NONPARAMETRIC ANALYSIS OF RANDOM UTILITY MODELS: TESTING

YUICHI KITAMURA* AND JÖRG STOYE**

ABSTRACT. This paper aims at formulating econometric tools for investigating stochastic rationality, using the Random Utility Models (RUM) to deal with unobserved heterogeneity nonparametrically. Theoretical implications of the RUM have been studied in the literature, and in particular this paper utilizes the axiomatic treatment by McFadden and Richter (McFadden and Richter, 1991, McFadden, 2005). A set of econometric methods to test stochastic rationality given a cross-sectional data is developed. This also provides means to conduct policy analysis with minimal assumptions. In terms of econometric methodology, it offers a procedure to deal with nonstandard features implied by inequality restrictions. This might be of interest on its own right, both theoretically and practically.

1. INTRODUCTION

This paper develops new tools for the analysis of Random Utility Models (RUM). The leading application is stochastic revealed preference theory, that is, the modeling of aggregate choice behavior in a population characterized by individual rationality and unobserved heterogeneity. We test the null hypothesis that a repeated cross-section of demand data was generated by such a population, without restricting unobserved heterogeneity in any form whatsoever. Equivalently, we empirically test McFadden and Richter's (1991) Axiom of Revealed Stochastic Preference (ARSP, to be defined later), using only nonsatiation and the Strong Axiom of Revealed Preference (SARP) as restrictions on individual level behavior. Doing this is computationally challenging. We provide various algorithms that can be implemented with reasonable computational resources. Also, new tools for statistical inference for inequality restrictions are introduced in order to deal with the high-dimensionality and non-regularity of the problem at hand.

Let

$$u : \mathbf{R}_+^K \to \mathbf{R}$$

denote a random utility function. Again, randomness of u represents unobserved heterogeneity across individuals due to cross-sectional variation.[1] Each consumer faces an income level and a price vector in $\mathbf{R}_+^K$. Normalizing income to 1, the budget set for each consumer can be denoted as $B(p)$, $p \in \mathbf{R}_+^K$, and the consumer's choice is determined as

$$y = \arg \max_{y \in B(p)} u(y).$$

The econometrician observes a random sample of (y, p). In other words, she observes (a sample analog of) choice probability

$$\Pr(y \in Y \,|\, \text{price is } p)$$

for each $Y \subset \mathbf{R}_+^K$. The question is whether the joint distribution of (y, p) (or the choice probabilities) can be rationalized as an outcome of RUM. Our approach can be very briefly described by the following steps:

- The first insight is that, although demand data are continuous, they can be discretized without any loss of information as long as the set of budgets is finite. Thus, a random utility model of demand on a finite set of budgets is really a model of discrete choice, though the number of distinct choice objects can be large (up to 67 in our empirical application in Section 8). The next steps of our approach immediately apply to choice problems that were discrete to begin with.

- If there is a finite list of discrete choice problems, then there is a finite list of rational "choice types." Each such type is uniquely characterized by a rationalizable nonstochastic choice pattern. In realistic problem sizes, there are many such types (up to 177352 in our application), and obtaining the list is computationally challenging. Some tricks for efficiently computing the list are an important part of our contribution.

- Think of every rational choice type as defining a vector of degenerate choice probabilities over discrete choice objects. Then a corresponding vector of nondegenerate choice probabilities is consistent with a random utility model iff it is a convex combination of the degenerate ones. We collect the latter into columns of a matrix; choice probabilities are then rationalizable

---

[1]Random utility models were originally developed in mathematical psychology, and in principle, our results apply to stochastic choice behavior by an individual as well. However, in these settings it would frequently be natural to impose much more structure than we do.

iff they are in the column cone spanned by this matrix. The same insight informs our test statistic, which will be weighted Euclidean distance of estimated choice probabilities from this cone. In particular, it is computationally convenient to work with this cone and not the polytope that is generated by explicitly imposing that choice probabilities must be proper probabilities.

- The limiting distribution of the test statistic is intricate and, more importantly, depends discontinuously on unknown parameters. Such a feature has been studied extensively in the literature on moment inequalities. However, our problem has characteristics which are not well studied in that literature. In particular, unlike the standard problem where inequality constraints are stated in terms of moments, our inequality constraint is formulated in terms of unidentified parameters which in turn are connected to sample moments. In this sense the standard problem is concerned with 'direct' moment inequalities, whereas this paper considers what might be described as 'indirect' moment inequalities. Moreover the unidentified parameters are very high dimensional. To overcome these challenges, we generate critical values by a modified bootstrap method that avoids potential inconsistency. The procedure, which we call the 'tightening' method, is related to some approaches in the literature on moment inequalities, such as Andrews and Soares (2010), Bugni (2010) and Canay (2010), though these papers consider 'direct' inequalities. In contrast, our tightening method resolves difficulties associated with indirect inequalities. It is easy to implement and has practical appeal, potentially in very broad contexts.

In summary, the present paper contributes to the literature by (i) showing that McFadden and Richter's ARSP, and in particular stochastic rationality of a population, can be tested nonparametrically, and developing computational tools to do so, and (ii) proposing a method based on 'inequality tightening' as a way to address statistical and computational problems in dealing with high-dimensional, indirect inequalities being tested by the nonparametric test in (i). We emphasize that (ii) is also a contribution to specification testing of partially identified moment inequality models more generally, namely in cases where the indirect moment inequalities are the natural description of a model. We will also briefly explain how to carry out counterfactual analysis under ARSP and intend to flesh out this part of the analysis in a companion paper.

2. Related Literature

This paper first and foremost builds on the classic literature on (deterministic and stochastic) revealed preference. Inspired by Samuelson's (1938) statement of the revealed preference paradigm, Houthakker (1950), Afriat (1967; see also Varian (1982)), and Richter (1966) delineated the precise content of utility maximization if one observes a single consumer's demand behavior. For our purposes, this content is embodied in the Strong Axiom of Revealed Preference (SARP): The transitive closure of directly revealed preference must be acyclical.[2] This approach was extended to random utility maximization by Block and Marschak (1960), Falmagne (1978), McFadden and Richter (1991), and McFadden (2005). In particular, McFadden and Richter show that the precise content of random utility maximization – or equivalently, of individual level utility maximization in the presence of unrestricted, unobservable heterogeneity – is expressed by a collection of inequalities collectively dubbed "Axiom of Revealed Stochastic Preference" (ARSP).

These findings resolve this paper's questions "in the limit" when all identifiable quantities are known. In this idealized setting, they allow one to decide with certainty whether a given demand system or distribution of demands is rationalizable. In reality, estimators of these quantities might fail to be rationalizable because of sampling variation, and one can merely test the hypothesis that data might have been generated by a rational individual or population. For testing individual level rationality, such a test was proposed by Epstein and Yatchew (1985).[3] To the best of our knowledge, we provide the first such test for ARSP.

Perhaps the closest paper to ours in spirit is Manski (2007). In a simple, very abstract discrete choice problem (the universal set of options is a finite and, in practice, small list), he analyzes essentially the same question as we do. In particular, he states the testing and extrapolation problems in the

---

[2]The aforecited papers subtly differ in their handling of indifference. SARP characterizes rationality in the absence of indifference; else, it is sufficient but not necessary. Richter (1966) characterizes rationality if indifference is revealed through set-valued choice. The Afriat inequalities, or equivalently Varian's (1982) Generalized Axiom of Revealed Preference (GARP), characterize rationality if indifference is permitted but cannot be revealed through set-valued choice.

These differences do not matter in our setting. Observed choice is always unique, hence Richter's (1966) axiom collapses to SARP. Because nonsatiation will be assumed, GARP differs from SARP only with respect to choice objects that lie on the intersections of budget planes. With continuous demand on finitely many budgets, this case has zero probability (and does not occur in our data). In cases where these subtleties do matter, it would be easy to adopt our approach to any of these variations.

[3]See also the survey by Cherchye et al. (2009) and recent work by Dean and Martin (2011) or Echenique et al. (2011) for other approaches to testing individual level rationality.

abstract, solves them in simple examples, and outlines an approach to exact finite sample inference. Further results for simple instances of the problem, including results on the degree of underidentification of choice types, were provided by Sher et al. (2011). While we start from a continuous problem and use asymptotic theory rather than exact inference, the settings become similar after our initial discretization step. Our main contribution relative to these papers is to provide the computational toolkit as well as asymptotic theory to handle problems of realistic size. Indeed, this toolkit was recently employed for choice extrapolation by Manski (2012).

In a series of highly influential papers, Blundell, Browning, and Crawford (2003, 2008; BBC henceforth) develop a nonparametric approach to conduct demand analysis based on variations in Engel curves. They assume the same observables as we do, and apply their method to the British FES data (we use the same data set in Section 8). BBC analyze similar questions to ours but with respect to one individual level demand system generated by nonparametric estimation of Engel curves from these data. This approach could be loosely summarized as revealed preference analysis of a representative consumer (and in practice, given their specific estimation technique, of average demand). One possible foundation for it was provided by Lewbel (2001), who gives conditions on the random part of a random utility model that ensure integrability of expected demand. Lewbel's (2001) assumptions therefore precisely delineate one possible bridge between BBC's assumptions and ours. Also, BBC exploit only the implications of WARP for demand system and therefore test a necessary but not sufficient condition for rationalizability of aggregate demand; see Blundell , Browning, Cherchye, Crawford, de Rock, and Vermeulen (2012). A different test of necessary conditions is proposed by Hoderlein (2011), who shows that certain features of rational individual demand, like adding up and standard properties of the Slutsky matrix, are inherited by aggregate demand under weak conditions. The resulting test is passed by the FES data.

Hoderlein and Stoye (2010) again use the same assumptions and the same data but ask different questions. They bound from above and below the fraction of the population who violate the weak axiom of revealed preference (WARP). As a corollary (namely, by exhibiting the conditions under which the lower bound is zero), they show what discipline is put on the data by WARP alone. In the very special case of two goods, their results are formally connected to ARSP because WARP implies SARP (Rose (1958)), thus their corollary and ARSP must have the same specialization. This specialization is developed, and some implications are pointed out, in Stoye and Hoderlein (2013); in particular, they show that in the two-good case, recovering invertibility of demand by assuming monotonicity of demand with respect to an individual fixed effect is without loss of generality. The

latter paper contains no asymptotic theory, and the asymptotic theory in Hoderlein and Stoye (2010) is not closely related to ours.

Finally, our approach can be usefully contrasted to the recently active literature on invertibility of demand, that is, on conditions under which individual demand can be backed out from observation of repeated cross-sections. See Beckert and Blundell (2007), Berry, Gandhi and Haile (2011), and references therein. Unsurprisingly, invertibility requires substantial assumptions on structural parameters, i.e. utility functions and/or their distributions, which we avoid. The paper in this literature that is perhaps closest to ours is Blundell, Kristensen, and Matzkin (2011), who investigate nonparametric extrapolation of demand and, compared to our setting, essentially add only invertibility. Other than by adding this assumption, their paper differs from ours by restricting attention to two goods. The extension of their approach to more than two goods is challenging.

3. METHODOLOGY

Following McFadden and Richter (1991) as well as many of the aforecited references, we presume a finite number of budgets parameterized by $p \in \{p_1, ..., p_J\}$.[4] Indexing the corresponding budget planes as $B_j = B(p_j) := \{y \geq 0 : p_j y = 1\}$, we can drop p from notation. Also anticipating that we will assume monotonicity, we can write choice probability functions as

$$\pi(y \in Y | B_j) := \Pr(y \in Y | \text{budget plane is } B_j), \quad Y \subset \mathcal{Y}.$$

The space of choices $\mathcal{Y} \subset \mathbf{R}_+^K$ is unrestricted. In particular, it can be – and is, in our application – continuous. However, the following, important observation much simplifies the problem. For every budget B_j, let the I_j elements of $\left\{x_{1|j}, \ldots, x_{I_j|j}\right\}$ form the coarsest partition of B_j s.t. no budget plane other than B_j intersects the interior of any one element of the partition. In other words, $\left\{x_{1|j}, \ldots, x_{I_j|j}\right\}$ has the property that its elements are disjoint, that $B_j = \bigcup_{i=1}^{I_j} x_{i|j}$ for every j, and also that for any $k = 1, ..., J$, any $i = 1, ..., I_j$, and any $y_1, y_2 \in x_{i|j}$, one has $(p_k y_1 - 1)(p_k y_2 - 1) \geq 0$; furthermore, $\left\{x_{1|j}, \ldots, x_{I_j|j}\right\}$ is the coarsest such partition. Henceforth, we use the word "patch" to denote a generic element $x_{i|j}$ of $\mathcal{X}$. For the simplest nontrivial example, let there be $J = 2$ budgets that intersect, then there is a total of four patches: two on B_1, where one is above B_2 and the other one is below it, and two patches on B_2, where one is above B_1 and one is below it.

---

[4]The theory was extended to the continuous case in McFadden (2005), and we discuss prospects for the analogous extension of our approach in the conclusion.

We only restrict individual consumers' behavior by monotonicity ("more is better") and the Strong Axiom of Revealed Preference (SARP); equivalently, we assume that each consumer maximizes some nondecreasing utility function.[5] The only implication of monotonicity in our setting is that choices must be on (and not below) budget planes, as already reflected in notation. The only implication of SARP is to restrict, in possibly very intricate ways, whether choice can simultaneously be above certain budgets and below others. Therefore, two consumers who exhibit different demand behavior but choose from the same patch of each budget plane are either both consistent with our assumptions or both fail them. In short, the model enforces that the probability distribution of demand is supported on the patches and furthermore constrains the implied distribution *over* patches but not the distributions *on* them.

We can, therefore, restrict the choice space to be $\mathcal{X}$, effectively discretizing the choice problem to one with a total of $I := \sum_{j=1}^{J} I_J$ choice objects and corresponding choice probabilities. This leaves us with a large (in applications of practical interest) but finite set of distinct nonstochastic choice types. We will now explain how to efficiently encode these. To do so, arrange $\mathcal{X}$ as a vector

$$\mathcal{X} = (x_{1|1}, x_{2|1}, ..., x_{I_J|J})'$$

and similarly write

$$\pi_{i|j} := \Pr(y \in x_{i|j} | B_j),$$
$$\pi_j := (\pi_{1|j}, ..., \pi_{I_j|j})'$$
$$\pi := (\pi_1', ..., \pi_J')' = (\pi_{1|1}, \pi_{2|1}, ..., \pi_{I_J|J})',$$

then the I-vector π contains all information that is relevant for testing RUM.

