

KU LEUVEN

DEPARTMENT OF ECONOMICS

Bounding counterfactual demand with unobserved heterogeneity and endogenous expenditures

Laurens CHERCHYE, Thomas DEMUYNCK and Bram DE ROCK

FACULTY OF ECONOMICS AND BUSINESS



DISCUSSION PAPER SERIES DPS17.14

NOVEMBER 2017

Bounding counterfactual demand with unobserved heterogeneity and endogenous expenditures

Laurens Cherchye* Thomas Demuynck[†] Bram De Rock[‡]

November 8, 2017

Abstract

We propose a novel method to predict rational counterfactual demand responses from an observed set of repeated cross-sections. Our method derives bounds on the distribution of counterfactual demands that are consistent with the Weak Axiom of Revealed Preferences without putting any restriction on the preference heterogeneity across consumers. In contrast to existing methods, our method also allows for endogeneity of total expenditures. In addition, the method can readily incorporate restrictions on the income elasticities of the consumption goods, which further enhances its identifying power (i.e. tighter bounds). The method is easy to implement and yields informative bounds on demand, which we illustrate through an application to data drawn from the U.S. Consumer Expenditure Survey (CEX).

*Department of economics, University of Leuven. E. Sabbelaan 53, B-8500 Kortrijk, Belgium. E-mail: laurens.cherchye@kuleuven-kortrijk.be. Laurens Cherchye thanks the European Research Council (ERC; ERC-2013-CoG-614221), the Fund for Scientific Research - Flanders (FWO) and the Research Fund of the KU Leuven for financial support.

[†]ECARES, Université Libre de Bruxelles. Avenue F.D. Roosevelt, CP 114, B-1050 Brussels, Belgium. E-mail: thomas.demuynck@ulb.ac.be

[‡]ECARES, Université Libre de Bruxelles and Department of Economics, University of Leuven. Avenue F. D. Roosevelt 50, CP 114, B-1050 Brussels, Belgium. E-mail: bderock@ulb.ac.be. Bram De Rock gratefully acknowledges FWO and FRNS for their financial support.

Keywords: unobserved heterogeneity, endogenous expenditures, WARP, counterfactual demand

JEL codes: C14, D12

1 Introduction

Unobserved preference heterogeneity means that different consumers (e.g. households) behave differently even when taking account of all observable characteristics (like household composition, household age profile, marital status, education, etc.). The implications of unobserved heterogeneity have received considerable attention in the recent literature on counterfactual demand analysis, which predicts the demand effects at new (i.e. non-observed) prices and incomes (e.g. due to a change in the indirect taxes on some good). The question is whether we can make informative statements regarding these demand effects that effectively account for the unobserved heterogeneity. In the present paper, we propose a novel method to do so under minimalistic prior assumptions.

More specifically, we consider a setting in which the empirical analyst can use repeated cross-sectional data on consumers' demands for a set of goods, with each cross-section corresponding to a different price regime. This is the type of data most frequently used in applied demand analysis. In such a setting, the empirical analyst typically has only a single consumption observation for each individual consumer. Therefore, we are faced with the problem that individual preferences and, hence, counterfactual choices are not identified, unless one is willing to impose some additional restrictions on the structure of the unobserved heterogeneity component and/or on the form of the underlying consumer preferences.

Our method abstains from imposing such restrictions. It derives nonparametric bounds on the *distribution* of counterfactual demands (instead of *individual* counterfactual de-

mands) for a population of heterogeneous consumers. It defines these bounds by exploiting the empirical implications of rational consumer behavior that are captured by the Weak Axiom of Revealed Preference (WARP). As we discuss next, compared to existing methods, we impose a minimal set of conditions. We also show that the method can easily incorporate restrictions on the income elasticities of the consumption goods, which further enhances its identifying power (i.e. tighter bounds).

Approaches to deal with unobserved heterogeneity. The literature has invoked alternative assumptions to deal with unobserved preference heterogeneity in applied demand analysis. These assumptions can be broadly classified into three categories: (i) assumptions on the heterogeneity component, (ii) independence assumptions regarding the data generating process and (iii) assumptions on the structure of consumer behavior (i.e. rationality assumptions). In what follows, we discuss these three types of assumptions in more detail. In turn, this will allow us to articulate our own contribution.

The first category of assumptions concerns the unobserved preference heterogeneity component. The most widely used assumption is that the unobserved heterogeneity component is single-dimensional and enters the demand functions in an additively separable way. If, in addition, a mean independency assumption is imposed, then demand behavior can be analyzed by focusing on the mean demand function of all consumers in the population. However, as discussed in detail by Brown and Walker (1989) and Lewbel (2001), this approach is only valid if individual preferences satisfy a number of quite stringent requirements.

In response to this critique, a recent strand of literature (including Blundell, Kristensen, and Matzkin (2014), Blundell, Horowitz, and Parey (2012) and Hoderlein and Vanhems (2017)) has relaxed the additivity assumption, and used the weaker assumption that the unobserved heterogeneity component has a dimension that equals at most the number of

goods minus one and, in addition, satisfies an invertibility requirement (basically stating that demand functions are invertible in the unobserved heterogeneity component). As shown by Matzkin (2003) (see also Beckert and Blundell (2008) and Matzkin (2008)), in a two-goods setting these assumptions identify each consumer with its own quantile in the demand distribution. Specifically, for two different price-income situations, the consumers situated on the same quantile of the demand distribution have the same preferences.

If we abstain from imposing such assumptions, the identification of individual preferences and demands must be abandoned. In such a case, the best we can do is identify the distribution of preferences or demands over the heterogeneous population. Notably, this still allows us, for example, to identify the median demand function in the population, but it will no longer be possible to recover the demand for the “median” consumer (in fact, this median consumer need no longer exist).

The second group of assumptions is of a statistical nature and imposes independence restrictions on the distribution of the unobserved heterogeneity component. Essentially, these restrictions impose that, when conditioning on observable consumer characteristics, the unobserved heterogeneity component is independent from the consumers’ budget set (i.e. prices and income). In a cross-sectional setting, with fixed prices within each cross-section, this assumption requires that the distribution of preferences remains unchanged when considering subpopulations with different expenditure levels (conditional on other observable characteristics). As an implication, it also imposes that the distributions of preferences for rich and poor consumers are the same. To the best of our knowledge, all papers in the literature impose such (conditional) independence (see also our literature review below).

In many cases, total expenditures are partially determined by preferences, which makes them endogenous. Usually, this endogeneity problem is tackled by using gross earnings as an instrumental variable (see, for example, Blundell and Duncan (1998), Blundell, Brown-

ing, and Crawford (2003) and Blundell, Chen, and Kristensen (2007b)). This instrument, however, is only valid if total earnings influence demand *only* through their effect on the total expenditures. For example, it also requires that the wage rate (which partially determines the gross earnings) is not correlated with consumers' preferences, and this is arguably a controversial assumption in many situations.

The third group of assumptions imposes rationality constraints on consumers' behavior. Such rationality conditions allow us to predict the demand distribution in counterfactual price-income situations.¹ We distinguish two types of rationality conditions that are used in the literature. The first type takes the form of differential conditions on consumer demands. Examples are the well-known Slutsky symmetry, Slutsky negativity, adding up and homogeneity conditions. The second type of conditions is expressed in terms of revealed preference axioms. The most frequently used axioms are the Weak Axiom of Revealed Preference (WARP), the Strong Axiom of Revealed Preference (SARP) and the Generalized Axiom of Revealed Preference (GARP). In a population setting, combining these revealed preference axioms with the "independence of budgets" restriction that we discussed above gives the so-called stochastic axioms of revealed preference (see McFadden (2005) for an overview). Importantly, the validity of the budget independence assumption is again crucial to justify the application of these stochastic axioms in empirical demand analysis.

Our contribution. Our main contribution is that we impose a minimal set of assumptions to analyze counterfactual demand distributions in a setting with unobserved preference heterogeneity. First of all, we impose no restrictions on the structure or distribution of the unobserved heterogeneity component. As an implication, our focus will be on re-

¹In principle, when the data show sufficient price and income variation, rationality conditions can be omitted for conducting demand analysis. However, even in such a case the conditions can be useful to increase estimation precision.

covering demand distributions rather than individual demands. Next, we maintain the assumption that unobserved preference heterogeneity is independent of commodity prices, which is rather uncontroversial for the type of setting that we analyze (i.e. repeated cross-sections with fixed prices per cross-section). Importantly, however, we no longer assume (conditional) independence between preference heterogeneity and total expenditures. Finally, we require that consumers satisfy WARP.²

A main motivation for our paper originates from a recent strand of literature that focuses on the issue of partial identification in treatment response models (see for example Manski (1997)). The starting point in this literature is what is known as the credibility problem. The credibility problem says that every empirical model faces a negative trade-off between the assumptions imposed on the model and the credibility of its outcomes: the stronger the assumptions imposed on the model, the less credible are the findings. To control this trade-off, we must start from a base model which imposes only minimal assumptions on the data generating process. Next, departing from this base scenario, additional assumptions are imposed. By comparing the results with and without these additional assumptions, their identifying power can be analyzed.

Our base model uses minimalistic (“uncontroversial”) assumptions: we only impose a very basic rationality constraint (WARP), we do not impose any restriction on the unobserved heterogeneity component, and we allow expenditure levels to be endogenous. We show that, in principle, this base model can define informative WARP-consistent bounds on counterfactual demand distributions, despite these very weak priors. These bounds provide a concrete insight into the identifying power of our minimal assumptions. However, not very surprisingly, these WARP-consistent bounds may often be quite wide in practice. Therefore, in a following step, we show that our method can readily incorporate specific

²Basically, WARP imposes that preferences are acyclic for all chains of length two (see Section 2). It does not impose transitivity of preferences (in contrast to SARP and GARP) and, in that sense, imposes a fairly weak condition of rational consumer behavior.

restrictions regarding the income elasticities of the consumption goods. We believe that such restrictions may be fairly easy to define in many practical situations. For example, we may often reasonably assume (or learn from existing studies) that a particular commodity is typically a normal good (i.e. positive income elasticity) or a necessary good (i.e. income elasticity between zero and one). Our empirical application will show that such additional restrictions can substantially enhance the identifying power of our method in empirical analysis.

Related literature. We mentioned above that the literature on demand analysis has devoted considerable attention to the issue of unobserved preference heterogeneity. To conclude this introductory section, we provide a (non-exhaustive) overview of some of the more recent proposals, together with a brief discussion of their main underlying assumptions. This will also highlight the differences between existing methods and our new proposal.

A first interesting study is Blundell, Horowitz, and Parey (2012), who use nonparametric regression techniques subject to Slutsky shape restrictions in order to compute the demand response and deadweight loss of a tax increase on the price of gasoline. These authors assume that the unobserved heterogeneity component is additively separable. In a follow-up study, Blundell, Horowitz, and Parey (2017) extend this analysis by weakening the additive separability assumption towards invertibility of the unobserved heterogeneity component.

A closely related study is Hoderlein and Vanhems (2017), who use nonparametric regression to compute money metric utility functions under the assumption that the unobserved heterogeneity component is invertible. Hausman and Newey (2016) use a similar framework but relax the heterogeneity assumption. They obtain bounds on the average deadweight loss by bounding the average income effect within the population.

A different strand of literature was initiated by Blundell, Browning, and Crawford (2003, 2007a, 2008). These authors combine nonparametric mean Engel curve estimates with WARP restrictions to bound money metric utility functions and counterfactual demand responses in a repeated cross-sectional setting. Blundell, Browning, Cherchye, Crawford, De Rock, and Vermeulen (2015) extend these authors' analysis by using SARP instead of WARP as a condition of consumer rationality. All these studies assume additive separability of the unobserved preference heterogeneity component.

