Revealed Preferences in a Heterogeneous Population

Stefan Hoderlein* Jörg Stoye†
Boston College Cornell University

March 1, 2013

Abstract

This paper explores the empirical content of the weak axiom of revealed preference (WARP) for repeated cross-sectional data. Specifically, in a heterogeneous population, think of the fraction of consumers violating WARP as the parameter of interest. This parameter depends on the joint distribution of choices over different budget sets. Repeated cross-sections do not reveal this distribution but only its marginals. Thus, the parameter is not point identified but can be bounded.

We frame this as a copula problem and use copula techniques to analyze it. The bounds, as well as some nonparametric refinements of them, correspond to intuitive behavioral assumptions in the two goods case. With three or more goods, the intuitions break down, and plausible assumptions can have counterintuitive implications. Inference on the bounds is an application of partial identification through moment inequalities. We implement our analysis with the British Family Expenditure Survey (FES) data. Upper bounds are frequently positive but lower bounds not significantly so, hence FES data are consistent with WARP in a heterogeneous population.

Keywords: Revealed Preference, Weak Axiom, Heterogeneity, Partial Identification, Moment Inequalities.

*Stefan Hoderlein, Department of Economics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA, email: stefan_hoderlein@yahoo.com.

†Jörg Stoye, Department of Economics, Cornell University, Uris Hall, Ithaca, NY 14853, stoye@cornell.edu.
Excellent research assistance by Jau-er Chen and Sonya Mihaleva is gratefully acknowledged. Parts of the paper were written while Stoye was visiting Mannheim respectively the Cowles Foundation, whose hospitality is gratefully acknowledged. Thanks for very helpful comments go to an editor and two anonymous referees, as well as to Don Andrews, Dan McFadden, Suyong Song, and conference audiences at Brown, Montréal, and Pompeu Fabra universities and at the Econometric Society North American Winter Meeting in Denver. Any and all errors are ours.

1 Introduction

Motivation. The weak axiom of revealed preference (WARP) is among the core elements of the theory of rational consumer behavior. In a heterogeneous population, agents differ in their specific demand behavior, yet economic theory predicts that they individually obey the weak axiom. This paper explores the empirical content of this prediction for repeated cross-sectional data, i.e. the ability of such data to reject, be consistent with, or even imply (up to sampling uncertainty) the weak axiom. We approach this question as an exercise in bounds or *partial identification*. In particular, we are interested in the fraction of consumers violating WARP. This parameter should be zero according to economic theory. Its empirical value depends on the joint distribution of choices over different budget sets. Repeated cross-sections do not reveal this joint distribution but do reveal the marginal distribution of demand on every single budget set. Bounds on the fraction of consumers who violate WARP are implied; indeed, they are closely related to the classic Fréchet-Hoeffding bounds. We develop these bounds, refine them using nonparametric assumptions, and apply them to the U.K. Family Expenditure Survey. One motivation for this exercise is to provide a complement to the nonparametric estimation of "revealed preference" bounds on behavior derived from similar data sets (e.g., Blundell, Browning, and Crawford (2003, 2008)). We provide some insight as to how much mileage can be gained from strict revealed preference assumptions alone, without additional aggregation assumptions and only invoking weak assumptions on the dependence structures.

To see the gist of the identification problem, suppose one is interested in the joint distribution of demand on just two different (intersecting) budget lines, but one only knows the marginal distribution of demand on each of these budget lines. The intended applications are situations where one faces repeated cross-sections of the population of interest, a setting that corresponds to many practical applications and data sets. Thus, one knows the marginal but not the joint distribution of choices; in other words, the aspect of the relevant distribution that is not identified is precisely the copula.

- figure 1 about here -.

The problem is illustrated in figure 1, which displays two intersecting budget lines and (as shaded areas) the marginal distributions of consumers on those lines. WARP is violated by those consumers whose choices lie in the emphasized segment of each budget line. The proportion of these consumers in the population depends on the displayed marginal distributions but also on the copula linking them. This copula has an intuitive interpretation: Consumers can be thought of as being ordered with respect to their revealed preference for good 2 in any given period. The copula describes how much realignment with respect to this ordering occurs as consumers get relocated from one budget to the other one. Two copulas stand out as extreme: The "best behaved" population might be one where this ordering is precisely maintained; the "worst behaved" (or at least most heterogeneous) one might correspond to its complete reversal. For this simple example and if distributions are continuous, these two dependence structures do indeed generate upper and lower worst-case bounds on $\mathbb{P}$(WARP violated). Furthermore, the two copulas just described correspond to the Fréchet-Hoeffding upper and lower limit copulas. In the continuous case, the problem thus becomes an application of a classic finding.

Our contribution is to observe this connection to the copula literature and to exploit it in numerous ways. First, we develop the result for mixed discrete-continuous distributions, with the above case as corollary, and also show how the resulting bounds can (under assumptions) be integrated over budget sets to bound $\mathbb{P}$(WARP violated) for populations that face heterogeneous budgets. Second, we use the existing literature on copulas, but also some novel ideas, to refine bounds from above and below. In particular, we impose some nonparametric dependence structure between demand in different budgets, i.e. we nonparametrically constrain unobserved heterogeneity, leading to tighter bounds. In the two-good case, it turns out that some such assumptions are both intuitively meaningful (and perhaps reasonable) and qualitatively affect bounds in the way that one might have expected. Third, we generalize the analysis to three and more goods. This generalization has some unpleasant features: While Fréchet-Hoeffding bounds still apply, the according worst-case copulas do not correspond to plausible, or at least easily comprehensible, restrictions on heterogeneity in the population. What is more, mathematically natural generalizations of the aforementioned, partially identifying assumptions fail to have clear intuitions and may have unexpected effects on the bounds. As one particular example, many assumptions which seemingly force the population to be well-behaved can actually induce spurious violations of WARP, that is, they can refine the lower bound of $\mathbb{P}$(WARP violated) away from zero for data that were generated by a rational population.

We finally bring the analysis to a practical application, estimating the bounds on data from the British Household Expenditure Survey. Inference on the bounds is an application of moment inequalities, a recently burgeoning literature in econometric theory that we apply and adapt. The empirical result is that point estimators of bounds indicate occasional violations of WARP but these are far from statistically significant. The data are consistent with WARP, either because consumers exhibit minimal rationality or because WARP is just too weak (or, of course, both).

Related Literature. This paper touches on a number of distinct issues, including the integrability of stochastic demand functions, the theory of copulas, and inference on parameters that are partially identified by moment inequalities. Consequently, there are points of contact to numerous literatures, some of which might be called classic and some of which are currently developing.

This is primarily a paper about revealed preference. The revealed preference approach to consumer demand was introduced and popularized by Afriat (1978), and Varian (1982). This work lays the economic foundations for our approach. However, all empirical applications at the time considered one (usually representative) consumer, and the notion of unobserved heterogeneity did not arise. The closest predecessor to our identification analysis is the literature on integrability of stochastic demand; see, in particular, Falmagne (1978), Barberà and Pattanaik (1986), McFadden and Richter (1991), and McFadden (2005). As we will explain below, some of our results could alternatively (if cumbersomely) be derived from there. One main difference is that we explicitly attack the problem as one of partial identification and consider upper and lower bounds on the fraction of rational consumers rather than just asking whether the lower bound is zero. Perhaps more importantly, by considering WARP rather than SARP we test an even more primitive notion of rationality and turn the problem into one that is precisely suited to the tools developed in the literature on copulas, as well as in the literature on moment inequalities. Also, WARP suffices as a foundation of consumer demand theory (Kihlstrom, Mas-Colell and Sonnenschein (1976)), hence seems natural as an object of interest for an approach that focuses solely on the core objects of consumer rationality.

In this setup, we provide an economic interpretation involving the behavior of individuals in a heterogeneous population, show how this relates to refinements proposed in the copula literature, and establish the sense in which these intuitions break down in the high dimensional case. As already mentioned, our work is also related to applications of revealed preference to consumer demand, in particular by Blundell, Browning, and Crawford (2003, 2008); see Cherchye, Crawford, de Rock, and Vermeulen (2009) for an overview. This literature tests revealed preference theory, and uses it to derive bounds on demand regression. However, their stochastic models of unobserved heterogeneity are limited. For instance, Blundell, Browning, and Crawford (2003) focus on revealed preference analysis using the mean regression, which comes close to imposing a representative agent assumption because additive deviations from the conditional mean cannot in general be generated by a structural model (due to well known aggregation problems of WARP; see, e.g., Mas-Colell, Whinston, and Green (1995), p.110).[1] Our contribution complements this line of research by adding unrestricted heterogeneity. Since we are employing a nonseparable model in a consumer demand setup, our contribution is also nested within the wider econometric literature on nonparametric identification of economic hypotheses using nonseparable models; see

[1]Like us, Blundell, Browning, and Crawford (2003, 2008) focus on the implications of WARP. In current work, Blundell, Browning, Cherchye, Crawford, de Bock, and Vermeulen (2012) extend the approach to SARP.

Matzkin (2006) for a lucid overview.

Formally, the present identification problem is also related to one that received attention in the treatment effects literature. Consider learning about the distribution of a treatment effect, $\Delta \equiv Y_1 - Y_0$ (or variations thereof), when a randomized experiment identifies the marginal distributions of potential outcomes Y_0 and Y_1. Clearly, a similar partial identification problem to ours emerges, namely, marginals are perfectly but copulas are not at all identified. The issue is commonly avoided by focusing attention on the expected value of the treatment, which does not depend on the copula. Researchers genuinely interested in the distribution of the treatment effect have, however, brought Fréchet-Hoeffding's and related bounds to this problem (Heckman, Smith, and Clemens (1997), Manski (1997), Fan and Park (2010)). While motivated by a very different question, this literature has some formal similarities to what we are doing. The technical difference is that we are interested in features of the joint distribution, notably $\mathbb{P}$(WARP violated), that do not correspond to interesting aspects of the distribution of Δ, thus the detail of our identification analysis is quite different. Also, both Heckman, Smith, and Clemens (1997) and Manski (1997) recognize that inference on the resulting bounds is nonstandard but do not focus on it; Fan and Park's (2010) results on inference do not apply here.

Finally, inference on our bounds is an application of moment inequalities, a currently very active literature. While we do not provide a conceptual innovation to this field, it is interesting to note that mechanical application of existing approaches, in particular of Andrews and Soares (2010), can be improved upon by exploiting the specific structure of our bounds. We expect that the same will hold true for many other applications of moment inequalities, and that this paper might accordingly be of interest as case study of such an application.

Structure of the Paper. The remainder of this paper is structured as follows. Section 2 is devoted to identification analysis: We describe and solve the identification problem, that is, we find bounds on the fraction of consumers that violate WARP under the assumption that all observable features of population distributions are known. We provide worst-case bounds as well as bounds that use partially identifying assumptions and conduct this analysis in two as well as more dimensions, with the latter analysis having a qualitatively different message. Section 3 develops the necessary tools for inference. Section 4 contains our empirical application, and section 5 concludes. In the online supplement to this paper, Appendix I contains auxiliary tables, and appendix II collects all proofs.

2 Identification Analysis

This section analyzes identification of the fraction of a population who violate WARP. Thus, we discover what we could learn about this fraction if the population distribution of observables were known. Estimation and inference will be considered later.

Consider, therefore, a population of agents who face an income process $(Y_t)_{t=1,..,T}$ and a consumption set $\mathbb{R}^m_+$, where $m \geq 2$ denotes the number of distinct goods. Individual demand is given by the time invariant function $\mathbf{Q}(y_t, \mathbf{p}_t, \mathbf{z}, \mathbf{a}) : \mathbb{R} \times \mathbb{R}^m_+ \times \mathcal{Z} \times \mathcal{A} \to \mathbb{R}^m_+$, $t = 1, .., T$, where $\mathbf{Q}(y_t, \mathbf{p}_t, \mathbf{z}, \mathbf{a}) \in \{\mathbf{x} \in \mathbb{R}^m_+ : \mathbf{p}_t\mathbf{x} \leq y_t\}$, $\mathbf{z} \in \mathcal{Z} \subseteq \mathbb{R}^l$ denotes observable covariates (assumed time invariant for simplicity) and $\mathbf{a} \in \mathcal{A}$ denotes time invariant unobservable covariates. Note that the function $\mathbf{Q}$ is nonstochastic and constant across consumers; without loss of generality, heterogeneity is absorbed by $\mathbf{A}$. In the repeated cross-section scenario that constitutes our leading application, one would think of $\mathbf{A} = \mathbf{a}_l$ as a consumer with preference $\mathbf{a}_l$. The distribution of $(Y_t, \mathbf{Q}(Y_t, \mathbf{p}_t, \mathbf{Z}, \mathbf{A}), \mathbf{Z})$ is identified for every t in the sample.[2] The sequence $(\mathbf{p}_t)_{t=1,..,T}$ is considered nonstochastic. With slight abuse of notation (by suppressing (y_s, y_t) as arguments), we will also define the random variables $\mathbf{Q}_s \equiv \mathbf{Q}(y_s, \mathbf{p}_s, \mathbf{z}, \mathbf{A})$ and $\mathbf{Q}_t \equiv \mathbf{Q}(y_t, \mathbf{p}_t, \mathbf{z}, \mathbf{A})$ in settings where realizations $(y_s, y_t, \mathbf{z})$ of $(Y_s, Y_t, \mathbf{Z})$ are conditioned upon.

To obtain implications that are testable from our data, we have to assume some structure within and across time periods.

Assumption 1. ***(i)*** *Budgets are exhausted, i.e.* $\mathbb{P}(\mathbf{p}_t'\mathbf{Q}(Y_t, \mathbf{p}_t, \mathbf{Z}, \mathbf{A}) = Y_t) = 1$.

(ii) *For any time periods s and t, $\Delta Y_{st} \equiv Y_s - Y_t$ is independent of $\mathbf{A}$ conditional on $(\mathbf{Z}, Y_t)$.*

Assumption 1(i) can be substantively motivated by nonsatiation or free disposal. It is a practical necessity because our data do not include independent observations on income, i.e. we have to equate income with expenditure. Assumption 1(ii) is standard in the related literature on nonseparable models (Matzkin (2006)). It states that preferences for the goods in question and income changes are independent conditional on current income and household characteristics. For an intuition, suppose there are two types of income shocks, say positive and negative, where the size depends on covariates $(\mathbf{Z}, Y_t)$ (think of this conditioning as allocating an individual to a cell defined by values $\mathbf{z}_0, y_{t,0}$). Suppose further that for a good k, there are two types of individuals, $\mathbf{a}_k$ and $\mathbf{a}_{k'}$ say, where type $\mathbf{a}_k$ idiosyncratically likes good k and type $\mathbf{a}_{k'}$ does not. Then assumption 1(ii) states that, conditional on covariates having a certain value and for both positive and negative income shocks, there must be equal proportions of $\mathbf{a}_k$ and $\mathbf{a}_{k'}$ in the

[2] We use the following conventions: Large letters denote random variables and small letters denote realizations (as well as nonrandom variables). Vectors are identified by bold typeface.

population. This implies identification of the distribution of $(Y_s, Y_t, \mathbf{Q}(Y_s, \mathbf{p}_s, \mathbf{Z}, \mathbf{A}), \mathbf{Z})$ as well as $(Y_s, Y_t, \mathbf{Q}(Y_t, \mathbf{p}_t, \mathbf{Z}, \mathbf{A}), \mathbf{Z})$; conditional on covariates, we can identify the joint distribution of consumption in the respective period, and income across periods, from the respective marginals. The assumption is *not* sufficient to identify the joint distribution of consumption across periods, i.e. of $(Y_s, Y_t, \mathbf{Q}(Y_s, \mathbf{p}_s, \mathbf{Z}, \mathbf{A}), \mathbf{Q}(Y_t, \mathbf{p}_t, \mathbf{Z}, \mathbf{A}), \mathbf{Z})$, from its marginals.

If assumption 1(ii) does not hold, then it is possible to introduce instruments into this framework. In particular, in classical consumer demand often total expenditure is used as an income concept, which is valid under an intertemporal separability assumption on preferences, see Lewbel (1999). In this case, we employ labor income as instrument in a control function fashion, which is the common instrument in the demand literature, see again Lewbel (1999). In our setup, this requires to add control function residuals as additional regressors, and assumption 1(ii) has to be modified to hold conditionally on these residuals, which would be implied if the increments in labor income were jointly independent of $\mathbf{A}$ conditionally on $\mathbf{Z}$ and labor income. Moreover, we could extend our setup to allow the prices to be stochastic. In this case, we would have to modify assumption 1(ii) to allow for price increments to be independent of $\mathbf{A}$ conditionally on $(\mathbf{Z}, \mathbf{Y}_t, \mathbf{P}_t)$. However, given that we have aggregate prices and only limited price variation, the assumption of those being nonrandom seems hardly restrictive.

Fix any two time periods s and t and initially condition on a realization $(y_s, y_t, \mathbf{z})$ of $(Y_s, Y_t, \mathbf{Z})$; integration of the resulting bounds will be considered at the end of this section. Conditional demand is, then, distributed as $\mathbf{Q}_s$ in period s and $\mathbf{Q}_t$ in period t. Recall that the marginal distributions of $\mathbf{Q}_t$ and $\mathbf{Q}_s$ are identified but their joint distribution is not. A given consumer's choices violate WARP if one choice would have been strictly affordable given the other budget, that is, if $\mathbf{p}_s'\mathbf{Q}_t \leq y_s$ and $\mathbf{p}_t'\mathbf{Q}_s \leq y_t$, with at least one inequality being strict. The fraction of consumers who violate WARP (or, equivalently from an identification point of view, the population probability of violating it) is

$$\mathbb{P}_{yz}(\text{WARP violated}) = \mathbb{P}_{yz}\left((\mathbf{p}_s'\mathbf{Q}_t \leq y_s, \mathbf{p}_t'\mathbf{Q}_s < y_t) \;\vee\; (\mathbf{p}_s'\mathbf{Q}_t < y_s, \mathbf{p}_t'\mathbf{Q}_s \leq y_t)\right), \tag{2.1}$$

for all $s, t \in \{1, .., T\}$, where $\mathbb{P}_{yz}(\cdot) \equiv \mathbb{P}(\cdot | Y_s = y_s, Y_t = y_t, \mathbf{Z} = \mathbf{z})$. This probability is a feature of the joint distribution of $(\mathbf{Q}_t, \mathbf{Q}_s)$ and hence, is not identified. We will initially develop bounds on it for the two-good case. This case turns out to be characterized by a tight relation between bounds and meaningful (if not necessarily reasonable) assumptions about the evolution of demand in the population. We then generalize the analysis to three and more goods, illustrating all concepts with an example in the three goods case, where a graphical intuition is still available. The multiple goods case qualitatively differs from the two goods one. Bounds are easily derived by generalizing previous concepts, but plausible conditions on individual behavior are harder to find. Conversely, natural generalizations of the behavioral interpretation of the two-dimensional case will fail to provide reasonable bounds. Assumptions that seem to impose a "more regular" behavior of the population may refine bounds from below but not from above and may spuriously indicate violations of WARP. The reason for this may be found in the difficulty of finding a natural ordering of goods.

