cemmap

centre for microdata methods and practice

Nonparametric analysis of random utility models

Yuichi Kitamura
Jörg Stoye

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP27/16

An ESRC Research Centre

NONPARAMETRIC ANALYSIS OF RANDOM UTILITY MODELS

YUICHI KITAMURA* AND JÖRG STOYE**

Abstract. This paper develops and implements a nonparametric test of Random Utility Models. The motivating application is to test the null hypothesis that a sample of cross-sectional demand distributions was generated by a population of rational consumers. We test a necessary and sufficient condition for this that does not rely on any restriction on unobserved heterogeneity or the number of goods. We also propose and implement a control function approach to account for endogenous expenditure. An econometric result of independent interest is a test for linear inequality constraints when these are represented as the vertices of a polyhedron rather than its faces. An empirical application to the U.K. Household Expenditure Survey illustrates computational feasibility of the method in demand problems with 5 goods.

1. Introduction

This paper develops new tools for the nonparametric analysis of Random Utility Models (RUM). We test the null hypothesis that a repeated cross-section of demand data might have been generated by a population of rational consumers, without restricting either unobserved heterogeneity or the number of goods. Equivalently, we empirically test McFadden and Richter's (1991) *Axiom of Revealed Stochastic Preference*.

Date: This Version: June 14, 2016.
Keywords: Stochastic Rationality.

JEL Classification Number: C14.

* ** Kitamura acknowledges financial support from the National Science Foundation via grants SES-0851759 and SES-1156266. Stoye acknowledges support from the National Science Foundation under grant SES-1260980 and through Cornell University's Economics Computer Cluster Organization, which was partially funded through NSF Grant SES-0922005. We thank Donald Andrews, Richard Blundell, Rahul Deb, Phil Haile, Chuck Manski, Daniel Martin, Rosa Matzkin, Francesca Molinari, participants at various seminars, and in particular, Elie Tamer and the four referees for helpful comments. We are grateful to Richard Blundell and Ian Crawford for their help with the FES data. We would like to thank Whitney Newey in particular for suggesting the use of a control function to account for endogeneity in total expenditure. All errors are our own. Ben Friedrich, Duk Gyoo Kim, Matthew Thirkettle, and Yang Zhang provided excellent research assistance.

Our contribution most directly connects to the recently burgeoning work on nonparametric demand estimation. Unobserved heterogeneity is a first-order concern in this literature and has received much renewed attention. It is generally believed to drive low goodness of fit in empirical demand estimation. Our test can help indicate whether this is true or whether more fundamental failures of economic assumptions are to blame.

Within this rich literature, there are numerous other, recent papers that nonparametrically test rationality, bound welfare, or bound demand responses in repeated cross-sections.[1] But all papers that we are aware of have at least one of the following features: (i) Unobserved heterogeneity is restricted; (ii) the approach is conceptually limited to an environment with two goods or practically limited to a very small choice universe; (iii) a necessary but not sufficient condition for rationalizability is tested. The present paper avoids all of these. It is meant to be the beginning of a research program: Estimation of demand distributions subject to rationalizability constraints, welfare analysis, and bounds on counterfactual random demand are natural next steps.

Testing at this level of generality is computationally challenging. We provide various algorithms that can be implemented with reasonable computational resources. Also, we establish uniform asymptotic validity of our test over a large range of parameter values in a setting related to moment inequalities but where existing methods (notably *Generalized Moment Selection*; see Andrews and Soares (2010), Bugni (2010), and Canay (2010)) do not apply. The method by which we ensure this is of independent interest. Finally, we leverage recent results on control functions (Imbens and Newey (2009); see also Blundell and Powell (2003)) to deal with endogeneity for unobserved heterogeneity of unrestricted dimension. All these tools are illustrated with one of the "work horse" data sets of the related literature. In that data, estimated demand distributions are not stochastically rationalizable, but the rejection is not statistically significant.

Our model can be briefly described as follows. Let

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

denote a utility function. Each consumer's choice problem is characterized by some u, some expenditure level W, and a price vector $p \in \mathbf{R}_+^K$. The consumer's demand is determined as

$$y \in \arg \max_{x \in \mathbf{R}_+^K : p'x \leq W} u(x), \tag{1.1}$$

[1] Blundell, Kristensen, and Matzkin (2014), Hausman and Newey (2016), Hoderlein and Stoye (2014), and Manski (2014) are just a few examples.

with arbitrary tie-breaking if the utility maximizing y is not unique. For simplicity, we restrict utility functions by monotonicity ("more is better"), but even this minimal restriction is not conceptually necessary. Also, imposing strict concavity of u would be easy.

We initially assume that W and p are nonrandom, while the utility function u is randomly drawn according to probability law P_u:

$$u \sim P_u.$$

Nonrandom W and p are the framework of McFadden and Richter (1991) and others but may not be realistic in applications. In the econometric analysis in Section 5 as well as in our empirical analysis in Section 6, we treat w as a random variable that may furthermore covary with u. For the moment, our assumptions allow us to normalize $w = 1$ and drop it from notation. The demand y in equation (3.2) is then indexed by the normalized price vector p. Denoting this by $y(p)$, we have a collection of distributions of demand

$$\Pr(y(p) \in Y), \ Y \subset \mathbf{R}^K_+ \tag{1.2}$$

indexed by $p \in \mathbf{R}^K_{++}$. Note that in (1.2) it is assumed that P_u is the same across $p \in \mathbf{R}^K_{++}$. Once W (hence p, after income normalization) is formulated as a random variable, this is essentially the same as imposing $W \perp\!\!\!\perp u$, an assumption we maintain in Section 5.1, then relax in Section 5.2 and our empirical application. Throughout the analysis, we do not at all restrict P_u. Thus, we allow for completely unrestricted, infinite dimensional unobserved heterogeneity across consumers.

