles Foundation Discussion Papers, No. 1650.
- Chernozhukov, V., I. Fernandez-Val and A. Galichon (2006): "Improving Point and Interval Estimators of Monotone Functions by Rearrangement," *Biometrika* 96, 559-575.
- Chernozhukov, V., G. Imbens and W.K. Newey (2007): "Instrumental Variable Estimation of Nonseparable Models," *Journal of Econometrics* 139, Pages 4-14
- Chernozhukov, V., H. Hong and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica* 75, 1243 - 1284.
- Chesher, A. (2003): "Identification in Nonseparable Models," *Econometrica* 71, 1405-1441.

- Gallant, A. R. & Golub, G. H. (1984): "Imposing Curvature Restrictions on Flexible Functional Forms," *Journal of Econometrics* 26, 295-321.
- Haag, B.R., S. Hoderlein and K. Pendakur (2009): "Testing and imposing Slutsky symmetry in nonparametric demand systems," *Journal of Econometrics* 153, 33-50.
- Hall, P. (1992): *The Bootstrap and Edgeworth Expansion*. New York: Springer Verlag.
- Hall, P. and J. Horowitz (2005): "Nonparametric Methods for Inference in the Presence of Instrumental Variables," *Annals of Statistics* 33, 2904-2929.
- Hausman, J.A., W.K. Newey & J.L. Powell (1995): "Nonlinear Errors in Variables: Estimation of Some Engel Curves," *Journal of Econometrics* 65, 205-233.
- Hoderlein, S. (2008): "How Many Consumers are Rational?" Manuscript, Dep. of Economics, Brown University.
- Hoderlein, S. and J. Stoye (2009) "Revealed Preferences in a Heterogeneous Population," Manuscript, Dep. of Economics, Brown University.
- Imbens, G. and W.K. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity," *Econometrica* 77, 1481-1512.
- Kiefer, J. (1982): "Optimum Rates for Non-parametric Density and Regression Estimates under Order Restrictions," in *Statistics and Probability: Essays in Honor of C. R. Rao* (eds. G. Kallianpur, P.R. Krishnaiah & J.K. Ghosh), 419-428. Amsterdam: North-Holland.
- Kim, M.-O. (2006) Quantile Regression with Shape-Constrained Varying Coefficients. *Sankhyā* 68, 369-391.
- Klepper, S. and E.E. Learner (1984): "Consistent Sets of Estimates for Regressions with Errors in All Variables," *Econometrica* 52, 163-184.
- Koenker, R. and G. Bassett, Jr. (1978): "Regression Quantiles," *Econometrica* 46, 33-50.
- Koenker, R. and P. Ng (2005): "Inequality Constrained Quantile Regression," *Sankhyā* 67, 418-440.
- Koenker, R., P. Ng and S. Portnoy (1994): "Quantile smoothing splines," *Biometrika* 81, 673-680.
- Leser, C. E. V. (1963) Forms of Engel Functions. *Econometrica*, 31, 694-703.
- Lewbel, A. (1991): "The Rank of Demand Systems: Theory and Nonparametric Estimation," *Econometrica*, 59, 711-730.
- Lewbel, A. (1995): "Consistent nonparametric hypothesis tests with an application to Slutsky symmetry," *Journal of Econometrics* 67, 379-401.

- Lewbel, A. (2001): "Demand Systems With and Without Errors," *American Economic Review* 91, 611-618.
- Lewbel, A. and K. Pendakur (2009): "Tricks With Hicks: The EASI Demand System," *American Economic Review* 99, 827-863.
- Mammen, E. (1991): "Estimating a Smooth Monotone Regression Function," *Annals of Statistics* 19, 724-740.
- Mammen, E., J.S. Marron, B.A. Turlach and M.P. Wand (2001): "A General Projection Framework for Constrained Smoothing," *Statistical Science* 16, 232-248.
- Mammen, E. and C. Thomas-Agnan (1999): "Smoothing Splines and Shape Restrictions," *Scandinavian Journal of Statistics* 26, 239-252.
- Matzkin, R.L. (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 1339-1375.
- Matzkin, R.L. (2005): "Identification in Nonparametric Simultaneous Equations," *Econometrica* 76, 945 - 978
- Mukerjee, H. (1988): "Monotone Nonparametric Regression," *Annals of Statistics* 16, 741-750.
- Newey, W.K. & J. Powell (2003): "Instrumental Variables Estimation for Nonparametric Models," *Econometrica*, 71, 1565-1578.
- Roehrig, C.S. (1988) Conditions for Identification in Nonparametric and Parametric Models. *Econometrica*, 56, 433-447.
- Van de Geer, S. (2000): *Empirical Processes in M-estimation*. Cambridge University Press.
- Van der Vaart, A. and J. Wellner (1996): *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer-Verlag, New York.
- Working, H. (1943): "Statistical Laws of Family Expenditure," *Journal of the American Statistical Association*, 38, 43-56.
- Wright, F.T. (1981): "The Asymptotic Behavior of Monotone Regression Estimates," *Annals of Statistics* 9, 443-448.
- Wright, F.T. (1984): "The Asymptotic Behavior of Monotone Percentile Regression Estimates," *Canadian Journal of Statistics* 12, 229-236.
- Yatchew, A. and Bos, L. (1997): "Nonparametric Regression and Testing in Economic Models," *Journal of Quantitative Economics* 13, 81-131.