2.1 The Two Good Case

2.1.1 The General Result

The two-good case will be developed separately for at least two reasons. First, in the two-good case only, WARP and SARP are equivalent (Rose (1958)), and hence this subsection is really about testing either. Second, the two good case has many intuitive features that do *not* generalize to three or more goods. Thus, fix any two time periods s and t and set $m = 2$, meaning that the consumers' problem is characterized by two time periods and two budget lines $\mathcal{B}_s$ and $\mathcal{B}_t$ in $\mathbb{R}^2_+$. Normalizing $\mathbf{p}_t = (1, p_t)$, we can use assumption 1(i) to write $\mathbf{Q}_t = (y_t - p_t Q_t, Q_t)$, thus it suffices to think in terms of scalar random variables (Q_s, Q_t). Budgets $\mathcal{B}_s$ and $\mathcal{B}_t$ intersect at $Q_s = Q_t = (y_t - y_s)/(p_t - p_s)$. Assume w.l.o.g. that $y_s/p_s > y_t/p_t$, i.e. $\mathcal{B}_s$ has the larger vertical intercept, then WARP is violated iff $Q_t \geq (y_t - y_s)/(p_t - p_s) > Q_s$ or $Q_t > (y_t - y_s)/(p_t - p_s) \geq Q_s$. (See again figure 1, where the distribution of Q_s corresponds to the lighter shaded probability.) The probability of this event is constrained by the marginal distributions of Q_t and Q_s but also depends on how consumers are re-ordered along the budget lines between periods s and t. It can be bounded as follows.

Proposition 1. *Suppose the model of individual demand as outlined above holds. Let assumption A1 hold and assume that the distributions of $\mathbf{Q}_s$ and $\mathbf{Q}_t$ are known. Finally, assume that all probabilities are well defined. Then,*

$$\max\left\{\begin{array}{c} \mathbb{P}_{yz}(Q_s \leq (y_t - y_s)/(p_t - p_s)) - \mathbb{P}_{yz}(Q_t < (y_t - y_s)/(p_t - p_s)) \\ - \min\left\{\mathbb{P}_{yz}(Q_s = (y_t - y_s)/(p_t - p_s)), \mathbb{P}_{yz}(Q_t = (y_t - y_s)/(p_t - p_s)\right\}, 0 \end{array}\right\}$$

$$\leq \mathbb{P}_{yz}(\textit{WARP violated}) \leq$$

$$\min\left\{\begin{array}{c} \max\left\{\mathbb{P}_{yz}(Q_s < (y_t - y_s)/(p_t - p_s)), \mathbb{P}_{yz}(Q_t > (y_t - y_s)/(p_t - p_s)\right\}, \\ \mathbb{P}_{yz}(Q_s \leq (y_t - y_s)/(p_t - p_s)), \mathbb{P}_{yz}(Q_t \geq (y_t - y_s)/(p_t - p_s) \end{array}\right\}.$$

These bounds are tight in the sense that in the absence of further information, both bounds as well as any intermediate value are attainable.

This result applies no matter whether $(\mathbf{Q}_s, \mathbf{Q}_t | Y_s, Y_t, \mathbf{Z})$ is distributed continuously, discretely, or as a mixture of the two. It provides bounds for the parameter of interest that can be determined from the marginals, using only observable, and hence estimable, quantities. The bounds are tight, meaning that they fully exploit the information contained in these quantities. In particular, we show in the proof (in appendix II) that there exist probability distributions that generate the relevant marginals for $\mathbf{Q}_s$ and $\mathbf{Q}_t$ and achieve the bounds. Hence, improving on these bounds is only possible at the price of introducing additional assumptions. Similar remarks apply to all bounds reported later.

2.1.2 Specialization to Continuous Demand

We now specialize proposition 1 to the case where $(\mathbf{Q}_s, \mathbf{Q}_t | Y_s, Y_t, \mathbf{Z})$ is distributed continuously. We will also work with this case, which leads to a rather simple and intuitive result, later on. Whether the continuity assumption is realistic depends on one's perspective. If one thinks of all British consumers as the population of interest, then the true population distribution is of course discrete, albeit so finely grained that the simplification gained from assuming continuity may be worth the price. Continuity is appropriate without any such caveats if one thinks of the U.K.'s populace as a (rather large) sample from a meta-population of interest.

In the continuous case, we can simplify (2.1) as follows:

$$\mathbb{P}_{yz}(\text{WARP violated}) = \mathbb{P}_{yz}\left(\mathbf{p}_s'\mathbf{Q}_t < y_s, \mathbf{p}_t'\mathbf{Q}_s < y_t\right). \tag{2.2}$$

This event has a simple geometric interpretation. Denote the boundary of the time t budget constraint by

$$\mathcal{B}_t \equiv \{\mathbf{q} \geq \mathbf{0} : \mathbf{p}_t'\mathbf{q} = y_t\}$$

and remember that $\mathbb{P}(\mathbf{Q}_t \in \mathcal{B}_t) = 1$ by assumption 1(i). Thus, the set of time t choices that are both consistent with assumption 1(i) and affordable given the time s budget is

$$\underline{\mathcal{B}}_{t,s} \equiv \{\mathbf{q} \geq \mathbf{0} : \mathbf{p}_t'\mathbf{q} = y_t, \mathbf{p}_s'\mathbf{q} < y_s\},$$

the intersection of $\mathcal{B}_t$ with the half-space below (the hyperplane containing) $\mathcal{B}_s$. Hence,

$$\mathbb{P}_{yz}(\text{WARP violated}) = \mathbb{P}_{yz}\left((\mathbf{Q}_s, \mathbf{Q}_t) \in \underline{\mathcal{B}}_{s,t} \times \underline{\mathcal{B}}_{t,s}\right),$$

the probability that both $\mathbf{Q}_s$ and $\mathbf{Q}_t$ are contained in the respective half-spaces; consider again figure 1. Given that the marginal probabilities $\left(\mathbb{P}_{yz}\left(\mathbf{Q}_s \in \underline{\mathcal{B}}_{s,t}\right), \mathbb{P}_{yz}\left(\mathbf{Q}_t \in \underline{\mathcal{B}}_{t,s}\right)\right)$ are known, the problem of bounding $\mathbb{P}_{yz}\left((\mathbf{Q}_s, \mathbf{Q}_t) \in \underline{\mathcal{B}}_{s,t} \times \underline{\mathcal{B}}_{t,s}\right)$ is the original Fréchet-Hoeffding bounding problem, and it can indeed be verified that proposition 1 simplifies to this classic result.

Corollary 2. ***(Fréchet-Hoeffding Bounds)*** *Let the conditions of Proposition 1 be satisfied and assume in addition that the distributions of* $\mathbf{Q}_s$ *and* $\mathbf{Q}_t$ *are continuous. Then*

$$\max\{\mathbb{P}_{yz}\left(\mathbf{Q}_s \in \underline{\mathcal{B}}_{s,t}\right) + \mathbb{P}_{yz}\left(\mathbf{Q}_t \in \underline{\mathcal{B}}_{t,s}\right) - 1, 0\} \\ \leq \mathbb{P}_{yz}(\textit{WARP violated}) \\ \leq \min\left\{\mathbb{P}_{yz}\left(\mathbf{Q}_s \in \underline{\mathcal{B}}_{s,t}\right), \mathbb{P}_{yz}\left(\mathbf{Q}_t \in \underline{\mathcal{B}}_{t,s}\right)\right\}.$$

These bounds are tight.

In the two-dimensional case, the crucial probabilities can be written in terms of $(Q_s, Q_t, p_s, p_t, y_s, y_t)$. Recalling that s is normalized to correspond to the budget line with higher intercept, we have

$$\begin{aligned} \mathbf{Q}_s &\in \underline{\mathcal{B}}_{s,t} \Longleftrightarrow Q_s < (y_t - y_s)/(p_t - p_s) \\ \mathbf{Q}_t &\in \underline{\mathcal{B}}_{t,s} \Longleftrightarrow Q_t > (y_t - y_s)/(p_t - p_s). \end{aligned}$$

Notice in particular that the lower half-space $\underline{\mathcal{B}}_{t,s}$ is an *upper* contour set of Q_t. The Fréchet-Hoeffding Bounds then become

$$\max\left\{\mathbb{P}_{yz}(Q_s < (y_t - y_s)/(p_t - p_s)) - \mathbb{P}_{yz}(Q_t < (y_t - y_s)/(p_t - p_s)), 0\right\} \\ \leq \mathbb{P}_{yz}(\text{WARP violated}) \leq \\ \min\left\{\mathbb{P}_{yz}(Q_s < (y_t - y_s)/(p_t - p_s)), 1 - \mathbb{P}_{yz}(Q_t < (y_t - y_s)/(p_t - p_s))\right\},$$

which is the expression we will work with later on.

We will restrict attention to continuous distributions henceforth. Before turning to extensions of our result, we elaborate further on some corollaries and the relation of this result to the literature.

(i) Empirical Content of WARP A corollary of our result is to identify the empirical content of WARP. Specifically, choice probabilities are consistent with WARP iff the lower bound on $\mathbb{P}_{yz}$(WARP violated) equals zero. This is the case if

$$\mathbb{P}_{yz}(Q_s \leq (y_t - y_s)/(p_t - p_s)) - \mathbb{P}_{yz}(Q_t < (y_t - y_s)/(p_t - p_s)) \\ - \min\left\{\mathbb{P}_{yz}(Q_s = (y_t - y_s)/(p_t - p_s)), \mathbb{P}_{yz}(Q_t = (y_t - y_s)/(p_t - p_s)\right\} \leq 0,$$

which in the continuous case simplifies to

$$\mathbb{P}_{yz}(Q_s \leq (y_t - y_s)/(p_t - p_s)) - \mathbb{P}_{yz}(Q_t < (y_t - y_s)/(p_t - p_s)) \leq 0$$

or equivalently to

$$\begin{aligned}&\mathbb{P}_{yz}\left(\mathbf{Q}_s \in \underline{\mathcal{B}}_{s,t}\right) + \mathbb{P}_{yz}\left(\mathbf{Q}_t \in \underline{\mathcal{B}}_{t,s}\right) \leq 1 \\ \Longleftrightarrow\ &\mathbb{P}_{yz}\left(\mathbf{Q}_s \in \underline{\mathcal{B}}_{s,t}\right) \leq \mathbb{P}_{yz}\left(\mathbf{Q}_t \notin \underline{\mathcal{B}}_{t,s}\right) \\ \Longleftrightarrow\ &\mathbb{P}_{yz}\left(\mathbf{Q}_t \in \underline{\mathcal{B}}_{t,s}\right) \leq \mathbb{P}_{yz}\left(\mathbf{Q}_s \notin \underline{\mathcal{B}}_{s,t}\right).\end{aligned} \tag{2.3}$$

This finding is not new; see (ii) below for previous appearances in the literature. The corollary fully applies to higher dimensions (with notational adaptations indicated later). Intuitively, it means that if a budget plane is rotated, the probability mass on the part of the plane that is rotated toward the origin must weakly shrink, whereas the mass on the part that is rotated away from the origin must weakly expand. There is no empirical content if the "before" and "after" budgets fail to intersect.

These restrictions certainly do not appear very strong. Furthermore, they are implied not only by WARP but also by a host of other restrictions on individual behavior. For example, they are easily derived from the assumption that consumers choose independently (across consumers and time periods) from uniform distributions over budget planes.[3] Nonparametric, assumption-free tests of WARP from cross-sectional data will have accordingly limited power, but this is simply due to the limited empirical content that WARP has on its own.

(ii) Relation to Bounds on $\mathbb{P}$(SARP violated). Our findings are related to classic work on stochastic revealed preference theory (Falmagne (1978), Barberà and Pattanaik (1986), McFadden and Richter (1991), McFadden (2005)). For an idealized problem where the set of possible budgets is finite at the population level, McFadden and Richter (1991) show that consistency of a collection of cross-sectional demand distributions with SARP can be checked by solving a linear programming problem.[4] Intuitively, the value of this problem is the maximal probability mass that can be assigned to rational (in the sense of fulfilling SARP) types within an unrestricted, underlying type space. The difference between this value and 1 is, therefore, the minimal probability mass that must be assigned to irrational (in the same sense) types, i.e. a lower bound on the probability that SARP is violated in the population. It seems clear that an upper bound on this probability could similarly be computed by solving the same problem except for probability mass assigned to irrational types.

[3] This observation resembles a classic discussion by Becker (1962).

[4] McFadden (2005) generalizes the analysis to the case of continuous families of budgets. While this adds many technical intricacies, the operational test of rationality suggested is to perform the same analysis on a (large) finite selection of budgets.

In the two-periods, two-goods case, it can be verified that explicit solution of these programming problems recovers corollary 2 (as it should, recalling that WARP implies SARP in this case). One could therefore derive expression (2.3) by McFadden and Richter's method, as is indeed done in Matzkin (2006). Analogies break down in more complex versions of the problem, including the case of more than two goods and the refinements presented later in this paper. While McFadden and Richter's findings are suggestive of how to bound $\mathbb{P}$(SARP violated) in these cases, the technical problem becomes quite a different one: It is not usefully phrased in terms of copulas, nonparametric refinements of the type discussed below appear elusive, and inference will not be tightly connected to the theory of moment inequalities. We therefore leave this question to future research.

(iii) Relation to the Weak Axiom of Stochastic Revealed Preference. This paper is about demand in a heterogeneous population, not stochastic individual demand. However, connections between these two models were investigated by Bandyopadhyay, Dasgupta, and Pattanaik (BDP henceforth), and some of their findings relate to ours. Consider a model where one individual draws a demand function at random from a set of latent demand functions whenever she faces a choice. This can be formally identified with our model by letting $\mathbf{A}$ index the latent demand functions. In this context, WARP is not a natural restriction on demand, but BDP (1999) proposed a stochastic analog, the weak axiom of stochastic revealed preference (WASRP). Some known results regarding the relation between the two are as follows. If budget sets are exploited, then WASRP is equivalent to a condition that BDP (2004) call "stochastic substitutability." It is easily verified that stochastic substitutability is equivalent to the lower bound from proposition 1 being zero. Furthermore, BDP (2002) show that imposing WARP on each demand function that the individual can draw is strictly stronger than imposing WASRP on her stochastic demand. In our setting, these findings jointly imply that the lower bound must be zero under the algebraic condition under which we show it to be zero (but could also be zero in other cases) and that there must exist examples in which the lower bound is 0 yet the upper bound is strictly positive (in fact, 1 upon examination of their particular example). Proposition 1 obviously improves on this. The improvement can be re-imported into Bandyopadhyay et al.'s setting. If budgets are exploited (which is assumed in their (2004) but not (2002)), then we show that WASRP precisely delineates the empirical content of imposing WARP on latent demand functions, and by providing the upper bound (which can be interpreted in the stochastic utility setting as maximal proportion of latent demand functions that violate WARP), we quantify the wedge between imposing WARP on individual functions and imposing WASRP on stochastic demand.

(iv) Unconditional Bounds Suppose one also knows (or can estimate) the joint distribution of $(Y_1, \ldots Y_T, \mathbf{Z})$; this is a realistic assumption as panel income data sets exist for many countries. Under assumption 1, one can then generate unconditional bounds on $\mathbb{P}$(WARP is violated) by integrating the preceding bounds over $(Y_s, Y_t, \mathbf{Z})$. For the worst-case bounds, this means the following.[5]

Lemma 2.1. *Let the conditions of corollary 2 hold and suppose that the distribution of $(Y_s, Y_t, \mathbf{Z})$ is known. Then*

$$\int \max\{\mathbb{P}_{yz}\left(\mathbf{Q}_s \in \underline{\mathcal{B}}_{s,t}\right) + \mathbb{P}_{yz}\left(\mathbf{Q}_t \in \underline{\mathcal{B}}_{t,s}\right) - 1, 0\} F(d(y_s, y_t, \mathbf{z})) \leq \mathbb{P}(\textit{WARP is violated})$$
$$\leq \int \min\left\{\mathbb{P}_{yz}\left(\mathbf{Q}_s \in \underline{\mathcal{B}}_{s,t}\right), \mathbb{P}_{yz}\left(\mathbf{Q}_t \in \underline{\mathcal{B}}_{t,s}\right)\right\} F(d(y_s, y_t, \mathbf{z})).$$

These bounds are tight.

2.1.3 Nonparametric Refinements

One upshot of the preceding section is that the identification problem is really about copulas, more specifically, about the copula connecting Q_s and Q_t. Recognizing this allows one to refine the above bounds by importing results about copulas. We now present some such assumptions and their exact implications.

The lower and upper bounds on $\mathbb{P}_{yz}$(WARP violated) correspond to measures of dependence between Q_s and Q_t that are extremal in an intuitive sense: They impose that this dependence is either perfectly positive (the α-quantile of Q_t is always realized jointly with the α-quantile of Q_s) or perfectly negative (the α-quantile of Q_t is always realized jointly with the $(1-\alpha)$-quantile of Q_s). Many nonparametric measures of dependence interpolate between these extremes. Restrictions on any of them may induce narrower bounds.

One nonparametric dependence concept that has gained popularity in the copula literature is quadrant dependence:

Definition 3. *The copula linking Q_s and Q_t exhibits positive [negative] quadrant dependence if*

$$\mathbb{P}_{yz}((Q_s, Q_t) \geqq (a, b)) \geq [\leq] \mathbb{P}_{yz}(Q_s \geq a)\mathbb{P}_{yz}(Q_t \geq b))$$

for all scalars a, b.

[5] To keep expressions simple, we here abuse notation: $\underline{\mathcal{B}}_{s,t}$ and $\underline{\mathcal{B}}_{t,s}$ depend on (Y_s, Y_t) and therefore vary over the integrals.

Positive quadrant dependence means that large (and small) values of the individual variables coincide more often than would be expected under independence. In our application, this is to say that consumers who reveal strong taste for good 2 at time s tend to do the same at time t. Negative quadrant dependence is the intuitive opposite.

Imposing quadrant dependence leads to the following refinement.

Lemma 2.2. ***(i)*** *Let the conditions of Proposition 1 hold and assume positive quadrant dependence, then*

$$\max\{\mathbb{P}_{yz}\left(\mathbf{Q}_s \in \underline{\mathcal{B}}_{s,t}\right) + \mathbb{P}_{yz}\left(\mathbf{Q}_t \in \underline{\mathcal{B}}_{t,s}\right) - 1, 0\} \\ \leq \mathbb{P}_{yz}(\textit{WARP is violated}) \leq \\ \mathbb{P}_{yz}\left(\mathbf{Q}_s \in \underline{\mathcal{B}}_{s,t}\right)\mathbb{P}_{yz}\left(\mathbf{Q}_t \in \underline{\mathcal{B}}_{t,s}\right).$$

(ii) *Let the conditions of Proposition 1 hold and assume negative quadrant dependence, then*

$$\mathbb{P}_{yz}\left(\mathbf{Q}_s \in \underline{\mathcal{B}}_{s,t}\right)\mathbb{P}_{yz}\left(\mathbf{Q}_t \in \underline{\mathcal{B}}_{t,s}\right) \\ \leq \mathbb{P}_{yz}(\textit{WARP is violated}) \leq \\ \min\left\{\mathbb{P}_{yz}\left(\mathbf{Q}_s \in \underline{\mathcal{B}}_{s,t}\right), \mathbb{P}_{yz}\left(\mathbf{Q}_t \in \underline{\mathcal{B}}_{t,s}\right)\right\}.$$

These bounds are tight.

Positive respectively negative quadrant dependence therefore neatly separate the worst-case bounds into two regions, one that is associated with positive and one that is associated with negative dependence. The boundary between the regions corresponds to independence. Substantively, it is certainly positive rather than negative quadrant dependence that we mean to suggest as interesting restriction on behavior across choice situations.