Any conceivable pattern of nonstochastic choice behavior can be identified with a binary I-vector $a = (a_{1|1}, ..., a_{I_J|J})'$, where a component $a_{i|j}$ of a equals 1 iff the corresponding component $x_{i|j}$ of $\mathcal{X}$ is chosen from budget B_j. We call a rationalizable if there exists $u^* \in \mathcal{U}$ such that $a_{i|j} = \mathbf{1}\{\arg\max_{y \in B_j} u^*(x) \in x_{i|j}\}$ for all $a_{i|j}$. Under our assumptions, this is the case iff behavior encoded in a fulfills SARP, but we emphasize that the approach could be modified to impose more or less structure on a and thereby on agents' behavior. Every rationalizable vector a can be thought of as characterizing a rational, nonstochastic choice type. Let $\{a_1, ..., a_H\}$ collect all rationalizable vectors and collect them in the $(I \times H)$-matrix $A := [a_1, ..., a_H]$. Then a vector of choice probabilities π is

---

[5]In a setting characterized by choice from linear budgets, this is furthermore equivalent to assuming maximization of strictly concave utility (Varian (1982)).

stochastically rational iff there exists $\nu \in \Delta^{H-1}$ such that $A\nu = \pi$. In words, π must be a convex combination of the columns of A. The weights ν can be interpreted as implied population distribution over rational choice types; remember, though, that each of these choice types represents an equivalence class of observationally distinct choice types.

To see how little structure is placed on the model, observe how invertibility of demand, i.e. recoverability of individual consumers' demand, fails on many levels. First, we consider finitely many budgets, so with unconstrained heterogeneity, any given demand pattern is rationalizable iff it is rationalizable by infinitely many utility functions. Second, every column a represents a continuum of observationally distinct, rationalizable demand patterns because of the lumping of continuous demand data into patches. Third, ν is not identified: A is typically very far from full column rank, and a π that is rationalizable at all will typically be rationalizable by many different vectors ν. We return to partial identification of ν in section 9.1.

McFadden and Richter (1991; see also McFadden (2005, theorem 3.3)) anticipated the possible discretization of choice space in our setting and also also noted various equivalent statements for the empirical content of RUM. For example, ν as required here exists iff the linear program $\min_{\nu \geq 0, s \geq 0} \mathbf{1}'_I s$ s.t. $A\nu + s \geq \pi, \mathbf{1}'_H \nu \leq 1$ has an optimal solution with $s = 0$. However, we employ the first statement verbalized above, thus we directly test:

($\mathbf{H}_A$):   There exist a $\nu \in \Delta^{H-1}$ such that $A\nu = \pi$,

where Δ^{H-1} denotes the $(H-1)$-dimensional unit simplex. To test this hypothesis, we transform it as follows. First, note that $\mathbf{1}'_I A = [J, ..., J]$ and $\mathbf{1}'_I \pi = J$ hold by definition. Therefore, $A\nu = \pi$ implies $\mathbf{1}'_H \nu = 1$. Thus ($\mathbf{H}_A$) is equivalent to

($\mathbf{H}_B$): There exist $\nu \geq 0$ such that $A\nu = \pi$.

It is easy to see that Hypothesis ($\mathbf{H}_B$) is, in turn, equivalent to

($\mathbf{H}_C$): $\min_{\eta \in C}[\pi - \eta]'\Omega[\pi - \eta] = 0$,

where Ω is a positive definite matrix and $C := \{A\nu | \nu \geq 0\}$. Note that the constraint set C is a cone in $\mathbf{R}^I$. The solution η_0 of ($\mathbf{H}_C$) is the projection of $\pi \in \mathbf{R}^I_+$ onto the cone under the weighted norm

$\|x\|_\Omega = \sqrt{x'\Omega x}$. The corresponding value of the objective function is the square of the length of the projection residual vector. The projection η_0 is unique, but the corresponding ν is not. Stochastic rationality holds if and only if the length of the residual vector is zero.

A natural sample counterpart of the objective function in ($\mathbf{H}_C$) would be $\min_{\eta \in C}[\hat{\pi}-\eta]'\Omega[\hat{\pi}-\eta]$, where $\hat{\pi}$ is an estimator of π, for example by the vector of sample choice frequencies. It is useful to normalize this sample counterpart by N to obtain an appropriate asymptotic distribution, so define

$$(3.1) \qquad \begin{aligned} J_N &:= N \min_{\eta \in C}[\hat{\pi} - \eta]'\Omega[\hat{\pi} - \eta] \\ &= N \min_{\nu \in \mathbf{R}_+^H}[\hat{\pi} - A\nu]'\Omega[\hat{\pi} - A\nu]. \end{aligned}$$

Once again, ν is not unique at the optimum, but $\eta = A\nu$ is. Call its optimal value $\hat{\eta}$, noting that $\hat{\eta}$ can also be thought of as rationality-constrained estimator of choice probabilities. Then $\hat{\eta} = \hat{\pi}$, and $J_N = 0$, iff the estimated choice probabilities $\hat{\pi}$ are stochastically rational; obviously, our null hypothesis will be accepted in this case. To determine an appropriate critical value for our test, we will have to estimate the distribution of J_N. This estimation problem is rather intricate and will be handled in section 5.

4. COMPUTATION

We now turn to computational implementation of our approach. The challenge is threefold: First, how to encode the choice set $\mathcal{X}$; second, how to generate the matrix A, and last, how to carry out the constrained minimization in the definition of J_N. We explain our response to these issues in this order.

4.1. **Encoding $\mathcal{X}$.** Geometrically, the patches $x_{i|j}$ are polyhedra with rather complex descriptions. However, the only information that matters for computing A is how any given patch relates to the different budgets. To implement this insight, we switch from the double index expression $\mathcal{X} = \{\{x_{i|j}\}_{i=1}^{I_j}\}_{j=1}^{J}$ to a a single index $i = 1, ..., I$ and write $\mathcal{X} = \{x_i\}_{i=1}^{I}$ in the remainder of this section. Now, let

$$X_{ij} := \begin{cases} -1 & \text{if } x_i \text{ is below } B_j \\ 0 & \text{if } x_i \text{ is on } B_j \qquad j = 1, ..., J, \\ +1 & \text{if } x_i \text{ is above } B_j \end{cases}$$

then we can represent each x_i by the vector

$$X_i := [X_{i1}, ..., X_{iJ}], \quad i = 1, ..., I.$$

The choice set is represented by the $(I \times J)$-matrix $X = [X_1', ..., X_I']'$ with typical cell X_{ij}. Our algorithm for generating X from budget data $(p_1, ..., p_J)$ is as follows: First, generate all possible vectors X_i, ignoring non-negativity constraints. This generates a matrix with J columns and $\sum_{j=1}^{J} 2^{I-I_j}$ rows. Next, test the validity of each row. To illustrate this step, suppose $J = 5$ with K goods, and we want to verify whether the patch encoded by $(0, -1, 1, 1, 1)$ exists. It is easy to see that this is the case iff the system of $(J + 1 + K)$ inequalities

$$p_1'y \le 1, p_1'y \ge 1, p_2'y \le 1, p_3'y \ge 1, p_4'y \ge 1, p_5'y \ge 1, y \ge 0$$

has a solution. This can in principle be checked by repeated application of Fourier-Motzkin elimination, but in practice this is feasible only for small problems. A more efficient approach is to use numerical solvers that can quickly verify consistency of a set of linear inequality constraints. Using the `CVX` optimization package (Grant and Boyd (2008, 2011)), we find that checking the above inequalities is computationally inexpensive even for rather high dimensional commodity spaces.[6]

4.2. **Computing A.** To compute the matrix A, write

$$A = \{a_{ih}\},$$

where a_{ih} equals 1 if choice type h picks patch x_i from budget B_j, where B_j contains x_i. (As with $\mathcal{X}$, it is convenient for this section's purpose to index rows of A by $i = 1, ..., I$ instead of the double index $i|j$.) The challenge is to ensure that a nonstochastic choice type is represented as a column of A iff she is rationalizable, that is, iff the incomplete preference ordering revealed by her choice behavior is acyclical. To do this, we must first extract the preferences revealed by a choice type, then check for acyclicity. The following fact is crucial for the first step: Choice of x_i from budget B_j reveals that $x_i \succ x_m$ for every patch $x_m \ne x_j$ s.t. $X_{mj} \in \{-1, 0\}$ (the symbol $\succ$ signifies the revealed preference relation). Here, $X_{mj} = 0 \Rightarrow x_i \succ x_m$ follows directly because x_i and x_m are on the same budget plane, hence was chosen over x_m; $X_{mj} = -1 \Rightarrow x_i \succ x_m$ follows by additional uses of "more is better" and transitivity because x_i was chosen over some patch that dominates x_m. In a second step, acyclicity of a given set of revealed preferences can be checked by the Floyd-Warshall algorithm, efficient implementations of which are readily available.

---

[6]If this step were a computational bottleneck, this algorithm could be refined along the lines of the decision tree crawling algorithm discussed in the next subsection.

We implemented three ways to generate A.[7] First, a "brute force" approach initially creates a matrix $A^{\max}$ that represents all possible choice combinations from J budgets, thus it is of size $I \times \prod_{j=1}^{J} I_j$. Every column of $A^{\max}$ is then checked for rationality. This approach has the benefit of also collecting irrational types, which can be handy for simulations, but gets unwieldy rapidly because the number of columns of $A^{\max}$ grows extremely rapidly as budgets are added.

Our second approach is a "crawling" algorithm. Here, the core insight is that all possible choice patterns can be associated with terminal nodes of a decision tree whose initial node corresponds to choice from B_1, the next set of nodes to choice from B_2, and so on. The algorithm exhaustively crawls this tree. The important idea is to check for choice cycles at every node that is visited and, if a cycle is detected, to abandon the entire branch, i.e. to not visit any successor of the current node. A column of A is discovered every time that a terminal node is visited without detecting a cycle. The abandoning of entire branches means that most nonrationalizable choice patterns are never visited. For example, if a cycle is detected after specifying behavior on 4 out of 10 budgets, then none of the many possible completions of this choice pattern are considered. The downside is more frequent (for any rational pattern that is detected) execution of the Floyd-Warshall algorithm, but this algorithm is cheap (it terminates in polynomial time). The net effect is to improve computation time by orders of magnitude in complicated problems. Indeed, this is our most powerful algorithm that is directly applicable "off the shelf" in every conceivable application of our approach.

Finally, a modest amount of problem-specific adjustment can lead to further, dramatic improvement in many cases, including our empirical application. The key to this is contained in the following proposition, which is established in appendix A.

Proposition 4.1. *Suppose that for some $M \geq 1$, none of $(B_1, ..., B_M)$ intersect B_J. Suppose also that choices from $(B_1, ..., B_{J-1})$ are jointly rationalizable. Then choices from $(B_1, ..., B_J)$ are jointly rationalizable iff choices from $(B_{M+1}, ..., B_J)$ are.*

This proposition is helpful whenever not all budgets mutually intersect. To exploit it in applications, one must manually check for such sets of budgets and possibly be willing to reorder budgets. The benefit is that if the proposition applies, all rationalizable choice patterns can be discovered by checking for rationalizable choice patterns on smaller domains and then combining the results. In particular, proposition 1 informs the following strategy: First collect all rationalizable choice patterns on $(B_{M+1}, ..., B_{J-1})$. Next, for each such pattern, find all rationalizable completions of it to choice on

---

[7]Appendix B contains a more detailed description, including pseudo-code for some algorithms.

$(B_1, ..., B_{J-1})$ as well as all rationalizable completions to $(B_{M+1}, ..., B_J)$. (For the first of these steps, one may further economize by computing all rationalizable patterns on $(B_1, ..., B_M)$ in a preliminary step.) Every combination of two such completions is itself a rationalizable choice pattern. Note that no step in this algorithm checks rationality on J budgets at once; furthermore, a Cartesian product structure of the set of rationalizable choice patterns is exploited. The potential benefit is substantial – in our application, the refinement frequently improves computation times by orders of magnitude.

4.3. **Computing** J_N. Computation of J_N is a quadratic programming problem subject to a possibly large number of linear inequality constraints. We have nothing to add to the theory of solving such problems, though we utilize modern numerical solvers that can handle high-dimensional quadratic programming problems. Our currently preferred implementation utilizes `CVX`. We also implemented computation of J_N with `fmincon` (using stepwise quadratic programming) and Knitro. All solvers agree on those problems that they can handle, with `fmincon` being practical only for rather small problem sizes.

4.4. **Summary.** The above algorithms for computation of X, A, and J_N are readily applicable to actual problems with reasonable dimensionality, therefore they work for empirically interesting and relevant applications as we demonstrate in Section 8. To ensure understanding of the procedure, reconsider the simplest nontrivial example, i.e. two budgets which intersect. In this case $J = 2$ and $I_1 = I_2 = 2$, yielding $I = 4$ patches, and we have

$$
X = \begin{bmatrix} 0 & -1 \\ 0 & 1 \\ -1 & 0 \\ 1 & 0 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}, \quad \pi = \begin{bmatrix} \Pr(x_1 \text{ is chosen}|\text{budget is } B_1) \\ \Pr(x_2 \text{ is chosen}|\text{budget is } B_1) \\ \Pr(x_3 \text{ is chosen}|\text{budget is } B_2) \\ \Pr(x_4 \text{ is chosen}|\text{budget is } B_2) \end{bmatrix}.
$$

Here, the first row of X represents the part of budget B_1 that is below B_2 and so on. Rows of A corresponds to rows of X, whereas its columns correspond to types. Thus, the first column of A indicates the rational choice type whose choice from B_1 is below B_2 but whose choice from B_2 is above B_1; as a comprehension check, note that this particular type is better off under budget B_2. There are four logically possible choice types, but one of them, namely $[1, 0, 1, 0]'$, would violate the weak axiom of revealed preference and is therefore not represented in A. Given an estimator $\widehat{\pi} = (\widehat{\pi}_1, \widehat{\pi}_2, \widehat{\pi}_3, \widehat{\pi}_4)'$ for π and setting $\Omega = I_2$, one can then easily verify that $J_N = N \cdot (\max\{\widehat{\pi}_1 + \widehat{\pi}_3 - 1, 0\})^2$. In particular, the test statistic J_N is zero if $\widehat{\pi}_1 + \widehat{\pi}_3 \leq 1$, and in that case we immediately conclude that the data is consistent with the random utility model.