A Proofs of Section 2 and 3

Proof of Theorem 1. We suppress the dependence on t since it is kept fixed in the following. Write the first demand equation as a quantile regression,

$$q_1 = d_1(x, \tau) + e(\tau), \quad (22)$$

where e is defined as the generalized residual, $e(\tau) := d_1(x, \varepsilon) - d_1(x, \tau)$. This formulation of the model corresponds to the quantile regression considered in Chen (2007, Section 3.2.2). We then wish to verify the conditions stated there. First, we show that the distribution of $e(\tau) | x$ is described by a density $f_{t, \tau}(e|x)$ that satisfies

$$0 < \inf_{x \in \mathcal{X}} f_{\tau}(e|x) \leq \sup_{x \in \mathcal{X}} f_{\tau}(e|x) < \infty, \quad (23)$$

$$\sup_{x \in \mathcal{X}} |f_{\tau}(e|x) - f_{\tau}(0|x)| \rightarrow 0, \quad |e| \rightarrow 0. \quad (24)$$

To see this, we first note that due to the independence between x and ε , the invertibility and differentiability of $\varepsilon \mapsto d_1(x, \varepsilon)$, and that $\varepsilon \sim U[0, 1]$,

$$f_{\tau}(e|x) = \mathbb{I}\{0 \leq d_1^{-1}(x, e - d_1(x, \tau)) \leq 1\} \left| \frac{\partial d_1^{-1}(x, e - d_1(x, \tau))}{\partial e} \right|.$$

>From this expression it is easily seen that Eq. (23) holds since $d_1(x, t, \varepsilon)$ and its derivative w.r.t. ε are continuous in x and \mathcal{X} is compact. Eq. (24) clearly holds pointwise due to the continuity of $\varepsilon \mapsto d_1(x, \varepsilon)$. This can be extended to uniform convergence since $\sup_{x \in \mathcal{X}, e \in [0, 1]} f_{\tau}(e|x) < \infty$.

Given the above results, we can conclude from Chen (2007, Theorem 3.2) that the first part of our theorem holds if the Conditions 3.6-3.8 in Chen (2007) are satisfied. However, in the Proof of Chen (2007, Proposition 3.4), it is demonstrated that these conditions are satisfied under Eqs. (23)-(24). Next, applying the results of Chen and Shen (1998, p. 311), we obtain the desired result under either Assumption (A.3.1) or (A.3.2). ■

Proof of Theorem 2. [TBC] Kim (2006) shows the result for quantile spline estimators in varying-coefficient models. We proceed as there... ■

Proof of Theorem 3. Let $x_0(t)$ be a given income expansion path generated from \mathbf{d} . We first note that the expansion path based on the unconstrained demand function satisfies

$$\hat{x}(T-1) := \mathbf{p}(T-1)' \hat{\mathbf{d}}(x(T), t, \tau) = x_0(T-1) + O_P\left(n^{-m/(2m+1)}\right).$$

By recursion, we easily extend this to $\max_{t=1, \dots, T} |\hat{x}(t) - x_0(t)| = O_P\left(n^{-m/(2m+1)}\right)$. Thus,

$$\begin{aligned} \hat{x}(t) - \mathbf{p}(t)' \hat{\mathbf{d}}(\hat{x}(s), s, \tau) &= \{\hat{x}(t) - x_0(t)\} + \mathbf{p}(t)' \left\{ \mathbf{d}_0(x_0(s), s, \tau) - \hat{\mathbf{d}}(\hat{x}(s), s, \tau) \right\} \\ &\quad + x_0(t) - \mathbf{p}(t)' \mathbf{d}_0(x_0(s), s, \tau) \\ &= O_P\left(n^{-m/(2m+1)}\right). \end{aligned}$$

Thus, as $n \rightarrow \infty$, $\hat{x}(t) - \mathbf{p}(t)' \hat{\mathbf{d}}(\hat{x}(s), s, \tau) \leq \epsilon$ with probability approaching one (w.p.a.1) as $n \rightarrow \infty$. This proves that $\hat{\mathbf{d}} \in \mathcal{D}_{C,n}^T(\epsilon)$ w.p.a.1 such that $\hat{\mathbf{d}}_C^\epsilon = \hat{\mathbf{d}}$ w.p.a.1 as $n \rightarrow \infty$. ■

B Proof of Theorem 4

We here prove a more general version of Theorem 4 since we believe this has independent interest. In particular, the general result takes as input any set of demand function estimators and derive the asymptotic properties of the corresponding bounds. The result is stated in such a fashion that it allows for both fully parametric, semi- and nonparametric first-step estimators. Let in the following $\hat{\mathbf{d}}(x, t) = (\hat{d}_1(x, t), \dots, \hat{d}_L(x, t))'$ be any given set of demand function estimators for $L \geq 2$ goods. We then assume that the following conditions are met:

C.1 $x(t) \mapsto \mathbf{d}_0(x(t), t)$ is monotonically increasing.

C.2 The matrix $\mathbf{P} = [\mathbf{p}(1), \dots, \mathbf{p}(T)]' \in \mathbb{R}_+^{T \times L}$ has rank L .

C.3 For some sequence $r_n \rightarrow \infty$, and for any vector $\underline{x} = (x(1), \dots, x(T)) \in \mathbb{R}_+^T$:

$$(i) \max_{t=1, \dots, T} \left\| \hat{\mathbf{d}}(x(t), t) - \mathbf{d}_0(x(t), t) \right\| = O_P(1/\sqrt{r_n}).$$

(ii)

$$\sqrt{r_n} \begin{pmatrix} \hat{\mathbf{d}}(x(1), 1) - \mathbf{d}_0((1), 1) \\ \vdots \\ \hat{\mathbf{d}}(x(T), T) - \mathbf{d}_0(x(T), T) \end{pmatrix} \rightarrow^d N(0, \Sigma(\underline{x})),$$

where $\Sigma(\underline{x}) \in \mathbb{R}^{LT \times LT}$ is the joint asymptotic variance of the sequence of demand function estimators.