Using a formally similar framework, Blundell, Kristensen, and Matzkin (2014) also construct bounds on the counterfactual demands in a repeated cross-sectional setting. A main difference with the above mentioned studies is that these authors estimate nonparametric quantile demand functions subject to WARP restrictions, while assuming that the unobserved heterogeneity component is invertible. Cosaert and Demuyck (2017) further relax this invertibility assumption and put no restriction on the unobserved preference heterogeneity. They use results from stochastic revealed preference theory in order to bound the distribution of money metric utilities and counterfactual demand correspondences.

A final class of studies does not explicitly consider counterfactual demand analysis, but instead focuses on checking whether and to what extent observed demand behavior is consistent with consumer rationality under unobserved heterogeneity. Hoderlein and Stoye (2014) compute bounds on the size of the population that violates WARP for repeated cross-sectional data. Kitamura and Stoye (2016) address the same question but use SARP as a rationality condition, and Kawaguchi (2017) develops tests to verify various other stochastic axioms of revealed preference. In a similar vein, Hoderlein (2011) and Dette, Hoderlein, and Neumeyer (2016) check whether a population of heterogeneous consumers satisfies the differential Slutsky symmetry or negativity restrictions. None of these studies imposes restrictions on the unobserved heterogeneity component.

Importantly, all of the papers listed above require that unobserved heterogeneity and

total expenditures are (conditionally) independent. In our opinion, this constitutes a principal difference between our new proposal and existing methods in the literature.

Structure of the paper. This paper is organized as follows. Section 2 introduces some necessary definitions and notation. Section 3 shows how to obtain WARP-consistent bounds on demand distributions under our minimalistic priors. Section 4 explains how we can derive tighter bounds by restricting the income elasticities of the consumption goods. Section 5 discusses estimation and inference. Section 6 presents an empirical illustration to data drawn from the U.S. Consumer Expenditure Survey (CEX). Finally, Section 7 summarizes and concludes.

2 Preliminaries

We follow Blundell, Kristensen, and Matzkin (2014) and focus on a setting with two goods, with quantities $q \in \mathbb{R}_+$ and $z \in \mathbb{R}_+$. We normalize the price of good z to one and denote the price of good q by $p \in \mathbb{R}_{++}$. Our focus on a two-goods settings allows us to graphically illustrate our main theoretical results. Importantly, however, it is fairly easy to generalize our analysis to settings with more than two goods.

Set-up. We consider a probability space (J, Ω, P) where J is the space of consumers, Ω a suitable σ -algebra and P the probability measure. A “random” consumer is denoted by $j \in J$, and we represent the associated demand for good q by the function $q_j(p, x_j)$, where $x_j \in \mathbb{R}_+$ stands for total expenditures of consumer j . Thus, the demand of consumer j for good q is defined as a continuous function $q_j : \mathbb{R}_{++} \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that, $0 \leq q_j(p, x_j) \leq x_j/p$ (i.e. the maximum expenditure on good q equals x_j). Correspondingly, the demand

of consumer j for good z is equal to

$$z_j(p, x_j) = x_j - p q_j(p, x_j).$$

As indicated above, we assume a repeated cross-sectional setting. Within each cross-section, the empirical analyst observes the demanded quantities and expenditure levels for a large number of consumers drawn i.i.d from (J, Ω, P) . The price is assumed to be fixed and identical for all consumers within the same cross-section, but prices may vary over cross-sections. Thus, we treat prices as non-stochastic, which also means that unobserved preference heterogeneity (captured by the random variable j) is distributed independently from the prices. If the distribution of consumer preferences remains stable over time (i.e. over the different cross-sections) and if each cross-section is randomly drawn from the same population, then this assumption will be satisfied. We denote the cross-sections by $\{1, \dots, t, \dots T\}$ and we represent the price in cross-section t by p_t .

We assume a sufficiently large number of independent observations for each cross-section t . Then, we can identify the joint distribution of the random variables $(x_{t,j}, q_{t,j}, z_{t,j})$, with, for consumer j , $x_{t,j}$ the total expenditures, $q_{t,j} = q_j(p_t, x_{t,j})$ the demanded quantity of good q , and $z_{t,j} = z_j(p_t, x_{t,j})$ the quantity of good z . Of course, these random variables satisfy the adding up condition, $x_{t,j} = p_t q_{t,j} + z_{t,j}$. For now, we will assume that for each cross section t , we know the joint distribution of $(x_{t,j}, q_{t,j}, z_{t,j})$. We will indicate how we deal with small sample and estimation issues in Section 5.

Importantly, we do not impose that the unobserved heterogeneity component is distributed independently of the expenditure level. For example, we do not require that $\Pr[q_j(p_t, x) \leq q | x_{t,j} = x] = \Pr[q_j(p_t, x) \leq q]$. As discussed in the Introduction, this constitutes a main difference between our method and other proposals in the literature.

WARP restrictions. We use the Weak Axiom of Revealed Preference (WARP; Samuelson (1938)) as our condition of rational consumer behavior.³ To formally define the WARP restrictions, let us consider two income levels $x, y \in \mathbb{R}_+$ and two prices $p_t, p_v \in \mathbb{R}_{++}$. Consumer j satisfies WARP if, for the associated consumption bundles $(q_j(p_t, x), z_j(p_t, x)) \neq (q_j(p_v, y), z_j(p_v, y))$,

$$y \geq p_v q_j(p_t, x) + z_j(p_t, x) \Rightarrow x < p_t q_j(p_v, y) + z_j(p_v, y).$$

In words, if the bundle $(q_j(p_t, x), z_j(p_t, x))$ could be bought with income y at price p_v , then the bundle $(q_j(p_v, y), z_j(p_v, y))$ should not have been affordable at income x and price p_t .

We can also rewrite this condition as

$$\text{if } p_t > p_v \text{ and } q_j(p_t, x) \geq \frac{x - y}{p_t - p_v}, \text{ then } q_j(p_v, y) > \frac{x - y}{p_t - p_v}, \quad (1)$$

$$\text{if } p_t < p_v \text{ and } q_j(p_t, x) \leq \frac{y - x}{p_v - p_t}, \text{ then } q_j(p_v, y) < \frac{y - x}{p_v - p_t}, \quad (2)$$

which distinguishes between two WARP conditions depending on $p_t > p_v$ or $p_t < p_v$. In our next analysis, we will often use this alternative formulation of WARP (see in particular our proof of Theorem 1).

3 Bounds on counterfactual demands

Consider some measurable set $S \in \Omega$. In what follows, we will use the notation

$$\Pr[S] = \int \mathbb{1}[j \in S] dP(j),$$

³Interestingly, in a two goods setting such as ours, WARP is equivalent to SARP; see Rose (1958) for a proof. Cherchye, Demuynck, and De Rock (2017) provide a recent discussion on the empirical equivalence of WARP and SARP.

where the indicator function $\mathbb{1}[\cdot]$ equals 1 if the expression between brackets is true and 0 otherwise. In words, the function $\Pr[S]$ gives the probability that the expression $j \in S$ holds for a randomly drawn consumer j . Or, putting it differently, it stands for the proportion of consumers that belong to S .

Counterfactual demand analysis wants to predict demand for some unobserved price p_0 and expenditure level x_0 . More specifically, we want to identify information on the distribution of the random variable $q_j(p_0, x_0)$. We address this question by defining upper and lower bounds on the value of $\Pr[q_j(p_0, x_0) \leq q_0]$ for any q_0 . Obviously, for $z_0 = x_0 - p_0 q_0$, we have that $\Pr[q_j(p_0, x_0) \leq q_0] = \Pr[z_j(p_0, x_0) \geq z_0]$.

To define our bounds on $\Pr[q_j(p_0, x_0) \leq q_0]$ we consider the sets $L = \{t \in T | p_t < p_0\}$ and $U = \{t \in T | p_t > p_0\}$ (using that p_0 is different from any observed price), i.e. two mutually exclusive subsets of the observed cross sections. In what follows, the set L will be used to define a WARP-consistent lower bound on $\Pr[q_j(p_0, x_0) \leq q_0]$, and the set U to define a WARP-consistent upper bound.

Lower WARP-consistent bounds. Our first result defines the tightest WARP-consistent lower bound on the value of $\Pr[q_j(p_0, x_0) \leq q_0]$.

Theorem 1. *For all q_0 , we have that*

$$\max_{t \in L} \Pr \left[\begin{array}{l} x_0 \geq p_0 q_{t,j} + z_{t,j} \quad \text{and} \\ x_{t,j} \geq z_0 + p_t q_0 \end{array} \right] \leq \Pr [q_j(p_0, x_0) \leq q_0].$$

This bound is the tightest WARP-consistent lower bound.

We illustrate the reasoning underlying the result in Theorem 1 in the left panel of Figure 1. The solid line depicts the budget line at income level x_0 and price level p_0 . Similarly, the dashed line gives a budget line for cross-section t with observed price $p_t < p_0$ (i.e. $t \in L$).

In particular, this dashed line corresponds to an expenditure level $x_0 + (p_t - p_0)q_0$ (i.e. the budget necessary to buy the bundle (q_0, z_0) at price level p_t). The solid and dashed budget lines intersect at the bundle (q_0, z_0) . For this situation, the lower WARP-consistent bound in Theorem 1 corresponds to the fraction of all observed bundles j in cross-section t that are situated in the shaded area of the figure. These are the bundles for which $x_0 \geq p_0 q_{t,j} + z_{t,j}$ (below the solid line) and $x_{t,j} \geq z_0 + p_t q_0$ (above the dashed budget line).

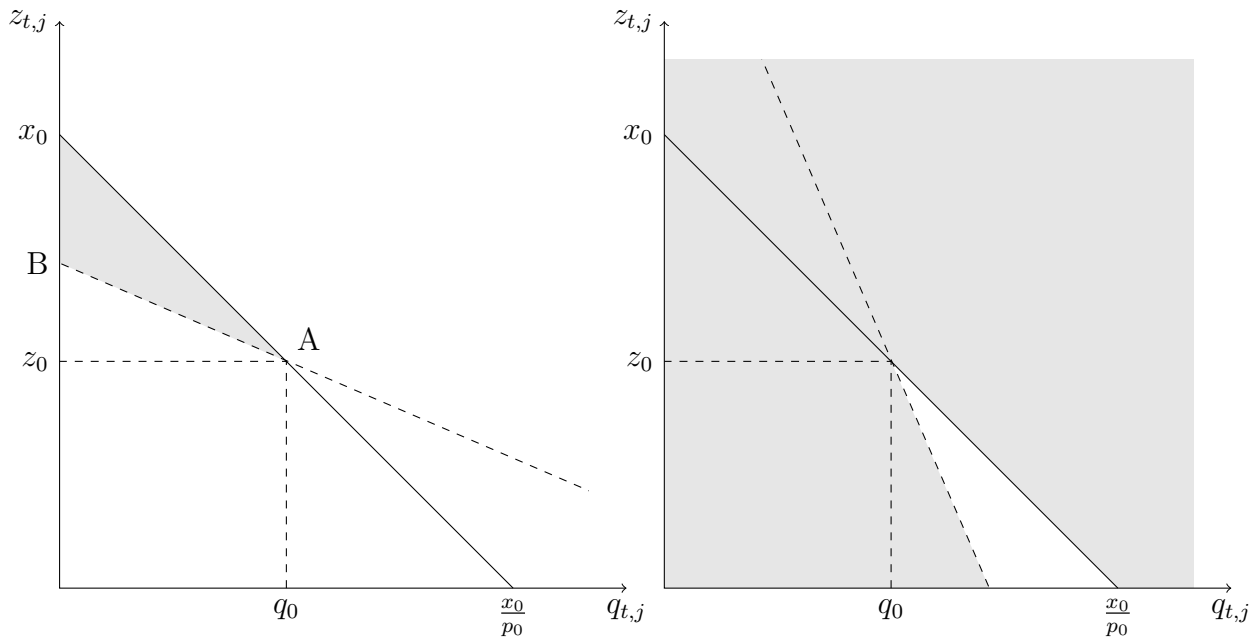
To grasp the intuition behind this lower bound, let us assume that some consumer j chooses a bundle $(q_{t,j}, z_{t,j})$ that lies in the shaded area of the left panel of Figure 1. Then, consistency with WARP requires that, under the new prices p_0 and income x_0 , this consumer cannot consume $(q_j(p_0, x_0), z_j(p_0, x_0))$ on the line segment $[A, x_0/p_0]$ of the solid budget line (as this would imply a violation of the WARP condition in Section 2). As a result, this consumer must consume $(q_j(p_0, x_0), z_j(p_0, x_0))$ on the segment $[x_0, A]$. This implies a consumption of $q_j(p_0, x_0)$ below q_0 , which can be used to define a WARP-consistent lower bound on $\Pr[q_j(p_0, x_0) \leq q_0]$.