Numerous nonparametric measures of dependence can be used to strengthen positive quadrant dependence. In particular, one could impose that the copula exhibit tail monotonicity, stochastic monotonicity, corner set monotonicity, or likelihood ratio dependence (known as affiliation in the auctions literature). See, for example, Nelsen (2006) for definitions of all of these, which are listed roughly in order of increasing stringency. Imposing any of them would lead to the same bounds identified above: All of them imply quadrant dependence, so the bounds cannot be wider; but all of them also allow for independence as boundary case as well as for the relevant one of the original worst-case bounds, so the bounds do not become tighter. Within this family, quadrant dependence therefore stands out as the weakest restriction that generates the above refinement.[6]

[6] Comparisons of nonparametric concepts of positive dependence in Yanamigoto (1972) support the same con-

The approach can be extended by importing other suggestions from the literature on copulas. For example, one could bound dependence between choice on $\mathcal{B}_s$ and choice on $\mathcal{B}_t$ in terms of the medial correlation coefficient (Blomquist's β), the rank correlation coefficient (Spearman's ρ), or Kendall's τ. The resulting bounds on joint c.d.f.'s, and hence on $\mathbb{P}_{yz}(\mathbf{Q}_s \in \underline{\mathcal{B}}_{s,t}, \mathbf{Q}_t \in \underline{\mathcal{B}}_{t,s})$, then follow from known results (Nelsen et al. (2001), Nelsen and Úbeda-Flores (2004)). Alternatively, one could limit the change in quantile position that any given consumer experiences as the budget changes from $\mathcal{B}_s$ to $\mathcal{B}_t$ (closed-form results for such assumptions are available from the authors). We omit elaborations because displaying any of these bounds involves much algebra, but note that numerical evaluation would in all cases be easy.

2.2 The Multiple Goods Case

We now analyze the multiple goods case, emphasizing differences to the two goods one. One can use integration of bounds as in the two good case, and we therefore condition on $(y_s, y_t, \mathbf{z})$ throughout this section. We also continue to restrict attention to continuous distributions. As a result, the Fréchet-Hoeffding bounds from corollary 2 still apply; note in particular that the definitions of $\mathcal{B}_t$ and $\underline{\mathcal{B}}_{t,s}$ did not restrict dimensionality of commodity space. The analysis could also be generalized to mixed continuous-discrete distributions, but the necessary bookkeeping regarding point masses becomes very tedious.

We will keep our discussion of the multiple goods case brief but use the following example to make two cautionary remarks.[7]

Example 1. *Let there be three goods, let* $\mathbf{p}_1 = (10, 6, 5)$, $\mathbf{p}_2 = (5, 10, 6)$, *and* $y_1 = y_2 = 30$. *The according budget hyperplanes are most easily described by their intercepts:* $\mathcal{B}_1$ *is spanned by* $((3,0,0),(0,5,0),(0,0,6))$, $\mathcal{B}_2$ *is spanned by* $((6,0,0),(0,3,0),(0,0,5))$. *Assume that* $\mathbf{Q}_1$ *is supported on* $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3) = ((5/7, 0, 32/7), (1, 0, 28/7), (9/7, 20/7, 0))$, *that* $\mathbf{Q}_2$ *is supported on* $(\mathbf{b}_1, \mathbf{b}_2) = ((0,0,5),(6,0,0))$, *and that the joint distribution of* $(\mathbf{Q}_1, \mathbf{Q}_2)$ *is characterized by the following population-level contingency table, where the bold row and column indicate marginal distributions.*

clusion, i.e. quadrant dependence is weakest among large classes of such concepts, none of which would lead to tighter bounds. Two concepts that are insufficient to generate the above bounds are positive correlation and a positive value of Kendall's τ.

[7] For easy verifiability, the example uses mass points, but it is not dependent on them – all statements made about the example also hold true if the mass points are "fudged" into uniform distributions on ε-balls around them.

	$\mathbf{b}_1$	$\mathbf{b}_2$	
$\mathbf{a}_1$	1/4	0	**1/4**
$\mathbf{a}_2$	0	1/4	**1/4**
$\mathbf{a}_3$	1/4	1/4	**1/2**
	1/2	**1/2**	

It is easily calculated that of the support of $\mathbf{Q}_1$*, only* $\mathbf{a}_2$ *lies on* $\underline{\mathcal{B}}_{1,2}$*. On the support of* $\mathbf{Q}_2$*, it is* $\mathbf{b}_1$ *that lies on* $\underline{\mathcal{B}}_{2,1}$*. Thus, the population does not violate WARP. Using only the marginals, one would find* $\mathbb{P}_{yz}(\mathbf{Q}_1 \in \underline{\mathcal{B}}_{1,2}) = 0.25$*,* $\mathbb{P}_{yz}(\mathbf{Q}_2 \in \underline{\mathcal{B}}_{2,1}) = 0.5$*, and hence* $0 \leq \mathbb{P}_{yz}(WARP\ violated) \leq 0.25$.

Remark 1: Extremal copulas do not have an easy economic interpretation any more. The copulas achieving the bounds in example 1 have a geometric interpretation. Order elements of $\mathcal{B}_s$ increasingly according to $\mathbf{p}_t'\mathbf{q}$, i.e. according to how expensive they would be given time t prices. Similarly, order elements of $\mathcal{B}_t$ according to $\mathbf{p}_s'\mathbf{q}$ but in decreasing order. These orderings have in common that their level sets are parallel to $\mathcal{B}_s \cap \mathcal{B}_t$; furthermore, they identify $\underline{\mathcal{B}}_{s,t}$ as a lower and $\underline{\mathcal{B}}_{t,s}$ as an upper contour set.[8] Then the lower Fréchet-Hoeffding bound is achieved by assuming that all consumers maintain their quantile position with respect to these orderings, and the upper bounds are achieved by inversion of the orderings.

In two dimensions, this procedure has the interpretation given before, i.e. the copulas correspond to minimal respectively maximal reordering of consumers with respect to revealed preference for good 1 over good 2. This interpretation is now lost. The ordering to be maintained or reversed is according to how much one would have to pay consumers (or tax them) so that they could just afford their previous consumption bundle. This ordering does not have much economic significance.[9] In particular, it is hard to see how an ordering of subjects on $\mathcal{B}_s$ that depends on time t prices (and vice versa) would arise from natural restrictions on behavior.

[8] Tie-breaking rules of the orderings matter only if distributions have mass points, which were assumed away.

[9] To see a vestige of economic meaning, assume for the moment that preferences exist and consider the problem of bounding the compensating variation for a consumer who moves from $\mathcal{B}_s$ to $\mathcal{B}_t$. With the few assumptions imposed here, a lower bound on this CV will always be zero because the consumer could be indifferent across all bundles. The upper bound is given by $(\mathbf{p}_t'\mathbf{q}_s - y_t)$, the payment needed so that the consumer can afford her previous bundle. The ordering over $\mathcal{B}_s$ therefore accords with the upper bound on compensating variation as consumers move from $\mathcal{B}_s$ to $\mathcal{B}_t$, whereas a similar reasoning reveals that the ordering on $\mathcal{B}_t$ accords with an upper bound on equivalent variation. Perfectly positive dependence then means that both orderings coincide, perfectly negative dependence maximizes their disagreement.

Remark 2: Restrictions on copulas are harder to justify and may have unexpected effects. While the copula that generates lower worst–case bounds has no clean interpretation, one might want to maintain quantile constancy with respect to "revealed preference for good k" anyway. The effects of such assumptions can still be computed numerically. We will now illustrate in example 1 that they can be rather counterintuitive; thus, example 1 is intended as somewhat of a cautionary tale. To this end, consider the following definitions:

Definition 4. *Let* $\left(\overline{\mathbf{Q}}_s, \overline{\mathbf{Q}}_t\right)$ *and* $(\widetilde{\mathbf{Q}}_s, \widetilde{\mathbf{Q}}_t)$ *be distributed independently according to the distribution of* $(\mathbf{Q}_s, \mathbf{Q}_t)$*. Then we can define the following properties of the copula joining* $\mathbf{Q}_s$ *and* $\mathbf{Q}_t$*.*

Co-ordinatewise quantile constancy:

$$\left(\overline{\mathbf{Q}}_s - \widetilde{\mathbf{Q}}_s\right) \odot \left(\overline{\mathbf{Q}}_t - \widetilde{\mathbf{Q}}_t\right) \geqq \mathbf{0} \text{ a.s.},$$

where $\odot$ *denotes the componentwise product. In words, any matching components of* $\left(\overline{\mathbf{Q}}_s - \widetilde{\mathbf{Q}}_s\right)$ *and* $\left(\overline{\mathbf{Q}}_t - \widetilde{\mathbf{Q}}_t\right)$ *are concordant (their product is non-negative) a.s.*

Association:

$$\mathbb{P}_{yz}(\mathbf{Q}_s \in A, \mathbf{Q}_t \in B) \geq \mathbb{P}_{yz}(\mathbf{Q}_s \in A)\mathbb{P}_{yz}(\mathbf{Q}_t \in B)$$

for all upper contour sets A *and* B*.*

Co-ordinatewise positive quadrant dependence:

$$\mathbb{P}_{yz}(Q_s^{[i]} \geq a, Q_t^{[i]} \geq b) \geq \mathbb{P}_{yz}(Q_s^{[m]} \geq a, Q_t^{[m]} \geq b)$$

for all scalars a, b *and orthants* $i \leq m$*; here,* $Q_t^{[i]}$ *is the i-th component of* $\mathbf{Q}_t$*.*

Co-ordinatewise quantile constancy is rather strong; it is not only point identifying but testable in the sense of generating cross-marginal restrictions that might be violated in the data.[10] We therefore also showcase two weakenings of it that are assumptions about copulas proper and, therefore, not testable from observation of marginals. Association is strictly stronger than co-ordinatewise positive quadrant dependence.

All of these assumptions seem to enforce some consistency of tastes and hence, appear optimistic in the sense of limiting the probability of violating WARP. This impression is misleading – the assumptions may actually refine bounds from below but not from above.

Example 2. ***(solution)*** *The population marginals are compatible with co-ordinatewise quantile constancy. They then imply that* $\mathbb{P}(\mathit{WARP\ violated}) = 0.25$*. If the location of* $\mathbf{b}_2$ *were changed to*

[10]Hence, despite our wording, it is not strictly an assumption about copulas only and accordingly not known in the copulas literature. Heterogeneous Cobb-Douglas preferences constitute an important special case where the assumption is fulfilled.

$(0,3,0)$, *co-ordinatewise quantile constancy would have to be violated. Imposing association implies* $0.125 \leq \mathbb{P}_{yz}(WARP\ violated) \leq 0.25$.[11] *Imposing co-ordinatewise positive quadrant dependence does not affect the bounds.*

Co-ordinatewise quantile constancy point identifies $\mathbb{P}_{yz}$(WARP violated) at its *highest* possible level, and association refines away the *lower* half of the identified set. In particular, the first assumptions imply spurious violations of WARP – the lower bound becomes strictly positive even though the data were generated by rational consumers. Our example indicates that in the multiple goods case, apparently reasonable assumptions may be at tension with WARP; we see this as cautionary advice for users who wish to use specifications of this type in nonparametric analysis of demand. The intuitive reason why the example can be constructed is that the ordering of consumers that is maintained – at least in some stochastic sense – is simply not relevant for the Fréchet-Hoeffding problem.

3 Hypothesis Tests and Confidence Intervals

Estimation of the bounds developed in this paper presents a relatively routine nonparametric estimation problem. Inference, however, raises a number of conceptual and technical issues that are the subject of a currently active literature, notably (for our purposes) Andrews and Soares (2010), Imbens and Manski (2004), and Stoye (2009). To tackle these, we continue to assume continuity of relevant population distributions, and we also focus on worst-case Fréchet bounds. Thus, we use corollary 2 to construct estimators and confidence regions for parameters of interest

$$\theta_{yz} \in \Theta_{yz} = [\max\{\pi_{yz} - \psi_{yz}, 0\}, \min\{\pi_{yz}, 1 - \psi_{yz}\}],$$

where

$$\begin{aligned} \pi_{yz} &= \mathbb{P}_{yz}(Q_s < (y_t - y_s)/(p_t - p_s)) \\ \psi_{yz} &= \mathbb{P}_{yz}(Q_t < (y_t - y_s)/(p_t - p_s)). \end{aligned}$$

We will estimate Θ_{yz} by the plug-in estimator

$$\widehat{\Theta}_{yz} = \left[\max\left\{\widehat{\pi}_{yz} - \widehat{\psi}_{yz}, 0\right\}, \min\left\{\widehat{\pi}_{yz}, 1 - \widehat{\psi}_{yz}\right\}\right],$$

where the estimators $\left(\widehat{\pi}_{yz}, \widehat{\psi}_{yz}\right)$ are defined below. The resulting inference problem is somewhat intricate and of interest in its own right for users who wish to apply Fréchet-Hoeffding bounds in

[11] This claim is established in the online appendix.

other contexts. Some issues are as follows. First, should a confidence region cover the identified set, i.e. the relevant coverage probability is $\mathbb{P}(\Theta_{yz} \subseteq CI_{1-\alpha})$, or should it cover the partially identified parameter, i.e. the relevant coverage probability is $\inf_{\theta_{yz}\in\Theta_{yz}} \mathbb{P}(\theta_{yz} \in CI_{1-\alpha})$? The answer plainly depends on what is conceived as the quantity of interest. Both approaches have been pursued in the literature. The empirical application will employ confidence regions for θ_{yz}, but we develop both types.

Second, some "naive" confidence regions will suffer from known problems with inference on moment inequalities. Identification through Fréchet bounds is an instance of identification through moment inequalities because

$$\max\{\pi_{yz} - \psi_{yz}, 0\} \leq \theta_{yz} \leq \min\{\pi_{yz}, 1 - \psi_{yz}\}$$

is equivalent to the conjunction of

$$\pi_{yz} - \psi_{yz} - \theta_{yz} \leq 0 \tag{3.1a}$$

$$-\theta_{yz} \leq 0 \tag{3.1b}$$

$$\theta_{yz} - \pi_{yz} \leq 0 \tag{3.1c}$$

$$\theta_{yz} - 1 + \psi_{yz} \leq 0. \tag{3.1d}$$

Confidence intervals will be lower contour sets of a test statistic that aggregates violations of sample versions of these inequalities. The limiting distribution of this statistic depends on which – if any – inequalities bind, and its distribution for a given, finite sample size can also be influenced by inequalities that are close to binding. It is not possible to pre-estimate the identities of the binding inequalities, or the slackness of the non-binding ones, with sufficient precision for such a "model selection" step to be ignorable. We resolve this by using conservative pre-tests. Conceptually, this method is by now well understood, including results for very general settings. By exploiting the specific structure of the present problem – e.g., the knowledge that at least one of (3.1a-3.1b) and one of (3.1c-3.1d) bind and the fact that (3.1b) is a nonnegativity constraint –, we can however improve on mechanical application of existing approaches. In addition, confidence regions for θ_{yz} (as opposed to Θ_{yz}) may encounter specific problems if Θ_{yz} is short, that is, if the binding one of (3.1a-3.1b) and the binding one of (3.1c-3.1d) are close to each other. This issue received close attention in Imbens and Manski (2004) and Stoye (2009) and will be taken care of.

Finally, (π_{yz}, ψ_{yz}) are probabilities, hence asymptotic normality of estimators will not hold in a uniform sense if at least one of $\{\pi_{yz}, \psi_{yz}\}$ approaches 0 or 1. This problem necessitates numerous case distinctions, making for confidence regions that appear quite involved. However, inference is in some sense easier in these boundary cases because estimators of (near) degenerate probabilities are superefficient. We are, therefore, able to deal with these issues without making confidence regions larger than is standard for nonparametrically estimated parameters.

3.1 Estimation

We first establish some properties of estimators $\left(\widehat{\pi}_{yz}, \widehat{\psi}_{yz}\right)$ of marginal probabilities (π_{yz}, ψ_{yz}). In particular, these will be Nadaraya-Watson estimators applied to the model

$$W_i \equiv 1\{Q_{is} < (y_t - y_s)/(p_t - p_s)\} = \pi_{yz} + \eta_i,$$

where $1\{.\}$ denotes the indicator function, or of course the same model with (ψ_{yz}, t) replacing (π_{yz}, s). Thus, the estimator as a function of (y_s, y_t, z) is

$$\widehat{\pi}_{yz} = \frac{\sum_i K\left((Y_{is} - y_s, \mathbf{Z}_i - \mathbf{z})/h\right) \cdot 1\{Q_{is} < (y_t - y_s)/(p_t - p_s)\}}{\sum_i K\left((Y_{is} - y_s, \mathbf{Z}_i - \mathbf{z})/h\right)}$$

and similarly for (ψ_{yz}, t). In this model, $Var(\eta_i|\mathbf{Z}_i = \mathbf{z}, Y_i = y) = \Xi_{yz} = \pi_{yz}(1 - \pi_{yz})$. If π_{yz} is not close to $\{0, 1\}$, asymptotic normality of these estimators is standard. However, it will be crucial to understand the behavior of this estimator also in situations where π_{yz} is close to 0. Consider, therefore, the possibility that $\pi_{yz} = \pi_n \to 0$ as $n \to \infty$; note that to limit the number of subscripts, we here drop the conditioning variables from notation. We will focus on the case where $\pi_n = c_{yz} n^{-\beta}$ with $0 < \underline{c} \leq c_{yz} \leq \overline{c} < \infty$ for all $(y, \mathbf{z})$ and $0 \leq \beta \leq 1 - \varepsilon$. (The case of $\pi_n \to 1$ is covered analogously. Also, recall that ψ_{yz} is just π_{yz} evaluated at a different point.) We will impose the following primitive assumptions on the d.g.p. and on our kernel K.

Assumption 2. ***(i)*** $(Y_i, \mathbf{Z}_i)_{i=1,\ldots,n}$ *is an independent and identical sequence of d-dimensional random vectors drawn from* F_{YZ}.

(ii) $(Y_i, \mathbf{Z}_i)$ *is continuously distributed with common density* f_{YZ}*, where* f_{YZ} *is bounded and* $f_{YZ}(z) > 0\ \forall (y, z)$.

(iii) *The function* c_{yz} *(hence,* π_n *or* ψ_n*) is twice continuously differentiable with uniformly bounded Hessian* $H(c_{yz})$.

(iv) K *is symmetric about* 0 *(hence,* $\int uK(u)du = 0$*),* $K \geq 0$, $\int K(u)du = 1$*, and* $\kappa_2 \equiv \int K^2(u)du < \infty$.

(v) $h_n \to 0$, $n^{1-\beta}h^d \to \infty$.

(vi) $n^{1-\beta}h^{d+4} \to 0$.

Note that the boundedness restriction (iii) is not as restrictive in this scenario. Also, the below result requires that $E(|\eta|_i^{2+\delta} |Y_i = y, \mathbf{Z}_i = \mathbf{z})$ is finite for some $\delta > 0$, but this obtains trivially because the dependent variable takes values in $\{0, 1\}$. We then have:

Lemma 3.1. *Let the model be as defined above, and let Assumption 2 hold. Then:*

$$\phi_n \sigma_n^{-1} \left(\hat{\pi}_{yz} - \pi_n - h^2 n^{-\beta} \frac{\kappa_2}{2} \iota' H\left(c_{yz}\right) \iota \right) \xrightarrow{d} \mathcal{N}(0,1),$$

where $\phi_n = \sqrt{nh^d}$, ι *is a conformable vector of ones,* $\kappa_2 = \int K(t)^2 dt$, *and* $\sigma_n = \sqrt{c_{yz} n^{-\beta}}$. *In addition, let* $\widehat{\sigma}_{yz} = \sqrt{\hat{\pi}_{yz}(1-\hat{\pi}_{yz})}$, *then* $\widehat{\sigma}_{yz}/\sigma_n \xrightarrow{p} 1$.