We then show the following result which includes our sieve estimator as a special case:

Theorem 6 *Assume that (C.1)-(C.3) hold. Then the conclusions of Theorem 4 hold for the general demand estimator in (C.3).*

To show this result, we first define b and \hat{b} by

$$b_t = \mathbf{p}(t)' \mathbf{d}_0(x(t), t), \quad \hat{b}_t = \mathbf{p}(t)' \hat{\mathbf{d}}(\hat{x}(t), t), \quad t = 1, \dots, T,$$

where

$$\mathbf{p}_0' \mathbf{d}_0(t, x(t)) = \mathbf{p}_0' \hat{\mathbf{d}}_0(t, \hat{x}(t)) = x_0, \quad t = 1, \dots, T,$$

while $\hat{A} = \mathbf{P}$ where \mathbf{P} is given in Assumption C.2. The support set now takes the form considered in Appendix C, and we verify Assumptions 1–2 and 4 of Theorem 10 stated there. It is easily seen that Assumptions 1–2 hold, so we only have to verify Assumptions 4: We have

$$\sqrt{r_n}(\hat{b} - b) = \bar{\mathbf{P}} \sqrt{r_n}(\hat{D} - D),$$

where $\hat{D} = (\hat{\mathbf{d}}(1, \hat{x}(1))', \dots, \hat{\mathbf{d}}(T, \hat{x}(T))')' \in \mathbb{R}^{LT}$, $D = (\mathbf{d}_0(1, x(1))', \dots, \mathbf{d}_0(T, x(T))')' \in \mathbb{R}^{LT}$ and

$$\bar{\mathbf{P}} = \begin{bmatrix} \mathbf{p}(1)' & 0 & \dots & 0 \\ 0 & \mathbf{p}(2)' & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \mathbf{p}(T)' \end{bmatrix} \in \mathbb{R}^{T \times LT}.$$

We write

$$\hat{D} - D = \left\{ \hat{D} - \tilde{D} \right\} + \left(\tilde{D} - D \right),$$

where $\tilde{D} = (\hat{\mathbf{d}}(1, x(1)), \dots, \hat{\mathbf{d}}(T, x(T)))'$. We then wish to derive the joint asymptotic distribution of $(\hat{D} - \tilde{D}, \tilde{D} - D)$. We first derive the asymptotics of $\hat{x}(t)$ which is characterized by

$$\begin{aligned} 0 &= \mathbf{p}'_0 \hat{\mathbf{d}}(t, \hat{x}(t)) - x_0 \\ &= \mathbf{p}'_0 \hat{\mathbf{d}}(t, x(t)) - x_0 + \mathbf{p}'_0 \left. \frac{\partial \hat{\mathbf{d}}(t, x)}{\partial x} \right|_{x=\hat{x}(t)} (\hat{x}(t) - x(t)) \\ &= \mathbf{p}'_0 \hat{\mathbf{d}}(t, x(t)) - x_0 + \mathbf{p}'_0 \left\{ \left. \frac{\partial \mathbf{d}(t, x)}{\partial x} \right|_{x=\hat{x}(t)} + o_P(1) \right\} (\hat{x}(t) - x(t)). \end{aligned}$$

Thus, $\underline{\hat{x}} = (\hat{x}(1), \dots, \hat{x}(T))'$ satisfies

$$\sqrt{r_n}(\underline{\hat{x}} - \underline{x}) = Q_0^{-1} \bar{\mathbf{P}}_0 \sqrt{r_n}(\tilde{D} - D) + o_P(1),$$

where

$$\bar{\mathbf{P}}_0 = \begin{bmatrix} \mathbf{p}'_0 & 0 & \dots & 0 \\ 0 & \mathbf{p}'_0 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \mathbf{p}'_0 \end{bmatrix} \in \mathbb{R}^{T \times LT}, \quad Q_0 = \begin{bmatrix} \mathbf{p}'_0 \frac{\partial \mathbf{d}(x(1), 1)}{\partial x(1)} & 0 & \dots & 0 \\ 0 & \mathbf{p}'_0 \frac{\partial \mathbf{d}(x(2), 2)}{\partial x(2)} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \mathbf{p}'_0 \frac{\partial \mathbf{d}(x(T), T)}{\partial x(T)} \end{bmatrix} \in \mathbb{R}^{T \times T}.$$

Next, use this result to conclude that

$$\sqrt{r_n}(\hat{\mathbf{d}} - \tilde{\mathbf{d}}) = \left[\frac{\partial D}{\partial x} + o_P(1) \right] \sqrt{r_n}(\hat{x} - x) = \frac{\partial D}{\partial x} Q_0^{-1} \bar{\mathbf{P}}_0 \sqrt{r_n}(\tilde{D} - D) + o_P(1),$$

where

$$\frac{\partial D}{\partial x} = \begin{bmatrix} \frac{\partial \mathbf{d}(x(1), 1)}{\partial x(1)} & 0 & \dots & 0 \\ 0 & \frac{\partial \mathbf{d}(x(2), 2)}{\partial x(2)} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \frac{\partial \mathbf{d}(x(T), T)}{\partial x(T)} \end{bmatrix} \in \mathbb{R}^{LT \times T}.$$

In total,

$$\sqrt{r_n} \begin{pmatrix} \hat{D} - \tilde{D} \\ \tilde{D} - D \end{pmatrix} = \Psi \sqrt{r_n}(\tilde{D} - D) + o_P(1),$$

with

$$\Psi = \begin{pmatrix} \frac{\partial \mathbf{d}}{\partial x} Q_0^{-1} \bar{\mathbf{P}}_0 \\ I_{LT} \end{pmatrix} \in \mathbb{R}^{2LT \times LT}. \quad (25)$$

In conclusion,

$$\sqrt{r_n}(\hat{b} - b) = \bar{\mathbf{P}}(I_{LT}, I_{LT}) \sqrt{r_n} \begin{pmatrix} \hat{D} - \tilde{D} \\ \tilde{D} - D \end{pmatrix} = \bar{\mathbf{P}}(I_{LT}, I_{LT}) \Psi \sqrt{r_n}(\tilde{D} - D) + o_P(1),$$

which is $O_P(1)$ under C.3(i) while under C.3(ii),

$$\sqrt{r_n}(\hat{b} - b) \rightarrow^d N(0, \bar{\mathbf{P}}(I_{LT}, I_{LT}) \Psi \Sigma(\underline{x}) \Psi'(I_{LT}, I_{LT})' \bar{\mathbf{P}}').$$

C Proof of Theorem 5

As with Theorem 4, we here prove a more general version of the theorem that takes as input any estimate of the demand system. As before, let $\hat{\mathbf{d}}(x, t) = (d_1(x, t), \dots, d_L(x, t))'$ be any given set of demand function estimators for $L \geq 2$ goods, and let $\hat{\mathbf{q}}(\mathbf{p}_0, x_0)$ be the implied demands computed as in eqs. (10)-(11). We then show that Theorem 5 holds under the general assumptions C.1-C.3 stated in the previous section.