The example reproduces the known finding (Matzkin (2006), Hoderlein and Stoye (2009)) that with two budgets, the content of stochastic rationality is exhausted by the restriction "$\pi_1 + \pi_3 \leq 1$" on population choice probabilities. However, the tools proposed here allow one to perform similar computations for very complicated examples. Think of the size of A, that is, "number of patches $\times$ number of rational choice types," as an indicator of problem complexity. Then the example just given has size (4×3). For our empirical application, we successfully computed A, and computed J_N 2000 times, in cases where A has size (67×149570) respectively (64×177352). The computational bottlenecks in implementation are twofold: Computation of A took several hours in some cases (although reduced to minutes using proposition 4.1), and computation of J_N (which must be iterated over to compute critical values) took up to a minute.

5. Inference

This section discusses inferential procedures to deal with the hypothesis ($\mathbf{H}_A$) or its equivalent forms. We initially assume that choice probabilities were estimated by sample frequencies. For each budget j, denote the choices of N_j individuals, indexed by $n = 1, ..., N_j$, from the budget set B_j, by

$$d_{i|j,n} = \begin{cases} 1 \text{ if individual } n \text{ chooses } x_{i|j} \\ \\ 0 \text{ otherwise} \end{cases} \quad n = 1, ..., N_J.$$

Assume that one observes J random samples $\{\{d_{i|j,n}\}_{i=1}^{I_j}\}_{n=1}^{N_j}$, $j = 1, 2, ..., J$. For later use define

$$d_{j,n} := \begin{bmatrix} d_{1|j,n} \\ \vdots \\ d_{I_j|j,n} \end{bmatrix}, \quad N = \sum_{j=1}^{J} N_J.$$

An obvious way to estimate the vector π is to use choice frequencies

(5.1) $$\hat{\pi}_{i|j} = \sum_{n=1}^{N_j} d_{i|j,n}/N_j, i = 1, ..., I_j, j = 1, ..., J.$$

Our main task here is to estimate the sampling distribution of J_N under the null hypothesis that the true π is rationalizable. This problem is closely related to the literature on inequality testing, which is often concerned with hypotheses of the form

$$H_0 : B\theta \geq 0 \quad B \in \mathbf{R}^{p \times q} \text{ is known.}$$

With an (asymptotically) normal estimator so that $\sqrt{N}(\hat{\theta} - \theta) \approx N(0, S)$, one can form

$$T_N := \min_{\eta \in \mathbf{R}_+^q} N[B\hat{\theta} - \eta]' S^{-1}[B\hat{\theta} - \eta].$$

This type of problem has been studied by Gourieroux, Holly and Monfort (1982) and Wolak (1991). See also Chernoff (1954), Kudo (1963), Perlman (1969), Shapiro (1988), Kuriki and Takemura (2000), Andrews (2001), Rosen (2008), and Guggenberger, Hahn and Kim (2007).[8] A common way to get a critical value for T_N is to consider the least favorable case, which is $\theta = 0$. This strategy is inappropriate in the present context for a number of reasons. First, the vector π is a collection of probability vectors, so its least favorable value cannot be at the origin. Second, as discussed later, direct application of the above results to our problem is not computationally feasible. Third, even in the absence of the two obstacles, the suggested least favorable distribution can lead to an unnecessarily conservative procedure. Alternatively, one might consider a resampling method. As noted in the literature of set-identified models, the main difficulty in calculating a valid critical value for J_N is the treatment of points where more than one inequality is (almost) binding. This invalidates the naive bootstrap and like methods. For this and related issues, see, for example, Imbens and Manski (2004), Chernozhukov, Hong and Tamer (2007), Stoye (2009), Andrews and Soares (2010), Bugni (2010), and Romano and Shaikh (2010).

There are three methods for inference that can be potentially used to calculate a critical value for the statistic J_N. They are: (1) Regularization-based method, (2) Inequality selection method, and (3) "Inequality tightening." We now discuss them in turn. It is trivial that $\sqrt{N}(\hat{\pi} - \pi) \to_d N(0, S)$ holds, where S is the asymptotic covariance matrix. Let $\hat{S}$ denote a consistent estimator for S. The first approach, the method based on regularization, is easy to implement. This is basically the idea behind subsampling and the m-out-of-n bootstrap. Let $\tilde{\eta}_h := \hat{\eta} + \sqrt{\frac{h}{N}} N(0, \hat{S})$. Recall that $\hat{\eta}$ is the projection of the choice frequency vector $\hat{\pi}$ onto the cone C. The factor h is a sequence that goes to infinity slowly. The random variable $\tilde{\eta}_h$ is essentially a subsampled or m-out-of-n bootstrapped version of $\hat{\eta}$. Define $\tilde{J}_N(h) := \frac{N}{h} \min_{\nu \in \mathbf{R}_+^H} [\tilde{\eta}_h - A\nu]' \Omega [\tilde{\eta}_h - A\nu]$. The distribution of $\tilde{J}_N(h)$ can be evaluated by simulation. It provides a valid approximation of the distribution of J_N asymptotically, regardless

---

[8]More precisely, Guggenberger, Hahn and Kim (2007) consider specification testing of linear moment inequality models of the form $C\theta \leq \mathbb{E}[x], \theta \in \mathbf{R}^m$ where C is a known conformable matrix, and propose to test its equivalent form $R\mathbb{E}[x] \geq 0$ where R is another known matrix. This equivalence follows from the Weyl-Minkowski Theorem, as discussed later in this section. Our test of ARSP can be formally regarded as a specification test of this kind, though our testing procedure is quite different. In particular, we avoid resorting to Weyl-Minkowski equivalence because doing so is computationally infeasible.

of the position of η_0, the population version of $\hat{\eta}$, on the cone C. This is convenient computationally, though Andrews and Guggenberger (2009, 2010) forcefully argue that this approach can suffer from potential conservatism, often very severely. The second approach is based on inequality selection. This is essentially the so-called moment selection procedure proposed for moment inequality models (see, e.g., Andrews and Soares, 2010, Bugni, 2010). While the problem at hand is not exactly a moment inequality model, it does apply to the statistic J_N at least in theory. The idea of this procedure is that when $\hat{\eta}$ is close to the region where multiple constraints bind, it forces them to bind. The constraint set $C = \{A\nu|\nu \geq 0\}$ in the definition of J_N is expressed in terms of ν. Recall that ν corresponding to the projection η_0 is unidentified. Let $\{\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_H\}$ be the column vectors of an matrix A. Define

$$\text{cone}(A) = \{\nu_1\mathbf{a}_1 + ... + \nu_H\mathbf{a}_H : \nu_h \geq 0\},$$

then $C = \text{cone}(A)$. This is called the $\mathcal{V}$-representation of the cone C in the literature of convex geometry (see, e.g., Ziegler, 1995). The $\mathcal{V}$-representation is not useful in detecting whether $\hat{\eta}$ is close to an irregular point or not, and an alternative form expressed in terms of η has to be obtained. A result known as Weyl's Theorem guarantees that that is possible *theoretically*: it says that if a set C in $\mathbf{R}^J$ is represented as $\text{cone}(A), A \in \mathbf{R}^{I \times H}$, we can write

$$C = \{t \in \mathbf{R}^I | Bt \leq 0\} \text{ for some } B \in \mathbf{R}^{m \times I}.$$

The last line is called the $\mathcal{H}$-representation of C. (The converse is also true and known as Minkowski's Theorem.) Applying Weyl to the definition of J_N, obtain

$$J_N = \min_{t \in \mathbf{R}^I : Bt \leq 0} N[\hat{\pi} - t]'\Omega[\hat{\pi} - t].$$

The optimal t $(:= \hat{t})$ is unique, and it is now possible to apply the inequality selection procedure. Let κ be a sequence that diverges slowly to infinity. Let $b_1, ..., b_m$ be the row vectors of B. Suppose (wlog) $-\sqrt{N}b_1\hat{t} \leq \kappa, ..., -\sqrt{N}b_f\hat{t} \leq \kappa$ and $-\sqrt{N}b_{f+1}\hat{t} \geq \kappa, ..., -\sqrt{N}b_m\hat{t} \geq \kappa$ hold. Let $B_1 = [b_1', ..., b_f']'$ and $B_2 = [b_{f+1}', ..., b_m']'$. Redo the above minimization, but this time

$$\min_{t \in \mathbf{R}^I : B_1 t = 0, B_2 t \leq 0} N[\hat{\pi} - t]'\Omega[\hat{\pi} - t].$$

Let $\hat{t}_{select}$ denote the minimizer and define $\tilde{\eta}_\kappa := \hat{t}_{select} + \frac{1}{\sqrt{N}}N(0, \hat{S})$. The distribution of

$$\tilde{J}_n(\kappa) = \min_{t \in \mathbf{R}^I : Bt \leq 0} N[\tilde{\eta}_\kappa - t]'\Omega[\tilde{\eta}_\kappa - t].$$

offers a valid approximation to the distribution of J_N. This is expected to work well in finite samples, though it is computationally infeasible in the present context. To our knowledge, implementing Weyl's

Theorem to obtain an $\mathcal{H}$-representation out of the $\mathcal{V}$-representation is done by repeated application of (some variation of) the Fourier-Motzkin elimination algorithm, which is notoriously difficult to apply when the dimension of ν is high. This is exactly the case for the problem considered in this subsection, even for a small number of budgets.

The third approach, dubbed "inequality tightening" above, sidesteps the need for an $\mathcal{H}$-representation of C, and therefore it is simple to implement and applicable to problems of a realistic size. The idea is to "tighten" the constraint by replacing

$$\nu \geq 0$$

with

$$\nu \geq \tau_N \mathbf{1}_H$$

for some positive scalar τ_N that declines to zero slowly. The vector $\mathbf{1}_H$ can be some other fixed H-vector. Now solve

$$
\begin{aligned}
J_N(\tau_N) \quad &:= \quad \min_{\eta \in C_{\tau_N}} N[\hat{\pi} - \eta]'\Omega[\hat{\pi} - \eta] \\
&= \quad \min_{[\nu - \tau_N \mathbf{1}_H] \in \mathbf{R}_+^H} N[\hat{\pi} - A\nu]'\Omega[\hat{\pi} - A\nu]
\end{aligned}
$$

where $C_{\tau_N} := \{A\nu | \nu \geq \tau_N \mathbf{1}_H\}$, and let $\hat{\eta}_{\tau_N}$ denote the solution. The constraints that are almost binding at the solution (i.e. the ones with little slack) in the original statistic should be binding with zero slack after tightening. Let $\tilde{\eta}_{\tau_N} := \hat{\eta}_{\tau_N} + \frac{1}{\sqrt{N}}N(0, \hat{S})$. Notice that, as in the inequality selection procedure, no regularization (or subsampling/m-out-of-n bootstrapping) is necessary at this stage. Finally, define

$$
\begin{aligned}
(5.2) \qquad \tilde{J}_N(\tau_N) \quad &: \quad = \min_{\eta \in C_{\tau_N}} N[\tilde{\eta}_{\tau_N} - \eta]'\Omega[\tilde{\eta}_{\tau_N} - \eta] \\
&= \quad \min_{[\nu - \tau_N \mathbf{1}_H] \in \mathbf{R}_+^H} N[\tilde{\eta}_{\tau_N} - A\nu]'\Omega[\tilde{\eta}_{\tau_N} - A\nu],
\end{aligned}
$$

and use its distribution to approximate that of J_N. This has the same theoretical justification as the inequality selection procedure. Unlike the latter, however, it avoids the use of $\mathcal{H}$-representation, thus it does not require Fourier-Motzkin type algorithms. It offers a computationally feasible empirical testing procedure for stochastic rationality.

In the above procedure we tighten the constraint set C using its $\mathcal{V}$-representation. The next lemma shows that this, in effect, tightens the $\mathcal{H}$-representation of C as well, though directly tightening the $\mathcal{H}$-representation is computationally infeasible for the present problem. For a matrix B, let $\mathrm{col}(B)$ denote its column space.

Lemma 5.1. *For a $\tau > 0$ and a matrix $A \in \mathbf{R}^{I \times H}$, define*

$$C_\tau := \{A\nu | \nu \geq \tau \mathbf{1}_H\}.$$

Then it can be alternatively written as

$$C_\tau = \{t : Bt \leq -\tau\phi\},$$

where $B \in \mathbf{R}^{m \times I}$, $\phi \in \mathbf{R}^m_{++}$ and $\phi \in \text{col}(B)$.

This is different from the Minkowski-Weyl theorem for polyhedra, which would provide an expression of C_τ as the Minkowski sum of a convex hull and a non-negative hull. Lemma 5.1 is useful for proving the asymptotic validity of our procedure. Note that Canay (2010) proposes an approach that that can be regarded as a method that tightens the $\mathcal{H}$-representation of a moment inequality model. We think this is a potentially very important contribution, but its nature is quite distinct from that of ours. Like other papers in this area such as Andrews and Soares (2010), Bugni (2010) and Chernozhukov, Hong and Tamer (2007), Canay (2010) analyzes inequalities that are stated in terms of moments, the type of inequalities we called 'direct moment inequalities' in the overview in Section 1. Obtaining 'direct inequalities' in our problem would amount to transforming our original $\mathcal{V}$-representation to a $\mathcal{H}$-representation. If this were possible, then we could simply apply standard methods such as those proposed in Andrews and Soares (2010) or Bugni (2010). But the very point of our approach is to avoid this transformation because it is impossible to implement in practice. Our tightening procedure applies to an indirect inequality model, which, in our problem, is defined in terms of the numerically convenient $\mathcal{V}$-representation only. As we shall see in Section 8, this enables us to deal with very high-dimensional problems that appear in our empirical analysis.

The following assumption is used for our asymptotic theory.

Assumption 5.1. *For all $j = 1, ..., J$, $\frac{N_j}{N} \to \rho_j$ as $N \to \infty$, where $\rho_j \in (0, 1)$.*

Let $b_{k,i}$, $k = 1, ..., m$, $i = 1, ..., I$ denote the (k, i) element of B, then define

$$b_k(j) = [b_{k,N_1+\cdots N_{j-1}+1}, b_{k,N_1+\cdots N_{j-1}+2}, ..., b_{k,N_1+\cdots N_j}]'$$

for $1 \leq j \leq J$ and $1 \leq k \leq m$. Consider the following requirement:

Condition 5.1. *For some $\epsilon > 0$ and some $j = 1, ..., J$, $var(b_k(j)'d_{j,n}) \geq \epsilon$ holds for all $1 \leq k \leq m$.*

Note that the distribution of observations is uniquely characterized by the vector π. Let $\mathcal{P}$ denote the set of all π's that satisfy Condition 5.1 for some (common) value of ϵ.