Inference will be based on this lemma applied to both $\hat{\pi}_{yz}$ and $\widehat{\psi}_{yz}$ and the observation that asymptotically, these two estimators are distributed independently of each other. An important implication of the lemma is that the asymptotic variance of $\hat{\pi}_{yz}$ (i.e., σ_n^2) is of order $O(\phi_n n^{-\beta})$. Hence, $\hat{\pi}_{yz}$ is superefficient relative to the usual nonparametric rate ϕ_n; furthermore, if $\pi_n \to 0$ and $\psi_n \to 0$ but the latter vanishes at a faster rate, then $\widehat{\psi}_{yz}$ is superefficient relative to $\hat{\pi}_{yz}$. At the same time, as long as they do not vanish too quickly, all of these estimators will be asymptotically normal. In particular, for $\pi_n \to 0$ but slowly enough, one has both superefficiency and asymptotic normality. Intuitively, the former holds because the constant term of the asymptotic variance vanishes as the probability in question converges toward zero; the latter holds because this convergence is slow enough so that the number of successes sampled diverges.[12] Since we are concerned with small and moderate values of the probabilities in our application (as opposed to values that are almost exactly zero), we feel that this is the relevant asymptotic approximation (which also contains nonvanishing probabilities case as special case). A useful feature of this first order asymptotic is that it preserves the standard nonparametric asymptotics with the exception of the factor ϕ_n. While the rate is not a fundamental issue for deriving bootstrap version of the confidence bands, we have to place particular emphasis on removing the bias, i.e., perform undersmoothing as the mean square optimal bandwidth is excluded; see Horowitz (2001) for a lucid discussion as well as application. Note finally that because $\widehat{\sigma}_{yz}/\sigma_n \xrightarrow{p} 1$, estimation of σ_n^2 effectively allows for rate-adaptive inference in cases of superefficiency. Indeed, because we will undersmooth our estimators, we can base inference on roots that are asymptotically pivotal.

3.2 Inference

This subsection also makes prominent use of moving parameters, which will continue to be denoted with subscripts n. We also omit conditioning on $(Y, \mathbf{Z})$ from notation throughout for readability.

[12] For a slightly more formal intuition, consider simple sample probabilities as estimators of corresponding population probabilities π. Then superefficiency (relative to root-n-convergence) for moving parameters $\pi_n \to 0$ follows immediately from inspection of the sample variance $\pi_n(1-\pi_n)/n$. At the same time, inspection of the Berry-Esseen bounds or the conditions for triangular array Central Limit Theorems reveals that for the same simple probabilities, asymptotic normality will obtain as long as $n\pi_n \to \infty$. Lemma 3.1 essentially states that our nonparametric estimators behave sufficiently similarly to sample probabilities in the limit for this observation to carry over.

Following the moment inequalities literature, we use θ_0 and Θ_0 to denote the true (partially identified) value of θ and its true identified set.

The inference problem would be relatively straightforward if estimators of the upper and lower bounds were jointly normal. In this case, a confidence region for Θ_0 can be defined by projecting a joint confidence region for the upper and lower bound. If one also knows that the interval Θ_0 is long relative to standard errors, one could furthermore use Imbens and Manski's (2004) confidence region for θ_0. Specifically, a 90%-confidence region for Θ_0 would also be a 95%-confidence region for θ_0 because one effectively encounters one-sided testing problems or confidence intervals at each end of the interval Θ_0. Things are not as simple here for a number of reasons:

(i) Estimators $\left(\widehat{\psi}, \hat{\pi}\right)$ need not be uniformly asymptotically normal, specifically not if one of the true values (ψ, π) is local to zero. This problem is much ameliorated by the fact that whenever it occurs, the corresponding probability is estimated superefficiently.

(ii) Even if $\left(\widehat{\psi}, \hat{\pi}\right)$ are uniformly asymptotically normal, this property is not inherited by estimators of upper and lower bounds because these are maxima respectively minima between two other estimators. Thus, their limit distribution is not normal if both of those estimators converge to the same value, and they will fail to be uniformly asymptotically normal even if these cases (but not neighborhoods around them) are excluded.

(iii) Θ_0 might be of the same order or smaller than sampling error, in which case Imbens and Manski's (2004) approach does not in general work. This problem has been analyzed in depth by Stoye (2009), ideas of whom are used in the following.

All of these cases can be taken care of by conservative pre-tests, i.e. pre-tests whose size approaches 1 as sample size expands. On an abstract level, this applies ideas that are by now well understood (e.g., Andrews and Soares (2010)), but we can substantially improve on mechanical application of these ideas by exploiting the specific structure of our problem. A particularly important insight is that the length of the identified set Θ_0 equals

$$\Delta = \min\{\pi, 1-\psi\} - \max\{\pi - \psi, 0\} = \min\{\pi, 1-\pi, \psi, 1-\psi\}.$$

This means that Δ is local to zero iff $\min\{\pi, 1-\pi, \psi, 1-\psi\}$ is local to zero, in which case $\min\{\pi, 1-\pi, \psi, 1-\psi\}$, and therefore Δ, is (implicitly) estimated superefficiently. Regarding problem (iii), we are therefore in a situation quite similar to the favorable case of local superefficiency of $\widehat{\Delta}$ discussed in Stoye (2009).

That said, taking care of all of the above difficulties requires numerous case distinctions. Thus, let

$$(Z_1, Z_2) \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \widehat{\sigma}_\pi^2 & 0 \\ 0 & \widehat{\sigma}_\psi^2 \end{bmatrix}\right).$$

The cutoff values that define the confidence intervals will be specified in terms of (Z_1, Z_2). Restatements in terms of integrals of standard normal c.d.f.'s would be possible but tedious. Also, let a_n be a pre-specified sequence s.t. $a_n \to 0$ but $\phi_n a_n \to \infty$. Also, set the sequences b_n and c_n s.t. $c_n = o(b_n)$, $b_n = o(\phi_n^{-1})$, and lemma 3.1 applies if $\pi_n = O(c_n)$.[13] Then:

Definition 5. ***Confidence Region for*** Θ_0

If $\min\left\{\widehat{\pi}, 1-\widehat{\psi}\right\} < c_n$, *then*

$$CI_{1-\alpha}(\Theta) = CI^1_{1-\alpha}(\Theta) \equiv [0, c_n + \sqrt{c_n}/\phi_n \cdot \Phi^{-1}(1-\alpha/2)]. \tag{3.2}$$

Else, if $\min\left\{\widehat{\pi}, 1-\widehat{\psi}\right\} > 1 - b_n$, *then*

$$CI_{1-\alpha}(\Theta) = CI^2_{1-\alpha}(\Theta) \equiv [1 - b_n - \sqrt{2b_n}/\phi_n \cdot \Phi^{-1}(1-\alpha/2), 1]. \tag{3.3}$$

Else, if $\widehat{\psi} < c_n$, *then*

$$CI_{1-\alpha}(\Theta) = CI^3_{1-\alpha}(\Theta) \equiv \left[\widehat{\pi} - \widehat{\psi} - c_{1-\alpha/2}\widehat{\sigma}_\pi \phi_n^{-1}, \widehat{\pi} + c_{1-\alpha/2}\widehat{\sigma}_\pi \phi_n^{-1}\right], \tag{3.4}$$

where $c_{1-\alpha/2} = \Phi^{-1}(1-\alpha/2)$.

Else, if $\widehat{\pi} > 1 - c_n$, *then*

$$CI_{1-\alpha}(\Theta) = CI^4_{1-\alpha}(\Theta) \equiv \left[\widehat{\pi} - \widehat{\psi} - c_{1-\alpha/2}\widehat{\sigma}_\psi \phi_n^{-1}, 1 - \widehat{\psi} + c_{1-\alpha/2}\widehat{\sigma}_\psi \phi_n^{-1}\right]. \tag{3.5}$$

Else, if $\widehat{\pi} - \widehat{\psi} > -a_n$, *let* (c^l, c^u) *minimize* $(c^l + c^u)$ *subject to the constraint that*

$$\begin{array}{ll} \mathbb{P}(Z_1 \geq -c^u, Z_1 - Z_2 \leq c^l) \geq 1-\alpha & \text{if } \widehat{\pi} < 1 - \widehat{\psi} - a_n, \\ \mathbb{P}(Z_2 \leq c^u, Z_1 - Z_2 \leq c^l) \geq 1-\alpha & \text{if } \widehat{\pi} > 1 - \widehat{\psi} + a_n \\ \mathbb{P}(Z_1 \geq -c^u, Z_2 \leq c^u, Z_1 - Z_2 \leq c^l) \geq 1-\alpha & \text{otherwise.} \end{array} \tag{3.6}$$

Then

$$CI_{1-\alpha}(\Theta) = CI^5_{1-\alpha}(\Theta) \equiv \left[\widehat{\pi} - \widehat{\psi} - \phi_n^{-1} c^l, \min\left\{\widehat{\pi}, 1-\widehat{\psi}\right\} + \phi_n^{-1} c^u\right] \cap [0,1]. \tag{3.7}$$

If none of the above apply, define c^u *by*

$$\begin{array}{ll} \mathbb{P}(Z_1 \geq -c^u) \geq 1-\alpha & \text{if } \widehat{\pi} < 1 - \widehat{\psi}, \\ \mathbb{P}(Z_2 \leq c^u) \geq 1-\alpha & \text{if } \widehat{\pi} > 1 - \widehat{\psi}, \\ \mathbb{P}(Z_1 \geq -c^u, Z_2 \leq c^u) \geq 1-\alpha & \text{otherwise} \end{array} \tag{3.8}$$

and let

$$CI_{1-\alpha}(\Theta) = CI^6_{1-\alpha}(\Theta) \equiv \left[0, \min\left\{\widehat{\pi}, 1-\widehat{\psi}\right\} + \phi_n^{-1} c^u\right]. \tag{3.9}$$

[13] All of these sequences are tuning parameters that play the role of the tuning parameter κ in Andrews and Soares (2010). Optimal choice for such parameters is an area of current research.

The construction is probably best understood by inspecting the last (simplest) case first, which projects a simultaneous confidence set for the upper and lower bound, whose size will then extend to Θ_0 by convexity. Problem (ii) is resolved by conservative pre-tests reflected in the case distinctions in (3.6), (3.8), and the case distinction separating those cases. Problem (i) is handled by the earlier case distinctions. Problem (iii) does not apply to this confidence set. The confidence region for θ_0 uses the above ideas but also handles problem (iii) by using ideas from Stoye (2009) and the above observation about Δ.

Definition 6. ***Confidence Region for*** θ_0

If $\min\left\{\widehat{\pi}, 1-\widehat{\psi}\right\} < c_n$, *then* $CI_{1-\alpha}(\theta) = CI^1_{1-\alpha}(\Theta)$.

Else, if $\min\left\{\widehat{\pi}, 1-\widehat{\psi}\right\} > 1-b_n$, *then* $CI_{1-\alpha}(\theta) = CI^2_{1-\alpha}(\Theta)$.

Else, if $\widehat{\psi} < c_n$, *then*

$$CI_{1-\alpha}(\theta) = CI^3_{1-\alpha}(\theta) \equiv \left[\widehat{\pi} - \widehat{\psi} - c_{1-\alpha}\widehat{\sigma}_\pi \phi_n^{-1}, \widehat{\pi} + c_{1-\alpha}\widehat{\sigma}_\pi \phi_n^{-1}\right], \tag{3.10}$$

where $c_{1-\alpha}$ *fulfils*

$$\Phi(c_{1-\alpha}) - \Phi\left(-c_{1-\alpha} - \widehat{\sigma}_\pi^{-1}\phi_n \widehat{\psi}\right) = 1-\alpha.$$

Else, if $\widehat{\pi} > 1 - c_n$, *then*

$$CI_{1-\alpha}(\theta) = CI^4_{1-\alpha}(\theta) \equiv \left[\widehat{\pi} - \widehat{\psi} - c_{1-\alpha}\widehat{\sigma}_\psi \phi_n^{-1}, 1 - \widehat{\psi} + c_{1-\alpha}\widehat{\sigma}_\psi \phi_n^{-1}\right],$$

where $c_{1-\alpha}$ *fulfils*

$$\Phi(c_{1-\alpha}) - \Phi\left(-c_{1-\alpha} - \widehat{\sigma}_\pi^{-1}\phi_n (1-\widehat{\pi})\right) = 1-\alpha. \tag{3.11}$$

Else, let $\widehat{\Delta} = \min\left\{\widehat{\pi}, 1-\widehat{\psi}\right\} - \max\left\{\widehat{\pi} - \widehat{\psi}, 0\right\}$. *If* $\widehat{\pi} - \widehat{\psi} > -a_n$, *let* (c^l, c^u) *minimize* $(c^l + c^u)$ *subject to the constraint that*

$$\mathbb{P}\left(-c^u - \phi_n \widehat{\Delta} \leq Z_1 - Z_2 \leq c^l\right) \geq 1-\alpha$$

and that

$$\begin{array}{ll} \mathbb{P}\left(-c^u \leq Z_1 \leq c^l + \phi_n \widehat{\Delta}\right) \geq 1-\alpha & \text{if } \widehat{\pi} < 1 - \widehat{\psi} - a_n \\ \mathbb{P}\left(-c^l - \phi_n \widehat{\Delta} \leq Z_2 \leq c^u\right) \geq 1-\alpha & \text{if } \widehat{\pi} > 1 - \widehat{\psi} + a_n \\ \mathbb{P}\left(-c^u \leq Z_1 \leq c^l + \phi_n \widehat{\Delta}, -c^l - \phi_n \widehat{\Delta} \leq Z_2 \leq c^u\right) \geq 1-\alpha & \text{otherwise} \end{array}.$$

Then

$$CI_{1-\alpha}(\theta) = CI^5_{1-\alpha}(\theta) \equiv \left[\widehat{\pi} - \widehat{\psi} - \phi_n^{-1} c^l, \min\left\{\widehat{\pi}, 1-\widehat{\psi}\right\} + \phi_n^{-1} c^u\right] \cap [0,1]. \tag{3.12}$$

If $\widehat{\pi} - \widehat{\psi} \leq -a_n$, *then* $CI_{1-\alpha}(\theta) = CI_{1-\alpha}(\Theta)$.

The interval is calibrated to have size $(1-\alpha)$ if θ_0 coincides with either the lower or the upper bound; it will be larger in between. Note the interval is not calibrated to, and will generally fail to, cover both bounds simultaneously with that probability. That is precisely the difference between $CI_{1-\alpha}(\Theta)$ and $CI_{1-\alpha}(\theta)$, and is why the latter is shorter.

We then have:

Lemma 3.2. *Let assumption 2 hold. Then:*

$$\lim_{n\to\infty} \mathbb{P}(\Theta_0 \subseteq CI_{1-\alpha}(\Theta)) = \lim_{n\to\infty} \inf_{\theta_0\in\Theta_0} \mathbb{P}(\theta_0 \in CI_{1-\alpha}(\theta)) = 1-\alpha$$

uniformly over $(\pi, \psi) \in [0,1]^2$.

4 Empirical Application: WARP and the British Household Expenditure Survey

We now employ this paper's framework on real world data, namely the Family Expenditure Survey (FES), which was the basis for successful, recent applications of revealed preference approaches (Blundell, Browning, Crawford (2003, 2008)). This section is structured as follows: We first provide a description of the data. Then we present some econometric details. Finally, we display the empirical results.

4.1 Description of the Data

The FES reports a yearly cross section of labor income, expenditures, demographic composition, and other characteristics of about 7,000 households. We use the years 1974-1993, but exclude the respective Christmas periods as they contain too much irregular behavior. As is standard in the demand system literature, we focus on the subpopulation of two person households where both are adults, at least one is working, and the head of household is a white collar worker. This is to reduce the impact of measurement error; see Lewbel (1999) for a discussion. We provide a summary statistic of our data in table 1 in the appendix.

We form several expenditure categories. The first category is related to food consumption and consists of the subcategories food bought, food out (catering) and tobacco. The second category contains housing expenditures, namely rent or mortgage payments and household goods and services, excluding furniture. The last group consists of motoring and fuel expenditures. For brevity, we call these categories food, housing and energy. These broader categories are formed since more detailed accounts suffer from infrequent purchases (recall that the recording period is 14 days) and are thus often underreported. These three categories account for 20-30% of total expenditure on average, leaving a fourth residual category. Results actually displayed were generated by considering consumption of food versus nonfood items, but similar analyses were performed for all of the goods, and with similar results. We removed outliers by excluding the upper and lower 2.5% of the population in the three groups.

For the pairwise comparisons, we normalize prices by dividing all variables by the general price index excluding the good into consideration (in particular, for food we consider the price of food vs. the price of all nondurable goods except food). This removes both general inflation and transforms all prices to be relative to the price index. Quantities are defined by dividing the normalized expenditures by the respective normalized price, e.g. food by the food price index. We also divide total expenditure by the price index.

To account for possible endogeneities, i.e. violations of assumption 1(ii), we use labor income as an instrument. This is standard practise in the demand literature, see Lewbel (1999), and, assuming the existence of preferences, is satisfied under an assumption of separability of the labor supply from the consumer demand decision. Labor income is constructed as in the household below average income study (HBAI), that is, it is roughly defined as labor income after taxes and transfers. We include the remaining household covariates as regressors. Specifically, we use principal components to reduce the vector of remaining household characteristics to a few orthogonal, approximately continuous components, mainly because we require continuous covariates for nonparametric estimation. Since we already condition on a lot of household information by using the specific subgroup, we only use the first principal component. While this is arguably ad hoc, we perform some robustness checks like alternating the component or adding several others, and results do not change appreciably.

4.2 Econometric Specification and Empirical Results

We estimate conditional probabilities (π_{yz}, ψ_{yz}) via a locally linear estimator with a standard Epanechnikov kernel. The bandwidth is selected by cross validation. We checked the sensitivity of our results by varying the bandwidth; there was no material effect on results. Sampling distributions of $\phi_n^{1/2}\left(\widehat{\psi}_{yz} - \psi_{yz}\right)$ and $\phi_n^{1/2}\left(\widehat{\pi}_{yz} - \pi_{yz}\right)$ were simulated by a wild hybrid bootstrap (Shao and Tu (1995)). Specifically, we use the inverted c.d.f. of $F_{\hat{\psi}^*_{yz} - \hat{\psi}_{yz}}$, where $\hat{\psi}^*_{yz}$ denotes the bootstrap estimator and $\hat{\psi}_{yz}$ the original estimator, to derive a consistent estimator for $F_{\hat{\psi}_{yz} - \psi_{yz}}$, and similarly for $\widehat{\pi}_{yz}$. For the estimator of the bounds, the bootstrap is known to be consistent (see Hall (1992) or Horowitz (2001), for a general discussion). This is obvious in the case of non-vanishing probabilities, but also applies to the few cases (2 out of 100) of vanishing probabilities, as the standard asymptotics are retained with the exception of the modified rate. As elaborated before, we also assume that $\widehat{\psi}_{yz}$ and $\widehat{\pi}_{yz}$ are generated from independent samples.