Theorem 7 *Assume that (C.1)-(C.3) hold, and that $W_n \xrightarrow{P} W > 0$. Then the conclusions of Theorem 5 hold for the general demand estimator in (C.3).*

To prove this result, define $\ell_n(\mathbf{q}) = r_n MD_n(\mathbf{q}|\mathbf{p}_0, x_0)$ as the normalized version of the statistic, and let $\hat{\mathbf{q}}$ be the unrestricted estimator, $\hat{\mathbf{q}} = \arg \min_{\mathbf{q} \in \mathbb{R}_+^{TL}} \ell_n(\mathbf{q})$. Clearly, $\hat{\mathbf{q}} = \hat{\mathbf{q}}(\mathbf{p}_0, x_0)$ and $\ell_n(\hat{\mathbf{q}}) = 0$. Also define the restricted estimator $\hat{\mathbf{q}}_0$ by $\hat{\mathbf{q}}_0 = \arg \min_{\mathbf{q} \in \mathbb{S}} \ell_n(\mathbf{q})$ such that $\ell_n(\hat{\mathbf{q}}_0) = r_n MD_n^*(\mathbf{p}_0, x_0)$. Thus, our test can be written as

$$r_n MD_n^*(\mathbf{p}_0, x_0) = \left\{ \ell_n(\hat{\mathbf{q}}_0) - \ell_n(\mathbf{q}_0) \right\} - \left\{ \ell_n(\hat{\mathbf{q}}) - \ell_n(\mathbf{q}_0) \right\}, \quad (26)$$

where $\mathbf{q}_0 = \mathbf{q}_0(\mathbf{p}_0, x_0)$ is given in Eq. (13). For convenience, let $Z(x_0, \mathbf{p}_0) = r_n(\mathbf{q}_0 - \hat{\mathbf{d}}(x_0, \mathbf{p}_0))$ denote the normalized estimator such that $Z(x_0, \mathbf{p}_0) \sim N(0, \Sigma(x_0, \mathbf{p}_0))$ is its limit, $Z_n(x_0, \mathbf{p}_0) \xrightarrow{d} Z(x_0, \mathbf{p}_0)$, c.f. (C.3.ii). In the following, we suppress all dependence on (x_0, \mathbf{p}_0) since these are kept fixed.

We first note that the second term in Eq. (26) follows a standard distribution since the criterion function is smooth and \mathbf{q}_0 is situated in the interior of \mathbb{R}_+^{TL} : The first and second order derivatives are given by

$$\frac{\partial \ell_n(\mathbf{q})}{\partial \mathbf{q}} = 2W_n r_n(\mathbf{q} - \hat{\mathbf{q}}(\mathbf{p}_0, x_0)), \quad \frac{\partial^2 \ell_n(\mathbf{q})}{\partial \mathbf{q} \partial \mathbf{q}'} = 2r_n W_n,$$

and the third derivative being zero. Thus, by a third order Taylor expansion

$$\begin{aligned} \ell_n(\mathbf{q}_0) - \ell_n(\hat{\mathbf{q}}) &= \frac{\partial \ell_n(\hat{\mathbf{q}})}{\partial \mathbf{q}}(\mathbf{q}_0 - \hat{\mathbf{q}}) + \frac{1}{2}(\hat{\mathbf{q}} - \mathbf{q}_0)' \frac{\partial^2 \ell_n(\hat{\mathbf{q}})}{\partial \mathbf{q} \partial \mathbf{q}'}(\hat{\mathbf{q}} - \mathbf{q}_0) \\ &= \sqrt{r_n}(\hat{\mathbf{q}}(\mathbf{p}_0, x_0) - \mathbf{q}_0)' W_n \sqrt{r_n}(\hat{\mathbf{q}}(\mathbf{p}_0, x_0) - \mathbf{q}_0) \\ &= Z_n' W_n Z_n. \end{aligned} \quad (27)$$

The first term in Eq. (26) requires some more work since \mathbf{q}_0 will be on the boundary of the constrained set \mathbb{S} as defined in Eq. (12). In order to deal with this, we employ the general results of Andrews (2001) [A01]. Using his notation, the parameter of interest is $\theta = \mathbf{q} \in \mathbb{R}^{TL}$, the parameter space is given by $\Theta = \mathbb{S}$ and the objective function is given by $\ell_T(\theta) = \ell_n(\mathbf{q})$ (A01 uses T for sample size). A01's norming matrix B_n is chosen as $B_n = \sqrt{r_n}$. We now verify Assumptions 1, \mathcal{Q}^{2*} , 3^* , 5^* and 6 in A01 for the constrained estimator such that we can appeal to his Theorem 1.

Assumptions \mathcal{Q}^{2} :* We first note that $\mathbb{S} - \mathbf{q}_0$ is a union of linear inequality constraints. Thus Assumption \mathcal{Q}^{2*} (a) holds with $\mathbb{S}^+ = \mathbb{S} \cap B(\mathbf{q}_0, \delta)$ for some (small) $\delta > 0$. Assumption \mathcal{Q}^{2*} (b) and

\mathcal{Q}^{2*} (c) clearly hold since $\ell_n(\underline{\mathbf{q}})$ is three times differentiable with $B_n^{-2}\partial^2\ell_n(\underline{\mathbf{q}})/(\partial\underline{\mathbf{q}}\partial\underline{\mathbf{q}}') = 2W_n \rightarrow^P 2W^{-1} > 0$.

*Assumptions 3**: By (C.2.ii),

$$B_n^{-1} \frac{\partial \ell_n(\underline{\mathbf{q}}_0)}{\partial \underline{\mathbf{q}}} = 2W_n \sqrt{r_n} (\underline{\mathbf{q}} - \hat{\underline{\mathbf{q}}}(\mathbf{p}_0, x_0)) = 2W_n Z_n \rightarrow^d 2W^{-1} Z,$$

while $B_n^{-2}\partial^2\ell_n(\underline{\mathbf{q}})/(\partial\underline{\mathbf{q}}\partial\underline{\mathbf{q}}') \rightarrow^P 2W^{-1} > 0$ by (C.3).