Theorem 5.1. *Choose τ_N so that $\tau_N \downarrow 0$ and $\sqrt{N}\tau_N \uparrow \infty$. Also let $\Omega = \mathbf{I}_I$, the I-dimensional identity matrix. Then under Assumption 5.1*

$$\liminf_{N\to\infty} \inf_{\pi\in\mathcal{P}\cap C} \Pr\{J_N \le \hat{c}_{1-\alpha}\} = 1 - \alpha$$

where $\hat{c}_{1-\alpha}$ is the $1 - \alpha$ quantile of $\tilde{J}_N(\tau_N)$, $0 \le \alpha \le \frac{1}{2}$.

Proof: See Appendix A.

Technically, the boundedness of our observations $d_{j,n}$ — each element of $d_{i|j,n}$ is a Bernoulli($\pi_{i|j}$) random variable — and the above condition guarantee that the Lindeberg condition holds for $b_k'\hat{\pi}, k = 1,...,m$, which is important for the uniform size control result in the following theorem as it relies on a triangular array CLT. It is known that a triangular array $X_{iN}, i = 1,...,N \sim_{\text{iid}}$ Bernoulli(p_N), $N = 1, 2, ...$ obeys CLT iff $Np_N(1 - p_N) \to \infty$. Since the size control here is defined against a fixed class of distribution $\mathcal{F} \cap C$ as $N \to \infty$, it seems natural to impose a fixed lower bound ϵ in Condition 5.1. See the proof of Theorem 5.1 in Appendix for more on this point. In this theorem we use a particular choice of the weighting matrix Ω (i.e. the identity matrix); this suffices to ensure that the tightening method works appropriately when the empirical choice probability vector $\hat{\pi}$ is close to a nonregular point where multiple constraints (almost) bind. Note that we use identity matrices for weighting in our applications, in accordance with the limit theory presented here.

6. Bootstrap Algorithm with Tightening

This section details how to simulate the distribution of J_N with a bootstrap procedure that employs Theorem 5.1. First, we apply the standard nonparametric bootstrap to obtain resampled unrestricted choice probability vector estimates $\hat{\pi}^{*(r)}, r = 1,...,R$, where R denotes the number of bootstrap replications. This provides the bootstrap distribution estimate as the distribution of $\hat{\pi}^{*(r)} - \hat{\pi}, r = 1,...,R$, where, as before, $\hat{\pi}$ denotes the unrestricted choice probability vector. We need to generate the bootstrap samples under the null, however. A naive way to achieve this would be to center it around the restricted estimator $\hat{\eta}$, that is

$$\hat{\pi}_{naive}^{*(r)} = \hat{\pi}^{*(r)} - \hat{\pi} + \hat{\eta}, \quad r = 1,...,R.$$

Recall that $\hat{\eta}$ is the solution to

$$\begin{aligned} J_N &:= N \min_{\eta\in C}[\hat{\pi} - \eta]'\Omega[\hat{\pi} - \eta] \\ &= N \min_{\nu\in\mathbf{R}_+^H}[\hat{\pi} - A\nu]'\Omega[\hat{\pi} - A\nu]. \end{aligned}$$

But $\hat{\pi}^*_{naive}$ is invalid due to standard results about the failure of the bootstrap in discontinuous models (e.g., Andrews (2000)). The "tightening" remedy is to center it instead around the tightened restricted estimator. More precisely, our procedure is as follows:

(i) Obtain the τ_N-*tightened* restricted estimator $\hat{\eta}_{\tau_n}$, which solves

$$
\begin{aligned}
J_N &:= \min_{\eta \in C_{\tau_N}} N[\hat{\pi} - \eta]'\Omega[\hat{\pi} - \eta] \\
&= \min_{[\nu - \tau_N \mathbf{1}_H] \in \mathbf{R}^H_+} N[\hat{\pi} - A\nu]'\Omega[\hat{\pi} - A\nu]
\end{aligned}
$$

(ii) Calculate the bootstrap estimators under the restriction, using the recentering factor $\hat{\eta}_{\tau_n}$ obtained in (i):

$$
\hat{\pi}^{*(r)}_{\tau_n} := \hat{\pi}^{*(r)} - \hat{\pi} + \hat{\eta}_{\tau_n}, \quad r = 1, ..., R.
$$

(iii) Calculate the bootstrap test statistic by solving the following problem:

$$
\begin{aligned}
J_N^{*(r)}(\tau_N) &:= \min_{\eta \in C_{\tau_N}} N[\hat{\pi}^{*(r)}_{\tau_n} - \eta]'\Omega[\hat{\pi}^{*(r)}_{\tau_n} - \eta] \\
&= \min_{[\nu - \tau_N \mathbf{1}_H] \in \mathbf{R}^H_+} N[\hat{\pi}^{*(r)}_{\tau_n} - A\nu]'\Omega[\hat{\pi}^{*(r)}_{\tau_n} - A\nu],
\end{aligned}
$$

for $r = 1, ..., R$.

(iv) Use the empirical distribution of $J_N^{*(r)}(\tau_N), r = 1, ..., R$ to obtain the critical value for J_N.

This method relies on a tuning parameter τ_N which plays the role of a similar tuning parameter in the moment selection approach (namely, the parameter labeled κ in Andrews and Soares (2010)). In a simplified procedure in which the unrestricted choice probability estimate is obtained by simple sample frequencies, one reasonable choice would be

$$
\tau_N = \sqrt{\frac{\log \underline{N}}{\underline{N}}}
$$

where $\underline{N}$ is the minimum of the 'sample size' across budgets: $\underline{N} = \min_j N_j$ (N_j is the number of observations on Budget B_j: see (5.1)). The logarithmic penalization corresponds to the Bayes Information Criterion. The use of $\underline{N}$ can probably be improved upon, but suffices to ensure validity of our inference.

7. Test statistic with smoothing

In our empirical analysis in Section 8, the estimator $\hat{\pi}$ is a standard kernel estimator, so the above formula needs to be modified. Let $h_N(j)$ be the bandwidth applied to Budget B_j. Then an appropriate choice of τ_N is obtained by replacing N_j in the definition above with the "effective sample size" $N_j h_N(j)$. That is:

$$\tau_N = \sqrt{\frac{\log \underline{Nh}}{\underline{Nh}}}$$

where $\underline{Nh} = \min_j N_j h_N(j)$. Strictly speaking, asymptotics with nonparametric smoothing involve bias, and the bootstrap does not solve the problem. A standard procedure is to claim that one used undersmoothing and can hence ignore the bias. We follow this convention.

To formally state the asymptotic theory behind our procedure with smoothing, let $\mathbf{x}(j)$ denote the median (log) income levels for year j. We observe J random samples $\left\{ \left(\{d_{i|j}\}_{i=1}^{I_j}, x_n(j) \right) \right\}_{n=1}^{N_j}, j = 1, ..., J$. Instead of sample frequency estimators, we use

$$\hat{\pi}_{i|j} = \frac{\sum_{n=1}^{N_j} K\left(\frac{x_n(j) - \mathbf{x}(j)}{h_N(j)}\right) d_{i|j,n}}{\sum_{n=1}^{N_j} K\left(\frac{x_n(j) - \mathbf{x}(j)}{h_N(j)}\right)},$$

$$\hat{\pi}_j = (\hat{\pi}_{1|j}, ..., \hat{\pi}_{I_j|j})',$$

$$\hat{\pi} = (\hat{\pi}_1', ..., \hat{\pi}_J')',$$

(note the index j in $d_{i|j,n}$ now refers to year), where $x_n(j)$ is the log income of person n, $1 \leq n \leq N_j$ observed in year j. The kernel function K is assumed to be symmetric about zero, integrate to 1 and satisfy $\int |K(z)|dz < \infty$ and $\int z^2 K(z)dz < \infty$. Needless to say, $\hat{\pi}_{i|j}$ is a standard kernel regression estimator for

$$\pi_{i|j} := p_{i|j}(\mathbf{x}(j))$$

where

$$p_{i|j}(x) := \Pr(d_{i|j,n} = 1 | x_n(j) = x).$$

As before, we write $\pi_j := (\pi_{1|j}, ..., \pi_{I_j|j})'$ and $\pi := (\pi_1', ..., \pi_J')' = (\pi_{1|1}, \pi_{2|1}, ..., \pi_{I_J|J})'$. The smoothed version of J_N is computed using the above kernel estimator for $\hat{\pi}$ in (3.1). We also replace the normalizing factor N with $N \min_j h_N(j)$, which is convenient for asymptotic analysis. Likewise, $\tilde{J}_N(\tau_N)$ is obtained applying the same replacements to the formula (5.2), although generating $\tilde{\eta}_{\tau_N}$ requires a slight modification. Let $\hat{\eta}_{\tau_N}(j)$ be the j-th block of the vector $\hat{\eta}_{\tau_N}$, and $\hat{S}_j$ be a consistent estimator for the asymptotic variance of $\sqrt{N_j h_N(j)}(\hat{\pi}_j - \pi_j)$. We use $\tilde{\eta}_{\tau_N} = (\tilde{\eta}_{\tau_N}(1)', ..., \tilde{\eta}_{\tau_N}(J)')'$ for the smoothed version of $\tilde{J}_N(\tau_N)$, where $\tilde{\eta}_{\tau_N}(j) := \hat{\eta}_{\tau_N}(j) + \frac{1}{\sqrt{N_j h_N(j)}} N(0, \hat{S}_j), j = 1, ..., J$.

Let $\dot{h}$ ($\ddot{h}$) denote the first (second) derivative of a function h. Recall that the modulus of continuity of h defined on $\mathbf{R}$ at x_0 is given by

$$\omega(h, t, x_0) := \sup |h(x) - h(x_0)| \quad s.t. \quad |x - x_0| \leq t$$

for $t > 0$.

Condition 7.1. *Let $\epsilon > 0$ and $K < \infty$ be some constants, and $\rho : [0, \infty] \to [0, \infty]$ some function such that $\rho(t) \downarrow 0$ as $t \downarrow 0$. The following holds:*

(i) $\pi \in C$;

(ii) $var(b_k(j)'d_{j,n}|x_n(j) = \mathbf{x}(j)) \geq \epsilon$ holds for all $1 \leq k \leq m$ for some $j = 1, ..., J$;

(iii) $x_n(j)$ is continuously distributed with density f_j, and $f_j(\mathbf{x}(j)) \geq \epsilon$ for every $1 \leq j \leq J$;

(iv) $p_{i|j}$ is twice differentiable and $\omega\left(\ddot{p}_{i|j}, t, \mathbf{x}(j)\right) \leq \rho(t), t \geq 0$, $\sup |\dot{p}_{i|j}| \leq K$ and $\sup |\ddot{p}_{i|j}| \leq K$ for every $1 \leq i \leq I_j$ and $1 \leq j \leq J$;

(v) f_j is twice differentiable and $\omega\left(\ddot{f}_j, t, \mathbf{x}(j)\right) \leq \rho(t), t \geq 0$, $\sup f_j \leq K$, $\sup |\dot{f}_j| \leq K$ and $\sup |\ddot{f}_j| \leq K$ for every $1 \leq j \leq J$.

In what follows F_j signifies the joint distribution of $(d_{i|j,n}, x_n(j))$. Let $\mathcal{F}$ be the set of all $(F_1, ..., F_J)$ that satisfy Condition 7.1 for some $(\epsilon, K, \rho(\cdot))$.

Corollary 7.1. *Choose τ_N and $h_N(j), j = 1, ..., J$ so that $\tau_N \downarrow 0$, $h_N(j) = o(N^{-\frac{1}{5}})$ and $\sqrt{N \min_j h_N(j)}\tau_N \uparrow \infty$. Also let $\Omega = \mathbf{I}_I$, the I-dimensional identity matrix. Then under Assumption 5.1*

$$\liminf_{N \to \infty} \inf_{(F_1, ..., F_J) \in \mathcal{F}} \Pr\{J_N \leq \hat{c}_{1-\alpha}\} = 1 - \alpha$$

where $\hat{c}_{1-\alpha}$ is the $1 - \alpha$ quantile of $\tilde{J}_N(\tau_N)$, $0 \leq \alpha \leq \frac{1}{2}$.

The proof of Corollary 7.1 is almost the same as the proof of Theorem 5.1, except that the asymptotic normality proof in Bierens (1987, Section 2.2) and Condition 7.1 are used to obtain triangular array CLT's and uniform consistency results.

8. Empirical Application

We apply our methods to data from the U.K. Family Expenditure Survey, the same data used by Blundell, Browning, and Crawford (2008, BBC henceforth). To facilitate comparison of results, we use the same subset of these data as they do, namely the time periods from 1975 through 1999, households with a car and at least one child, and also the same composite goods, namely food,

nondurable consumption goods, and services (with total expenditure on the three standing in for income).

To test random utility models, our strategy is to extract one budget per year, which corresponds to that year's median expenditure. We then estimate the distribution of demand on that budget by kernel density estimation applied to budget shares, where the kernel is normal, the bandwidth was chosen according to Silverman's rule of thumb, and the underlying distance metric was $\log(income)$ (thus spending 10% more or 10% less than the median was considered equidistant from it). Like BBC, we assume that all consumers in one year face the same prices, and we use their price data. While budgets have a tendency to move outward over time, we find that there is substantial overlap of budgets. Thus, our analysis is not subject to the frequently reported problem that revealed preference tests are near vacuous because income gains dominate relative price changes. At the same time, the tendency of budgets to move outward means that budgets which are more than a few years apart do not tend to overlap, making the refinement of our crawling algorithm via Proposition 4.1 very powerful in this application.

It is computationally prohibitively difficult to test stochastic rationality on all 25 periods at once, so we rather work with all possible sets of eight consecutive periods. This leads to a collection of 18 matrices (X, A). Testing problems were of vastly different complexity, with the size of the matrix A ranging from (14×21) to (67×149570) and (64×177352); thus, there were up to 67 patches and up to 177352 rational choice types. Over this range of problem sizes, time required to compute A varied from split-second to several hours with the crawling algorithm. Time required to compute the larger matrices improves by a factor of about 100 using the refinement. Time required to compute J_N varied from split-second to about one minute, with problem size (67×149570) turning out to be the hardest. Thus, our informal assessment is that for computation of J_N, increasing I (and therefore the dimensionality of the quadratic programming problem's objective function) is more costly than increasing H (and therefore the number of linear constraints).