When applying our local polynomial estimators to the choice data, we first group the population into "bands" of three years, e.g., we collect all people surveyed in the years 1974-1976 into one group. This is done to increase the number of observations. As a consequence, our cross sections actually comprise 3 years, and we assume the individuals to face the mean price in this period.[14] We then compare every cross section with the two adjacent ones only because for groups that are many years apart, apparent violations of WARP could plausibly be driven by changing preferences.

Our first important finding is that for most such comparisons, the income change swamps the price effect, leading to upper bounds of zero. This is easy to explain: The order of magnitude of the relative price change is -0.05, while most quantities are around 10. Thus, the overall effect of a price change on quantities is $\Xi_{ts} = (p_t - p_s)\, Q_s \approx 0.5$. Figure 2 provides a graphical representation of the density of this effect in the second year (so the variable is $\Xi_{22} = (p_3 - p_2)\, Q_2$, corresponding to the years of 1978 and 1981).

Fig. 2 approx. here

The probability mass is highly concentrated between -1 and 0. In contrast, mean real income increased in the same two periods from 50.4 to 54.3, and median income increased less dramatically from 43.5 to 45.5. The typical case is that $\pi_{yz} = \psi_{yz} \cong 0$, so that $\mathbb{P}_{yz}(\text{WARP violated})$ is estimated as zero.[15] We focus on regions in the data that are at least potentially informative, as operationalized by a nonzero (estimated) upper bound on $\mathbb{P}_{yz}(\text{WARP violated})$. For instance, given the distribution of Ξ_{22}, we focus on income changes that are between 0 and -1, i.e. on subpopulations who become marginally poorer.[16] More specifically, we form a 10×10 grid that combines each of the $(9, 18, ..., 90)$-quantiles of the income distribution in the period (1977-1982) with all of the income changes in $(-1, -0.9, ..., -.1)$; see figure 2. This choice of subsample is certainly ad hoc, and we leave to future research a more systematic treatment of the choice of region for which the data is informative.

We start our analysis by forming confidence intervals. We set $c_n = 0.02$ and $b_n = 0.2$. The

[14]This induces a measurement error. Compared to the already incurred measurement error, and in light of the fact that all that matters is the change in prices, we feel that this is a minor issue.

[15]The according confidence regions will include some strictly positive numbers, but then, recall that a confidence region for a probability based on 0 successes in n trials will include some small but positive numbers for *any* n.

[16]To clarify, this des not amount to tracking specific individuals in our sample; after all, we do not have panel data. But assumption 1(ii) ensures that in making the comparison, we compare different cross-sectional samples from the same population. Of course, this illustrates that assumption 1(ii) is not innocuous; though, recall our use of an instrument in this application.

near degenerate cases that our modified confidence region guards against are exceedingly rare in the data: $\min\left\{\widehat{\pi}_{yz}, 1-\widehat{\psi}_{yz}\right\} < 0.02$ occurs in 2 out of 100 positions, and we use the appropriate CI; the cases $\min\left\{\widehat{\pi}_{yz}, 1-\widehat{\psi}_{yz}\right\} > 0.8$, $\widehat{\psi}_{yz} < 0.02$, and $1-\widehat{\pi}_{yz} > 0.98$ do not appear. Degenerate probabilities are therefore not much of an issue, and we essentially apply the confidence region defined in (3.12). Still, we have to distinguish two cases, one where we have reason to believe that the lower bound is exactly zero because $\widehat{\pi}_{yz} - \widehat{\psi}_{yz}$ is strongly negative, and one where we cannot rule out that $\pi_{yz} - \psi_{yz} \geq 0$ is the binding constraint at the lower bound. The cut off we take is $a_n = -.1$. Thus, if $\widehat{\pi}_{yz} - \widehat{\psi}_{yz} < -.1$, then we consider the lower bound to be exactly zero. Recall from section 3 that this leads to a simplification; intuitively, if the lower bound is nonstochastic (i.e., exactly 0), then all randomness arises from the upper bound being stochastic and we only have to control coverage at the upper end.

Rather than presenting 100 confidence intervals, we show densities for both the upper and lower bounds as well as densities for the associated confidence intervals. As we detail below, the reason for us to proceed in this fashion is that the point estimate of the lower bound, and hence also the lower end of the confidence interval, is mostly close to zero. In particular, high point estimates of upper bounds are only weakly associated with high estimates of lower bounds, hence not much is lost by looking the bounds in isolation. Figures 3-5 display our results for the upper bound in our subsample. We start with the distribution of the point estimate:

Fig. 3 approx. here

The estimated upper bound exceeds 1% in 97 out of the 100 points of support of the regressors and exceeds 5% (the threshold indicated by the vertical line) in 80 cases. To construct the upper ends of the confidence intervals for the parameter, we have to jump a final hurdle, namely to take precautions regarding whether one or both restrictions bind. Therefore we pre-test for equality of π_{yz} and $(1-\psi_{yz})$ with critical value $a_n = .1$. If equality is rejected, we form a standard 95% confidence band for whichever of π_{yz} and $(1-\psi_{yz})$ appears smaller; if it is not rejected, we form joint confidence sets, which are effectively 97.5% for either parameter because $\pi_{yz} \approx (1-\psi_{yz})$ in these cases. Figure 4 shows the distribution of the upper bound and that of the upper end of the appropriate bootstrap CI around the upper bound.

Fig. 4 approx. here

Thus, the data appear potentially informative about WARP, and as the confidence intervals for the parameters indicate, there might be rejections of rationality. Whether we can positively tell with any degree of certainty that there are violations remains a question that can only be answered by considering the lower bound. Before we do this, we want to clarify whether we have positive evidence that the data is informative. Therefore, to check whether the positive values are statistically significant, we use lemma 3.2 to compute a lower confidence bound for the upper bound, i.e., when constructing the confidence intervals we allow again for the possibility of both upper bound constraints binding simultaneously.

We indeed find that many of the positive upper bounds are statistically significant. In 84 out of 100 cases, the lower confidence interval is above 0.01. The mean and median of the upper bound are 0.177 respectively 0.146. More than a third of cases are above 0.25, while the 95% bootstrap CI usually has length between 0.10 and 0.15. Figure 5 shows the distribution of the upper bound and that of the lower end of the 95% bootstrap CI around the upper bound.

Fig. 5 approx. here

While these numbers should be interpreted with care due to our failure to control familywise error rates, they do suggest that the data are reasonably informative – in the sense of allowing for potential violation of WARP – in this selected subpopulation. They also allow for a sizable fraction of the population to violate rationality. Still, figure 6 illustrates that we find hardly any conclusive evidence against WARP.

Fig. 6 approx. here

The lower bound is typically close to zero even in this informative subpopulation; it exceeds 0.05 in only 1 out of 100 instances. What is more, one-sided 95% bootstrap confidence intervals for $\pi_{yz} - \psi_{yz}$, constructed in accordance with lemma 3.2, include zero at 97 of 100 positions on our grid. Hence, we cannot statistically distinguish the positive lower bounds from zero even with confidence regions that fail to control familywise error rates.

We now showcase the effect of one of the refinements discussed in section 2. Specifically, we impose positive quadrant dependence (PQD; see section 2.1.3). The refined bounds are illustrated in figure 7, and figure 8 shows through a comparison the effect of the introduction of PQD on the distribution of the upper bound.

Figs. 7, 8 approx. here

As explained in section 2, PQD will not induce refined upper bounds of zero unless worst-case upper bounds were zero, however the upper bounds are much reduced, typically by $50-75\%$. The highest possible proportion of "violators" is substantially reduced and exceeds 20% only at very select data points. Having said that, most of the positive bounds are statistically significant, with many p-values being small enough that controlling for familywise error rates would not overturn this conclusion. Hence, the data are still consistent with some violations of WARP. Lower bounds on $\mathbb{P}_{yz}$(WARP violated) remain zero because in the two-dimensional case, PQD refines worst-case bounds from above but not below; recall that this is not true for natural generalizations of it in the higher dimensional case. All in all, PQD substantially narrows the bounds, although in the two-good case, it cannot lead to rejection of rationality if worst-case bounds did not warrant that conclusion already.

None of these results change appreciably if we include a measure of household characteristics and/or correct for endogeneity using a control function approach. Moreover, they are stable across the large groups of goods we consider, for pairwise comparisons (e.g., energy vs. non-energy). In summary, we tend to think that at least as a reasonable approximation to behavior, WARP is more corroborated than questioned by these data, but we would like to emphasize the need for further research with other data.

5 Conclusion

This paper investigated exactly what power revealed preference assumptions have under realistic data constraints. The leading question was to bound the fraction of a population that violates WARP given repeated cross-section data. Side results were to elucidate the exact empirical content of WARP and to carry out an inference exercise that applies recent insights about inference under partial identification. The empirical result with respect to the U.K. Family Expenditure Survey is that even for those observations where budget planes meaningfully overlap, as reflected by large upper bounds on the probability of violating WARP, lower bounds are not significantly positive, i.e. WARP cannot be rejected. Furthermore, imposing a very weak, nonparametric limitation on heterogeneity (namely, positive quadrant dependence with respect to budget shares spent on different goods) leads to uniformly rather small, though not uniformly zero, upper bounds.

The core difference between this paper and existing work that estimates demand for applied purposes is that we consider the revealed preference paradigm on individual level in isolation, being careful to impose no or very weak homogeneity assumptions across individuals. This, of course, leads to less conclusive results. While the data may be interpreted to be mildly supportive of WARP, this could certainly be due not to the population being substantively rational, but to the weak axiom being, well, weak. To be sure, we do not mean to implicitly criticize other papers, but rather to augment them by showing how much mileage can be gained from revealed preference assumptions proper. Thus, our motivation is somewhat similar to early papers on partial identification of treatment effects, which frequently stress the conceptual value of understanding just how much one could learn from the data without identifying assumptions. Insofar as the result is somewhat negative, the substantive message might well corroborate approaches that use stronger assumptions. We hope, however, to illuminate the degree to which sharper conclusions than ours will depend on using sharper assumptions, whether or not these assumptions are formally semi- or nonparametric.

Appendix I - Summary Statistics of Data: Household Characteristics, Income and Normalized Expenditures

Variable	Minimum	1st Quartile	Median	Mean	3rd Quartile	maximum
number of female	0	1	1	1.073	1	2
number of retired	0	0	0	0.051	0	1
number of earners	0	1	2	1.692	2	2
Age of HHhead	19	31	49	46	58	90
Fridge	0	1	1	0.987	1	1
Washing Machine	0	1	1	0.882	1	1
Centr. Heating	0	1	1	0.804	1	1
TV	0	1	1	0.874	1	1
Video	0	0	0	0.407	1	1
PC	0	0	0	0.792	0	1
number of cars	0	1	1	1.351	2	10
number of rooms	1	4	5	5.455	6	26
HHincome	6.653	37.550	52.210	61.820	73.920	3981.000
Food	0	5.565	7.346	7.867	9.602	52.519
Housing	0	4.052	7.859	9.715	12.910	375.486
Energy	0	1.271	1.812	2.121	2.509	34.103

Appendix II - Proofs

Proposition 1 Recall the Fréchet-Hoeffding bounds: For any two random variables X_1 and X_2 and events (in the relevant algebras) A_1 and A_2, one has the tight bounds

$$\max\{\mathbb{P}(X_1 \in A_1) + \mathbb{P}(X_2 \in A_2) - 1, 0\} \leq \mathbb{P}(X_1 \in A_1, X_2 \in A_2) \leq \min\{\mathbb{P}(X_1 \in A_1), \mathbb{P}(X_2 \in A_2)\}.$$

To see validity of the lower bound, note that

$$\begin{aligned}
& \mathbb{P}_{yz}(\text{WARP violated}) \\
= \; & \mathbb{P}_{yz}(Q_s \leq (y_t - y_s)/(p_t - p_s), Q_t \geq (y_t - y_s)/(p_t - p_s), Q_s \neq Q_t) \\
= \; & \mathbb{P}_{yz}(Q_s \leq (y_t - y_s)/(p_t - p_s), Q_t \geq (y_t - y_s)/(p_t - p_s)) - \\
& \mathbb{P}_{yz}(Q_s \leq (y_t - y_s)/(p_t - p_s), Q_t \geq (y_t - y_s)/(p_t - p_s), Q_s = Q_t) \\
= \; & \mathbb{P}_{yz}(Q_s \leq (y_t - y_s)/(p_t - p_s), Q_t \geq (y_t - y_s)/(p_t - p_s)) - \mathbb{P}_{yz}(Q_s = Q_t = (y_t - y_s)/(p_t - p_s)) \\
\geq \; & \max\left\{\mathbb{P}_{yz}(Q_s \leq (y_t - y_s)/(p_t - p_s)) - \mathbb{P}_{yz}(Q_t < (y_t - y_s)/(p_t - p_s)), 0\right\} - \\
& \min\left\{\mathbb{P}_{yz}(Q_s = (y_t - y_s)/(p_t - p_s)), \mathbb{P}_{yz}(Q_t = (y_t - y_s)/(p_t - p_s))\right\},
\end{aligned}$$

where the first equality spells out the event that WARP is violated, the next two steps use basic probability calculus, and the last step uses the *lower* Fréchet-Hoeffding bound on $\mathbb{P}_{yz}(Q_s \leq (y_t - y_s)/(p_t - p_s)), Q_t \geq (y_t - y_s)/(p_t - p_s))$ as well as the *upper* Fréchet-Hoeffding bound on $\mathbb{P}_{yz}(Q_s = Q_t = (y_t - y_s)/(p_t - p_s))$. The expression in the lemma is generated by taking the maximum between the last expression and zero, observing that this renders redundant the max-operator in the preceding display.

The bound is tight because a joint distribution of (Q_s, Q_t) that achieves it can be constructed as follows: First, assign probability $\min\left\{\mathbb{P}_{yz}(Q_s = (y_t - y_s)/(p_t - p_s)), \mathbb{P}_{yz}(Q_t = (y_t - y_s)/(p_t - p_s))\right\}$ to the event $(Q_s = Q_t = (y_t - y_s)/(p_t - p_s))$. Second, remove this probability mass from the marginal distributions of (Q_s, Q_t) and rescale them so they integrate to 1. Third, the joint distribution of $(Q_s, Q_t | Q_s = Q_t = (y_t - y_s)/(p_t - p_s)$ does not hold) is characterized by those rescaled marginal distributions and the Fréchet-Hoeffding lower bound (perfectly positive dependence) copula.

To see validity of the upper bound, note that

$$\mathbb{P}_{yz}(\text{WARP violated}) \leq \min\left\{\mathbb{P}_{yz}(Q_s \leq (y_t - y_s)/(p_t - p_s)), \mathbb{P}_{yz}(Q_t \geq (y_t - y_s)/(p_t - p_s))\right\}$$

by the upper Fréchet-Hoeffding bounds and furthermore that

$$\begin{aligned}
& \mathbb{P}_{yz}(\text{WARP violated}) \\
= \; & \mathbb{P}_{yz}((Q_s \leq (y_t - y_s)/(p_t - p_s), Q_t > (y_t - y_s)/(p_t - p_s)) \\
& \text{or } \; (Q_s < (y_t - y_s)/(p_t - p_s), Q_t \geq (y_t - y_s)/(p_t - p_s))) \\
\leq \; & 1 - \mathbb{P}_{yz}(Q_s \geq (y_t - y_s)/(p_t - p_s), Q_t \leq (y_t - y_s)/(p_t - p_s)) \\
\leq \; & 1 - \min\left\{\mathbb{P}_{yz}(Q_s \geq (y_t - y_s)/(p_t - p_s)), \mathbb{P}_{yz}(Q_t \leq (y_t - y_s)/(p_t - p_s))\right\} \\
= \; & \max\left\{\mathbb{P}_{yz}(Q_s < (y_t - y_s)/(p_t - p_s)), \mathbb{P}_{yz}(Q_t > (y_t - y_s)/(p_t - p_s))\right\},
\end{aligned}$$

where all equalities and the first inequality use basic probability calculus and the second inequality utilizes a lower Fréchet-Hoeffding bound on $\mathbb{P}_{yz}(Q_s \geq (y_t - y_s)/(p_t - p_s), Q_t \leq (y_t - y_s)/(p_t - p_s))$.

To see that the bound is tight, note that it is achieved by the Fréchet-Hoeffding upper bound (perfectly negative dependence) copula. Intermediate values can be attained by mixing the two distributions that achieve the bounds.

Lemma 2.2 Assume positive quadrant dependence. Recall that $\underline{\mathcal{B}}_{s,t}$ is a lower contour set and $\underline{\mathcal{B}}_{t,s}$ an upper one, hence

$$\mathbb{P}_{yz}(\mathbf{Q}_s \notin \underline{\mathcal{B}}_{s,t}, \mathbf{Q}_t \in \underline{\mathcal{B}}_{t,s}) \geq \mathbb{P}_{yz}(\mathbf{Q}_s \notin \underline{\mathcal{B}}_{s,t})\mathbb{P}_{yz}(\mathbf{Q}_t \in \underline{\mathcal{B}}_{t,s}),$$

hence

$$\begin{aligned}
\mathbb{P}_{yz}(\mathbf{Q}_s \in \underline{\mathcal{B}}_{s,t}, \mathbf{Q}_t \in \underline{\mathcal{B}}_{t,s}) &= \mathbb{P}_{yz}(\mathbf{Q}_t \in \underline{\mathcal{B}}_{t,s}) - \mathbb{P}_{yz}(\mathbf{Q}_s \notin \underline{\mathcal{B}}_{s,t}, \mathbf{Q}_t \in \underline{\mathcal{B}}_{t,s}) \\
&\leq \mathbb{P}_{yz}(\mathbf{Q}_t \in \underline{\mathcal{B}}_{t,s}) - \mathbb{P}_{yz}(\mathbf{Q}_s \notin \underline{\mathcal{B}}_{s,t})\mathbb{P}_{yz}(\mathbf{Q}_t \in \underline{\mathcal{B}}_{t,s}) \\
&= (1 - \mathbb{P}_{yz}(\mathbf{Q}_s \notin \underline{\mathcal{B}}_{s,t}))\mathbb{P}_{yz}(\mathbf{Q}_t \in \underline{\mathcal{B}}_{t,s}) = \mathbb{P}_{yz}(\mathbf{Q}_s \in \underline{\mathcal{B}}_{s,t})\mathbb{P}_{yz}(\mathbf{Q}_t \in \underline{\mathcal{B}}_{t,s}).
\end{aligned}$$

The refined lower bound for (ii) is established similarly. The old lower and upper bounds are tight because the distributions that generate them are consistent with positive respectively negative quadrant dependence. The bounds at $\mathbb{P}_{yz}(\mathbf{Q}_s \in \underline{\mathcal{B}}_{s,t})\mathbb{P}_{yz}(\mathbf{Q}_t \in \underline{\mathcal{B}}_{t,s})$ are tight because independence of $\mathbf{Q}_s$ and $\mathbf{Q}_t$ cannot be excluded.