Next, we can use the same example to illustrate that Theorem 1 actually defines the tightest WARP-consistent lower bound. In particular, it is easy to see that, if some bundle $(q_{t,j}, z_{t,j})$ is not in the shaded area, then we can always draw an income expansion path for consumer j and price p_t that does not cross the shaded area.⁴ In this case, it is always possible to find an allocation $(q_j(p_0, x_0), z_j(p_0, x_0))$ on the solid budget line with $q_j(p_0, x_0) > q_0$, which implies no violation of WARP. As a result, the lower bound in Theorem 1 effectively produces the highest WARP-consistent lower bound that is possible for the set-up (with minimalistic priors) that we consider.

Upper WARP-consistent bounds. Following an analogous reasoning, we can define the tightest WARP-consistent upper bound on $\Pr[q_j(p_0, x_0) \leq q_0]$. In this case we make

⁴For consumer j and price p_t , the income expansion path defines the coordinates $(q_j(p_t, x), z_j(p_t, x))$ that correspond to varying values of the expenditures x .

Figure 1: lower and upper bounds for $\Pr[q_j(p_0, x_0) \geq q_0]$



use of the observed prices in the set U that we defined above.

Theorem 2. *For all q_0 , we have that*

$$\Pr [q_j(p_0, x_0) \leq q_0] \leq \min_{t \in U} \Pr \left[\begin{array}{l} x_0 \leq z_{t,j} + p_0 q_{t,j} \quad \text{or} \\ x_{t,j} \leq z_0 + p_t q_0 \end{array} \right].$$

This is the tightest WARP-consistent upper bound.

We use the right panel of Figure 1 to illustrate this result. The solid line has the same interpretation as before, and the dashed line now corresponds to a cross-section t with $p_t > p_0$ (i.e. $t \in U$). In this case, the upper WARP-consistent bound in Theorem 2 equals the fraction of observed bundles j with $x_0 \leq z_{t,j} + p_0 q_{t,j}$ or $x_{t,j} \leq z_0 + p_t q_0$, which are situated in the shaded area in the right panel of Figure 1. To see that this

defines an upper bound on $\Pr[q_j(p_0, x_0) \leq q_0]$, we note that this upper bound is actually the complement of the lower bound on the probability $\Pr[z_j(p_0, x_0) \leq z_0]$. Using this, the argument follows directly from the lower bound argument that we gave above (but now applied to $\Pr[z_j(p_0, x_0) \leq z_0]$ instead of $\Pr[q_j(p_0, x_0) \leq q_0]$).

Following a similar logic as before, we can also verify that the shaded area in the right panel of the figure effectively defines the lowest WARP-consistent upper bound that we can define. Specifically, for any consumer j in this shaded area, it is possible to draw an income expansion path (for prices p_t) that does not go through the non-shaded area. For such an expansion path, the corresponding bundle on the budget line for (p_0, x_0) will have q below q_0 , which guarantees consistency with WARP.

For the above examples, our argument that Theorems 1 and 2 define tightest WARP-consistent bounds uses that income expansion paths have not been restricted so far. This suggests that we may improve the identifying power of our method (i.e. obtain tighter demand bounds) by imposing specific structure on the shape of income expansion paths. We explore this further in the following section. In particular, we will consider shape restrictions on income expansion paths that take the form of restricted income elasticities for the consumption goods.

4 Restrictions on income elasticities

From our graphical example in the previous section, it should be clear that, if the cross-sectional prices p_t ($t \in T$) and the counterfactual price p_0 are close to each other, then the WARP-consistent bounds on the probability $\Pr[q_j(p_0, x_0) \leq q_0]$ will be close to the trivial bounds of zero and one. As a result, we can expect the bounds defined in Theorems 1 and 2 to be little informative if there is little cross-sectional price variation.

In such a case, it can be useful to add structure. As indicated above, an obvious

strategy is to impose shape restrictions on the income expansion paths. More specifically, let us consider again the situation in Figure 1 and now assume that we do know the exact income expansion path for each consumer j . Then, if $p_t < p_0$ (i.e. $t \in L$), the fraction of all expansion paths (over the set of consumers J) that go through the shaded area in the left panel of Figure 1 will define a lower bound on $\Pr[q_j(p_0, x_0) \leq q_0]$. Similarly, if $p_t > p_0$ (i.e. $t \in U$), then the fraction of all paths that do not enter the non-shaded area in the right panel of Figure 2 will define an upper bound on $\Pr[q(p_0, x_0) \leq q_0]$.

Of course, in practice we typically do not observe the expansion paths. However, in many cases it makes sense to place some restrictions on them. For example, we may often reasonably assume that goods are normal (i.e. positive income elasticities). In some cases, the empirical analyst may actually be able to define even more stringent bounds on the income elasticities of particular consumption goods (e.g. based on existing empirical studies). In what follows, we will consider these situations in more detail.

Normal goods. If q and z are both normal goods, then the income expansion paths will be increasing in x . For the situation in Figure 1, normality implies that any expansion path of some consumer j that goes through the shaded region in the left panel (i.e. $x_0 \geq p_0 q_j(p_t, x) + z_j(p_t, x)$ and $x \geq p_t q_0 + z_0$) will also go through the line segment BA (i.e. $q_j(p_t, p_t q_0 + z_0) \leq q_0$). We summarize this result in the following lemma.

Lemma 1. *Assume that the goods q and z are both normal and $p_t < p_0$ (i.e. $t \in L$). For any consumer j , there is an income level x such that $x_0 \geq p_0 q_j(p_t, x) + z_j(p_t, x)$ and $x \geq p_t q_0 + z_0$ if and only if $q_j(p_t, p_t q_0 + z_0) \leq q_0$.*

A direct implication of this lemma is that, under normality of the goods q and z , we only need information on $q_j(p_t, p_t q_0 + z_0)$ (i.e. consumption at the expenditure level $x = p_t q_0 + z_0$) to define our WARP-consistent bounds. In terms of Figure 1, these are the bundles that are situated on the dashed budget curve. However, we cannot directly use this result in

practice as, for consumer j and cross-section t , we only observe the consumption quantities $q_{t,j} = q_j(p_t, x_{t,j})$ and $z_{t,j} = z_j(p_t, x_{t,j})$ associated with the expenditure level $x_{t,j}$. We do not observe the counterfactual demand $q_j(p_t, x)$ for an income level $x \neq x_{t,j}$ (including $x = p_t q_0 + z_0$). In a sense, the structure of this problem is similar to the one of a treatment response model in which $q_j(p_t, \cdot)$ is the outcome function and $x_{t,j}$ the treatment (i.e. we only observe the outcome $q_j(p_t, x)$ for the treatment $x = x_{t,j}$).

Our following result builds further on Lemma 1 and circumvents the informational problem that we do not directly observe $q_{t,j} = q_j(p_t, p_t q_0 + z_0)$.

Theorem 3. *Assume that the goods q and z are both normal. For all q_0 , we have that*

$$\max_{t \in L} \Pr \left[\begin{array}{l} q_{t,j} \leq q_0 \quad \text{and} \\ z_{t,j} \geq z_0 \end{array} \right] \leq \Pr [q_j(p_0, x_0) \leq q_0] \leq \min_{t \in U} \Pr \left[\begin{array}{l} q_{t,j} \leq q_0 \quad \text{or} \\ z_{t,j} \geq z_0 \end{array} \right].$$

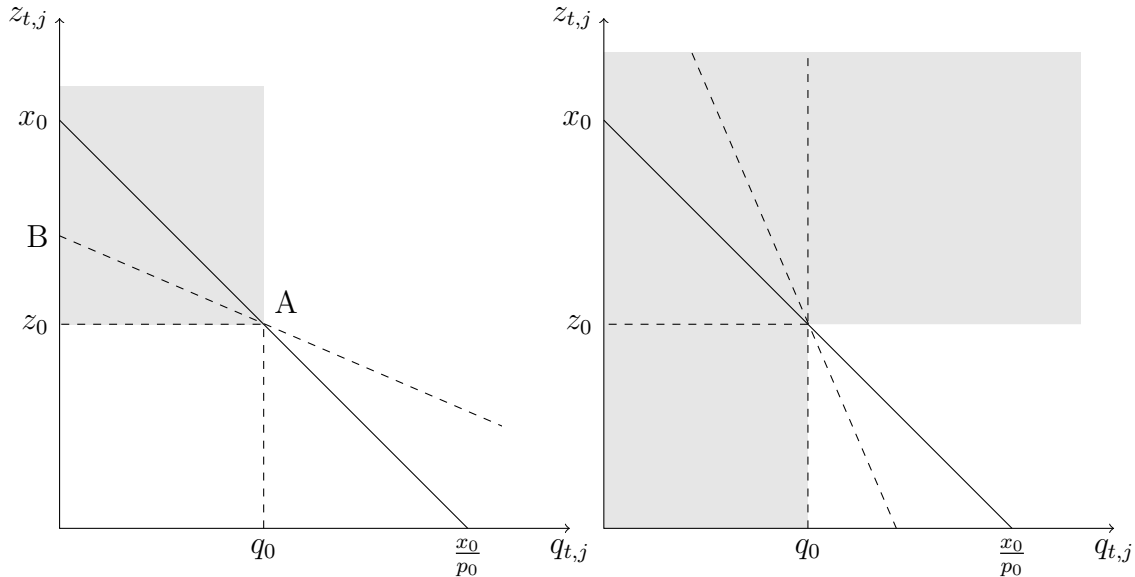
These bounds are the tightest WARP-consistent bounds.

This result is illustrated in Figure 2. Let us first consider the left panel, which presents the lower bound on $\Pr [q_j(p_0, x_0) \leq q_0]$. From Lemma 1, we know that this lower bound is defined by the fraction of expansion paths that go through the line segment BA . Given that both goods are normal, these expansion paths are everywhere increasing. Therefore, any expansion path associated with some observed consumption bundle in the shaded region necessarily crosses the line segment BA .

It is also easy to see that the lower bound in Theorem 3 is the highest WARP-consistent lower bound. In particular, for every consumer j that is not in the shaded region, we can construct a monotonically increasing expansion path that does not cross the line segment BA , which means that that the consumer is not informative for defining our WARP-based lower bound.

The right panel of Figure 2 illustrates the WARP-consistent upper bound that is defined

Figure 2: Lower and upper demand bounds for normal goods



in Theorem 3. The logic of the argument is readily similar to before. Comparing the shaded regions in Figures 1 and 2 clearly reveals that the assumption of normal goods can considerably enhance the identifying power of our method.