For each testing problem, we computed X, A, J_N, as well as critical values and p-values using the tightened bootstrap algorithm with $R = 2000$ resamples. Results are summarized in table 1. We find that $J_N = 0$ in only one case; that is, only in one case are the estimated choice probabilities rationalizable. However, violations of stochastic rationality are by and large not significant. We get one p-value below 10%, illustrating that in principle our test has some power, though this p-value must of course be seen in the context of the multiple hypothesis tests reported in table 1. Notice also

periods	I	H	N	J_N	p-value	10% c.v.	5% c.v.
75-82	50	42625	1229	.23	96%	3.1	4.0
76-83	67	149570	1209	.25	95%	2.5	3.3
77-84	55	97378	1160	.22	94%	2.0	2.5
78-85	49	61460	1108	.22	93%	1.7	2.3
79-86	43	38521	1033	.24	69%	.76	.94
80-87	35	7067	1033	.16	77%	2.1	3.4
81-88	27	615	995	.16	62%	.67	.82
82-89	14	21	987	0	100%	.24	.34
83-90	14	21	869	.00015	50%	2.1	3.5
84-91	16	63	852	.15	71%	2.6	3.8
85-92	27	1131	852	1.75	22%	3.1	4.3
86-93	43	10088	828	2.15	17%	2.9	3.8
87-94	49	55965	828	2.59	18%	3.5	4.6
88-95	52	75318	828	2.23	15%	2.8	3.6
89-96	64	17735	795	1.61	23%	2.7	3.6
90-97	46	22365	715	1.74	9.1%	1.69	2.2
91-98	39	9400	715	2.17	16%	2.8	3.5
92-99	21	225	715	2.58	10.2%	2.6	3.4

TABLE 1. Numerical Results of Testing Rationality Using FES data.

that the p-values seem to exhibit strong serial dependence. This is as expected – recall that any two consecutive p-values were computed from two subsets of data that overlap in seven of eight periods.

Figures 1 through 3 further illustrate the analysis, using the 1975-82 periods as example, i.e. they correspond to the first row in table 1. Figure 1 visualizes the relevant budget sets; figure 2 illustrates patches on the 1982 budget set; figure 3 depicts the bootstrapped distribution of J_N. There is substantial overlap across budgets, with the 1975-1982 periods generating 50 distinct "patches" and a total of 42625 choice types. The reason is that due to the 1970's recession, budgets initially contract and then recover over this time period, generating an intricate pattern of mutual overlaps. The test statistic for 1990-97 is close to the 10% critical value. An increase of the bootstrap size to 10000 (not reported) confirms a p-value of under 10% for this test statistic seen in isolation. However, a single

rejection at the 10% level is well within expectations for joint testing of many hypotheses. Our overall conclusion is that observed choice probabilities differ from rationalizable ones but not significantly so.

9. Further Applications and Extensions

9.1. **Partial Identification of ν.** Our setting gives rise to an identified set $\mathbf{H}_\nu$ for weights ν over rational choice types

$$\mathbf{H}_\nu = \{\nu \in \Delta^{H-1} : A\nu = \pi\},$$

and our main test could be interpreted as a specification test for the null hypothesis $H_0 : \mathbf{H}_\nu \neq \varnothing$. Estimation of $\mathbf{H}_\nu$ is, therefore, a natural issue. We focus on it somewhat less because the identified set is a collection of distributions over lists of choice behaviors and, at least in our application, is not immediately interpretable in terms of structural parameters of interest. This caveat might not apply to other applications, however.

If our model is misspecified, then $\mathbf{H}_\nu$ is empty. To generate an estimand that is nonempty by construction, define

$$\mathbf{H}_\nu(\Omega) = \arg \min_{\nu \in \Delta^{H-1}} \{(A\nu - \pi)' \Omega (A\nu - \pi)\},$$

This set is independent of the choice of the weighting matrix Ω (as long as Ω is positive definite), and then coincides with $\mathbf{H}_\nu$ as previously defined, iff the latter is nonempty. But it also defines a coherent notion of "pseudo-true (partially identified) distribution of rational types" even if the data are not rationalizable.

$\mathbf{H}_\nu$ is a singleton iff ν is point identified, which will rarely be the case in interesting applications of our framework. Indeed, in our application, $\mathbf{H}_\nu(\Omega)$ is a high-dimensional, convex polyhedron even in cases where $\widehat{\pi}$ is not rationalizable. That is, unlike in some applications of moment inequalities, failure of the sample criterion function to attain a minimal value of zero does not make the sample analog of the identified set a singleton.

Explicit computation of $\mathbf{H}_\nu(\Omega)$ is demanding. Our suggestion is to write

$$\mathbf{H}_\nu(\Omega) = \left\{ \nu \in \Delta^{H-1} : (A\nu - \pi)' \Omega (A\nu - \pi) - \min_{v \in \Delta^{H-1}} \{(Av - \pi)' \Omega (Av - \pi)\} = 0 \right\}$$

and compute a plug-in estimator that replaces π with $\widehat{\pi}$. Noting that we showed how to compute the inner minimum, computation of these estimators could utilize methods based on support vector machines as developed in current work by Bar and Molinari (2012). The estimator is clearly going to be consistent. An appropriately modified version of the method described in Sections 5 and 6 can be applied to inference on $\mathbf{H}_\nu$ or elements of $\mathbf{H}_\nu$, though we leave its detailed analysis for future research.

9.2. **Partial Identification of Counterfactual Choices.** The toolkit developed in this paper also allows for counterfactual analysis. At the most general level, to bound the value of any function $f(\nu)$ subject to the constraint that ν rationalizes the observed data, solve the program

$$\min_{\nu \in \mathbf{R}_+^H} / \max_{\nu \in \mathbf{R}_+^H} f(\nu) \quad \text{s.t.} \ A\nu = \hat{\eta},$$

recalling that $\hat{\eta} = \hat{\pi}$ whenever $\hat{\pi}$ is rationalizable.[9] Some interesting applications emerge by restricting attention to linear functions $f(\nu) = e'\nu$, in which case the bounds are furthermore relatively easy to compute because the program is linear. We briefly discuss bounding demand under a counterfactual budget, e.g. in order to measure the impact of policy intervention. This is close in spirit to bounds reported by Blundell et al. (2008; see also Cherchye et al. (2009) and references therein) as well as Manski (2007, 2012).

In this subsection only, assume that demand on budget B_J is not observed but is to be bounded. Write

$$A = \left[\begin{array}{c} A_{-J} \\ A_J \end{array} \right], \pi = \left[\begin{array}{c} \pi_{-J} \\ \pi_J \end{array} \right], \pi_J = \left[\begin{array}{c} \pi_{1|J} \\ \vdots \\ \pi_{I_J|J} \end{array} \right]$$

and let e_i signifies the i-th unit vector. We will begin by bounding components of the vector π_J.

Corollary 9.1. *The identified region for π_J is given by*

$$\mathbf{H}(\pi_J) = \{A_J\nu : A_{-J}\nu = \pi_{-J}, \quad \nu \geq 0\}.$$

In particular, $\pi_{i|J}$ is bounded by

$$\underline{\pi}_{i|J} \leq \pi_{i|J} \leq \overline{\pi}_{i|J}$$

where

$$\underline{\pi}_{i|J} = \min \left\{ e_i' A_J \nu \right\} \quad s.t. \quad A_{-J}\nu = \pi_{-J}, \quad \nu \geq 0$$
$$\overline{\pi}_{i|J} = \max \left\{ e_i' A_J \nu \right\} \quad s.t. \quad A_{-J}\nu = \pi_{-J}, \quad \nu \geq 0.$$

In this paper's application, the patches $x_{i|J}$, and hence the probabilities $\pi_{i|J}$, are not of intrinsic interest. However, they might be the object of interest in applications where the choice problem was

---

[9]Equivalently, one could use the previous subsection's notation to write

$$\min / \max f(v) \quad \text{s.t.} \ v \in \mathbf{H}(\nu, \Omega),$$

which is more similar to the way that similar problems are stated by Manski (e.g., Manski (2007)).

discrete to begin with. Indeed, the above is the bounding problem further analyzed in Sher at al. (2011).

Next, let $\delta(J) = \mathbb{E}[\arg\max_{y \in B_J} u(y)]$, thus the vector $\delta(J)$ with typical component $\delta_k(J)$ denotes expected demand in choice problem B_J. Define the vectors

$$\underline{d}_k(J) \quad : \quad = [\underline{d}_k(1|J), ..., \underline{d}_k(I_J|J)]$$

$$\overline{d}_k(J) \quad : \quad = [\overline{d}_k(1|J), ..., \overline{d}_k(I_J|J)]$$

with components

$$\underline{d}_k(i|J) \quad : \quad = \min\{y_k : y \in x_{i|J}\}, \quad 1 \le i \le I_J$$

$$\overline{d}_k(i|J) \quad : \quad = \max\{y_k : y \in x_{i|J}\}, \quad 1 \le i \le I_J,$$

thus these vectors list minimal respectively maximal consumption of good k on the different patches within B_J. Computing $(\underline{d}_k(i|J), \overline{d}_k(i|J))$ is a linear programming exercise. Then we have:

Corollary 9.2. *Expected demand for good k on budget B_J is bounded by*

$$\underline{\delta}_k(J) \le \delta_k(J) \le \overline{\delta}_k(J),$$

where

$$\underline{\delta}_k(J) \quad : \quad = \min \underline{d}_k(J) A_J \nu \quad s.t. \quad A_{-J}\nu = \pi_J, \quad \nu \ge 0$$

$$\overline{\delta}_k(J) \quad : \quad = \max \overline{d}_k(J) A_J \nu \quad s.t. \quad A_{-J}\nu = \pi_J, \quad \nu \ge 0.$$

Finally, consider bounding the c.d.f. of demand $F_k(z) = \Pr(y_k \le z)$. This quantity must be bounded in two steps. The event $(y_k \le z)$ will in general not correspond to a precise set of patches, that is, it is not measurable with respect to (the algebra generated by) $\{x_{1|J}, ..., x_{I_J|J}\}$. An upper bound on $F_k(z)$ will derive from an upper bound on the joint probability of all patches $x_{i|J}$ s.t. $y_k \le z$ holds for *some* $y \in x_{i|J}$. Similarly, a lower bound will derive from bounding the joint probability of all patches $x_{i|J}$ s.t. $y_k \le z$ holds for *all* $y \in x_{i|J}$.[10] We thus have:

Corollary 9.3. *For $k = 1, ..., K$ and $z \ge 0$, $F_k(z)$ is bounded from below by*

$$\min_{\nu \in \mathbf{R}_+^H} \sum_{\substack{i \in \{1, ..., I_J\}: \\ \overline{d}_k(i|J) \le z}} e_i' A_J \nu$$

$$s.t. \quad A_{-J}\nu \quad = \quad \pi_{-J}$$

---

[10]These definitions correspond to inner and outer measure, as well as to hitting and containment probability.

and from above by

$$\max_{\nu\in\mathbf{R}_+^H} \sum_{\substack{i\in\{1,...,I_J\}:\\ \underline{d}_k(i|J)\leq z}} e_i' A_J \nu$$

$$s.t. \quad A_{-J}\nu \quad = \quad \pi_{-J},$$

where $\left(\underline{d}_k(i|J), \overline{d}_k(i|J)\right)$ are defined as before.

While both the lower and the upper bound, seen as functions of z, will themselves be proper c.d.f.'s, they are not in general feasible distributions of demand for y_k. That is, the bounds are sharp pointwise but not uniformly. Also, bounds on a wide range of parameters such as the variance of demand follow from the above bounds on the c.d.f. through results in Stoye (2010). However, because the bounds on the c.d.f. are not uniform, these derived bounds will be valid but not necessarily sharp.

When trained on this paper's empirical application, these bounds are uncomfortably wide, motivating the search for nonparametric refinements that lead to narrower bounds without tightly constraining heterogeneity. This search, as well as the development of inference procedures for the bounds, are the subject of ongoing research. We also note that in his recent analysis of optimal taxation of labor, Manski (2012) uses the computational tools developed in this paper to find informative bounds.

9.3. **Choice from Binary Sets.** The methods developed in this paper, including the two extensions just discussed, immediately apply to nonparametric analysis of random discrete choice. Indeed, the initial discretization step that characterizes our analysis of a demand system is superfluous in this case. We briefly elaborate on one salient application that has received substantial attention in the literature, namely the case where choice probabilities for pairs of options,

$$\pi_{ab} := \Pr(a \text{ is chosen from } \{a, b\})$$

are observed for all pairs of choice objects $\{a, b\}$ drawn from some finite, universal set $\mathcal{A}$.

Finding abstract conditions under which a set of choice probabilities $\{\pi_{ab} : a, b \in \mathcal{A}\}$ is rationalizable has been the objective of two large, disjoint literatures, one in economics and one in operations research. See Fishburn (1992) for a survey of these literatures and Manski (2007) for a recent discussion of the substantive problem. There exists a plethora of necessary conditions, most famously Marschak's (1960) triangle condition, which can be written as

$$\pi_{ab} + \pi_{bc} + \pi_{ca} \leq 2, \forall a, b, c \in \mathcal{A}.$$

This condition is also sufficient for rationalizability if $\mathcal{A}$ contains at most 5 elements (Dridi (1980)). Conditions that are both necessary and sufficient in general have proved elusive. We do not discover abstract such conditions either, but provide the toolkit to numerically resolve the question in complicated cases, including a statistical test if probabilities are estimated rather than perfectly observed. To see this, define $J = (\#\mathcal{A})(\#\mathcal{A} - 1)/2$ "budgets" that correspond to distinct pairs $a, b \in \mathcal{A}$, and let the vector X (of length $I = 2J$) stack these budgets, where the ordering of budgets is arbitrary and options within a budget are ordered according to a preassigned ordering on $\mathcal{A}$. Each rational type (and thus, column of the matrix A) then corresponds to an ordering of the elements of $\mathcal{A}$ and can be characterized by a binary I-vector with just the same interpretation as before. As before, an I-vector of choice probabilities π whose components correspond to components of X is rationalizable iff $A\nu = \pi$ for some $\nu \in \Delta^{H-1}$. All methods developed in this paper apply immediately.