Example 3 The first and third claim are easy to see, we will establish the one regarding association. To see the lower bound, let $U \equiv \{\mathbf{q} : (1, 2, 1) \cdot \mathbf{q} \geq 5.1\}$, then U contains $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{b}_2$

but not $\mathbf{a}_2$ or $\mathbf{b}_1$. Now write

$$\begin{aligned}
\mathbb{P}_{yz}(\mathbf{Q}_1 \in U, \mathbf{Q}_2 \in U) &\geq \mathbb{P}_{yz}(\mathbf{Q}_1 \in U)\mathbb{P}_{yz}(\mathbf{Q}_2 \in U) \\
\Longleftrightarrow \mathbb{P}_{yz}(\mathbf{Q}_2 \in U | \mathbf{Q}_1 \in U) &\geq \mathbb{P}_{yz}(\mathbf{Q}_2 \in U) \\
\Longleftrightarrow \mathbb{P}_{yz}(\mathbf{Q}_2 \notin U | \mathbf{Q}_1 \in U) &\leq \mathbb{P}_{yz}(\mathbf{Q}_2 \notin U) \\
\Longleftrightarrow \mathbb{P}_{yz}(\mathbf{Q}_2 \notin U | \mathbf{Q}_1 \notin U) &\geq \mathbb{P}_{yz}(\mathbf{Q}_2 \notin U) \\
\Longleftrightarrow \mathbb{P}_{yz}(\mathbf{Q}_2 = \mathbf{b}_1 | \mathbf{Q}_1 = \mathbf{a}_2) &\geq \mathbb{P}_{yz}(\mathbf{Q}_2 = \mathbf{b}_1) = 1/2,
\end{aligned}$$

implying the claim. The bound is tight because association allows for independence.

Lemma 3.1 Throughout this proof, we denote by $\mathbb{Y} = (Y_1 ..., Y_n)$, $\mathbb{Z} = (Z_1 ..., Z_n)$ and y, z denote a fixed position. Following standard arguments for local polynomials (e.g., Fan and Gijbels (1996)), we obtain for the bias

$$\mathbb{E}\left[\hat{\pi}_{yz} - \pi_n | \mathbb{Y}, \mathbb{Z}\right] = h^2 n^{-\beta} \frac{\kappa_2}{2} \iota' H\left(c_{yz}\right) \iota + o_p(h^2).$$

The variance requires a bit more care. We decompose the estimator into bias and variance part, i.e.: $\hat{\pi}(y,z) - \pi(y,z) = \sum_i W_{in}(\eta_i + bias_i)$, where W_{in} are weights, see Fan and Gijbels (1999). Next, consider

$$\begin{aligned}
Var\left[\hat{\pi}_{yz}(y,z) | \mathbb{Y}, \mathbb{Z}\right] &= \mathbb{E}\left[\left(\hat{\pi}_{yz} - \pi_n\right)^2 | \mathbb{Y}, \mathbb{Z}\right] \\
&= \mathbb{E}\left[\sum_j \sum_i W_{in} W_{jn} \eta_i \eta_j | \mathbb{Y}, \mathbb{Z}\right] \\
&\quad + 2\mathbb{E}\left[\sum_j \sum_i W_{in} W_{jn} \eta_i bias_j | \mathbb{Y}, \mathbb{Z}\right] \\
&\quad + \mathbb{E}\left[\sum_j \sum_i W_{in} W_{jn} bias_i bias_j | \mathbb{Y}, \mathbb{Z}\right] \\
&= T_1 + T_2 + T_3
\end{aligned}$$

Observe that $T_2 = 0$ by iterated expectations, and $T_3 = o_p(T_1)$. Finally,

$$\mathbb{E}\left[\sum_j \sum_i W_{in} W_{jn} \eta_i \eta_j | \mathbb{Y}, \mathbb{Z}\right] = \sum_i W_{in}^2 \mathbb{E}\left(\eta_i^2 | Y_i = y, Z_i = z\right) = n^{-\beta} \sum_i W_{in}^2 c_{y_i z_i}.$$

Then, by standard arguments, $\sum_i W_{in}^2 c_{y_i z_i} = n^{-1} h^{-d} \kappa_2 c_{yz} + o_p((nh)^{-1})$, and the statement follows by a CLT for triangular arrays, see again Fan and Gijbels (1996).

To see that $\widehat{\pi}_{yz}/\pi_n \xrightarrow{p} 1$ (and hence $\widehat{\sigma}_{yz}/\sigma_n \xrightarrow{p} 1$), observe that

$$\frac{\widehat{\pi}_{yz}-\pi_n}{\pi_n}=\frac{\sqrt{nh^d n^\beta}}{\pi_n\sqrt{nh^d n^\beta}}\left(\widehat{\pi}_n-\pi_n\right)=\frac{1}{c_{yz}\sqrt{nh^d n^{-\beta}}}\underbrace{\sqrt{nh^d n^\beta}\left(\widehat{\pi}_{yz}-\pi_n\right)}_{\equiv A},$$

where the second step follows by substituting for $\pi_n = c_{yz}n^{-\beta}$. By the lemma's main claim, A is stochastically bounded, thus $n^{1-\beta}h^d \to \infty$ implies $\left(\widehat{\pi}_{yz}-\pi_n\right)/\pi_n \xrightarrow{p} 0$ and hence the claim.

Lemma 3.2 We establish the uniform result by showing a pointwise one but in moving parameters (π_n, ψ_n), implying the uniform result because the pointwise finding can be applied to a least favorable sequence. Also, we will make a finite number of case distinction depending on whether parameters are "large" or "small" in senses that will be defined. Every sequence can be partitioned into finitely many subsequences s.t. each subsequence conforms to one case below.

We first establish a number of lemmas showing that the different versions of $CI_{1-\alpha}(\Theta)$ are valid under different sets of conditions.

Lemma A.1. Assume that $\min\left\{\pi_n, 1-\psi_n\right\}/c_n \to 0$. Then $\lim_{n\to\infty}\mathbb{P}(\Theta_0 \subseteq CI^1_{1-\alpha}(\Theta)) = 1$.

Proof.

$$\Theta_0 = [\max\{\pi_n-\psi_n,0\},\min\{\pi_n,1-\psi_n\}] \subseteq [0,\min\{\pi_n,1-\psi_n\}] \subseteq [0,c_n] \subset [0, CI^1_{1-\alpha}(\Theta)].$$

Lemma A.2. Assume that $\left(1-\pi_n\right)/2b_n \to 0$ and $\psi_n/2b_n \to 0$. Then $\lim_{n\to\infty}\mathbb{P}(\Theta_0 \subseteq CI^2_{1-\alpha}(\Theta)) = 1$.

Proof. Noting that $\pi_n - \psi_n = 1-(1-\pi_n)-\psi_n \geq 1-b_n$ for n large enough, write

$$\Theta_0 = [\max\{\pi_n-\psi_n,0\},\min\{\pi_n,1-\psi_n\}] \subseteq [1-b_n,1] \subset CI^2_{1-\alpha}(\Theta).$$

Lemma A.3. Assume that $\phi_n\sigma_\pi\left(\hat{\pi}-\pi_n\right) \xrightarrow{d} \mathcal{N}(0,1)$, that $\phi_n\sigma_\pi\left(\widehat{\psi}-\psi_n\right) \xrightarrow{p} 0$, and that $\widehat{\sigma}_\pi/\sigma_\pi \to 1$. Then $\lim_{n\to\infty}\mathbb{P}(\Theta_0 \subseteq CI^3_{1-\alpha}(\Theta)) = 1-\alpha$.

Proof. Write

$$\begin{aligned}
& \lim_{n\to\infty} \mathbb{P}(\Theta_0 \subseteq CI^3_{1-\alpha}(\Theta)) \\
= & \lim_{n\to\infty} \mathbb{P}\left(\widehat{\pi} - \widehat{\psi} - c_{1-\alpha}\widehat{\sigma}_\pi \phi_n^{-1} \leq \pi_n - \psi_n, \pi_n \leq \widehat{\pi}_n + c_{1-\alpha}\widehat{\sigma}_\pi \phi_n^{-1}\right) \\
= & \lim_{n\to\infty} \mathbb{P}\left(\phi_n \widehat{\sigma}_\pi \left(\widehat{\pi} - \pi_n - \left(\widehat{\psi} - \psi_n\right)\right) \leq c_{1-\alpha}, \phi_n \widehat{\sigma}_\pi \left(\widehat{\pi}_n - \pi_n\right) \geq -c_{1-\alpha}\right) \\
= & \lim_{n\to\infty} \mathbb{P}\left(-c_{1-\alpha} \leq \phi_n \widehat{\sigma}_\pi \left(\widehat{\pi} - \pi_n\right) \leq c_{1-\alpha}\right) \\
= & 1 - \alpha.
\end{aligned}$$

where the last step uses the definition of $c_{1-\alpha}$ and this lemma's assumptions.

Lemma A.4. Assume that $\phi_n \sigma_\psi \left(\widehat{\psi} - \psi_n\right) \xrightarrow{d} \mathcal{N}(0,1)$, that $\phi_n \sigma_\psi \left(\widehat{\pi} - \pi_n\right) \xrightarrow{p} 0$, and that $\widehat{\sigma}_\psi / \sigma_\psi \to 1$. Then $\lim_{n\to\infty} \mathbb{P}(\Theta_0 \subseteq CI^4_{1-\alpha}(\Theta)) = 1 - \alpha$.

Proof. This mimics lemma A.4.

Lemma A.5. Assume that $\left[\phi_n \sigma_\pi \left(\hat{\pi} - \pi_n\right), \phi_n \sigma_\psi \left(\widehat{\psi} - \psi_n\right)\right] \xrightarrow{d} \mathcal{N}(0, I_2)$, that $\widehat{\sigma}_\pi / \sigma_\pi \to 1$, and that $\widehat{\sigma}_\psi / \sigma_\psi \to 1$. Then $\lim_{n\to\infty} \mathbb{P}(\Theta_0 \subseteq CI_{1-\alpha}(\Theta)) = 1 - \alpha$.

Proof. This case requires two sub-distinctions amounting to four distinct sub-cases, according to which of (3.1a-3.1d) must be presumed to (almost) bind. Let ε_n be a sequence s.t. $\varepsilon_n \to 0$, $\varepsilon_n / \phi_n \to \infty$, but $\varepsilon_n / a_n \to 0$. First, assume that $\pi_n - \psi_n \geq -\varepsilon_n$, meaning that (3.1a) must be taken into account. Then $CI_{1-\alpha}(\Theta)$ will be constructed according to (3.6-3.7) with probability approaching 1, thus it suffices to show validity of this construction. Assume first that $|\pi + \psi - 1| \leq \varepsilon_n$, thus $\left|\widehat{\pi} + \widehat{\psi} - 1\right| \leq \varepsilon_n$ with probability approaching 1. We can then write

$$\lim_{n\to\infty} \mathbb{P}(\Theta_0 \subseteq CI_{1-\alpha}(\Theta)) = \lim_{n\to\infty} \mathbb{P}\left(\begin{array}{c} \max\left\{\widehat{\pi} - \widehat{\psi} - \phi_n^{-1} c^l, 0\right\} \leq \max\{\pi_n - \psi_n, 0\}, \\ \min\{\pi_n, 1 - \psi_n\} \leq \min\left\{\widehat{\pi}, 1 - \widehat{\psi}\right\} + \phi_n^{-1} c^u \end{array}\right).$$

We bound the r.h. probability from below by observing some logical implications. First,

$$\widehat{\pi} - \widehat{\psi} - \phi_n^{-1} c^l \leq \pi_n - \psi_n \Longrightarrow \max\left\{\widehat{\pi} - \widehat{\psi} - \phi_n^{-1} c^l, 0\right\} \leq \max\{\pi_n - \psi_n, 0\}.$$

To see this, note that if $\pi_n - \psi_n \geq 0$, then the two inequalities are equivalent except if $\widehat{\pi} - \widehat{\psi} - \phi_n^{-1} c^l \leq 0$, in which case they are both fulfilled. If $\pi_n - \psi_n < 0$, then the l.h. inequality implies the r.h. one because whenever the l.h. inequality holds, both sides of the r.h. inequality equal 0.

Second,

$$
\begin{aligned}
& \min\left\{\widehat{\pi}, 1-\widehat{\psi}\right\} - \min\{\pi_n, 1-\psi_n\} \\
= & \min\left\{\widehat{\pi} - \min\{\pi_n, 1-\psi_n\}, 1-\widehat{\psi} - \min\{\pi_n, 1-\psi_n\}\right\} \\
\geq & \min\left\{\widehat{\pi} - \pi_n, \psi_n - \widehat{\psi}\right\}.
\end{aligned}
$$

Together, these implications yield

$$
\begin{aligned}
& \lim_{n\to\infty} \mathbb{P}(\Theta_0 \subseteq CI_{1-\alpha}(\Theta)) \\
\geq & \lim_{n\to\infty} \mathbb{P}\left(\widehat{\pi} - \widehat{\psi} - \phi_n^{-1}c^l \leq \pi_n - \psi_n, \min\left\{\widehat{\pi} - \pi_n, \psi_n - \widehat{\psi}\right\} \geq -\phi_n^{-1}c^u\right) \\
= & \lim_{n\to\infty} \mathbb{P}\left(\widehat{\pi} - \widehat{\psi} - \phi_n^{-1}c^l \leq \pi_n - \psi_n, \widehat{\pi} - \pi_n \geq -\phi_n^{-1}c^u, \widehat{\psi} - \psi_n \leq \phi_n^{-1}c^u\right) \\
= & \lim_{n\to\infty} \mathbb{P}\left(\phi_n\left(\widehat{\pi} - \pi_n - \left(\widehat{\psi} - \psi_n\right)\right) \leq c^l, \phi_n\left(\widehat{\pi} - \pi_n\right) \geq -c^u, \phi_n\left(\widehat{\psi} - \psi_n\right) \leq c^u\right) \\
= & \lim_{n\to\infty} \mathbb{P}\left(\sigma_\pi Z_1 - \sigma_\psi Z_2 \leq c^l, \sigma_\pi Z_1 \geq -c^u, \sigma_\psi Z_2 \leq c^u\right) \\
= & \lim_{n\to\infty} \mathbb{P}\left(\widehat{\sigma}_\pi Z_1 - \widehat{\sigma}_\psi Z_2 \leq c^l, \widehat{\sigma}_\pi Z_1 \geq -c^u, \widehat{\sigma}_\psi Z_2 \leq c^u\right) \\
\geq & 1-\alpha,
\end{aligned}
$$

where the last steps use this lemma's assumptions and condition (3.6).

Now, let $\pi_n - \psi_n < -\varepsilon_n$. In this case, $CI_{1-\alpha}(\Theta)$ will be constructed according to (3.6-3.7) with some probability and according to (3.8-3.9) with the remaining probability (which goes to 1 as $\pi_n - \psi_n$ becomes very small). In any case, construction (3.6-3.7) is by construction larger than (3.8-3.9), thus it suffices to show the claim under the premise that construction (3.8-3.9) applies with probability 1. The argument is similar to the above.

Proof of main result. Every sequence (π_n, ψ_n) can be decomposed into subsequences s.t. one of the above lemmas applies to each subsequence. The pre-tests are designed to use the appropriate procedure depending on features of (π_n, ψ_n). If (π_n, ψ_n) is far away from the benchmark sequences specified in the pre-tests, this match will be perfect and one of the above lemmas will apply directly. If $CI_{1-\alpha}(\Theta)$ might oscillate between different procedures in the limit, some additional argument is needed. To keep track of the 49 potential case distinctions, categorize the possible subsequences as in the following table.

		$\mathbf{1}-\psi_n = \ldots$			$\psi_n = \ldots$			
		$\mathbf{o}(\mathbf{c}_n)$	$\mathbf{O}(\mathbf{c}_n)$	**(other)**	$\mathbf{O}(\mathbf{b}_n)$	$\mathbf{o}(\mathbf{b}_n)$	$\mathbf{O}(\mathbf{c}_n)$	$\mathbf{o}(\mathbf{c}_n)$
$\pi_n = \ldots$	$\mathbf{o}(\mathbf{c}_n)$	2	2	2	2	2	2	2
	$\mathbf{O}(\mathbf{c}_n)$	2	4	4	4	4	10	8
	(other)	2	4	1	1	1	11	6
$\mathbf{1}-\pi_n = \ldots$	$\mathbf{O}(\mathbf{b}_n)$	2	4	1	5	5	12	6
	$\mathbf{o}(\mathbf{b}_n)$	2	4	1	5	3	3	3
	$\mathbf{O}(\mathbf{c}_n)$	2	13	14	15	3	3	3
	$\mathbf{o}(\mathbf{c}_n)$	2	9	7	7	3	3	3

In this table, "other" refers to all sequences s.t. $\psi_n > O(b_n)$ and $1 - \psi_n > O(c_n)$ (respectively the same for π_n). In cases labelled 1, the baseline construction is valid and will be used with probability approaching 1. The same is true for $CI^1_{1-\alpha}(\Theta)$ in the cases labelled 2 and for $CI^2_{1-\alpha}(\Theta)$ in the cases labelled 3. In cases labelled 4, $CI_{1-\alpha}(\Theta)$ may oscillate between constructions $CI^1_{1-\alpha}(\Theta)$ and $CI^5_{1-\alpha}(\Theta)$. Note, though, that in these cases one will have $CI^5_{1-\alpha}(\Theta) \subseteq CI^1_{1-\alpha}(\Theta)$ by construction and furthermore that lemma A.5 applies, thus $CI_{1-\alpha}(\Theta)$ is valid (if potentially conservative). In case 5, an analogous argument applies but with $CI^2_{1-\alpha}(\Theta)$ and $CI^5_{1-\alpha}(\Theta)$. In case 6, one can directly apply lemma A.3, and in case 7, the same holds for lemma A.4. In case 8, $CI_{1-\alpha}(\Theta)$ may oscillate between $CI^1_{1-\alpha}(\Theta)$ and $CI^3_{1-\alpha}(\Theta)$, but $CI^3_{1-\alpha}(\Theta) \subseteq CI^1_{1-\alpha}(\Theta)$ by construction and lemma A.3 applies. A similar argument applies to case 9. In all of cases 10-12, $CI^3_{1-\alpha}(\Theta)$ and $CI^5_{1-\alpha}(\Theta)$ are asymptotically equivalent. Validity in case 11, where $CI_{1-\alpha}(\Theta) \in \left\{CI^3_{1-\alpha}(\Theta), CI^5_{1-\alpha}(\Theta)\right\}$ with probability approaching 1, follows from lemma A.5. In cases 10 and 12, where the probability of $CI_{1-\alpha}(\Theta) = CI^1_{1-\alpha}(\Theta)$ (case 10) or $CI_{1-\alpha}(\Theta) = CI^2_{1-\alpha}(\Theta)$ (case 12) fails to vanish, additional argument along previous lines is needed. The analog argument holds for cases 13-15.

Consider now the claim that $\lim_{n\to\infty} \inf_{\theta_{yz}\in\Theta_{yz}} \mathbb{P}(\theta_{yz} \in CI_{1-\alpha}(\theta)) = 1-\alpha$. The proof plan for this is similar to the above, and we only elaborate steps that differ. In particular, lemmas A.1 and A.2 immediately imply the analogous result here. It remains to demonstrate the following.

Lemma B.3. Assume that $\phi_n\sigma_\pi (\hat{\pi} - \pi_n) \xrightarrow{d} \mathcal{N}(0,1)$, that $\phi_n\sigma_\pi \left(\widehat{\psi} - \psi_n\right) \xrightarrow{p} 0$, and that $\widehat{\sigma}_\pi/\sigma_\pi \to 1$. Then $\lim_{n\to\infty} \inf_{\theta_{yz}\in\Theta_{yz}} \mathbb{P}(\theta_{yz} \in CI^3_{1-\alpha}(\theta)) = 1-\alpha$.