Additional restrictions. Let us then turn to the case where the goods q and z are again normal, but now the empirical analyst can use even more information on possible values for the income elasticities of the consumption goods. In particular, we consider parameters $\eta_\ell, \eta_h, \delta_\ell$ and δ_h such that,

$$0 \leq \eta_\ell \leq \frac{\partial q_j(p_t, x)}{\partial x} \frac{x}{q_j(p_t, x)} \leq \eta_h,$$

$$0 \leq \delta_\ell \leq \frac{\partial z_j(p_t, x)}{\partial x} \frac{x}{z_j(p_t, x)} \leq \delta_h.$$

Obviously, normality of q and z can be imposed by setting $0 = \eta_\ell, \delta_\ell$ and η_h, δ_h arbitrarily high. We consider elasticity bounds that are independent of t and income x . In principle,

we can easily allow for different $\eta_\ell, \delta_\ell, \eta_h$ and δ_h associated with different cross-sections or income levels. We choose not to do so here to facilitate our discussion, but the extension is rather straightforward.

The next lemma uses the information on $\eta_\ell, \delta_\ell, \eta_h$ and δ_h to bound the counterfactual values of $q_j(p_t, x)$ and $z_j(p_t, x)$.

Lemma 2. *Assume that the goods q and z are both normal with income elasticities in the intervals $[\eta_\ell, \eta_h]$ and $[\delta_\ell, \delta_h]$. For any consumer j , we have that, for $x \leq x_{t,j}$,*

$$\begin{aligned} q_{t,j} \left(\frac{x}{x_{t,j}} \right)^{\eta_h} &\leq q_j(p_t, x) \leq q_{t,j} \left(\frac{x}{x_{t,j}} \right)^{\eta_\ell}, \\ z_{t,j} \left(\frac{x}{x_{t,j}} \right)^{\delta_h} &\leq z_j(p_t, x) \leq z_{t,j} \left(\frac{x}{x_{t,j}} \right)^{\delta_\ell}, \end{aligned}$$

and, for $x \geq x_{t,j}$,

$$\begin{aligned} q_{t,j} \left(\frac{x}{x_{t,j}} \right)^{\eta_\ell} &\leq q_j(p_t, x) \leq q_{t,j} \left(\frac{x}{x_{t,j}} \right)^{\eta_h}, \\ z_{t,j} \left(\frac{x}{x_{t,j}} \right)^{\delta_\ell} &\leq z_j(p_t, x) \leq z_{t,j} \left(\frac{x}{x_{t,j}} \right)^{\delta_h}. \end{aligned}$$

As we will explain below, this result allows us to define the following WARP-consistent bounds on $\Pr [q_j(p_0, x_0) \leq q_0]$.

Theorem 4. *Assume that the goods q and z are both normal with income elasticities in the intervals $[\eta_\ell, \eta_h]$ and $[\delta_\ell, \delta_h]$. For all q_0 , we have that*

$$\max_{t \in L} \ell_t(q_0) \leq \Pr(q_j(p_0, x_0) \leq q_0) \leq \min_{t \in U} h_t(q_0),$$

where

$$\begin{aligned} \ell_t(q_0) = & \Pr \left[\left(q_{t,j} \left(\frac{p_t q_0 + z_0}{x_{t,j}} \right)^{\eta_h} \leq q_0 \text{ or } z_{t,j} \left(\frac{p_t q_0 + z_0}{x_{t,j}} \right)^{\delta_\ell} \geq z_0 \right) \right. \\ & \left. \text{and } p_t q_0 + z_0 \geq x_{t,j} \right] \\ & + \Pr \left[\left(q_{t,j} \left(\frac{p_t q_0 + z_0}{x_{t,j}} \right)^{\eta_\ell} \leq q_0 \text{ or } z_{t,j} \left(\frac{p_t q_0 + z_0}{x_{t,j}} \right)^{\delta_h} \geq z_0 \right) \right. \\ & \left. \text{and } p_t q_0 + z_0 \leq x_{t,j} \right], \end{aligned}$$

and

$$\begin{aligned} h_t(q_0) = & \Pr \left[q_{t,j} \left(\frac{p_t q_0 + z_0}{x_{t,j}} \right)^{\eta_\ell} \leq q_0 \text{ and } z_{t,j} \left(\frac{p_t q_0 + z_0}{x_{t,j}} \right)^{\delta_h} \geq z_0 \right. \\ & \left. \text{and } p_t q_0 + z_0 \geq x_{t,j} \right] \\ & + \Pr \left[q_{t,j} \left(\frac{p_t q_0 + z_0}{x_{t,j}} \right)^{\eta_h} \leq q_0 \text{ and } z_{t,j} \left(\frac{p_t q_0 + z_0}{x_{t,j}} \right)^{\delta_\ell} \geq z_0 \right. \\ & \left. \text{and } p_t q_0 + z_0 \leq x_{t,j} \right]. \end{aligned}$$

These are the tightest WARP-consistent bounds.

Let us clarify the reasoning behind this result for the first term of $\ell_t(q_0)$, which defines the lower bound on $\Pr [q_j(p_t, x_0) \leq q_0]$. The explanation for the other term of $\ell_t(q_0)$ and for $h_t(q_0)$ proceeds analogously. We have $p_t < p_0$ (i.e. $t \in L$) with $x_{t,j} \leq p_t q_0 + z_0$ for consumer j , i.e. the bundle $(q_{t,j}, z_{t,j})$ is below the dashed budget line in the left panel of Figure 2. Following our above reasoning, we want to know whether the expansion path through this bundle crosses the dashed budget line through the segment BA , which corresponds to $q_j(p_t, p_t q_0 + z_0) \leq q_0$ or, equivalently, $z_j(p_t, p_t q_0 + z_0) \geq z_0$. Because $x_{t,j} \leq p_t q_0 + z_0$, Lemma 2 implies

$$q_j(p_t, p_t q_0 + z_0) \leq q_{t,j} \left(\frac{p_t q_0 + z_0}{x_{t,j}} \right)^{\eta_h},$$

and, thus, we have $q_j(p_t, p_t q_0 + z_0) \leq q_0$ if

$$q_{t,j} \left(\frac{p_t q_0 + z_0}{x_{t,j}} \right)^{\eta_h} \leq q_0.$$

Similarly, Lemma 2 also tells us that

$$z_{t,j} \left(\frac{p_t q_0 + z_0}{x_{t,j}} \right)^{\delta_\ell} \leq z_j(p_t, p_t q_0 + z_0),$$

which implies $z_j(p_t, p_t q_0 + z_0) \geq z_0$ if

$$z_0 \leq z_{t,j} \left(\frac{p_t q_0 + z_0}{x_{t,j}} \right)^{\delta_\ell}.$$

Taken together, for consumer j (with $x_{t,j} \leq p_t q_0 + z_0$) to be informative for our WARP-based lower bound, we need

$$q_{t,j} \left(\frac{p_t q_0 + z_0}{x_{t,j}} \right)^{\eta_h} \leq q_0 \text{ or } z_0 \leq z_{t,j} \left(\frac{p_t q_0 + z_0}{x_{t,j}} \right)^{\delta_\ell},$$

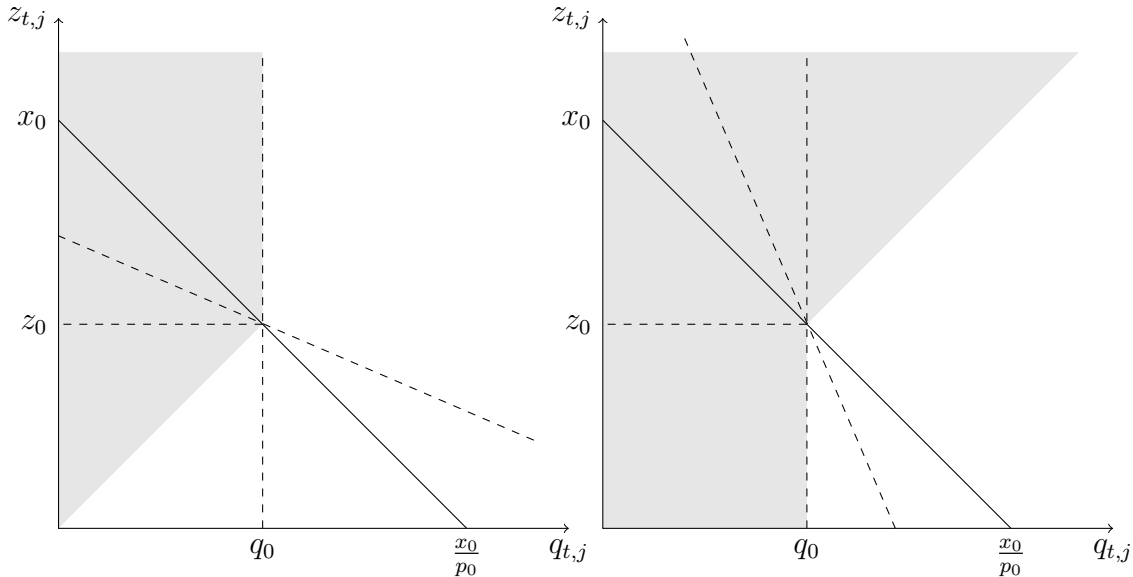
which is the condition that appears in Theorem 4.

To conclude, Figure 3 provides an example for the case where q is a necessary good ($\eta_\ell = 0$ and $\eta_h = 1$) and z a normal good ($\delta_\ell = 0$ and $\delta_h = \infty$). Then, Theorem 4 tells us that

$$\begin{aligned} \ell_t &= \Pr \left[\begin{array}{c} \frac{z_0}{q_0} \leq \frac{z_{t,j}}{q_{t,j}} \\ \text{and } p_t q_0 + z_0 \geq x_{t,j} \end{array} \right] + \Pr \left[\begin{array}{c} q_{t,j} \leq q_0 \\ \text{and } p_t q_0 + z_0 \leq x_{t,j} \end{array} \right], \text{ and} \\ h_t &= \Pr \left[\begin{array}{c} q_{t,j} \leq q_0 \\ \text{and } p_t q_0 + z_0 \geq x_{t,j} \end{array} \right] + \Pr \left[\begin{array}{c} \frac{z_{t,j}}{q_{t,j}} \geq \frac{z_0}{q_0} \\ \text{and } p_t q_0 + z_0 \leq x_{t,j} \end{array} \right], \end{aligned}$$

which defines the shaded areas in Figure 3. Comparing these areas with the ones in Figure

Figure 3: Illustration bounds if q is income inelastic



2 clearly shows the identifying power of using specific elasticity bounds for the consumption goods.

5 Estimation and inference

So far, we have assumed that the joint distribution of $(x_{t,j}, q_{t,j}, z_{t,j})$ is known to the empirical analyst. In practice, we can only use (repeated) cross-sectional data drawn from this distribution, which raises small sample and estimation issues. In our particular case, the parameter of interest is $\theta = \Pr[q_j(p_0, x_0) \leq q_0]$, which is partially identified with lower bound lb and upper bound ub , i.e.

$$lb \leq \theta \leq ub.$$

For example, for the case in Theorem 4 we have

$$lb \equiv \max_{t \in L} \{\ell_t\} \text{ and } ub \equiv \min_{t \in U} \{h_t\},$$

where $\ell_t, h_k \in [0, 1]$ for all $t \in L$ and $k \in U$.

The empirical application of our method requires us to deal with two related issues. First, we need to define the properties of our finite sample estimates of the bounds lb and ub . Second, we have to construct confidence intervals for our partially identified parameter θ .

Bounds estimators. Let us assume an empirical setting in which each cross-section contains n observations. It is fairly easy to obtain finite sample estimators $\hat{\ell}_{t,n}$ and $\hat{h}_{k,n}$ ($t \in L, k \in U$) for the lower and upper bounds that are consistent (and even unbiased), by using the sample analogues of the population statistics. Further, from standard large sample theory, we have that $\sqrt{n}(\hat{\ell}_{t,n} - \ell_t) \xrightarrow{d} \mathcal{N}(0, \ell_t(1 - \ell_t))$ and $\sqrt{n}(\hat{h}_{t,n} - h_t) \xrightarrow{d} \mathcal{N}(0, h_t(1 - h_t))$. Also, by the continuous mapping theorem, the maximum of the lower bound estimates $\hat{lb}_n = \max_{t \in L} \{\hat{\ell}_{t,n}\}$ is a consistent estimator of the population analogue lb and the minimum of the upper bound estimates $\hat{ub}_n = \min_{t \in U} \{\hat{h}_{t,n}\}$ is a consistent estimator of the population analogue ub .