*Assumptions 5** and 6: We wish to find a cone that is locally equal to $\mathbb{S} - \underline{\mathbf{q}}_0$. To this end, we wish to write \mathbb{S} on the form

$$\mathbb{S} = \{ \underline{\mathbf{q}} | V_t^*(\underline{\mathbf{q}}) - V_s^*(\underline{\mathbf{q}}) \geq \lambda_t^*(\underline{\mathbf{q}}) \mathbf{p}(t)' (\mathbf{q}(s) - \mathbf{q}(t)), \quad s, t = 1, \dots, T \},$$

for some functions $V^* : \mathbb{R}_+^{LT} \mapsto \mathbb{R}_+^T$ and $\lambda^* : \mathbb{R}_+^{LT} \mapsto \mathbb{R}_+^T$ that takes any set of demands, $\underline{\mathbf{q}} = (\mathbf{q}(1), \dots, \mathbf{q}(T))$ and maps it into an associated set of *unique* utility levels and marginal utilities. For $\underline{\mathbf{q}} \in \mathbb{S}$, the mapping can be constructed by, for example, Varian/Arifat's algorithm. Alternatively, we may define the two mappings as

$$\begin{aligned} (V(\underline{\mathbf{q}}), \lambda(\underline{\mathbf{q}})) &= \arg \min_{V, \lambda \in \mathbb{R}^T} V'V + \lambda' \lambda \\ \text{s.t. } V_t - V_s &\geq \lambda_t \mathbf{p}(t)' (\mathbf{q}(s) - \mathbf{q}(t)), \quad \lambda_t \geq 1, \quad V_t > 0, \quad t = 1, \dots, T. \end{aligned}$$

This convex optimization problem has a unique solution for any given value of $\underline{\mathbf{q}} \in \mathbb{S}$. However, this mapping is only defined for $\underline{\mathbf{q}} \in \mathbb{S}$ - not outside of the set. And we need the mapping to be well-defined and differentiable in a small neighbourhood of $\underline{\mathbf{q}}_0$. If $\underline{\mathbf{q}}_0$ lies on the boundary of \mathbb{S} , this implies that we need to extend the mapping to also be well-defined outside of \mathbb{S} . For any $\underline{\mathbf{q}}$ and (V, λ) satisfying with $\|\underline{\mathbf{q}} - \underline{\mathbf{q}}^*\| < \delta$ and $\|(V, \lambda) - (V(\underline{\mathbf{q}}^*), \lambda(\underline{\mathbf{q}}^*))\| < \delta$, for some small $\delta > 0$, define $e(\underline{\mathbf{q}}, V, \lambda)$ as

$$\begin{aligned} e(\underline{\mathbf{q}}, V, \lambda) &= \arg \min_{e \geq 0} e^2 \\ \text{s.t. } V_t - V_s + e &\geq \lambda_t \mathbf{p}(t)' (\mathbf{q}(s) - \mathbf{q}(t)), \quad t = 1, \dots, T. \end{aligned}$$

This is a well-defined function, which is continuously differentiable in a small neighbourhood of $(\underline{\mathbf{q}}^*, V(\underline{\mathbf{q}}^*), \lambda(\underline{\mathbf{q}}^*))$. For any $\underline{\mathbf{q}}$ with $\|\underline{\mathbf{q}} - \underline{\mathbf{q}}^*\| < \delta$, we now redefine $(V(\underline{\mathbf{q}}), \lambda(\underline{\mathbf{q}}))$ as

$$\begin{aligned} (V(\underline{\mathbf{q}}), \lambda(\underline{\mathbf{q}})) &= \arg \min_{V, \lambda} V'V + \lambda' \lambda \\ \text{s.t. } V_t - V_s + e(\underline{\mathbf{q}}, V, \lambda) &\geq \lambda_t \mathbf{p}(t)' (\mathbf{q}(s) - \mathbf{q}(t)), \quad \lambda_t \geq 1, \quad V_t > 0, \quad t = 1, \dots, T. \end{aligned}$$

The solution mapping $(V(\underline{\mathbf{q}}), \lambda(\underline{\mathbf{q}}))$ is differentiable for $\underline{\mathbf{q}}$ with $\|\underline{\mathbf{q}} - \underline{\mathbf{q}}^*\| < \delta$, and we can restrict our attention to this set since under the null, $\hat{\underline{\mathbf{q}}}^* \rightarrow \underline{\mathbf{q}}^*$. With $B(\underline{\mathbf{q}})$ defined in Section 7, we can then write the constrained set as $\mathbb{S} = \{ \underline{\mathbf{q}} | B(\underline{\mathbf{q}}) \leq 0 \}$, and employ Andrews (1997, Lemma 2) to obtain that the cone Λ is given as in Eq. (15).

It now follows by A01, (proof of) Theorem 1, that

$$\ell_n(\underline{\mathbf{q}}_0) - \ell_n(\hat{\underline{\mathbf{q}}}) = Z_n' \hat{W} Z_n - \inf_{\lambda \in \Lambda} (\lambda - Z_n)' W_n (\lambda - Z_n) + o_P(1). \quad (28)$$

Substituting Eqs. (27) and (28) back into Eq. (26) yields the desired result:

$$MD_n^*(x_0) = \inf_{\lambda \in \Lambda} (\lambda - Z_n)' \hat{W} + o_P(1) \rightarrow^d \inf_{\lambda \in \Lambda} (\lambda - Z)' W^{-1} (\lambda - Z).$$

D Auxiliary Results

We here state a general result for a family of set estimators involving linear constraints. These should have independent interest outside of demand bounds estimation. In particular, this seems to be the first general treatment of linear two-step set estimators. We consider a compact and convex parameter space $\Theta \subset \mathbb{R}^d$ and define the set of identified parameters by

$$\Theta_I = \{\theta \in \Theta | A\theta \leq b\},$$

for some known function $\phi : \Theta \mapsto \mathbb{R}^q$ and some unknown matrix $A \in \mathbb{R}^{p \times q}$ and unknown vector $b \in \mathbb{R}^p$. Since A and b are unknown, we cannot compute Θ_I , but suppose we are given preliminary estimators, $\hat{A} = \hat{A}_n \in \mathbb{R}^{p \times q}$ and $\hat{b} = \hat{b}_n \in \mathbb{R}^p$ for some sample of size $n \geq 1$. A natural estimator of Θ_I is then the following plug-in one,

$$\hat{\Theta}_I = \{\theta \in \Theta | \hat{A}\theta \leq \hat{b}\}.$$

However, in some situations, a different estimator has to be used to ensure that the estimator is consistent and in order to construct confidence sets, c.f. Section 3.