To illustrate, let $\mathcal{A} = \{a, b, c\}$, then one can write

$$X = \begin{bmatrix} a \\ b \\ b \\ c \\ c \\ a \end{bmatrix}, \pi = \begin{bmatrix} \pi_{ab} \\ \pi_{ba} \\ \pi_{bc} \\ \pi_{cb} \\ \pi_{ca} \\ \pi_{ac} \end{bmatrix}, A = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix},$$

and one can readily verify that $A\nu = \pi$ for some $\nu \in \Delta^5$ iff both $\pi_{ab} + \pi_{bc} + \pi_{ca} \leq 2$ and $\pi_{cb} + \pi_{ba} + \pi_{ac} \leq 2$. More generally, the matrix A has $H = (\#\mathcal{A})!$ columns. This limits computational feasibility of our approach, but note that the set of orderings of elements of $\mathcal{A}$ is easily characterized, so that computation time per column of A will be low.

10. CONCLUSION

This paper presented asymptotic theory and computational tools for completely nonparametric testing of Random Utility Models. Again, the null to be tested was that data were generated by a RUM, interpreted as describing a heterogeneous population, where the only restrictions imposed on individuals' behavior were "more is better" and SARP. In particular, we allowed for unrestricted, unobserved heterogeneity and stopped far short of assumptions that would recover invertibility of demand. As a result, the distribution over utility functions in the population is left (very) underidentified. We showed that specification testing of the model is possible nonetheless. The method is easily

adapted to choice problems that are discrete to begin with, and one can easily impose more (or also fewer) restrictions at the individual level.

Possibilities for extensions and refinements abound. We close by mentioning some salient issues.

(1) The methods discussed in this section are computationally intensive. The proposed algorithms work for a reasonably sized problem, though it is important to make further improvements in the algorithms if one wishes to deal with a problem that is large, say in terms of the number of budgets.

(2) We formulated the J_N statistic for the case with a finite set of budgets. In certain applications, however, it would be necessary to deal with a continuum of budgets. Theoretically, this can be handled by considering an appropriate discretization argument (McFadden (2005)). For the proposed projection-based econometric methodology, such an extension requires evaluating choice probabilities locally over points in the space of p via nonparametric smoothing, then use the choice probability estimators in the calculation of the J_N statistic. The asymptotic theory then needs to be modified. Another approach that can mitigate the computational constraint is to consider a partition of the space of p such that $\mathbf{R}_+^K = \mathcal{P}_1 \cup \mathcal{P}_2 \cdots \cup \mathcal{P}_M$. Suppose we calculate the J_N statistic for each of these partitions. Given the resulting M statistics, say $J_N^1, \cdots, J_N^M$, we can consider their function such as $J_N^{\max} := \max_{1 \leq m \leq M} J_N^m$. Or, alternatively, their weighted average may be used. These extensions and their formal statistical analysis are of practical interest.

(3) It might frequently be desirable to control for observable covariates to guarantee the homogeneity of the distribution of unobserved heterogeneity. Once again, this requires incorporating nonparametric smoothing in estimating choice probabilities, then averaging the corresponding J_N statistics over the covariates. This extension will be pursued.

(4) The econometric techniques outlined here can be potentially useful in much broader contexts. Again, our proposed hypothesis test can be regarded as specification test for a moment inequalities model. The proposed statistic J_N is an inequality analogue of goodness-of-fit statistics such as Hansen's (1982) overidentifying restrictions test statistic. While there are other proposals for specification testing in moment inequality models (see, e.g., Andrews and Guggenberger (2009), Andrews and Soares (2010), Bugni (2010), and Guggenberger, Hahn and Kim (2008)), the current approach provides an alternative and convenient procedure. As we discussed in Section 1, our procedure applies to 'indirect moment inequality models' as well. This can be useful in other applications, for example

in any setting where theoretical restrictions inform the $\mathcal{V}$-representation of a cone, meaning that its $\mathcal{H}$-representation may not be available in practice.

11. APPENDIX A: PROOFS

Proof of Proposition 4.1. We begin with some preliminary observations. Throughout this proof, $c(B_i)$ denotes the object actually chosen from budget B_i.

(i) If there is a choice cycle of any finite length, then there is a cycle of length 2 or 3 (where a cycle of length 2 is a WARP violation). To see this, assume there exists a length N choice cycle $c(B_i) \succ c(B_j) \succ c(B_k) \succ ... \succ c(B_i)$. If $c(B_k) \succ c(B_i)$, then a length 3 cycle has been discovered. Else, there exists a length $N-1$ choice cycle $c(B_i) \succ c(B_j) \succ c(B_k) \succ ... \succ c(B_i)$. The argument can be iterated until $N = 4$.

(ii) Call a length 3 choice cycle *irreducible* if it does not contain a length 2 cycle. Then a choice pattern is rationalizable iff it contains no length 2 cycles and also no irreducible length 3 cycles. (In particular, one can ignore reducible length 3 cycles.) This follows trivially from (i).

(iii) Let $J = 3$ and $M = 1$, i.e. assume there are three budgets but two of them fail to intersect. Then any length 3 cycle is reducible. To see this, assume w.l.o.g. that B_1 is below B_3, thus $c(B_3) \succ c(B_1)$ by monotonicity. If there is a choice cycle, we must have $c(B_1) \succ c(B_2) \succ c(B_3)$. $c(B_1) \succ c(B_2)$ implies that $c(B_2)$ is below B_1, thus it is below B_3. $c(B_2) \succ c(B_3)$ implies that $c(B_3)$ is below B_2. Thus, there is a length 2 cycle in (B_2, B_3).

We are now ready to prove the main result. The nontrivial direction is "only if," thus it suffices to show the following: If choice from $(B_1, ..., B_{J-1})$ is rationalizable but choice from $(B_1, ..., B_J)$ is not, then choice from $(B_{M+1}, ..., B_J)$ cannot be rationalizable. By observation (ii), if $(B_1, ..., B_J)$ is not rationalizable, it contains either a 2-cycle or an irreducible 3-cycle. Because choice from all triplets within $(B_1, ..., B_{J-1})$ is rationalizable by assumption, it is either the case that some (B_i, B_J) constitutes a 2-cycle or that some triplet (B_i, B_k, B_J), where $i < k$ w.l.o.g., reveals an irreducible choice cycle. In the former case, B_i must intersect B_J, hence $i > M$, hence the conclusion. In the latter case, if $k \leq M$, the choice cycle must be a 2-cycle in (B_i, B_k), contradicting rationalizability of $(B_1, ..., B_{J-1})$. If $i \leq M$, the choice cycle is reducible by observation (iii). Thus, $i > M$, hence the conclusion. $\qquad\square$

Proof of Lemma 5.1. By the Minkowski-Weyl theorem

$$C = \{t \in \mathbf{R}^I : Bt \leq 0\}.$$

Letting $\nu_\tau = \nu - \tau \mathbf{1}_H$, we have

$$
\begin{aligned}
C_\tau &= \{A\nu | \nu_\tau \geq 0\} \\
&= \{A[\nu_\tau + \tau \mathbf{1}_H] | \nu_\tau \geq 0\} \\
&= C \oplus \tau A \mathbf{1}_H \\
&= \{t : t - \tau A \mathbf{1}_H \in C\}
\end{aligned}
$$

where $\oplus$ signifies Minkowski sum. Define

$$
\phi = -BA\mathbf{1}_H.
$$

Using the $\mathcal{H}$-representation of C,

$$
\begin{aligned}
C_\tau &= \{t : B(t - \tau A\mathbf{1}_H) \leq 0\} \\
&= \{t : Bt \leq -\tau\phi\}.
\end{aligned}
$$

Note that the above definition of ϕ implies $\phi \in \mathrm{col}(B)$. Also define

$$
\begin{aligned}
\Phi &:= -BA \\
&= -\begin{bmatrix} b_1' \\ \vdots \\ b_m' \end{bmatrix} [a_1, \cdots, a_H] \\
&= \{\Phi_{kh}\}
\end{aligned}
$$

where $\Phi_{kh} = b_k' a_h, 1 \leq k \leq m, 1 \leq h \leq H$ and let e_h be the h-th standard unit vector in $\mathbf{R}^H$. Since $e_h \geq 0$, the $\mathcal{V}$-representation of C implies that $Ae_h \in C$, and thus

$$
BAe_h \leq 0
$$

by its $\mathcal{H}$-representation. Therefore

(11.1) $$\Phi_{kh} = -e_k' BAe_h \geq 0, \quad 1 \leq k \leq m, 1 \leq h \leq H.$$

Write $\phi = (\phi_1, ..., \phi_m)'$. We now show that $\phi_k \neq 0$ for all $1 \leq k \leq m$. To see it, note

$$
\begin{aligned}
\phi &= \Phi \mathbf{1}_H \\
&= \begin{bmatrix} \sum_{h=1}^H \Phi_{1h} \\ \vdots \\ \sum_{h=1}^H \Phi_{mh} \end{bmatrix}
\end{aligned}
$$

But unless we have the case of $\text{rank}(B) = 1$ (in which case the proof is trivial) it cannot be that

$$a_j \in \{x : b'_k x = 0\} \quad \text{for all } j$$

Therefore for each k, $\Phi_{kh} = b'_k a_h$ is nonzero at least for one $h, 1 \le h \le H$. Since (11.1) implies that all of $\{\Phi_{kh}\}_{h=1}^H$ are non-negative, we conclude that

$$\phi_k = \sum_{h=1}^H \Phi_{kh} > 0$$

for all k. We now have

$$C_\tau = \{t : Bt \le -\tau\phi\}, \quad \phi > 0$$

where the strict vector inequality is meant to hold element-by-element. $\square$

Proof of Theorem 5.1. By applying the Minkowski-Weyl theorem and Lemma 5.1 to J_N and $\tilde{J}_N(\tau_N)$, we see that our procedure is equivalent to comparing

$$J_N = \min_{t \in \mathbf{R}^I : Bt \le 0} N[\hat{\pi} - t]'\Omega[\hat{\pi} - t]$$

with the $1 - \alpha$ quantile of the distribution of

$$\tilde{J}_N(\tau_N) = \min_{t \in \mathbf{R}^I : Bt \le -\tau_N\phi} N[\tilde{\eta}_{\tau_N} - t]'\Omega[\tilde{\eta}_{\tau_N} - t]$$

with $\phi \in \mathbf{R}_{++}^m$, where

$$\tilde{\eta}_{\tau_N} = \hat{\eta}_{\tau_N} + \frac{1}{\sqrt{N}} N(0, \hat{S}),$$

$$\hat{\eta}_{\tau_N} = \underset{t \in \mathbf{R}^I : Bt \le -\tau_N\phi}{\text{argmin}} N[\hat{\pi} - t]'\Omega[\hat{\pi} - t].$$

In the rest of the proof we use the following definition and notation. For m dimensional vectors $a = (a_1, ..., a_m)'$ and $b = (b_1, ..., b_m)'$, we use the notation $\min(a, b)$ to signify the vector $(\min(a_1, b_1), ..., \min(a_m, b_m))'$. Suppose B is $m \times I$ dimensional and $\text{rank}(B) = \ell$. Define an $\ell \times p$ matrix K such that KB is a matrix whose rows consist of a basis of the row space $\text{row}(B)$. Also let M be an $(I - \ell) \times I$ matrix whose rows form an orthonormal basis of $\ker B = \ker(KB)$, and define $P = \binom{KB}{M}$. Finally, let $\hat{g} = B\hat{\pi}$ and $\hat{h} = M\hat{\pi}$. Then

$$\begin{aligned}
J_N &= \min_{Bt \le 0} N\left[\binom{KB}{M}(\hat{\pi} - t)\right]' P^{-1'}\Omega P^{-1}\left[\binom{KB}{M}(\hat{\pi} - t)\right] \\
&= \min_{Bt \le 0} N\binom{K[\hat{g} - Bt]}{\hat{h} - Mt}' P^{-1'}\Omega P^{-1}\binom{K[\hat{g} - Bt]}{\hat{h} - Mt}.
\end{aligned}$$

Let

$$\mathcal{U}_1 = \left\{\binom{K\gamma}{h} : \gamma = Bt, h = Mt, Bt \le 0, t \in \mathbf{R}^I\right\}$$

then writing $\alpha = KBt$ and $h = Mt$,

$$J_N = \min_{\binom{\alpha}{h} \in \mathcal{U}_1} N \begin{pmatrix} K\hat{g} - \alpha \\ \hat{h} - h \end{pmatrix}' P^{-1'} \Omega P^{-1} \begin{pmatrix} K\hat{g} - \alpha \\ \hat{h} - h \end{pmatrix}$$

Also define

$$\mathcal{U}_2 = \left\{ \begin{pmatrix} K\gamma \\ h \end{pmatrix} : \gamma \leq 0, \gamma \in \mathrm{col}(B), h \in \mathbf{R}^{I-\ell} \right\}$$

where $\mathrm{col}(B)$ denotes the column space of B. Obviously $\mathcal{U}_1 \subset \mathcal{U}_2$. Moreover, $\mathcal{U}_2 \subset \mathcal{U}_1$ holds. To see this, let $\binom{K\gamma^*}{h^*}$ be an arbitrary element of $\mathcal{U}_2$. We can always find $t^* \in \mathbf{R}^I$ such that $K\gamma^* = KBt^*$. Define

$$t^{**} := t^* + M'h^* - M'Mt^*$$

then $Bt^{**} = Bt^* = \gamma^* \leq 0$ and $Mt^{**} = Mt^* + MM'h^* - MM'Mt^* = h^*$, therefore $\binom{K\gamma^*}{h^*}$ is an element of $\mathcal{U}_1$ as well. Consequently,

$$\mathcal{U}_1 = \mathcal{U}_2.$$

We now have

$$\begin{aligned} J_N &= \min_{\binom{\alpha}{h} \in \mathcal{U}_2} N \begin{pmatrix} K\hat{g} - \alpha \\ \hat{h} - h \end{pmatrix}' P^{-1'} \Omega P^{-1} \begin{pmatrix} K\hat{g} - \alpha \\ \hat{h} - h \end{pmatrix} \\ &= N \min_{\binom{\alpha}{y} \in \mathcal{U}_2} \begin{pmatrix} K\hat{g} - \alpha \\ y \end{pmatrix}' P^{-1'} \Omega P^{-1} \begin{pmatrix} K\hat{g} - \alpha \\ y \end{pmatrix}. \end{aligned}$$