Unfortunately, the limiting distributions of these estimates are not straightforward to obtain. One obvious option would be to use a bootstrap estimator. However, the usual plug-in bootstrap estimators are not valid as the limiting distribution is discontinuous when several of the lower or upper bounds coincide (Andrews, 2000). An alternative that gives asymptotically valid results is based on a subsampling procedure which uses samples of size m ($\ll n$) (without replacement), with $m/n \rightarrow 0$.⁵ We refer to Politis, Romano, and Wolf (1999) for an in-depth presentation of this procedure. If $n \rightarrow \infty$, $m \rightarrow \infty$

⁵In practice, we take $m = n^{0.6}$.

and $m/n \rightarrow 0$, then the asymptotic distribution of $\sqrt{n}(\widehat{lb}_n - lb)$ will coincide with the asymptotic distribution of $\sqrt{m}(\widehat{lb}_m^* - \widehat{lb}_n)$, where \widehat{lb}_m^* is the estimator of lb based on a subsample of size m . The same applies to the asymptotic distributions of $\sqrt{n}(\widehat{ub}_n - ub)$ and $\sqrt{m}(\widehat{ub}_m^* - \widehat{ub}_n)$.

Confidence intervals. We are dealing with a partially identified estimator. This implies a difference between confidence intervals for the set $\Theta = [lb, ub]$ and confidence intervals for the parameter $\theta \in [lb, ub]$ itself.⁶ Here, we focus on confidence intervals for the parameter θ .

Let $\theta = a \cdot lb + (1 - a) \cdot ub$ for $a \in [0, 1]$ and $\Delta = ub - lb$. Then, for given α the confidence interval for θ is defined by c_L and c_U that satisfy

$$\begin{aligned} & \lim_{n \rightarrow \infty} \inf_{\theta \in [lb, ub]} \Pr \left(\widehat{lb}_n - \frac{c_L}{\sqrt{n}} \leq \theta \leq \widehat{ub}_n + \frac{c_U}{\sqrt{n}} \right) = 1 - \alpha \\ \iff & \lim_{n \rightarrow \infty} \inf_{a \in [0, 1]} \Pr \left(\widehat{lb}_n - \frac{c_L}{\sqrt{n}} \leq a \cdot lb + (1 - a) \cdot ub \leq \widehat{ub}_n + \frac{c_U}{\sqrt{n}} \right) = 1 - \alpha \\ \iff & \lim_{n \rightarrow \infty} \inf_{a \in [0, 1]} \Pr \left(\begin{array}{l} \sqrt{n}(\widehat{lb}_n - lb) \leq c_L + (1 - a)\sqrt{n}\Delta \text{ and} \\ \sqrt{n}(\widehat{ub}_n - ub) \geq -c_U - a\sqrt{n}\Delta \end{array} \right) = 1 - \alpha \\ \iff & \lim_{n \rightarrow \infty} \inf_{a \in [0, 1]} \Pr \left(\sqrt{n}(\widehat{lb}_n - lb) \leq c_L + (1 - a)\sqrt{n}\Delta \right) \Pr \left(\sqrt{n}(\widehat{ub}_n - ub) \geq -c_U - a\sqrt{n}\Delta \right) = 1 - \alpha, \end{aligned}$$

where the last step uses the assumption that different cross-sections have independent samples.

Next, if $\sqrt{n}(\widehat{lb}_n - lb) \xrightarrow{d} Z_1$ and $\sqrt{n}(\widehat{ub}_n - ub) \xrightarrow{d} Z_2$, where both Z_1 and Z_2 are random

⁶Constructing confidence intervals for partially identified estimators is a very active area of research. See, among others, Imbens and Manski (2004), Chernozhukov, Hong, and Tamer (2007), Stoye (2009), Bugni (2010) and Chernozhukov, Lee, and Rosen (2013) for various approaches to deal with statistical inference for partially identified parameters.

variables with non-degenerate distributions, we obtain

$$\Pr\left(\sqrt{n}(\widehat{lb}_n - lb) \leq c_L + (1-a)\sqrt{n}\Delta\right) \Pr\left(\sqrt{n}(\widehat{ub}_n - ub) \geq -c_U - a\sqrt{n}\Delta\right)$$

$$\xrightarrow{d} \begin{cases} 1 & \text{if } \Delta > 0, a \in]0, 1[\\ \Pr(Z_1 \leq c_L) & \text{if } \Delta > 0, a = 1 \\ \Pr(Z_2 \geq -c_U) & \text{if } \Delta > 0, a = 0 \\ \Pr(Z_1 \geq c_L) \Pr(Z_2 \geq -c_U) & \text{if } \Delta = 0 \end{cases}.$$

Observe that this limiting probability depends on the value of Δ (i.e. whether $\Delta = 0$ or $\Delta > 0$). Moreover, for n large enough, the confidence interval for the case where $\Delta = 0$ (i.e. point identification) will be larger than for the case where Δ is small but strictly positive.⁷

The subsampling plug-in version of the confidence interval, for $m \rightarrow \infty$ and $m/n \rightarrow 0$, is given by

$$\inf_{a \in [0,1]} \Pr\left(\sqrt{m}(\widehat{lb}_m^* - \widehat{lb}_n) \leq \widehat{c}_L + (1-a)\sqrt{m}\widehat{\Delta}_n\right) \Pr\left(\sqrt{m}(\widehat{ub}_m^* - \widehat{ub}_n) \geq -\widehat{c}_U - a\sqrt{m}\widehat{\Delta}_n\right) = 1 - \alpha,$$

where $\widehat{\Delta}_n = (\widehat{ub}_n - \widehat{lb}_n)$. Because $(\widehat{lb}_n - lb)$ and $(\widehat{ub}_n - ub)$ are both $O(\sqrt{n})$, we have that $(\widehat{\Delta}_n - \Delta)$ is also $O(\sqrt{n})$. As such, for $m \rightarrow \infty$ and $m/n \rightarrow 0$, we obtain

$$\begin{aligned} \sqrt{m}\widehat{\Delta}_n &= \sqrt{m}\left(\widehat{\Delta}_n - \Delta\right) + \sqrt{m}\Delta \\ &= \sqrt{\frac{m}{n}}\sqrt{n}\left(\widehat{\Delta}_n - \Delta\right) + \sqrt{m}\Delta \\ &= o(1) + \sqrt{m}\Delta. \end{aligned}$$

⁷See, among others, Imbens and Manski (2004) and Stoye (2009) for a discussion of this feature when constructing confidence intervals for partially identified parameters.

Therefore,

$$\begin{aligned}
& \Pr\left(\sqrt{m}(\widehat{lb}_m^* - \widehat{lb}_n) \leq \widehat{c}_L + (1-a)\sqrt{m}\widehat{\Delta}_n\right) \Pr\left(\sqrt{m}(\widehat{ub}_m^* - \widehat{ub}_n) \geq -\widehat{c}_U - a\sqrt{m}\widehat{\Delta}_n\right) \\
&= \Pr\left(\sqrt{m}(\widehat{lb}_m^* - \widehat{lb}_n) \leq \widehat{c}_L + (1-a)\sqrt{m}\Delta + o(1)\right) \Pr\left(\sqrt{m}(\widehat{ub}_m^* - \widehat{ub}_n) \geq -\widehat{c}_U - a\sqrt{m}\Delta + o(1)\right) \\
&\xrightarrow{d} \begin{cases} 1 & \text{if } \Delta > 0, a \in]0, 1[\\ \Pr(Z_1 \leq \widehat{c}_L) & \text{if } \Delta > 0, a = 1 \\ \Pr(Z_2 \geq -\widehat{c}_U) & \text{if } \Delta > 0, a = 0 \\ \Pr(Z_1 \geq \widehat{c}_L) \Pr(Z_2 \geq -\widehat{c}_U) & \text{if } \Delta = 0 \end{cases}.
\end{aligned}$$

The result follows from the fact that $\sqrt{m}(\widehat{lb}_m^* - \widehat{lb}_n)$ and $\sqrt{n}(\widehat{lb}_n - lb)$ have the same limiting distribution, which also applies to $\sqrt{m}(\widehat{ub}_m^* - \widehat{ub}_n)$ and $\sqrt{n}(\widehat{ub}_n - ub)$. The reason why the size is correct in the limit is that the estimator $\widehat{\Delta}_n$ converges to Δ at a rate of $O(n^{-1/2})$, which is of order $o(m^{-1/2})$ whenever $m/n \rightarrow 0$.

The asymptotic confidence interval is constructed by finding the values of \widehat{c}_L and \widehat{c}_U that satisfy

$$\inf_{a \in [0,1]} \Pr\left(\sqrt{m}(\widehat{lb}_m^* - \widehat{lb}_n) \leq \widehat{c}_L + (1-a)\sqrt{m}\widehat{\Delta}_n\right) \Pr\left(\sqrt{m}(\widehat{ub}_m^* - \widehat{ub}_n) \geq -\widehat{c}_U - a\sqrt{m}\widehat{\Delta}_n\right) = 1 - \alpha,$$

which allows us to define the $(1 - \alpha) \times 100\%$ confidence interval as

$$\left[\widehat{lb}_n - \widehat{c}_L, \widehat{ub}_n + \widehat{c}_U \right].$$

This produces valid limiting pointwise inference. It should be noticed, however, that this subsampling procedure is not uniformly valid. In particular, the procedure does not converge uniformly to the limiting distribution for values of $\Delta > 0$ close to 0.⁸ We refer to Andrews and Guggenberger (2009a,b) for more discussion.

⁸This can be modeled by considering Δ as a drifting sequence that converges to 0 at a rate of \sqrt{n} .

6 Application

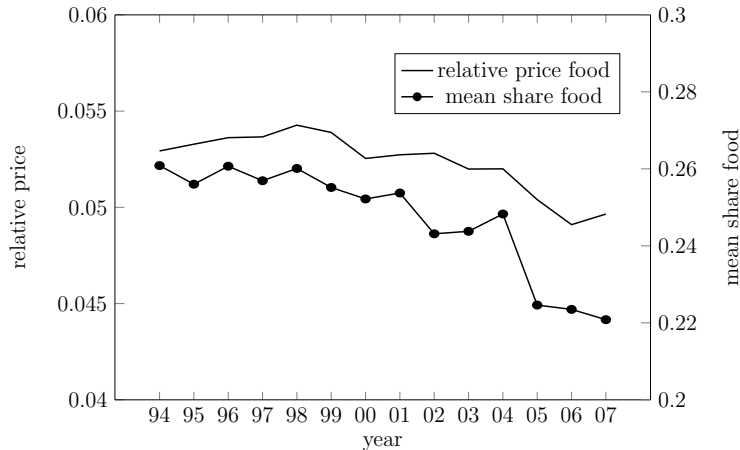
We demonstrate the practical usefulness of our method by means of an empirical illustration using data drawn from the Consumer Expenditure Survey (CEX). We have cross-sectional household consumption data for U.S. households from 1994 until 2007 (14 years), which are obtained from the diary survey contained in the CEX.

Set-up. Following our theoretical discussion, we focus on a setting with two goods, i.e. food (good q) and other (nondurable) goods and services (good z). Analyzing price responsiveness of food consumption relative to other goods and services can be interesting, for example, when debating a possible change of the indirect taxes (VAT) on food (which affects the relative price of food relative to other consumption). More generally, our analysis shows the possible use of our method to anticipate demand effects of rising or falling food prices (e.g. because of political intervention or changed conditions on the food world market). In our following analysis, we will also consider these price effects for the four quartiles of the observed expenditure distribution, so investigating whether these effects vary over income groups.