We note that by defining

$$m(\theta) = A\theta - b, \quad \hat{m}(\theta) = \hat{A}\theta - \hat{b},$$

we can express Θ_I and $\hat{\Theta}_I$ in terms of a set of moment inequalities: $\Theta_I = \{\theta \in \Theta | m(\theta) \leq 0\}$ and $\hat{\Theta}_I = \{\theta \in \Theta | \hat{m}(\theta) \leq 0\}$.

As already demonstrated in Section 7, the demand bound estimator fits into this framework. Some other problems are:

Example 8 (OLS with Errors-in-Variables) *Consider the regression model*

$$Y = \beta_0 + \beta_1' X + \varepsilon,$$

where we only have contaminated observations of $X \in \mathbb{R}^d$,

$$Z = X + u.$$

Here, u_t are the errors-in-variables. Assuming that u and X are independent with $E[u] = 0$ and $E[uu'] = D$, we obtain by following e.g. Klepper and Leamer (1984) that β_1 satisfies

$$r'b_1 \leq r'\beta_1 \leq \text{Var}(Y),$$

where

$$b_1 = \text{Var}(Z)^{-1} \text{Cov}(Z, Y) \in \mathbb{R}^d, \quad r = \text{Cov}(Z, Y) \in \mathbb{R}^d.$$

Here,

$$A = [r, -r]' \in \mathbb{R}^{2 \times d}, \quad b = [\text{Var}(Y), -r'b_1]' \in \mathbb{R}^2.$$

Example 9 (Regression with Interval-Censored Outcomes) Consider the regression model

$$Y = \beta'X + \varepsilon,$$

where $X \in \{x_1, \dots, x_L\}$ is discrete and we don't observe Y but only Y_{\min} and Y_{\max} where $Y \in [Y_{\min}, Y_{\max}]$. Then

$$E[Y_{\min}|X = x_i] \leq \beta'x_i \leq E[Y_{\max}|X = x_i], \quad i = 1, \dots, L.$$

Here,

$$A = [x'_1, \dots, x'_L] \in \mathbb{R}^{L \times d}, \quad b = (E[Y_{\min}|X = x_1], \dots, E[Y_{\min}|X = x_L])' \in \mathbb{R}^L.$$

We derive the asymptotic properties of $\hat{\Theta}_I$ and construct confidence sets for Θ_I by utilizing the general set of results found in Chernozukhov, Tamer and Hong (2007) [CHT07]. We will do so under the following high-level assumptions:

Assumption 1 The parameter space $\Theta \subset \mathbb{R}^d$ is compact and convex.

Assumption 2 The matrix $A = [a'_1, \dots, a'_p]' \in \mathbb{R}^{p \times d}$, $a_i \in \mathbb{R}^d$, has rank p .

Assumption 3 The set

$$\Theta_I^{-\varepsilon} = \{\theta \in \Theta_I | a_i\theta \leq b_i - \varepsilon, \quad i = 1, \dots, p\} \subset \Theta_I$$

has a non-empty interior for some (sufficiently small) $\varepsilon > 0$.

Assumption 4 For some sequence $r_n \rightarrow \infty$:

$$(i) \quad \|\hat{A} - A\| = O_P(1/\sqrt{r_n}) \text{ and } \|\hat{b} - b\| = O_P(1/\sqrt{r_n}).$$

(ii)

$$\sqrt{r_n} \begin{pmatrix} \text{vec}(\hat{A}) - \text{vec}(A) \\ \hat{b} - b \end{pmatrix} \rightarrow^d N \left(0, \begin{pmatrix} \Omega_{AA} & \Omega_{Ab} \\ \Omega_{bA} & \Omega_{bb} \end{pmatrix} \right),$$

where $V \in \mathbb{R}^{p(d+1) \times p(d+1)}$ is the joint asymptotic variance.

Depending on whether Assumption 3 holds or not, we will use two different estimators of Θ_I . These can both be written on the form

$$\hat{\Theta}_I(c) = \left\{ \theta \in \Theta \mid \sup_{\theta \in \Theta_I} \|\hat{m}(\theta)\|_+^2 \leq c/r_n \right\},$$

for some sequence $c = c_n \geq 0$, and where $\|x\|_+^2 = \|\max\{x, 0\}\|^2$. If Assumption 3 does not hold, then $\text{int}\Theta_I = \emptyset$, and we need to choose $\tau \propto \log(n)$ to obtain consistency. If it does hold, we may choose $\tau = 0$ such that $\hat{\Theta}_I = \hat{\Theta}_I(0) = \{\theta \in \Theta | \hat{A}\theta \leq \hat{b}\}$.

In order to do inference, we wish to choose a (possibly data dependent) sequence $\hat{\tau}$ such that

$$P(\Theta_I \subseteq \hat{\Theta}_I(\hat{\tau})) \rightarrow 1 - \alpha,$$

for some given confidence level $1 - \alpha$. Here, we need to choose $\hat{\tau} > 0$ to obtain the correct coverage, even if $\hat{\Theta}_I(0)$ is consistent.