Proof. Again, with probability approaching 1 we have $\Theta_0 = [\pi_n - \psi_n, \pi_n]$, $\Delta = \psi_n$, $\widehat{\Theta} = \left[\widehat{\pi} - \widehat{\psi}, \widehat{\pi}\right]$, and $\widehat{\Delta} = \min\left\{\widehat{\pi}, 1-\widehat{\pi}, \widehat{\psi}, 1-\widehat{\psi}\right\} = \widehat{\psi}$, thus $\widehat{\Delta}$ is superefficient relative to the rate of convergence of $\hat{\pi}$. Parameterizing the true parameter value as $\theta = \pi_n - a\psi_n$ for $a \in [0,1]$,[17] one

[17]Strictly speaking we should allow a to be a moving parameter as well, but obviously any sequence $\{a_n\}$ will have finitely many accumulation points in $[0,1]$ and the argument can be conducted separately along the according

can then write

$$
\begin{aligned}
& \lim_{n\to\infty} \mathbb{P}(\theta \in CI^3_{1-\alpha}(\theta)) \\
= \ & \lim_{n\to\infty} \mathbb{P}\left(\widehat{\pi} - \widehat{\psi} - c_{1-\alpha}\widehat{\sigma}_\pi {\phi_n}^{-1} \leq \pi_n - a\psi_n \leq \widehat{\pi} + c_{1-\alpha}\widehat{\sigma}_\pi {\phi_n}^{-1}\right) \\
= \ & \lim_{n\to\infty} \mathbb{P}\left(\phi_n \widehat{\sigma}_\pi^{-1}\left(a\psi_n - \widehat{\psi}\right) - c_{1-\alpha} \leq \phi_n \widehat{\sigma}_\pi^{-1}\left(\pi_n - \widehat{\pi}\right) \leq c_{1-\alpha} + \phi_n \widehat{\sigma}_\pi^{-1} a\psi_n\right) \\
= \ & \lim_{n\to\infty} \mathbb{P}\left(\phi_n \sigma_\pi^{-1}(a-1)\psi_n - c_{1-\alpha} \leq \phi_n \sigma_\pi^{-1}\left(\pi_n - \widehat{\pi}\right) \leq c_{1-\alpha} + \phi_n \sigma_\pi^{-1} a\psi_n\right) \\
= \ & \lim_{n\to\infty} \left(\Phi\left(c_{1-\alpha} + \phi_n \sigma_\pi^{-1} a\psi_n\right) - \Phi\left(\phi_n \sigma_\pi^{-1}(a-1)\psi_n - c_{1-\alpha}\right)\right).
\end{aligned}
$$

Direct evaluation of derivatives shows that this limit is concave in a and is minimized when $a \in \{0, 1\}$, in which case it equals $1 - \alpha$.

Lemma B.4. Assume that $\phi_n \sigma_\psi \left(\widehat{\psi} - \psi_n\right) \xrightarrow{d} \mathcal{N}(0,1)$, that $\phi_n \sigma_\psi \left(\widehat{\pi} - \pi_n\right) \xrightarrow{p} 0$, and that $\widehat{\sigma}_\psi / \sigma_\psi \to 1$. Then $\lim_{n\to\infty} \inf_{\theta_{yz}\in\Theta_{yz}} \mathbb{P}(\theta \in CI^4_{1-\alpha}(\theta)) = 1 - \alpha$.

Proof. This mimics lemma B.3.

Lemma B.5. Let the assumptions of lemma A.5 hold. Then $\mathbb{P}(\theta \in CI^5_{1-\alpha}(\theta)) = 1 - \alpha$.

Proof. In view of the fact that if $\Delta \to 0$, then $\phi_n \left(\widehat{\Delta} - \Delta\right) \to 0$, this follows by minimal adaptation of arguments in Stoye (2009, proposition 1).

Proof of main result. This is now analogous to the proof of the main result in the preceding proof.

References

[1] Afriat, S.N., 1973. On a System of Inequalities in Demand Analysis: An Extension of the Classical Method. *International Economic Review*, 14, 460-472.

[2] Andrews, D.W.K., and G. Soares, 2010. Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection. *Econometrica*, 78, 119–157.

[3] Bandyopadhyay, T., I. Dasgupta, and P.K. Pattanaik, 1999. Stochastic Revealed Preference and the Theory of Demand. *Journal of Economic Theory*, 84, 95-110.

subsequences.

[4] —, 2002. Demand Aggregation and the Weak Axiom of Stochastic Revealed Preference. *Journal of Economic Theory*, 107, 483-489.

[5] —, 2004. A General Revealed Preference Theorem for Stochastic Demand Behavior. *Economic Theory*, 23, 589-599.

[6] Barberà, S., and P.K. Pattanaik, 1986. Falmagne and the Rationalizability of Stochastic Choices in Terms of Random Orderings. *Econometrica*, 54, 707-715.

[7] Becker, G.S., 1962. Irrational Behavior and Economic Theory. *Journal of Political Economy*, 70, 1-13.

[8] Blundell, R., M. Browning, L. Cherchye, I. Crawford, B. de Rock, and F. Vermeulen, 2012. Sharp for SARP: Nonparametric Bounds on the Behavioural and Welfare Effects of Price Changes. Dicussion paper, Tilburg University.

[9] Blundell, R., M. Browning, and I. Crawford, 2003. Nonparametric Engel Curves and Revealed Preference. *Econometrica*, 71, 205-240.

[10] —, 2008. Best Nonparametric Bounds on Demand Responses. *Econometrica*, 76, 1227-1262.

[11] Cherchye, L., Crawford, I., De Rock, B., and F. Vermeulen, 2009. The Revealed Preference Approach to Demand. In D. Slottje (ed.), *Quantifying Consumer Preferences*. Emerald Books.

[12] Deaton, A., and J. Muellbauer, 1980. An Almost Ideal Demand System. *American Economic Review*, 70, 312-26.

[13] Falmagne, J.-C. 1978. A Representation Theorem for Finite Random Scale Systems. *Journal of Mathematical Psychology*, 18, 52–72.

[14] Fan, J., and I. Gijbels 1996. Local Polynomial Modeling and its Application. *Monographs on Statistics and Applied Probability*, Chapman and Hall.

[15] Fan, Y., and S.S. Park, 2010. Sharp Bounds on the Distribution of Treatment Effects and their Statistical Inference. *Econometric Theory*, 26, 931-951.

[16] Hall, P., 1992. On Bootstrap Confidence Intervals in Nonparametric Regression. *Annals of Statistics* 20, 695-711.

[17] Heckman, J.J., Smith, J.A., and N. Clements, 1997. Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts. *Review of Economic Studies*, 64, 487–536

[18] Horowitz, J. L., 2001. The Bootstrap. In J.J. Heckman & E.E. Leamer (eds.), *Handbook of Econometrics*, edition 1, volume 5, chapter 52, 3159-3228. Elsevier.

[19] Imbens, G., and C.F. Manski, 2004. Confidence Intervals for Partially Identified Parameters. *Econometrica*, 72, 1845-1857.

[20] Kihlstrom, R., Mas-Colell, A. and H. Sonnenschein, 1976, The Demand Theory of the Weak Axiom of Revealed Preference. *Econometrica*, 44, 971-978.

[21] Lewbel, A., 1999. Consumer Demand Systems and Household Expenditure. In H. Pesaran and M. Wickens (eds.), *Handbook of Applied Econometrics*. Basil Blackwell.

[22] Manski, C.F., 1997. The Mixing Problem in Programme Evaluation. *Review of Economic Studies*, 64, 537–553.

[23] —, 2003. *Partial Identification of Probability Distributions*. Springer Verlag.

[24] —, 2007. Partial Identification of Choice Probabilities. *International Economic Review*, 48, 1393-1410.

[25] Matzkin, R., 2006. Heterogeneous Choice. In R. Blundell, W. Newey, and T. Persson (eds.), *Advances in Econometrics: Proceedings of the 9th World Congress*. Cambridge University Press.

[26] McFadden, D., 2005. Revealed Stochastic Preference: A Synthesis. *Economic Theory*, 26, 245-264.

[27] McFadden, D., and M. Richter, 1991. Stochastic Rationality and Revealed Stochastic Preference. In J. Chipman, D. McFadden, and M. Richter (eds.), *Preferences, Uncertainty and Optimality: Essays in Honour of Leo Hurwicz.* Westview Press.

[28] Nelsen, R.B., 2006. An Introduction to Copulas (2^{nd} Edition). Springer Verlag.

[29] Nelsen, R.B., J.J. Quesada-Molina, J.A. Rodríguez-Lallena, and M. Úbeda-Flores, 2001. Bounds on Bivariate Distribution Functions with Given Margins and Measures of Association. *Communications in Statistics – Theory and Methods*, 30, 1155-1162.

[30] Nelsen, R.B. and M. Úbeda-Flores, 2004. A Comparison of Bounds on Sets of Joint Distribution Functions Derived from Various Measures of Association. *Communications in Statistics – Theory and Methods*, 33, 2299-2305.

[31] Rose, H., 1958. Consistency of Preferences: The Two-Commodity Case. *Review of Economic Studies*, 25, 124-125.

[32] Shao, J., and D. Tu, 1995. The Jackknife and Bootstrap. Springer Verlag.

[33] Stoye, J., 2009. More on Confidence Intervals for Partially Identified Parameters. *Econometrica*, 77, 1299-1315.

[34] Varian, H., 1982. The Nonparametric Approach to Demand Analysis. *Econometrica*, 50, 945-974.

[35] Yanagimoto, T., 1972. Families of Positively Dependent Random Variables. *Annals of the Institute of Statistical Mathematics*, 24, 559-573.

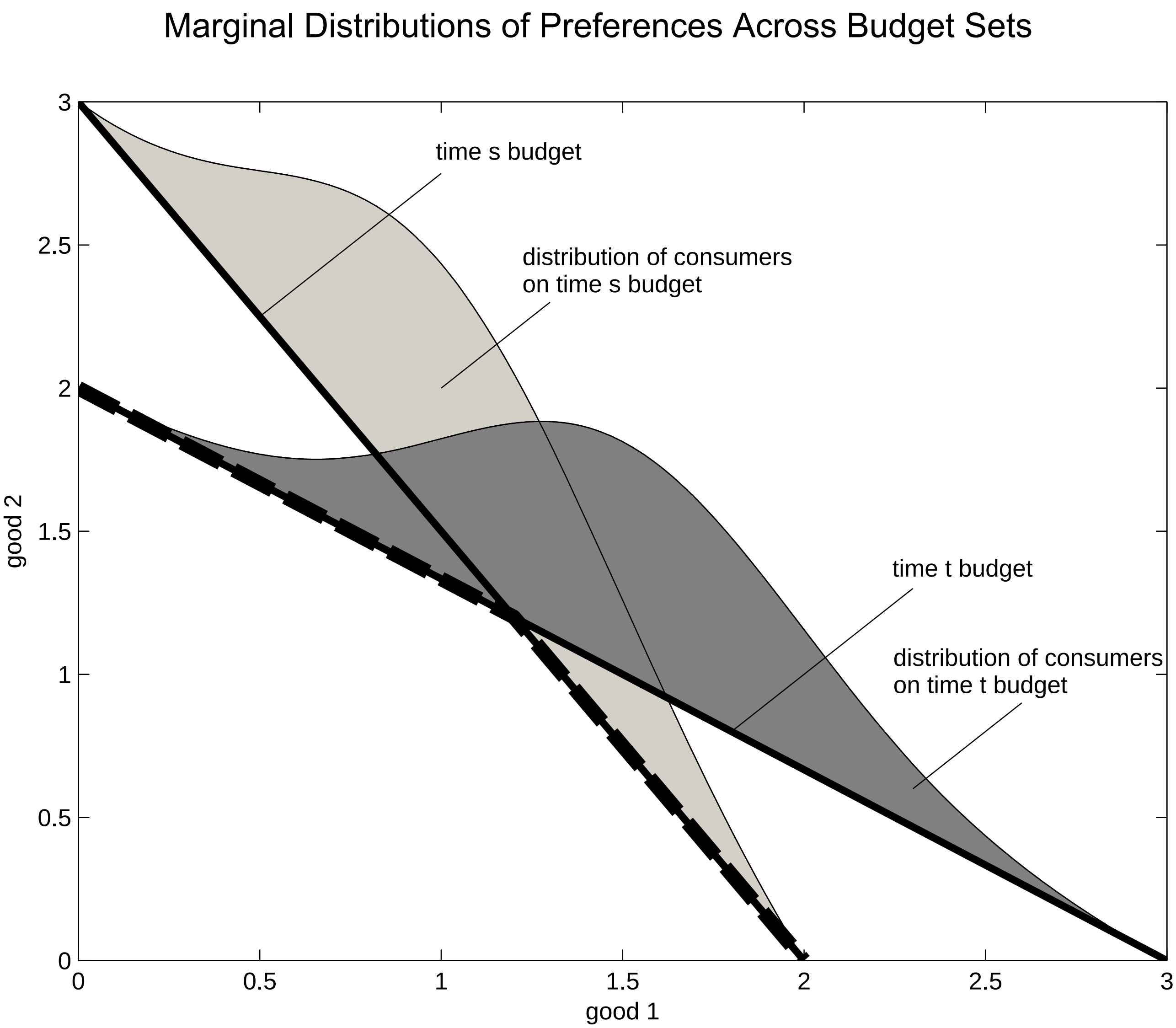
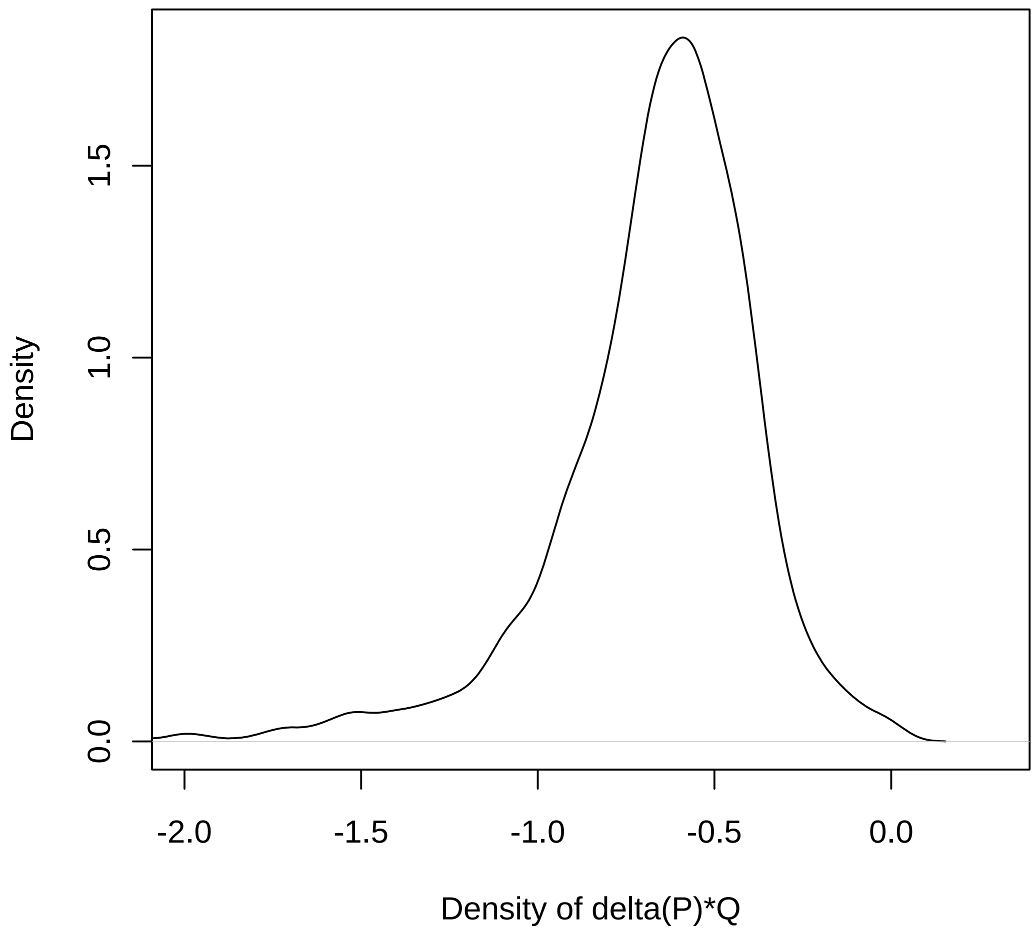
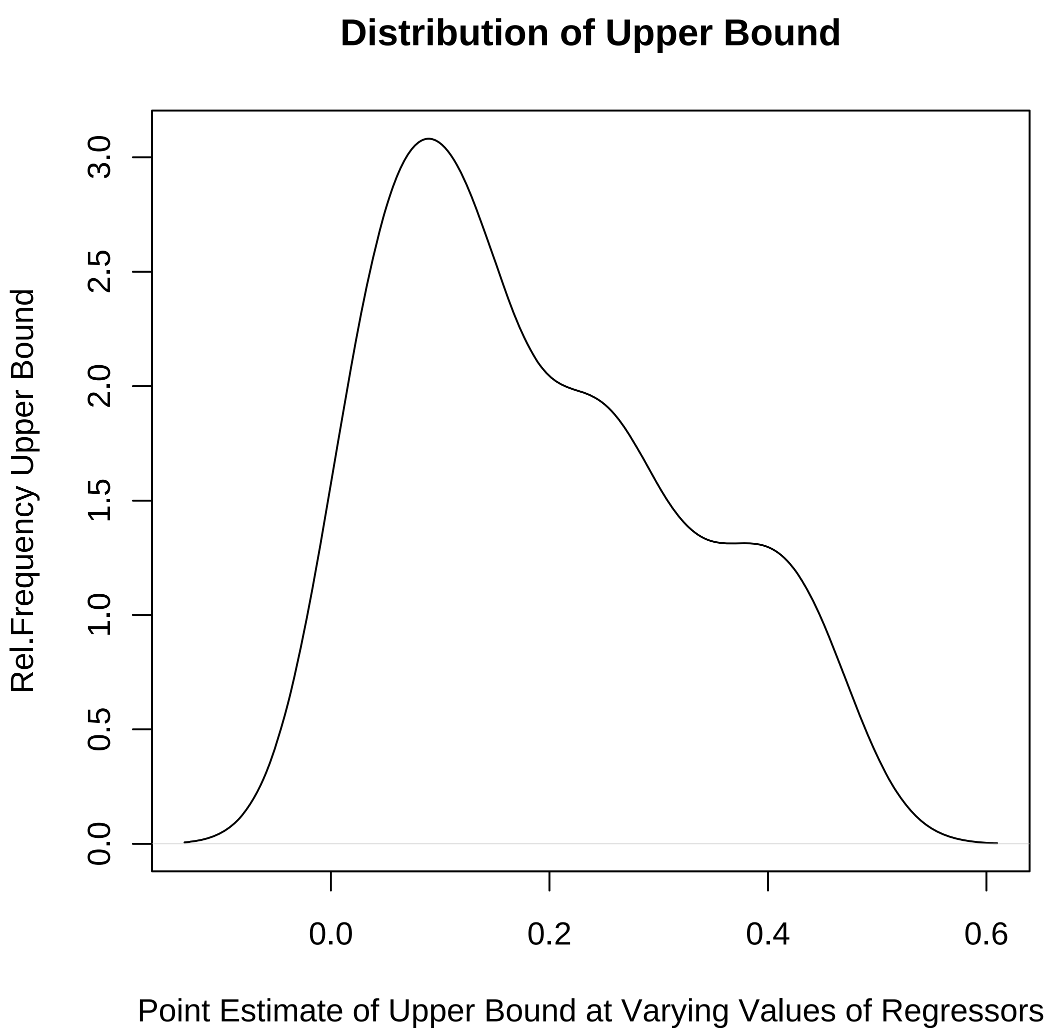