Our consumers are households, for which we deflate total expenditures by OECD equivalence scales to correct for household size.⁹ We control for observable characteristics by restricting our sample to households (i) that have completed the two-week diary, (ii) that are not living in student housing, (iii) that are vehicle owners, (iv) where both members work at least 17 hours, (v) in which both members are not self-employed, (vi) in which the age of the reference person is at least 21, and (vii) that consist of a husband, a wife and possibly children. By conditioning on housing status and vehicle ownership, we also

⁹In principle, we could explicitly account for differences in household size by adopting a so-called collective model of household consumption (after Chiappori (1988)). See Cherchye, De Rock, and Vermeulen (2007) and Cherchye, De Rock, Lewbel, and Vermeulen (2015) for associated axioms of revealed preference. We see the extension of our method towards collective consumption models as an interesting avenue for follow-up research.

Figure 4: Evolution of relative price and consumption share of food



mitigate the potential impact of our implicit assumption of separability between durable and non-durable consumption.

We deseasonalized our data by using a dummy regression approach. Next, we also removed a number of outlier observations.¹⁰ This gives us an average of 2163 observations (households) per cross-section, with a minimum of 1775 in 1994 and a maximum of 2379 in 1995. The full data set is available from the authors upon request. Figure 4 plots the evolution of the relative price of food (left axis) and the mean consumption share of food (right axis) over the considered time period. This shows that there is not so much price variation, but there is quite some variation in the average share of food consumption. The latter motivates the empirical relevance of our exercise, while the former suggests that we will need to impose extra assumptions in order to obtain meaningful bounds.

Counterfactual demand distribution. As a first exercise, we study the demand distribution for a specific counterfactual price-income scenario. In particular, we set the new

¹⁰In particular, we removed households for which rescaled total expenditures or expenditure shares are not within 3 standard deviations from the mean, households for which rescaled total expenditures are among the 5 percent lowest or 5 percent highest expenditures, and households for which expenditure shares are very close to 0.

price p_0 equal to the mean of the cross-sectional prices p_t over all periods, and we use as reference expenditure level x_0 the median household expenditures in 2007. For our sample of households, we compute bounds on the probability $\Pr[q_j(p_0, x_0) \leq q_0]$ for different values of q_0 . Specifically, we consider alternative budget shares of food $\frac{p_0 q_0}{x_0}$ for the given p_0 and x_0 .

Table 1 contains our main results. It reports bounds and 95% confidence intervals (CIs) for the upper and lower bounds on $\Pr[q_j(p_0, x_0) \leq q_0]$ when the food share $\frac{p_0 q_0}{x_0}$ varies between 0.05 and 0.8. We consider three types of income elasticity restrictions: (i) no restrictions (i.e. the basic scenario of Section 3), (ii) both goods are normal (i.e. positive elasticities), and (iii) food is a necessary good (i.e. income elasticity between zero and one) and other consumption is normal. We believe these scenarios involve very reasonable assumptions for the goods under study.

A few interesting observations emerge from Table 1. If we do not restrict the income elasticities of the goods, our method does not really deliver informative bounds for the current sample of households. In this case, the lower and upper bounds on $\Pr[q_j(p_0, x_0) \leq q_0]$ are everywhere very close (or even equal) to the trivial bounds of, respectively, zero and one. As indicated above, the intuition is that our data set contains too little relative price variation to yield empirically useful results without additional assumptions.

Importantly, however, the bounds do become substantially more informative when we put restrictions on the shape of the income expansion paths. Both the normality assumption and the assumption that food is a necessary good contribute substantially to the identifying power of our method. In fact, for many values of $\frac{p_0 q_0}{x_0}$ our bounds on the counterfactual demand distribution become remarkably tight under the mere (and in our opinion fairly weak) prior that food is a necessary good while the demand for other goods and services is normal. This clearly demonstrates the empirical usefulness of our method. This conclusion is all the more attractive given that, apart from WARP as a minimal

condition of rational consumer behavior, we impose no structure on the unobserved heterogeneity component and do not assume independency between preference heterogeneity and total expenditures.

Bounds on quantiles. We next consider the demand effects of changing food prices for different quantiles of the demand distribution. We maintain the assumption that food is a necessary good and remaining consumption is normal. For food, we will also consider the additional scenario in which the income elasticity is bounded below by 0.5, i.e. $[\eta_\ell, \eta_h] = [0.5, 1]$. Like before, this will give us insight into the identifying power of more stringent elasticity bounds.

Before presenting our empirical results, we recall that, for a fixed price p_0 and expenditure level x_0 , the π th quantile of the distribution of $q_j(p_0, x_0)$ is given by the value $q(\pi)$ that defines

$$\pi = \Pr(q_j(p_0, x_0) \leq q(\pi)).$$

In order to construct bounds on the value of $q(\pi)$, we first parametrize the upper and lower bounds in terms of q_0 , i.e.

$$lb(q_0) \leq \Pr(q_j(p_0, x_0) \leq q_0) \leq ub(q_0).$$

It is easy to show that both $lb(q_0)$ and $ub(q_0)$ are monotone in q_0 . Now, define the functions $q_\ell(\pi)$ and $q_h(\pi)$ such that

$$\pi = lb(q_h(\pi)),$$

$$\pi = ub(q_\ell(\pi)).$$

Table 1: Bounds on $\Pr[q_j(p_0, x_0) \leq q_0]$

$\frac{p_0 q_0}{x_0}$		no elasticity bounds		two normal goods		food as necessity	
		lb	ub	lb	ub	lb	ub
0.05	bounds	0	0.9934	0.0111	0.3758	0.0324	0.0411
	95% CI	0	0.9971	0.0044	0.4004	0.0218	0.0500
0.10	bounds	0.0004	0.9946	0.0473	0.4682	0.0861	0.1567
	95% CI	0	0.9988	0.0349	0.4928	0.0714	0.1734
0.15	bounds	0.0012	0.9952	0.1084	0.5876	0.1871	0.3562
	95% CI	0	0.9980	0.0913	0.6111	0.1667	0.3785
0.20	bounds	0.0016	0.9976	0.1941	0.7177	0.3211	0.5763
	95% CI	0	1	0.1728	0.7395	0.2955	0.6004
0.25	bounds	0.0019	0.9982	0.2975	0.8294	0.4573	0.7552
	95% CI	0	1	0.2707	0.8488	0.4251	0.7799
0.30	bounds	0.0043	0.9988	0.4090	0.9060	0.6019	0.8659
	95% CI	0.0012	1	0.3811	0.9209	0.5715	0.8848
0.35	bounds	0.0055	1	0.4909	0.9518	0.7094	0.9301
	95% CI	0.0028	1	0.4617	0.9635	0.6818	0.9447
0.40	bounds	0.0070	1	0.5649	0.9711	0.7957	0.9614
	95% CI	0.0034	1	0.5401	0.9794	0.7738	0.9720
0.45	bounds	0.0101	1	0.6325	0.9849	0.8595	0.9789
	95% CI	0.0061	1	0.6070	0.9918	0.8418	0.9866
0.50	bounds	0.0129	1	0.6921	0.9922	0.9050	0.9855
	95% CI	0.0083	1	0.6685	0.9970	0.8903	0.9925
0.55	bounds	0.0156	1	0.7464	0.9946	0.9407	0.9916
	95% CI	0.0105	1	0.7245	0.9988	0.9299	0.9968
0.60	bounds	0.0172	1	0.7919	0.9964	0.9643	0.9952
	95% CI	0.0111	1	0.7741	0.9997	0.9558	0.9990
0.65	bounds	0.0191	1	0.8318	0.9986	0.9790	0.9982
	95% CI	0.0137	1	0.8150	1	0.9728	1
0.70	bounds	0.0199	1	0.8710	0.9988	0.9853	0.9982
	95% CI	0.0148	1	0.8547	1	0.9799	1
0.75	bounds	0.0211	1	0.9105	0.9994	0.9920	0.9994
	95% CI	0.0149	1	0.8963	1	0.9891	1
0.80	bounds	0.0234	1	0.9512	1	0.9958	1
	95% CI	0.0165	1	0.9427	1	0.9943	1

Then, we have that

$$\Pr(q_j(p_0, x_0) \leq q(\pi)) = \pi = ub(q_\ell(\pi)) \geq \Pr(q_j(p_0, x_0) \leq q_\ell(\pi)),$$

$$\Pr(q_j(p_0, x_0) \leq q(\pi)) = \pi = lb(q_h(\pi)) \leq \Pr(q_j(p_0, x_0) \leq q_h(\pi)).$$

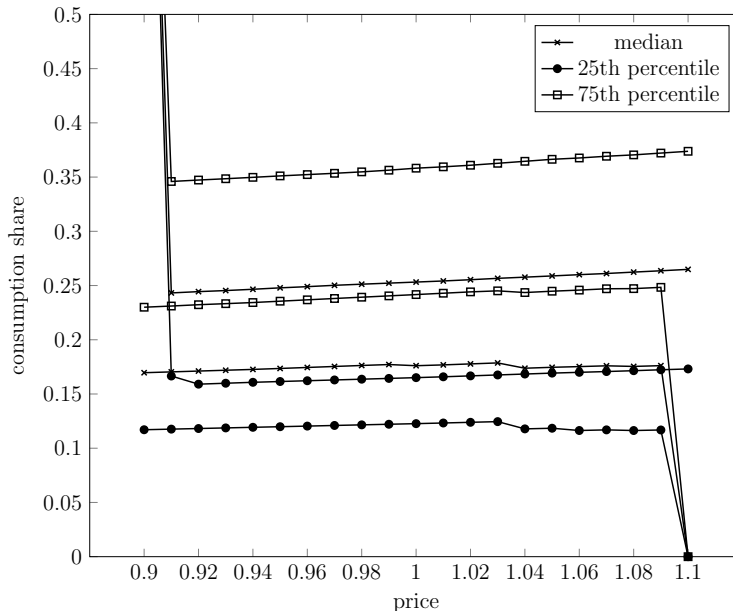
As such, we see that $q_\ell(\pi) \leq q(\pi) \leq q_h(\pi)$.

Figures 5 and 6 present these lower and upper bounds for the two scenarios that we described above, with the elasticity intervals $[\eta_\ell, \eta_h] = [0, 1]$ and $[\eta_\ell, \eta_h] = [0.5, 1]$, respectively. Like before, we set x_0 equal to the median income in 2007, but now we vary the price of food p_0 (measured as a percentage change from the mean price level in our sample). The figures show bounds for the median ($\pi = 0.5$), the 25th percentile ($\pi = 0.25$) and the 75th percentile ($\pi = 0.75$) of the counterfactual distribution of the food share $pq(\pi)/x_0$.

We find that the quantile bounds are fairly tight for price values close to 1 (i.e. prices close to the mean food price in our sample), while they become uninformatively wide for price values around 0.9 or 1.1 (i.e. away from the bulk of observed prices). As expected, we also observe that the budget share for the 25th percentile of the counterfactual demand distribution is significantly below that for the 75th percentile. In Figure 5, there is no very pronounced difference between the bounds for the median quantile demand and the 25th or 75th percentile demand. By contrast, in Figure 6 we do see a clear difference, with the median demand situated distinctively between the 25th and 75th percentile. Generally, we find that the quantile bounds are substantially tighter in Figure 6 than in Figure 5, which again demonstrates the enhancing power effect of imposing specific restrictions on the income elasticity of food.