Define

$$T(x,y) = \begin{pmatrix} x \\ y \end{pmatrix}' P^{-1'} \Omega P^{-1} \begin{pmatrix} x \\ y \end{pmatrix}, \quad x \in \mathbf{R}^\ell, y \in \mathbf{R}^{I-\ell},$$

and

$$t(x) := \min_{y \in \mathbf{R}^{I-\ell}} T(x,y), \quad s(g) := \min_{\gamma \leq 0, \gamma \in \mathrm{col}(B)} t(K[g - \gamma]).$$

It is easy to see that $t : \mathbf{R}^\ell \to \mathbf{R}_+$ is a positive definite quadratic form. We can write

$$\begin{aligned} J_N &= N \min_{\gamma \leq 0, \gamma \in \mathrm{col}(B)} t(K[\hat{g} - \gamma]) \\ &= N s(\hat{g}) \\ &= s(\sqrt{N}\hat{g}). \end{aligned}$$

Let $\hat{\gamma}_{\tau_N}^k$, $\hat{g}^k$ and ϕ^k denote the k-th elements of $\hat{\gamma}_{\tau_N} = B\hat{\eta}_{\tau_N}$, $\hat{g}$ and ϕ. Since $\hat{\eta}_{\tau_N}$ solves

$$\begin{aligned} &\min_{B\eta \leq -\tau_N \phi} [\hat{\pi} - \eta]' \Omega [\hat{\pi} - \eta] \\ &= \min_{B\eta \leq -\tau_N \phi} T(K[\hat{g} - B\eta], \hat{h} - M\eta), \end{aligned}$$

it is sufficient to choose the I-dimensional identity matrix for Ω so that $\hat{\gamma}_{\tau_N}^k$ can be written as

$$\hat{\gamma}_{\tau_N}^k = \mathbb{I}\{\hat{g}^k \leq -\tau_N\phi\}(\hat{g}^k + d_{\tau_N}(\hat{g})) + \mathbb{I}\{\hat{g}^k > -\tau_N\phi\}(-\tau_N\phi^k)$$

for $1 \leq k \leq m$, where $d_{\tau_N}(\cdot) = [d_{\tau_N}^1(\cdot), ..., d_{\tau_N}^m(\cdot)]'$ is a vector valued function from $\mathbf{R}^m$ to $\mathrm{col}(B)$, and satisfies $\|d_{\tau_N}(g)\| \leq K\tau_N$ for some finite K for every g in $\mathbf{R}^m$. Using this and defining $\xi \sim \mathrm{N}(0, \hat{S})$ and $\zeta = B\xi$,

$$
\begin{aligned}
\tilde{J}_N(\tau_N) &\sim \min_{Bt \leq -\tau_N\phi} N\left[\begin{pmatrix} KB \\ M \end{pmatrix}(\hat{\eta}_{\tau_N} + N^{-1/2}\xi - t)\right]' P^{-1'}\Omega P^{-1}\left[\begin{pmatrix} KB \\ M \end{pmatrix}(\hat{\eta}_{\tau_N} + N^{-1/2}\xi - t)\right] \\
&= N\min_{\gamma \leq -\tau_N\phi, \gamma \in \mathrm{col}(B)} t\left(K\left[\hat{\gamma}_{\tau_N} + N^{-1/2}\zeta - \gamma\right]\right).
\end{aligned}
$$

In what follows, for an element-by-element vector inequality $a \leq b$, $a = (a_1, ..., a_m)'$, $b = (b_1, ..., b_m)'$, we use $\vec{\mathbb{I}}\{a \leq b\}$ to signify the m-vector of indicator functions whose k-th element is given by $\mathbb{I}\{a_k \leq b_k\}$. Defining $\gamma^\tau = \gamma + \tau_N\phi$ in the above,

$$
\begin{aligned}
\tilde{J}_N(\tau_N) &\sim N\min_{\gamma^\tau \leq 0} t\left(K\left[\hat{\gamma}_{\tau_N} + \tau_N\phi + N^{-1/2}\zeta - \gamma^\tau\right]\right) \\
&= N\min_{\gamma^\tau \leq 0} t\left(K\left[\vec{\mathbb{I}}\{\hat{g} \leq \tau_N\phi\} \odot (\hat{g} + d_{\tau_N}(\hat{g}) + \tau_N\phi) + N^{-1/2}\zeta - \gamma^\tau\right]\right) \\
&= s\left(\sqrt{N}\tau_N\vec{\mathbb{I}}\{\tau_N^{-1}\hat{g} \leq \phi\} \odot [\tau_N^{-1}\hat{g} + \tau_N^{-1}d_{\tau_N}(\hat{g}) + \phi] + \zeta\right)
\end{aligned}
$$

where $\odot$ denotes the Hadamard product. Define functions

$$\varphi_l(\xi) := \vec{\mathbb{I}}\{\xi \leq -\phi\} \odot [\xi - K\iota + \phi]$$

and

$$\varphi_u(\xi) := (+\infty) \cdot \mathbb{I}\{\xi \leq -\phi\}$$

where ι is the m-vector of ones, with the convention that $0 \cdot +\infty = 0$. These functions satisfy Assumptions GMS 1, 3 and 7 of AS10. Without loss of generality, let $\tau_1 = 1$ and suppose $N\tau_N$ increases monotonically in N. Let

$$\tilde{J}_{N,l} := s(-\varphi_l(-\tau_N^{-1}\hat{g}) + \zeta)$$

and

$$\tilde{J}_{N,u} := s(-\varphi_u(-\tau_N^{-1}\hat{g}) + \zeta),$$

then by construction $\tilde{J}_{N,l} \leq \tilde{J}_N(\tau_N) \leq \tilde{J}_{N,u}$. Moreover, $\hat{c}_{l,1-\alpha}$ and $\hat{c}_{u,1-\alpha}$ be the $1 - \alpha$ quantiles of $\tilde{J}_{l,N}$ and $\tilde{J}_{u,N}$, respectively. Then we have

(11.2) $$\hat{c}_{l,1-\alpha} \leq \hat{c}_{1-\alpha} \leq \hat{c}_{u,1-\alpha}.$$

Now we invoke Theorem 1 of AS10. As noted before, the function t is a positive definite quadratic form on $\mathbf{R}^\ell$, and so is its restriction on $\mathrm{col}(B)$. Then Assumptions 1-3 of AS10 hold for the function s defined above if signs are adjusted appropriately as our formulae deal with negativity constraints whereas AS10 is formulated for positivity constraints (note that Assumption 1(b) does not apply here since we use a fixed weighting matrix). AS10 imposes a set of restrictions on the parameter space (see their Equation (2.2) on page 124). Their condition (2.2) (vii) is a Lyapunov condition for a triangular array CLT. Following AS10, consider a sequence of distributions $\pi_N = [\pi'_{1N}, ..., \pi'_{JN}]', N = 1, 2, ...$ in $\mathcal{P} \cap C$ such that (1) $\sqrt{N}B\pi_N \to h$ for a non-positive h as $N \to \infty$ and (2) $\mathrm{Cov}_{\pi_N}(\sqrt{N}B\hat{\pi}) \to \Sigma$ as $N \to \infty$ where Σ is positive semidefinite. As $\pi_N \in \mathcal{P}$, Condition 5.1 demands that $\mathrm{var}_{\pi_{jN}}(b_k(j)'d_{j,n}) \geq \epsilon$ holds for some j for each N. Therefore we have $\lim_{N\to\infty} \mathrm{var}_{\pi_{jN}}(b_k(j)'d_{j,n}) \geq \epsilon$ for some j. For such j, the Lindeberg condition for CLT holds for $\sqrt{N(j)}[b_1(j)'[\hat{\pi}_j - \pi_{jN(j)}], ..., b_m(j)'[\hat{\pi}_j - \pi_{jN(j)}]]'$ where $\hat{\pi}_j = [\hat{\pi}_{1|j}, ..., \hat{\pi}_{I_j|j}]'$, and we have $\sqrt{N}B(\hat{\pi} - \pi_N) \overset{\pi_N}{\leadsto} N(0, \Sigma)$. The other conditions (2.2)(i)-(vi) hold trivially. Finally, Assumptions GMS 2 and GMS 4 of AS10 are concerned with their thresholding parameter κ for the k-th moment inequality, and by letting $\kappa_N = N^{1/2}\tau_N\phi_k$, the former holds by the condition $\sqrt{N}\tau_N \uparrow \infty$ and the latter by $\tau_N \downarrow 0$. Therefore we conclude

$$(11.3) \qquad \liminf_{N\to\infty} \inf_{\pi\in\mathcal{P}\cap C} \Pr\{J_N \leq \hat{c}_{1,1-\alpha}\} = \liminf_{N\to\infty} \inf_{\pi\in\mathcal{P}\cap C} \Pr\{J_N \leq \hat{c}_{+\infty,1-\alpha}\} = 1 - \alpha$$

Combining (11.2) and (11.3) the desired conclusion follows. $\qquad\square$

Before stating the proof of Corollary 7.1, let us introduce some notation.

Notation. Let $B^{(j)} := [b_1(j), ..., b_k(j)]' \in \mathbf{R}^{k\times I_j}$. For $F \in \mathcal{F}$ and $1 \leq j \leq J$, define

$$p_F^{(j)}(x) := E_F[d_{j,n}|x_n(j) = x], \quad \pi_F^{(j)} = p_F^{(j)}(\mathbf{x}(j)), \qquad \pi_F = [\pi_F^{(1)'}, ..., \pi_F^{(J)'}]'$$

and

$$\Sigma_F^{(j)}(x) := \mathrm{Cov}_F[d_{j,n}|x_n(j) = x].$$

Also let $f_F^{(j)}$ denote the density of $x_n(j)$ under F.

Note that $\Sigma_F^{(j)}(x) = \mathrm{diag}\left(p_F^{(j)}(x)\right) - p_F^{(j)}(x)p_F^{(j)}(x)'$. The following is used in the proof of Corollary 7.1:

Proposition 11.1. *If Conditions 7.1(iv) and (v) hold for some $K < \infty$ and ρ such that $\rho(t) \to 0$ as $t \to 0$, then $\{p_{F\in\mathcal{F}}^{(j)}\}$, $\{\dot{p}_{F\in\mathcal{F}}^{(j)}\}$, $\{\ddot{p}_{F\in\mathcal{F}}^{(j)}\}$, $\{f_{F\in\mathcal{F}}^{(j)}\}$, $\{\dot{f}_{F\in\mathcal{F}}^{(j)}\}$, and $\{\ddot{f}_{F\in\mathcal{F}}^{(j)}\}$ are equicontinuous at $\mathbf{x}(j)$ for all $1 \leq j \leq J$.*

The proof is straightforward and thus omitted.

Proof of Corollary 7.1. The proof follows the same steps as those in the proof of Theorem 5.1, except for the treatment of $\hat{\pi}$. Therefore, instead of the sequence $\pi_N, N = 1, 2, \dots$ in $\mathcal{P} \cap C$, consider a sequence of distributions $F_N = [F_{1N}, \dots, F_{JN}], N = 1, 2, \dots$ in $\mathcal{F}$ such that (1) $\sqrt{Nh_N} B \pi_{F_N} \to h$ for a non-positive h as $N \to \infty$; (2) $\Sigma_{F_N}^{(j)}(\mathbf{x}(j)) \to \Sigma^{(j)}(\mathbf{x}(j))$ as $N \to \infty, 1 \le j \le J$ where $\Sigma^{(j)}$ is positive semidefinite; and (3) $f_{F_N}^{(j)}(\mathbf{x}(j)) \to f^{(j)}(\mathbf{x}(j))$ as $N \to \infty, 1 \le j \le J$.

Define

$$v_{N,n}(j) = \frac{1}{\sqrt{h_N}} K\left(\frac{x_n(j) - \mathbf{x}(j)}{h_N}\right) B^{(j)}[d_{j,n} - p_{F_N}^{(j)}(x_n(j))],$$

then

(11.4) $$E_{F_N}[v_{N,n}(j)] = 0, \qquad 1 \le j \le J.$$

By the standard change of variable argument

$$\mathrm{Cov}_{F_N}(v_{N,n}(j)) = \int B^{(j)} \Sigma_{F_N}^{(j)}(\mathbf{x}(j) + h_N z) B^{(j)'} f_{F_N}^{(j)}(\mathbf{x}(j) + h_N z) K^2(z) dz$$

for $1 \le j \le J$. Let $\Delta := \mathrm{Cov}_{F_N}(v_{N,n}(j)) - B^{(j)} \Sigma^{(j)}(\mathbf{x}(j)) B^{(j)'(j)}(\mathbf{x}(j)) \int K^2(z) dz$, then

$$\begin{aligned}
\Delta &= \int B^{(j)} \left[\Sigma_{F_N}^{(j)}(\mathbf{x}(j) + h_N z) - \Sigma^{(j)}(\mathbf{x}(j))\right] B^{(j)'} f_{F_N}^{(j)}(\mathbf{x}(j) + h_N z) K^2(z) dz \\
&\quad + B^{(j)} \Sigma_{F_N}^{(j)}(\mathbf{x}(j)) B^{(j)'} \int \left[f_{F_N}^{(j)}(\mathbf{x}(j) + h_N z) - f^{(j)}(\mathbf{x}(j))\right] K^2(z) dz,
\end{aligned}$$

but as $N \to \infty$

$$\begin{aligned}
&\left\|\Sigma_{F_N}^{(j)}(\mathbf{x}(j) + h_N z) - \Sigma^{(j)}(\mathbf{x}(j))\right\| \\
&\le \left\|\Sigma_{F_N}^{(j)}(\mathbf{x}(j) + h_N z) - \Sigma_{F_N}^{(j)}(\mathbf{x}(j))\right\| + \left\|\Sigma_{F_N}^{(j)}(\mathbf{x}(j)) - \Sigma^{(j)}(\mathbf{x}(j))\right\| \\
&\le \sup_{F \in \mathcal{F}} \left\|\Sigma_{F}^{(j)}(\mathbf{x}(j) + h_N z) - \Sigma_{F}^{(j)}(\mathbf{x}(j))\right\| + \left\|\Sigma_{F_N}^{(j)}(\mathbf{x}(j)) - \Sigma^{(j)}(\mathbf{x}(j))\right\| \\
&\to 0 \quad \text{for every } z
\end{aligned}$$

where the convergence holds by Proposition 11.1. Similarly, as $N \to \infty$

$$\left|f_{F_N}^{(j)}(\mathbf{x}(j) + h_N z) - f^{(j)}(\mathbf{x}(j))\right| \to 0 \quad \text{for every } z.$$

By the Bounded Convergence Theorem conclude that

(11.5) $$\Delta \to 0.$$

This means that a version of equation (2.2.10) in Bierens (1987) still holds even under the sequence of distributions $\{F_N\}_{N=1}^{\infty}$ as specified above.