Theorem 10 *Assume that Assumptions 1-2 and 4(i) hold.*

1. *If Assumption 3 does not hold, we choose $c \propto \log(n)$, and obtain:*

$$d_H(\hat{\Theta}_I(\tau), \Theta_I) = O_P(\sqrt{\log(n)/r_n}).$$

2. *If Assumption 3 does hold, we choose $c = 0$ and obtain:*

$$d_H(\hat{\Theta}_I(0), \Theta_I) = O_P(\sqrt{1/r_n}).$$

3. *If additionally Assumption 4(ii) hold then*

$$\sup_{\theta \in \Theta_I} r_n Q_n(\theta) \rightarrow^d \mathcal{Q} := \sup_{\theta \in \Theta_I} \|Z(\theta) + \xi(\theta)\|_+^2,$$

where $\theta \mapsto Z(\theta)$ is a zero-mean Gaussian process with covariance matrix

$$\Sigma(\theta_1, \theta_2) = \begin{pmatrix} \theta'_1 \otimes I_p & O_{p \times p} \\ O_{p \times pd} & -I_p \end{pmatrix} \begin{pmatrix} \Omega_{AA} & \Omega_{Ab} \\ \Omega_{bA} & \Omega_{bb} \end{pmatrix} \begin{pmatrix} \theta_2 \otimes I_p & O_{pd \times p} \\ O_{p \times p} & -I_p \end{pmatrix},$$

while $\xi(\theta) = (\xi_1(\theta), \dots, \xi_p(\theta))$ is given by

$$\xi_i(\theta) = \begin{cases} -\infty, & a_i\theta < b_i \\ 0, & a_i\theta = b_i \end{cases}, \quad i = 1, \dots, p.$$

Let consistent estimator $\hat{c}_{1-\alpha}$ of the $(1 - \alpha)$ th quantile of \mathcal{Q} . We then have

$$P(\Theta_I \subseteq \hat{\Theta}_I(\hat{\tau})) \rightarrow 1 - \alpha,$$

where $\hat{\tau}$ is given by

$$\text{Assumption 3 does not hold} : \hat{\tau} = \hat{c}_{1-\alpha} + O_P(\log(n));$$

$$\text{Assumption 3 does hold} : \hat{\tau} = \hat{c}_{1-\alpha}.$$

Proof. We verify conditions C.1-C.5 in CHT07 with $Q_n(\theta) = \|\hat{m}(\theta)\|_+^2$, $\hat{m}(\theta) = \hat{A}\theta - \hat{b}$, and $Q(\theta) = \|m(\theta)\|_+^2$, $\hat{m}(\theta) = A\theta - b$. This will yield the stated results.

Condition C.1: Given the definitions of $Q_n(\theta)$ and $Q(\theta)$, this holds with $a_n = r_n$ and $b_n = \sqrt{r_n}$ as an immediate consequence of Assumption 1 and Assumption 3(i).

Condition C.2: We show that this holds with $\gamma = 1$. Consider any $\theta \in \Theta \setminus \Theta_I$: Let $\theta^* = \arg \min_{\theta' \in \Theta_I} \|\theta - \theta'\|$ be the unique point in Θ_I which has minimum distance to θ . Let $\delta^* = \theta - \theta^*$ be the difference such that $\|\delta^*\| = \rho(\theta, \Theta_I)$. We can decompose the rows of (A, b) into binding and non-binding constraints respectively of θ^* . We write $A = [a'_1, \dots, a'_p]'$, where $a_i \in \mathbb{R}^d$ is an individual row of A . Let $(A^{(1)}, b^{(1)})$ and $(A^{(2)}, b^{(2)})$, with $A^{(1)} = [a_1^{(1)'}, \dots, a_{p_1}^{(1)'}]'$ and $b^{(1)} = (b_1^{(1)}, \dots, b_{p_1}^{(1)})'$, denote the set of rows which contain the binding and non-binding constraints respectively. That is, $m^{(1)}(\theta^*) := A^{(1)}\theta^* - b^{(1)} = 0$ while $m^{(2)}(\theta^*) := A^{(2)}\theta^* - b^{(2)} < 0$. Let λ_{\min} be the minimum eigenvalue of AA' . Due to Assumption 1, $\lambda_{\min} > 0$ and we have

$$\lambda_{\min} \|\delta^*\| \leq \|A^{(1)}\delta^*\| \leq \max_{i=1, \dots, p_1} |a_i^{(1)}\delta^*|.$$

Thus, there exists at least one $i_0 \in \{1, \dots, p_1\}$ such that either $a_{i_0}^{(1)}\delta^* \leq -\lambda_{\min} \|\delta^*\|$ or $\lambda_{\min} \|\delta^*\| \leq a_{i_0}^{(1)}\delta^*$. We then obtain

$$\begin{aligned} \|m(\theta)\|_+^2 &= \sum_{i=1}^p |a_i\theta - b_i|_+^2 \geq \left| a_{i_0}^{(1)}\theta - b_{i_0}^{(1)} \right|_+^2 = \left| a_{i_0}^{(1)}(\theta^* + \delta^*) - b_{i_0}^{(1)} \right|_+^2 \\ &= \left| a_{i_0}^{(1)}\delta^* \right|_+^2 \geq \lambda_{\min}^2 \|\delta^*\|^2 = \lambda_{\min}^2 \rho^2(\theta, \Theta_I) \end{aligned}$$

This establishes that $Q(\theta) = \|m(\theta)\|_+^2 \geq \lambda_{\min}^2 \rho^2(\theta, \Theta_I)$. Following the same steps as in CHT07's Proof of Theorem 4.2(Step 1) this verifies (C.2).

Condition C.3: Assume that Assumption 3 holds. Consider for any (small) $\varepsilon > 0$, the contraction $\Theta_I^{-\varepsilon}$ defined in Assumption 3. Note that we here do not use the same contraction as the one considered in CHT07, Section 4.2. We then wish to verify their (C.3) with $\Theta_n = \Theta_I^{-\varepsilon_n}$ where $\varepsilon_n = O_P(1/\sqrt{a_n})$. For any $\theta \in \Theta_n$,

$$\begin{aligned} nQ_n(\theta) &= n \|\hat{m}(\theta) - m(\theta) + m(\theta)\|_+^2 \\ &\leq \sum_{i=1}^p |\{\sqrt{n}\hat{m}_i(\theta) - m(\theta)\} + \sqrt{n}m(\theta)|_+^2 \\ &= \sum_{i=1}^p |O_P(1) + \sqrt{n}m(\theta)|_+^2 \\ &\leq \sum_{i=1}^p |O_P(1) + \sqrt{a_n}\varepsilon_n|_+^2 \\ &= 0. \end{aligned}$$