Marginal Distributions of Preferences Across Budget Sets
time s budget
distribution of consumers
on time s budget
time t budget
distribution of consumers
on time t budget
good 2
good 1
0
0.5
1
1.5
2
2.5
3
0
0.5
1
1.5
2
2.5
3

Figure 1



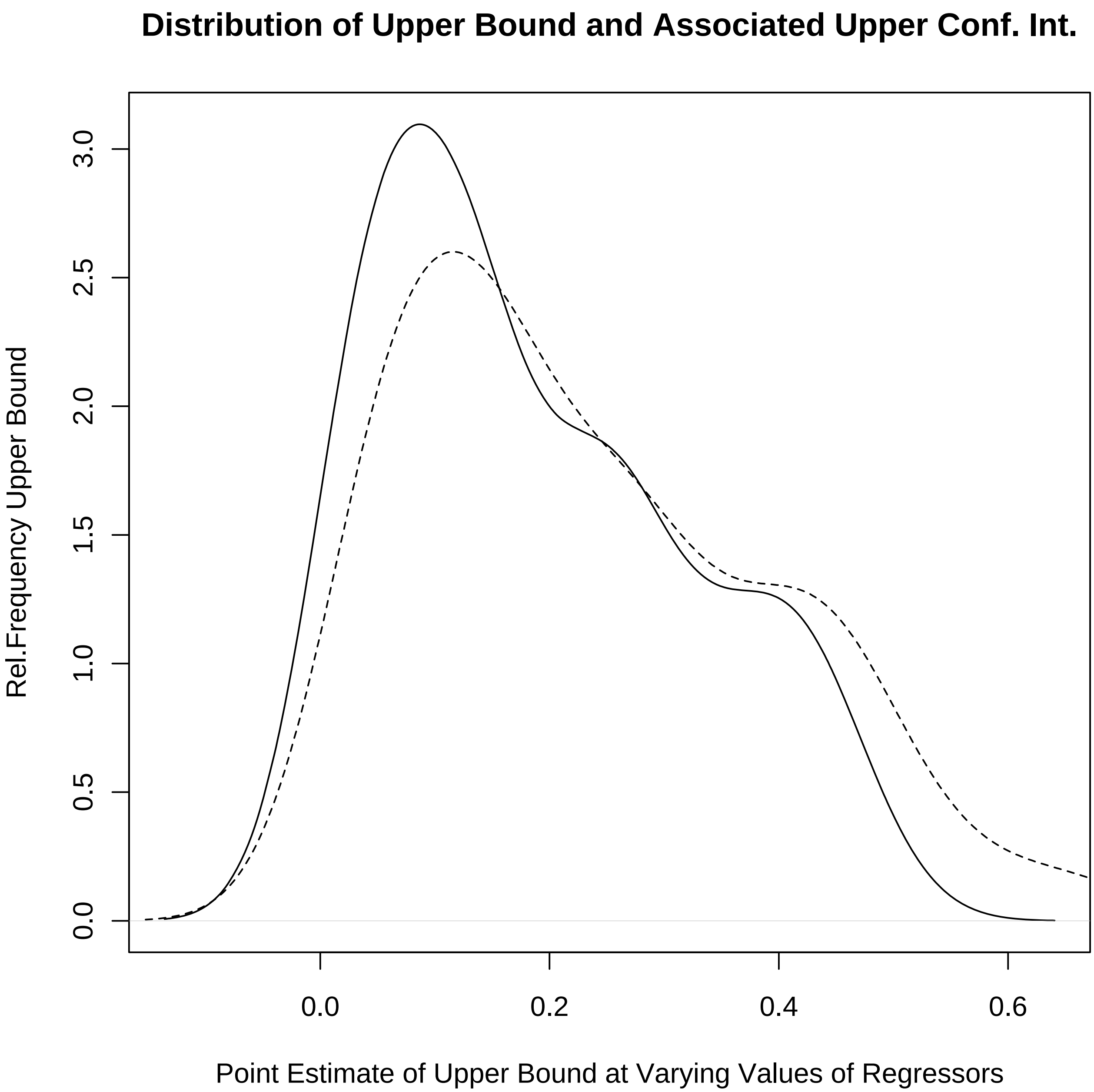
Graph showing density of quantity weighted price changes to determine region where data is informative.
Only income changes in the same order of magnitude can expose violations of rationality.
Subpopulations with large positive income shifts cannot be used to determine WARP.

Figure 2



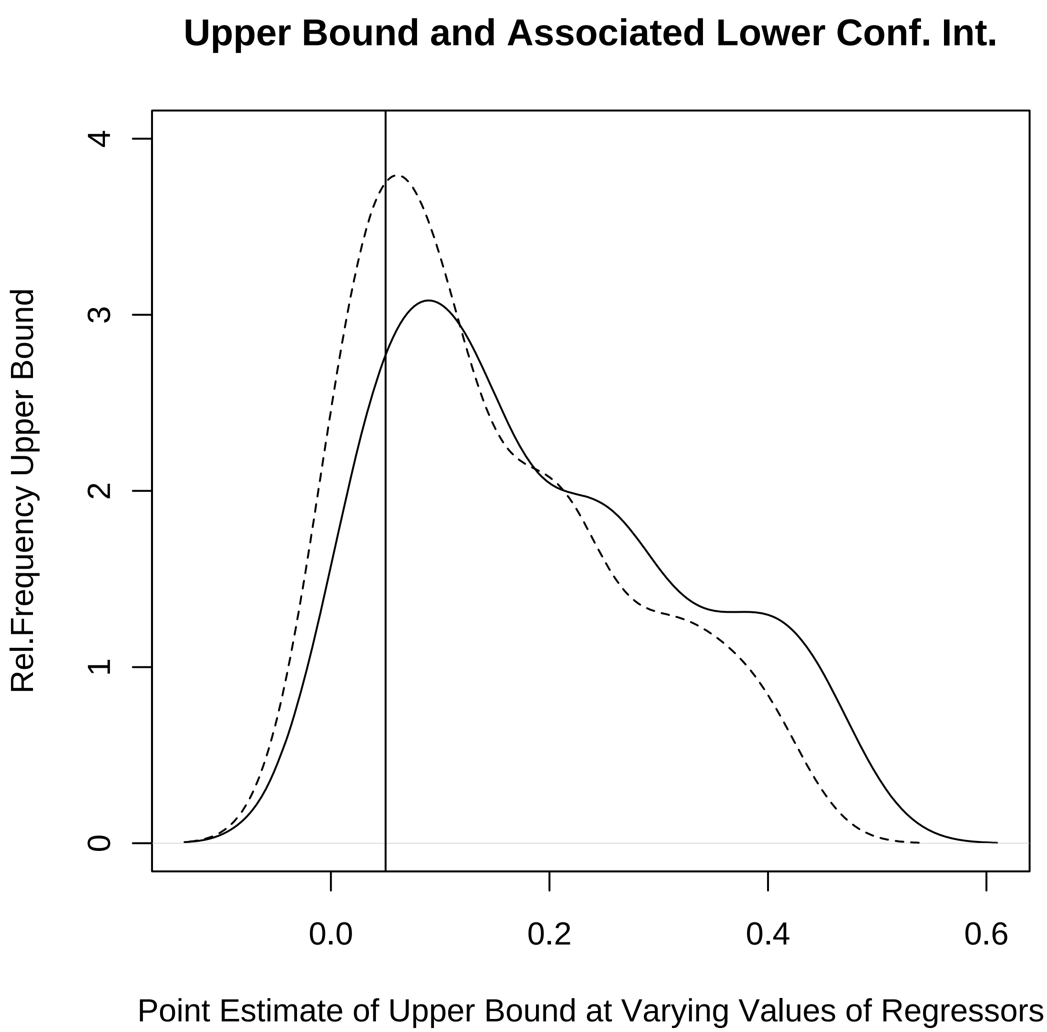
Graph showing the density of upper bound for violations of WARP over 100 subpopulations. In the subpopulations considered, the data are informative, as there could be between 10% and 60 % violations of rationality, and only few subpopulations show smaller upper bounds.

Figure 3



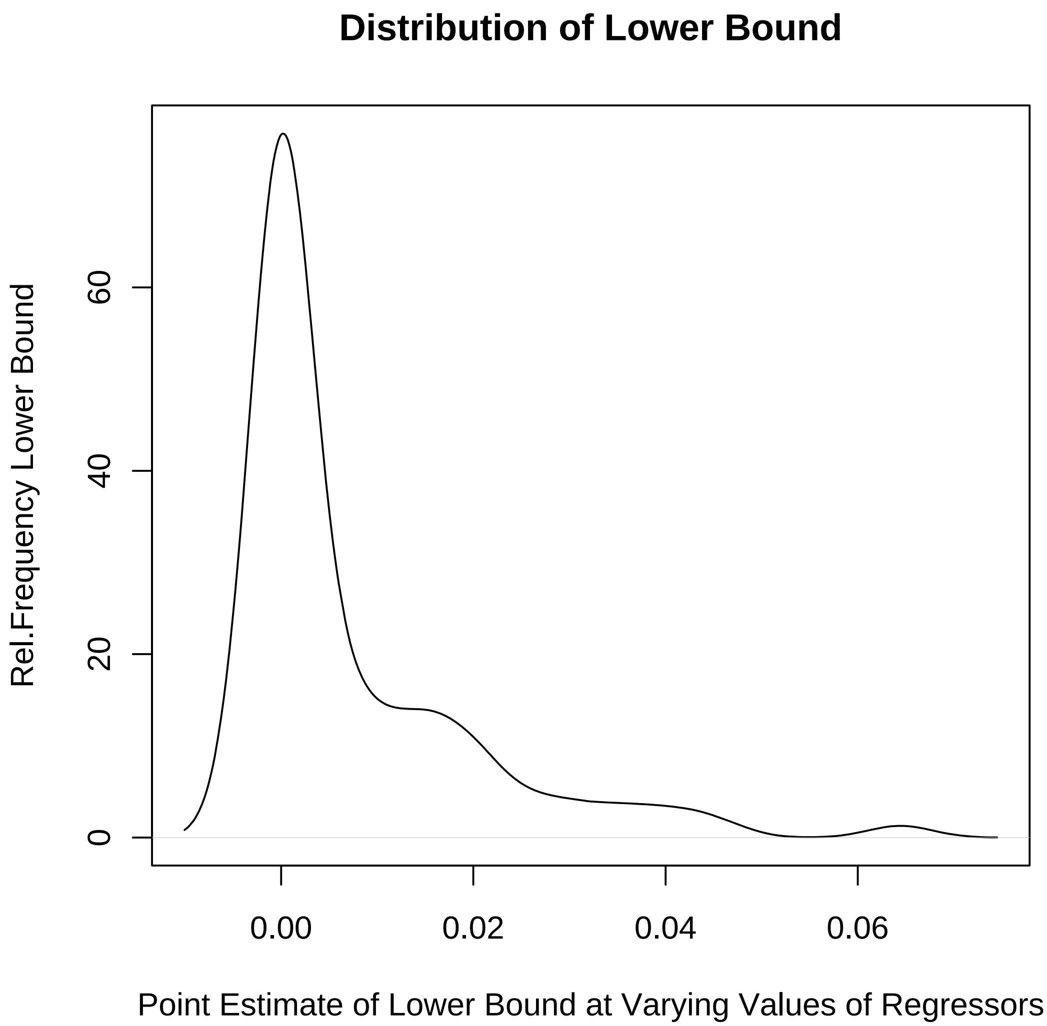
Same graph as figure 3, but with the density of upper confidence intervals, to account for sampling uncertainty. Compared to previous graph, data could be more informative, though not very much.

Figure 4



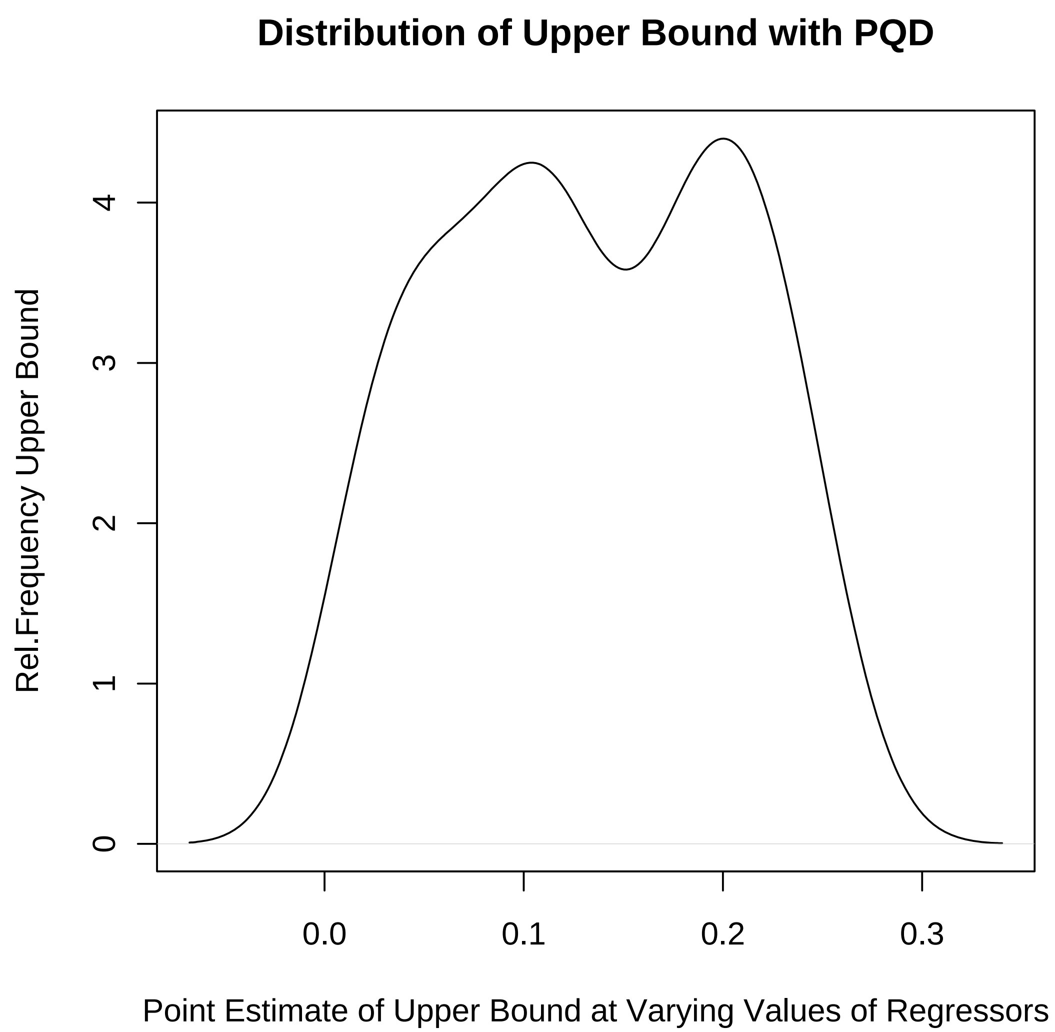
Same graph as figure 3, but with the density of lower confidence intervals, to account for sampling uncertainty. Vertical line = most subpopulations could have at least 5% violations with 95% Probability.

Figure 5



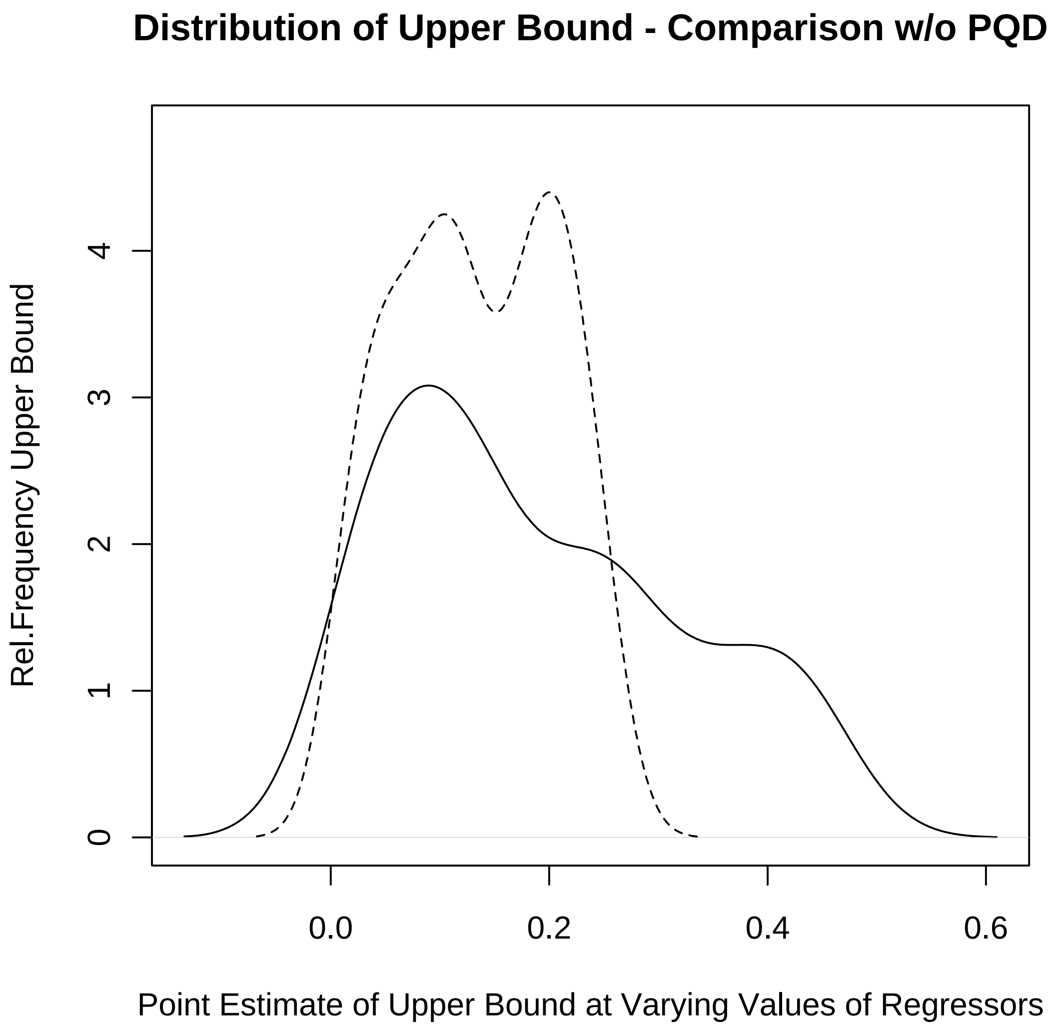
Density of point estimate for lower bound for subpopulations. Given the previous results that between 5% and 60% of the individuals within the subpopulations considered could violate rationality, we find little evidence that they do. Indeed, most point estimates are very close to 0.00.

Figure 6



Same graph as figure 3, but with the positive quadrant dependence refinement. Compared to figure 3, data are still informative, even after ruling out violations due to economically implausible behavior.

Figure 7



Direct comparison between figure 3 and figure 7, illustrating tightening of upper bound.

Figure 8