Different income groups. As a final exercise, we analyze whether demand effects vary over income groups. To this end, we stratified our sample into four groups according to

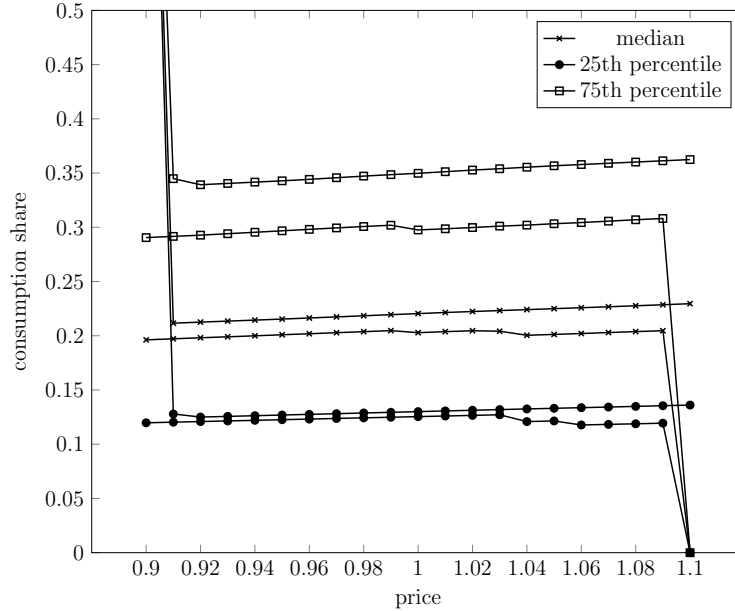
Figure 5: Quantiles bounds for $[\eta_\ell, \eta_h] = [0, 1]$ (food as necessary good)



their quartile in the observed distribution of total expenditures (as a proxy for disposable income). For each of these subgroups, we redid our previous exercise but now we (only) looked at the median budget share of food as a function of varying food prices (again measured as a percentage change from the mean price level in our sample). In doing so, we fix the expenditure x_0 at the subgroup’s median level in 2007. Finally, in contrast to our previous exercises, we no longer impose that food should be a necessary good. Specifically, we use the elasticity interval $[\eta_\ell, \eta_h] = [0, 1.5]$, which means that food may well be a luxury (but not “too luxury”) good for some consumers.

Figure 7 summarizes our results. First, and not really surprisingly, we see clear differences in the (median) budget shares of the different quartiles. In general, the lower the income (quartile), the higher the budget share of food. Also, for each quartile we find that the budget share of food increases with its price. The extent to which the budget share increases is roughly the same for the different income groups; from the figure, we cannot

Figure 6: Quantiles bounds for $[\eta_\ell, \eta_h] = [0.5, 1]$

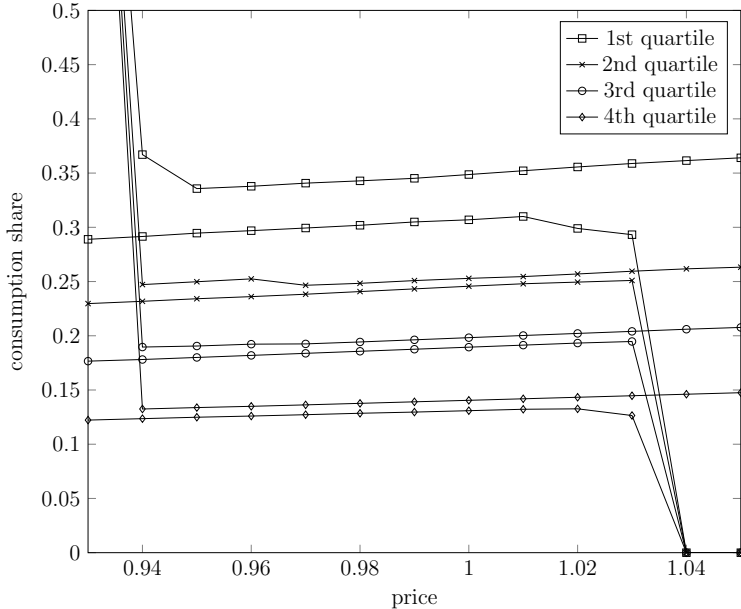


identify an obvious difference in price responsiveness over subgroups. Also, within income subgroups, we do not see a lot of variation in the price responses when varying the food price, at least not for our sample of households.

A notable observation is that the bounds in Figure 7 are again fairly tight. For example, for each of our four subgroups, they are considerably tighter than the median demand bounds in Figure 5. This is all the more remarkable because we actually use a weaker prior for the bounds in Figure 7 (i.e. elasticity interval $[\eta_\ell, \eta_h] = [0, 1.5]$) than for the bounds in Figure 5 (i.e. elasticity interval $[\eta_\ell, \eta_h] = [0, 1]$). An intuitive explanation for this finding is that, by conditioning on income levels, we actually define more homogeneous subgroups (i.e. less unobservable heterogeneity), which in turn yields tighter bounds on the demand distribution.

Summarizing, our different exercises clearly suggest the empirical usefulness of our method to address alternative questions related to counterfactual demand analysis. The

Figure 7: Median demand bounds for income quartiles and $[\eta_\ell, \eta_h] = [0, 1.5]$



method does allow for informative demand predictions under minimalistic prior assumptions. Our results also indicate that it may often be useful to impose specific restrictions on the income elasticities of the consumption goods (e.g. regarding the normal, necessary or luxury nature of goods). This makes our method particularly attractive for applied demand analysis in settings where such assumptions are fairly easy to justify.

7 Conclusion

We have proposed a novel method to derive WARP-consistent bounds on the distribution of demands for counterfactual income and price situations. Our method does not impose any condition on the unobserved heterogeneity component, and it does not assume (conditional) independence between total expenditures and unobserved preference heterogeneity. In addition, the method can readily incorporate restrictions on the income elasticities of the consumption goods, which further enhances its identifying power (i.e. tighter bounds).

We see this as a useful feature from an applied point of view, as we believe that such assumptions are often fairly easy to justify in real-life applications (e.g. based on existing empirical work). Finally, we also discussed small sample and estimation issues related to the practical implementation of our method.

The method is easy to implement and yields informative bounds on demand, which we have demonstrated through an empirical illustration to data drawn from the U.S. Consumer Expenditure Survey (CEX). In this application, we focused on the responsiveness of household demand for food relative to other (nondurable) goods and services as a function of food prices. First, we have shown that our method can deliver meaningfully tight bounds on the counterfactual demand distribution for a fixed new price. Next, we have illustrated that the method can yield informative results on the demand effects of changing prices, for specific quantiles of the demand distribution. Finally, we considered the median demand effect of changing prices for the four quartiles of the observed expenditure distribution in our sample. This last exercise showed the usefulness of our method to investigate differential demand effects over (observable) groups of consumers (in our case income groups).

References

- Andrews, D. W. K., 2000. Inconsistency of the bootstrap when a parameter is on the bound of the parameter space. *Econometrica* 68, 399–405.
- Andrews, D. W. K., Guggenberger, P., 2009a. Hybrid and size-corrected subsampling methods. *Econometrica* 77, 721–762.
- Andrews, D. W. K., Guggenberger, P., 2009b. Validity of subsampling and “plug-in asymp-

- totic inference for parameters defined by moment inequalities. *Econometric Theory* 25, 669–709.
- Beckert, W., Blundell, R., 2008. Heterogeneity and the non-parametric analysis of consumer choice: Conditions for invertibility. *Review of Economic Studies* 75, 1069–1080.
- Blundell, R., Browning, M., Cherchye, L., Crawford, I., De Rock, B., Vermeulen, F., 2015. Sharp for SARP: Nonparametric bounds on the behavioral and welfare effects of price changes. *American Economic Journal: Microeconomics* 7, 43–60.
- Blundell, R., Browning, M., Crawford, I., 2003. Nonparametric engel curves and revealed preference. *Econometrica* 71, 205–240.
- Blundell, R., Browning, M., Crawford, I., 2007a. Improving revealed preference bounds on demand responses. *International Economic Review* 48, 1227–1244.
- Blundell, R., Browning, M., Crawford, I., 2008. Best nonparametric bounds on demand responses. *Econometrica* 76, 1227–1262.
- Blundell, R., Chen, X., Kristensen, D., 2007b. Semi-nonparametric IV estimation of shape-invariant engel curves. *Econometrica* 75, 1613–1669.
- Blundell, R., Duncan, A., 1998. Kernel regression in empirical micro-economics. *Journal of Human Resources* 33, 62–87.
- Blundell, R., Horowitz, J., Parey, M., 2017. Nonparametric estimation of a nonseparable demand function under the Slutsky inequality restriction. *Review of Economics and Statistics* 99, 291–304.
- Blundell, R., Horowitz, J. L., Parey, M., 2012. Measuring the price responsiveness of gasoline demand: Economic shape restrictions and nonparametric demand estimation. *Quantitative Economics* 3, 29–51.

- Blundell, R., Kristensen, D., Matzkin, R., 2014. Bounding quantile demand functions using revealed preference inequalities. *Journal of Econometrics* 179, 112–127.
- Brown, B. W., Walker, M. B., 1989. The random utility hypothesis and inference in demand systems. *Econometrica* 57, 815–829.
- Bugni, F. A., 2010. Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set. *Econometrica* 78, 735–753.
- Cherchye, L., De Rock, B., Lewbel, A., Vermeulen, F., 2015. Sharing rule identification for general collective consumption models. *Econometrica* 83, 2001–20041.
- Cherchye, L., De Rock, B., Vermeulen, F., 2007. The collective model of household consumption: a nonparametric characterization. *Econometrica* 75, 553–574.
- Cherchye, L., Demuynck, T., De Rock, B., 2017. Transitivity of preferences: when does it matter? *Theoretical Economics* forthcoming.
- Chernozhukov, V., Hong, H., Tamer, E., 2007. Estimation and confidence regions for parameter sets in econometric models. *Econometrica* 75, 1243–1284.
- Chernozhukov, V., Lee, S., Rosen, A. M., 2013. Intersection bounds: Estimation and inference. *Econometrica* 81, 667–737.
- Chiappori, P.-A., 1988. Rational household labor supply. *Econometrica* 56, 63–89.
- Cosaert, S., Demuynck, T., 2017. Nonparametric welfare and demand analysis with unobserved individual heterogeneity. *Review of Economic and Statistics* forthcoming.
- Dette, H., Hoderlein, S., Neumeyer, N., 2016. Testing multivariate economic restrictions using quantiles: The example of Slutsky negative semidefiniteness. *Journal of Econometrics* 191, 129–144.

- Hausman, J. A., Newey, W. K., 2016. Individual heterogeneity and average welfare. *Econometrica* 84, 1225–1248.
- Hoderlein, S., 2011. How many consumers are rational. *Journal of Econometrics* 164, 294–309.
- Hoderlein, S., Stoye, J., 2014. Revealed preference in a heterogeneous population. *Review of Economics and Statistics* 96, 197–213.
- Hoderlein, S., Vanhems, A., 2017. Estimating the distribution of welfare effects using quantiles. *Journal of Applied Econometrics* forthcoming.
- Imbens, G. W., Manski, C. F., 2004. Confidence interval for partially identified parameters. *Econometrica* 72, 1845–1857.
- Kawaguchi, K., 2017. Testing rationality without restricting heterogeneity. *Journal of Econometrics* 197, 153–171.
- Kitamura, Y., Stoye, J., 2016. Nonparametric analysis of random utility models. CeMMaP Working Paper CWP36/.13, version June 2016.
- Lewbel, A., 2001. Demand systems with and without errors. *American Economic Review* 91, 611–618.
- Manski, C. F., 1997. Monotone treatment response. *Econometrica* 65, 1311–1334.
- Matzkin, R. L., 2003. Nonparametric estimation of nonadditive random functions. *Econometrica* 71, 1339–1375.
- Matzkin, R. L., 2008. Identification in nonparametric simultaneous equations models. *Econometrica* 76, 945–978.