We next verify the Lyapunov condition. As $F_N \in \mathcal{F}$, Condition 7.1(ii) demands that

$$\text{var}_{F_{jN}}(b_k(j)'d_{j,n}|x_n = \mathbf{x}(j)) \geq \epsilon$$

holds for some j for each N. Thus we have $\lim_{N \to \infty} \text{var}_{F_{jN}}(b_k(j)'d_{j,n}|x_n = \mathbf{x}(j)) \geq \epsilon$ for some j. Define $s_N^2 := \text{Cov}_{F_N}\left(\sum_{n=1}^N v_{N,n}(i|j)\right)$ for such j where $v_{N,n}(i|j)$ denotes the i-th element of the I_j random vector $v_{N,n}(j)$. Note that $s_N \asymp N_j$, and therefore proceeding as in the derivation of (11.5), then using Equation (2.2.10) in Bierens (1987) we obtain

$$(11.6) \qquad \frac{1}{s_{N_j}^{2+\delta}} \sum_{n=1}^{N_j} E_{F_N}[|v_{N,n}(i|j)|^{2+\delta}] \to 0$$

for $1 \leq i \leq I_j$. By (11.4), (11.5) and (11.6), we have

$$\frac{1}{\sqrt{N_j}} \sum_{n=1}^{N_j} v_{N_j,n}(j) \overset{F_N}{\leadsto} N(0, \Omega_1), \quad \Omega_1 = B^{(j)}\Sigma^{(j)}(\mathbf{x}(j))B^{(j)'}f^{(j)}(\mathbf{x}(j))\int K^2(z)dz.$$

Finally, the convergence results corresponding (2.2.4) and (2.2.5) in Bierens (1987) hold under the sequence of distributions $\{F_N\}_{N=1}^{\infty}$ as specified above, once again by proceeding as in the derivation of (11.5). It follows that

$$\sqrt{N_j h_N}B^{(j)}[\hat{\pi}_j - \pi_{F_N}] \overset{F_N}{\leadsto} N(0, \Omega_2), \quad \Omega_2 = \frac{1}{f^{(j)}(\mathbf{x}(j))}B^{(j)}\Sigma^{(j)}(\mathbf{x}(j))B^{(j)'}\int K^2(z)dz.$$

The rest is the same as the proof of Theorem 5.1. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

12. Appendix B: Algorithms for Computing A

This appendix algorithms for computation of A. The first algorithm is a brute-force approach that generates all possible choice patterns and then verifies which of these are rationalizable. The second one avoids the construction of the vast majority of possible choice patterns because it checks for rationality along the way as choice patterns are constructed. The third algorithm uses proposition 1. To give a sense of the algorithms' performance, the matrix A corresponding to the 1975-1982 data, which is of size $[50 \times 42625]$ and cannot be computed with our implementation of algorithm 1, computes in about 2 hours with our implementation of algorithm 2, and (after suitable rearrangement of budgets) in about 2 minutes with our implementation of algorithm 3. All implementations are in MATLAB and are available from the authors. The instruction to FW-test a sequence refers to use of the Floyd-Warshall algorithm to detect choice cycles. We use the FastFloyd implementation due to...

Algorithm 1: Brute Force

This algorithm is easiest described verbally. First generate a matrix $A^{\max}$ that contains all logically possible choice patterns. To do so, let E^i denote the set of unit vectors in R^i and observe that a stacked vector $(a'_1, ..., a'_J)'$ is a column of $A^{\max}$ iff $(a_1, ..., a_J) \in E^{I_1} \times ... \times E^{I_J}$. It is then easy to construct $A^{\max}$ by looping. Next, generate A by FM-testing every column of $A^{\max}$ and retaining only columns that pass.

Algorithm 2: Decision Tree Crawling

An intuition for this algorithm is as follows. All possible choice patterns can be arranged on one decision tree, where the first node refers to choice from B_1 and so forth. The tree is systematically crawled. Exploration of any branch is stopped as soon as a choice cycle is detected. Completion of a rationalizable choice pattern is detected when a terminal node has been reached.

Pseudo-code for this algorithm follows.

1. Initialize $m_1 = ... = m_J = 1$.

2. Initialize $l = 2$.

3. Set $c(B_1) = m_1, ..., c(B_l) = m_l$. FW-test $(c(B_1), ..., c(B_l))$.

4. If no cycle is detected, move to step 5. Else:

4a. If $m_l < I_l$, set $m_l = m_l + 1$ and return to step 3.

4b. If $m_l = I_l$ and $m_{l-1} < I_{l-1}$, set $m_l = 1$, $m_{l-1} = m_{l-1} + 1$, $l = l - 1$, and return to step 3.

4c. If $m_l = I_l$, $m_{l-1} = I_{l-1}$, and $m_{l-2} < I_{l-2}$, set $m_l = m_{l-1} = 1$, $m_{l-2} = m_{l-2} + 1$, $l = l - 2$, and return to step 3.

(...)

4z. Terminate.

5. If $l < J$, set $l = l + 1$, $m_l = 1$, and return to step 3.

6. Extend A by the column $[m_1, ..., m_J]'$. Also:

6a. If $m_J < I_J$, set $m_J = m_J + 1$ and return to step 3.

6b. If $m_J = I_J$ and $m_{J-1} < I_{J-1}$, set $m_J = 1$, $m_{J-1} = m_{J-1} + 1$, $l = J - 1$, and return to step 3.

6c. If $m_l = I_l$, $m_{l-1} = I_{l-1}$, and $m_{l-2} < I_{l-2}$, set $m_l = m_{l-1} = 1$, $m_{l-2} = m_{l-2} + 1$, $l = l - 2$, and return to step 3.

(...)

```
6z.  Terminate.
```

Algorithm 3: Refinement using Proposition 4.1

Let budgets be arranged s.t. $(B_1, ..., B_M)$ do not intersect B_J; for exposition of the algorithm, assume B_J is above these budgets. Then pseudo-code for an algorithm that exploits proposition 1 (calling either of the preceding algorithms for intermediate steps) is as follows.

```
    1.  Use brute force or crawling to compute a matrix $A_{M+1 \to J-1}$ corresponding to
budgets $(B_{M+1}, ..., B_J)$, though using the full $X$ corresponding to budgets $(B_1, ..., B_J)$.[11]
    2.  For each column $a_{M+1 \to J-1}$ of $A_{M+1 \to J-1}$, go through the following steps:
    2.1 Compute (by brute force or crawling) all vectors $a_{1 \to M}$ s.t.
$(a_{1 \to M}, a_{M+1, J-1})$ is rationalizable.
    2.2 Compute (by brute force or crawling) all vectors
$a_J$ s.t.  $(a_{M+1, J-1}, a_J)$ is rationalizable.
    2.3 All stacked vectors $(a'_{1 \to M}, a'_{M+1, J-1}, a'_J)'$ are valid columns of $A$.
```

REFERENCES

Afriat, S.N. (1967): "The Construction of Utility Functions from Expenditure Data," International Economic Review, 8, 67-77

Andrews, D.W.K., and P. Guggenberger (2009): "Validity Of Subsampling And "Plug-In Asymptotic" Inference For Parameters Defined By Moment Inequalities," Econometric Theory, 25, 669–709.

Andrews, D.W.K., and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," Econometrica, 78, 119–157.

Bar, H., and F. Molinari (2012): "Computation and Inference in Partially Identified Models via Data Augmentation and SVM," Manuscript.

Berry, S., A. Gandhi and P. Haile (2011): "Connected Substitutes and Invertibility of Demand," Manuscript.

Bierens, H.J. (1987) "Kernel Estimators for Regression Functions,"in Advances in Econometrics: Volume 1: Fifth World Congress, ed. by T. F. Bewley. Cambridge University Press, pp. 99-144.

---

[11](This matrix has more rows than an A matrix that is only intended to apply to choice problems $(B_{M+1}, ..., B_J)$.)

Block, H., and J. Marschak (1960): "Random Orderings and Stochastic Theories of Responses," in Contributions to Probability and Statistics, ed. by I. Olkin, S. Ghurye, H. Hoeffding, H. Madow, and H. Mann. California: Stanford University Press, pp. 97-132.

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," Manuscript.

Blundell, R., M. Browning, and I. Crawford (2003): "Nonparametric Engel Curves and Revealed Preference," Econometrica, 71, 205–240.

— (2008): "Best Nonparametric Bounds on Demand Responses," Econometrica, 76, 1227–1262.

Blundell, R., D. Kristensen, and R.L. Matzkin (2011): "Bounding quantile demand functions using revealed preference inequalities," Manuscript.

Bugni, F. (2010): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," Econometrica, 78, 735–753.

Canay, I.A. (2010): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," Journal of Econometrics, 156, 408-425.

Cherchye, L., Crawford, I., De Rock, B., and F. Vermeulen (2009): "The Revealed Preference Approach to Demand," Quantifying Consumer Preferences, ed. by D. Slottje. Emerald Books.

Chernoff, H. (1954): "On the Distribution of the Likelihood Ratio," Annals of Mathematical Statistics, 25, 573–578.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Model," Econometrica, 75, 1243–1284.

Dridi, T. (1980): "Sur les Distributions Binaires Associées à des Distributions Ordinales," Mathématiques et Sciences Humaines, 69, 15-31.

Epstein, L., and A. Yatchew (1985): ""Non-parametric hypothesis testing procedures and applications to demand analysis," Journal of Econometrics, 30, 149–169.

Falmagne, J.-C. (1978): "A Representation Theorem for Finite Random Scale Systems," Journal of Mathematical Psychology, 18, 52–72.

Fishburn, P.C. (1992): "Induced Binary Probabilities and the Linear Ordering Polytope: A Status Report," Mathematical Social Sciences, 23, 67-80.

Gourieroux, C., A. Holly, and A. Monfort (1982): "Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters," Econometrica, 50, 63–80.

Grant, M., and S. Boyd (2008): "Graph implementations for nonsmooth convex programs, Recent Advances in Learning and Control (a tribute to M. Vidyasagar)," in Lecture Notes in Control and Information Sciences,ed. by V. Blondel, S. Boyd, and H. Kimura. Springer, pp. 95–110, http://stanford.edu/~boyd/graph dcp.html.

— (2011): "CVX: Matlab software for disciplined convex programming, version 1.21," http://cvxr.com/cvx.

Guggenberger, P., J. Hahn, and K. Kim (2008): "Specification Testing under Moment Inequalities," Economics Letters, 99, 375–378.

Hansen, L.P. (1982): "Large Sample Properties of Generalized Methods of Moments Estimators," Econometrica, 50, 1029–1054.

Hoderlein, S. (2011): "How Many Consumers are Rational?" Journal of Econometrics 164, 294–309.

Hoderlein, S., and J. Stoye (2010): "Revealed Preferences in a Heterogeneous Population," Manuscript.

Houthakker, H.S. (1950): "Revealed Preference and the Utility Function," Economica, 17, 159-174.

Imbens, G.W., and C.F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," Econometrica, 72, 1845–1857.

Lewbel, A. (2001): "Demand Systems with and without Errors," American Economic Review, 91, 611-618.

Manski, C.F. (2007): "Partial Identification of Counterfactual Probabilities," International Economic Review, 48, 1393–1410.

— (2012): "Identification of Preferences and Evaluation of Income Tax Policy," Manuscript.

Marschak, J. (1960): "Binary Choice Constraints on Random Utility Indicators," in Stanford Symposium on Mathematical Methods in the Social Sciences, ed. by K. Arrow. Stanford University Press.

Matzkin, R.L. (2006): "Heterogeneous Choice," in Advances in Econometrics: Proceedings of the 9th World Congress, ed. by R. Blundell, W. Newey, and T. Persson. Cambridge University Press.

McFadden, D.L., and K. Richter (1991): "Stochastic Rationality and Revealed Stochastic Preference," in Preferences, Uncertainty and Rationality, ed. by J. Chipman, D. McFadden, and M.K. Richter. Boulder: Westview Press, pp. 161–186.

McFadden, D.L. (2005): "Revealed Stochastic Preference: A Synthesis," Economic Theory, 26, 245–264.

Perlman, M.D. (1969): "One-sided Testing Problems in Multivariate Analysis," Annals of Mathematical Statistics, 40, 549–567.

Richter, M.K. (1966): "Revealed Preference Theory," Econometrica, 34, 635-645.

Romano, J., and A. Shaikh (2010): "Inference for the Identified Set in Partially Identified Econometric Models," Econometrica, 78, 169–211.

Rose, H. (1958): "Consistency of Preferences: The Two-Commodity Case," Review of Economic Studies, 25, 124-125.

Rosen, A.M. (2008): "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities," Journal of Econometrics, 146, 107–117.

Samuelson, P. (1938): "A Note on the Pure Theory of Consumer's Behaviour," Economica, 5, 61–71.

Shapiro, A. (1988): "Towards a Unified Theory of Inequality Constrained Testing in Multivariate Analysis," International Statistical Review, 56, 49–62.

Sher, I., J.T. Fox, K. il Kim, and P. Bajari (2011): "Partial Identification of Heterogeneity in Preference Orderings Over Discrete Choices," NBER working paper 17346.

Stoye, J. (2009): "More on Confidence Intervals for Partially Identified Parameters," Econometrica, 77, 1299–1315.

— (2010): "Partial Identification of Spread Parameters," Quantitative Economics, 1, 323-357, 2010.

Stoye, J., and S. Hoderlein (2013): "Testing Stochastic Rationality and Predicting Stochastic Demand: The Case of Two Goods," Manuscript.

Takemura, A., and S. Kuriki (1997): "Weights of $\bar{\chi}^2$-distribution for smooth or piecewise smooth cone alternatives," Annals of Statistics, 25, 2368–2387.

Varian, H. (1982): "The Nonparametric Approach to Demand Analysis," Econometrica, 50, 945–972.

Wolak, F.A. (1991): "The Local Nature of Hypothesis Tests Involving Inequality Constraints in Nonlinear Models," Econometrica, 59, 981–995.

Ziegler, G. (1995): Lectures on Polytopes. Springer: New York.

Cowles Foundation for Research in Economics, Yale University, New Haven, CT-06520.

E-mail address: yuichi.kitamura@yale.edu

Department of Economics, Cornell University, Ithaca, NY-14853.

E-mail address: stoye@cornell.edu