Next, observe that, due to $\Theta_I^{-\varepsilon} \subset \Theta_I$ with both being compact,

$$d_H(\Theta_I^{-\varepsilon}, \Theta_I) = \sup_{\theta \in \Theta_I} \rho(\theta, \Theta_I^{-\varepsilon}) = \sup_{\theta \in \Theta_I} \inf_{\theta^* \in \Theta_I^{-\varepsilon}} \|\theta^* - \theta\| = \|\bar{\theta}^* - \bar{\theta}\|,$$

for some $\bar{\theta} \in \Theta_I$ and $\bar{\theta}^* \in \Theta_I^{-\varepsilon}$. These optimizers must satisfy $\bar{\theta} \in \partial\Theta_I$ and $\bar{\theta}^* \in \partial\Theta_I^{-\varepsilon}$; in particular, there exists $i \in \{1, \dots, p\}$ such that $m_i(\bar{\theta}^*) = 0$. Thus,

$$\|\bar{\theta}^* - \bar{\theta}\| \leq \lambda_{\max} |m_i(\bar{\theta}^*) - m_i(\bar{\theta})| = \lambda_{\max} |m_i(\bar{\theta}^*)| \leq \lambda_{\max} \varepsilon,$$

where λ_{\max} is the maximum eigenvalue of AA' .

Condition C.4-C.5: We here follow the same arguments as in CHT07, Proof of Theorem 4.2 (Step 2-3), and verify their high-level conditions in M.2 for our specific choice of m . Define

$$\Theta_{\mathcal{I}} = \{\theta \in \Theta | a_i \theta = b_i, i \in \mathcal{I}, a_i \theta < b_i, i \in \mathcal{I}^c\},$$

for any subindex $\mathcal{I} \subseteq \{1, \dots, p\}$ with $\mathcal{I}^c = \{1, \dots, p\} \setminus \mathcal{I}$, which is convex as the intersection of two convex sets. Also, define

$$\begin{aligned} \mathcal{V}_{n,\mathcal{I}}^\delta &= \left\{ (\theta, \lambda) \in \Theta_{\mathcal{I}} \times \overline{B(0, \delta)} | \lambda = \sqrt{a_n} (\Theta_{\mathcal{I}} - \theta) \right\}, \\ \mathcal{V}_{\infty,\mathcal{I}}^\delta &= \left\{ (\theta, \lambda) \in \Theta_{\mathcal{I}} \times \overline{B(0, \delta)} | \lambda = \sqrt{a_{n'}} (\Theta_{\mathcal{I}} - \theta) \text{ for some } n' \geq 1 \right\}. \end{aligned}$$

Since $\Theta_{\mathcal{I}} \times \overline{B(0, \delta)}$ is convex, it follows that $\mathcal{V}_{n,\mathcal{I}}^\delta \uparrow \mathcal{V}_{\infty,\mathcal{I}}^\delta$ monotonically in the set-theoretic sense. Since $\mathcal{V}_{n,\mathcal{I}}^\delta, \mathcal{V}_{\infty,\mathcal{I}}^\delta \subseteq \overline{\Theta_{\mathcal{I}} \times B(0, \delta)}$ where $\overline{\Theta_{\mathcal{I}} \times B(0, \delta)}$ is compact, this in turn implies convergence in the Hausdorff distance.

Re. convergence of m : We have

$$\begin{aligned} \sqrt{a_n}(\hat{m}(\theta) - m(\theta)) &= \sqrt{a_n}(\hat{A} - A)\theta - \sqrt{a_n}(\hat{b} - b) \\ &= (\theta' \otimes I_p) \sqrt{a_n} \text{vec}(\hat{A} - A) - \sqrt{a_n}(\hat{b} - b) \\ &= \begin{pmatrix} \theta' \otimes I_p & O_{p \times p} \\ O_{p \times pd} & -I_p \end{pmatrix} \sqrt{a_n} \begin{pmatrix} \text{vec}(\hat{A} - A) \\ \hat{b} - b \end{pmatrix}. \end{aligned}$$

Under Assumption 2(i), the last expression is $O_P(1)$ while under Assumption 2(ii) it converges towards $N(0, \Sigma(\theta, \theta))$. This shows pointwise weak convergence, while the following inequality demonstrates that $\theta \mapsto \sqrt{a_n} \{\hat{m}(\theta) - m(\theta)\}$ is stochastically equicontinuous:

$$\begin{aligned} \|\sqrt{a_n} \{\hat{m}(\theta_1) - m(\theta_1)\} - \sqrt{a_n} \{\hat{m}(\theta_2) - m(\theta_2)\}\| &= \|\sqrt{a_n} (\hat{A} - A) (\theta_1 - \theta_2)\| \\ &\leq \|\sqrt{a_n} (\hat{A} - A)\| \|\theta_1 - \theta_2\| \\ &= O_P(1) \times \|\theta_1 - \theta_2\|. \end{aligned}$$

In total, the stochastic process $\theta \mapsto \sqrt{a_n} \{\hat{m}(\theta) - m(\theta)\}$ weakly converges on the compact set Θ towards a zero-mean Gaussian process $Z(\theta)$ with $\text{Cov}(Z(\theta_1), Z(\theta_2)) = \Sigma(\theta_1, \theta_2)$, c.f. Van der Vaart and Wellner (2000, Example 1.5.10). We also note that $\nabla_\theta m(\theta) = A$ is well-defined. In conclusion,

$$\ell_n(\theta, \lambda) = r_n Q_n(\theta + \lambda/\sqrt{r_n}) \rightarrow^d \ell_\infty(\theta, \lambda) = \|Z(\theta) + A\lambda\|_+^2.$$

In particular,

$$\sup_{\theta \in \Theta_I} r_n Q_n(\theta) = \sup_{(\theta, \lambda) \in \mathcal{V}_\infty^0} \ell_n(\theta, \lambda) \rightarrow^d \bar{Q} = \sup_{\theta \in \Theta_I} \|Z(\theta) + \xi(\theta)\|_+^2.$$

■