McFadden, D. L., 2005. Revealed stochastic preference: a synthesis. *Economic Theory* 26, 245–264.

Politis, D. N., Romano, J. P., Wolf, M., 1999. *Subsampling*. Springer.

Rose, H., 1958. Consistency of preference: The two-commodity case. *Review of Economic Studies* 25, 124–125.

Samuelson, P. A., 1938. A note on the pure theory of consumer's behavior. *Economica* 5, 61–71.

Stoye, J., 2009. More on confidence intervals for partially identified parameters. *Econometrica* 77, 1299–1315.

A Proofs

A.1 Proof of Theorem 1

Take a cross-section for which $t \in L$, i.e. $p_t < p_0$. Then, from the WARP condition (1), we have that

$$q_{t,j} \leq \frac{x_0 - x_{t,j}}{p_0 - p_t} \Rightarrow q_j(p_0, x_0) \leq \frac{x_0 - x_{t,j}}{p_0 - p_t},$$

In particular,

$$\begin{aligned} q_{t,j} \leq \frac{x_0 - x_{t,j}}{p_0 - p_t} \text{ and } \frac{x_0 - x_{t,j}}{p_0 - p_t} \leq q_0 &\Rightarrow q_j(p_0, x_0) \leq q_0. \\ \Leftrightarrow \mathbb{1} \left[q_{t,j} \leq \frac{x_0 - x_{t,j}}{p_0 - p_t} \text{ and } \frac{x_0 - x_{t,j}}{p_0 - p_t} \leq q_0 \right] &\leq \mathbb{1} [q_j(p_0, x_0) \leq q_0]. \end{aligned}$$

Integrating both sides with respect to j gives

$$\begin{aligned} & \Pr \left[\begin{array}{l} q_{t,j} \leq \frac{x_0 - x_{t,j}}{p_0 - p_t} \text{ and} \\ x_{t,j} \geq x_0 + (p_t - p_0)q_0 \end{array} \right] \leq \Pr [q_j(p_0, x_0) \leq q_0], \\ \Leftrightarrow & \Pr \left[\begin{array}{l} x_0 \geq p_0 q_{t,j} + z_{t,j} \text{ and} \\ x_{t,j} \geq z_0 + p_t q_0 \end{array} \right] \leq \Pr [q_j(p_0, x_0) \leq q_0]. \end{aligned}$$

A.2 Proof of Theorem 2

Consider a cross-section $t \in U$, i.e. $p_t > p_0$. Then, from the WARP condition (2), we have that

$$q_{t,j} \geq \frac{x_{t,j} - x_0}{p_t - p_0} \Rightarrow q_j(p_0, x_0) \geq \frac{x_{t,j} - x_0}{p_t - p_0},$$

In particular,

$$\begin{aligned} & q_{t,j} \geq \frac{x_{t,j} - x_0}{p_t - p_0} \text{ and } \frac{x_{t,j} - x_0}{p_t - p_0} \geq q_0 \Rightarrow q_j(p_0, x_0) \geq q_0, \\ \Leftrightarrow & \mathbb{1} \left[q_{t,j} \geq \frac{x_{t,j} - x_0}{p_t - p_0} \text{ and } \frac{x_{t,j} - x_0}{p_t - p_0} \geq q_0 \right] \leq \mathbb{1} [q_j(p_0, x_0) \geq q_0]. \end{aligned}$$

Integrating both sides with respect to j gives

$$\begin{aligned}
& \Pr \left[\begin{array}{c} q_{t,j} \geq \frac{x_{t,j} - x_0}{p_t - p_0} \\ x_{t,j} \geq x_0 + (p_t - p_0)q_0 \end{array} \text{ and } \right] \leq \Pr[q_j(p_0, x_0) \geq q_0], \\
\iff & \Pr[q_j(p_0, x_0) \leq q_0] \leq \Pr \left[\begin{array}{c} q_{t,j} \leq \frac{x_{t,j} - x_0}{p_t - p_0} \\ x_{t,j} \leq x_0 + (p_t - p_0)q_0 \end{array} \text{ or } \right], \\
\iff & \Pr[q_j(p_0, x_0) \leq q_0] \leq \Pr \left[\begin{array}{c} x_0 \leq z_{t,j} + p_0 q_{t,j} \\ x_{t,j} \leq z_0 + p_t q_0 \end{array} \text{ or } \right],
\end{aligned}$$

A.3 Proof of Lemma 1

(Sufficiency:) Set $x = p_t q_0 + z_0$. Then, evidently, $x \geq p_t q_0 + z_0$ and

$$\begin{aligned}
p_0 q_j(p_t, x) + z_j(p_t, x) &= p_0 q_j(p_t, p_t q_0 + z_0) + z_j(p_t, p_t q_0 + z_0), \\
&= p_0 q_j(p_t, p_t q_0 + z_0) + p_t q_0 + z_0 - p_t q_j(p_t, p_t q_0 + z_0), \\
&= (p_0 - p_t) q_j(p_t, p_t q_0 + z_0) + p_t q_0 + z_0, \\
&\leq (p_0 - p_t) q_0 + p_t q_0 + z_0, \\
&= p_0 q_0 + z_0 = x_0.
\end{aligned}$$

The first equality follows from the fact that $p_t q_j(p_t, p_t q_0 + z_0) + z_j(p_t, p_t q_0 + z_0) = p_t q_0 + z_0$.

(Necessity:) We prove by contradiction. Assume that there is an income level x such that

$x_0 \geq p_0 q_j(p_t, x) + z_j(p_t, x)$, $x \geq p_t q_0 + z_0$ and $q_j(p_t, p_t q_0 + z_0) > q_0$. Then, we have that

$$\begin{aligned}
x_0 &\geq p_0 q_j(p_t, x) + z_j(p_t, x), \\
&\geq p_0 q_j(p_t, p_t q_0 + z_0) + z_j(p_t, p_t q_0 + z_0), \\
&= p_0 q_j(p_t, p_t q_0 + z_0) + p_t q_0 + z_0 - p_t q_j(p_t, p_t q_0 + z_0), \\
&= (p_0 - p_t) q_j(p_t, p_t q_0 + z_0) + p_t q_0 + z_0, \\
&> (p_0 - p_t) q_0 + p_t q_0 + z_0, \\
&= p_0 q_0 + z_0 = x_0,
\end{aligned}$$

which gives the wanted contradiction. In the above argument, the second inequality uses that both goods are normal.

A.4 Proof of Theorem 3

Consider a cross-section $t \in L$, i.e. $p_t < p_0$. Then,

$$\begin{aligned}
\Pr \left[\begin{array}{l} q_{t,j} \leq q_0 \quad \text{and} \\ z_{t,j} \geq z_0 \end{array} \right] &= \Pr \left[\begin{array}{l} q_{t,j} \leq q_0 \quad \text{and} \quad z_{t,j} \geq z_0 \\ \text{and} \quad x_{t,j} \leq p_t q_0 + z_0 \end{array} \right] + \Pr \left[\begin{array}{l} q_{t,j} \leq q_0 \quad \text{and} \quad z_{t,j} \geq z_0 \\ \text{and} \quad x_{t,j} \geq p_t q_0 + z_0 \end{array} \right], \\
&\leq \Pr \left[\begin{array}{l} z_0 \leq z_j(p_t, p_t q_0 + z_0) \\ \text{and} \quad x_{t,j} \leq p_t q_0 + z_0 \end{array} \right] + \Pr \left[\begin{array}{l} q_j(p_t, p_t q_0 + z_0) \leq q_0 \\ \text{and} \quad x_{t,j} \geq p_t q_0 + z_0 \end{array} \right], \\
&\leq \Pr [q_j(p_0, x_0) \leq q_0]
\end{aligned}$$

The first inequality follows from the assumption that both goods are normal. The last inequality follows from Lemma 1.

By a symmetry argument, we have, for $t \in U$,

$$\begin{aligned} \Pr \left[\begin{array}{l} q_{t,j} \geq q_0 \text{ and} \\ z_{t,j} \leq z_0 \end{array} \right] &\leq \Pr [z_j(p_0, x_0) \leq z_0], \\ \iff \Pr \left[\begin{array}{l} q_{t,j} \leq q_0 \text{ or} \\ z_{t,j} \geq z_0 \end{array} \right] &\geq \Pr [z_j(p_0, x_0) \geq z_0] = \Pr [q_j(p_0, x_0) \leq q_0] \end{aligned}$$

A.5 Proof of Lemma 2

Take the case where $x \leq x_{t,j}$. Then,

$$\begin{aligned} \ln(q_{t,j}) - \ln(q_j(p_t, x)) &= \int_{\ln(x)}^{\ln(x_{t,j})} \frac{\partial \ln(q_j(p_t, z))}{\partial \ln(z)} \, d \ln(z), \\ &\leq \int_{\ln(x)}^{\ln(x_{t,j})} \eta_h \, d \ln(z), \\ &= \eta_h [\ln(x_{t,j}) - \ln(x)]. \end{aligned}$$

This gives

$$q_j(p_t, x) \geq q_{t,j} \left(\frac{x}{x_{t,j}} \right)^{\eta_h}.$$

The other inequalities in Lemma 2 are obtained in a readily similar fashion.

A.6 Proof of Theorem 4

Using that $\eta_\ell, \delta_\ell \geq 0$, Lemma 1 implies, for $t \in L$,

$$\begin{aligned}
P(q_j(p_0, x_0) \leq q_0) &\geq \Pr(q_j(p_t, p_t q_0 + z_0) \leq q_0 \text{ or } z_j(p_t, p_t q_0 + z_0) \geq z_0), \\
&\geq \Pr \left[\begin{array}{l} q_{t,j} \left(\frac{p_t q_0 + z_0}{x_{t,j}} \right)^{\eta_h} \leq q_0 \text{ or } z_{t,j} \left(\frac{p_t q_0 + z_0}{x_{t,j}} \right)^{\delta_\ell} \geq z_0 \\ \text{and } p_t q_0 + z_0 \geq x_{t,j} \end{array} \right] \\
&+ \Pr \left[\begin{array}{l} q_{t,j} \left(\frac{p_t q_0 + z_0}{x_{t,j}} \right)^{\eta_\ell} \leq q_0 \text{ or } z_{t,j} \left(\frac{p_t q_0 + z_0}{x_{t,j}} \right)^{\delta_h} \geq z_0 \\ \text{and } p_t q_0 + z_0 \leq x_{t,j} \end{array} \right], \\
&\equiv \ell_t.
\end{aligned}$$

Similarly, for $t \in U$, we get

$$\begin{aligned}
\Pr(q_j(p_0, x_0) \leq q_0) &\leq 1 - \Pr(q_j(p_t, p_t q_0 + z_0) \geq q_0), \\
&= \Pr(q_j(p_t, p_t q_0 + z_0) \leq q_0). \\
&\leq \Pr \left[\begin{array}{l} q_{t,j} \left(\frac{p_t q_0 + z_0}{x_{t,j}} \right)^{\eta_\ell} \leq q_0 \text{ and } z_{t,j} \left(\frac{p_t q_0 + z_0}{x_{t,j}} \right)^{\delta_h} \geq z_0 \\ \text{and } p_t q_0 + z_0 \geq x_{t,j} \end{array} \right] \\
&+ \Pr \left[\begin{array}{l} q_{t,j} \left(\frac{p_t q_0 + z_0}{x_{t,j}} \right)^{\eta_h} \leq q_0 \text{ and } z_{t,j} \left(\frac{p_t q_0 + z_0}{x_{t,j}} \right)^{\delta_\ell} \geq z_0 \\ \text{and } p_t q_0 + z_0 \leq x_{t,j} \end{array} \right], \\
&\equiv h_t
\end{aligned}$$