We henceforth refer to (1.2) as a Random Utility Model (RUM).[2] A RUM is completely parameterized by P_u, but it only partially identifies P_u because many distinct distributions will be observationally equivalent in the sense of inducing the same distributions of demand.

Next, consider a finite list of budgets $(p_1, ..., p_J)$, and suppose observations of demand y from repeated cross-sections over J periods are available to the econometrician. In particular, for each $1 \le j \le J$, suppose N_j random draws of y distributed according to

$$P_j(Y) := \Pr(y(p_j) \in Y), \ Y \subset \mathbf{R}^K_+ \tag{1.3}$$

are observed by the econometrician. Define $N = \sum_{j=1}^J N_j$ for later use. Then $P_j(Y)$ can be estimated consistently as $N_j \uparrow \infty$ for each j, $1 \le j \le J$. The question is whether the estimated distributions

[2] Thus, we interpret randomness of u as arising from unobserved heterogeneity across individuals. Random utility models were originally developed in mathematical psychology, and in principle, our results also apply to stochastic choice behavior by an individual. However, in these settings it would frequently be natural to impose much more structure than we do.

may, up to sampling uncertainty, have arisen from a RUM. In the idealized setting in which there are truly J budgets, we show how to test this without any further assumptions.[3]

The main goals of our paper are as follows. We first show how to test stochastic rationalizability of a given data set and demonstrate this using the UK Household Expenditure Survey. The adequacy of the rationality assumption is undoubtedly a fundamental question. Even if we were to eventually proceed to counterfactual analysis and policy evaluation that often assumes rationality at the basic level, testing rationality without introducing ad hoc conditions would be a natural and important first step. Second, once a practical procedure for rationality testing is developed, it is straightforward to use it to obtain counterfactuals and carry out inference about them, once again under minimal assumptions. Some of this will be shown in Section 7.1. Third, this paper aims to offer a new statistical test with broad applicability. Though our procedure is motivated by the standard revealed preference axioms and that is also the application we report, the method has been used for other models as well. For example, our test has been applied to a nonparametric game theoretic model with strategic complementarity by Lazzati, Quah, and Shirai (2015) and to a novel, nonparametric model of "price preference" by Deb, Kitamura, Quah, and Stoye (2016). These are economically very different from standard revealed preference, yet our methods have been applied to them successfully. Uses of our algorithms for choice extrapolation in Manski (2014) and Adams (2016) are closer to our original motivation, but the former demonstrates that a restriction to linear budgets is not necessary. Indeed, the only way that the economics of a model affects our inference procedure is through the matrix A defined later. The rest of the algorithm remains the same if model elements such the standard revealed preference axiom or linear budget sets are replaced by other specifications.

The remainder of this paper is organized as follows. Section 2 discusses the related literature. Section 3 develops a simple geometric characterization of the empirical content of a RUM. This is a "population level" (all identifiable quantities are known) analysis that is related to classic work by McFadden and Richter (1991). Section 4 explains our test and its implementation under the assumption that one has an estimator of demand distributions and an approximation of the estimator's sampling distribution. Section 5 explains how to get the estimator, and a bootstrap approximation to its distribution, by both smoothing over income and adjusting for endogeneous expenditure. Section 6 contains our empirical application. Section 7 presents important additional applications and

[3]In the empirical application, there is a finite list of prices but continuous expenditure. We estimate demand at certain expenditure levels and therefore make some smoothness assumptions.

extensions. Section 8 concludes. All proofs are collected in appendix 9, and pseudocode for some algorithms is in appendix 10.

2. Related Literature

Our framework for testing Random Utility Models is built from scratch in the sense that it only presupposes classic results on nonstochastic revealed preference, notably the characterization of individual level rationalizability through SARP (Houthakker (1950)) and GARP (Afriat (1967)). (See also especially Samuelson (1938), Richter (1966), and Varian (1982).) At the population level, stochastic rationalizability was analyzed in classic work by McFadden and Richter (1991) updated by McFadden (2005). This work was an important inspiration for ours, and for the purpose of theoretical revealed preference analysis, the development in Section 3 clarifies and modestly extend theirs. Indeed, our test can be interpreted as statistical test of their Axiom of Revealed Stochastic Preference (ARSP). They did not consider statistical testing nor attempted to make the test operational (and could not have done so with computational constraints even of 2005).

An influential related research project is embodied in a sequence of papers by Blundell, Browning, and Crawford (2003, 2007, 2008; BBC henceforth). They assume the same observables as we do and apply their method to the same data. The core difference is that BBC analyze one individual level demand system generated by nonparametric estimation of Engel curves. This could be loosely characterized as revealed preference analysis of a representative consumer and in practice of average demand, where the 2003 paper focuses on testing rationality and bounding welfare and later papers focus on bounding counterfactual demand. Lewbel (2001) gives conditions on Random Utility Models that ensure integrability of average demand, so BBC can be thought of as adding those assumptions to ours. Also, the nonparametric estimation step in practice limits their approach to low dimensional commodity spaces, whereas we present an application to 5 goods.[4]

Manski (2007) analyzes stochastic choice from subsets of an abstract, finite choice universe. He states the testing and extrapolation problems in the abstract, solves them explicitly in simple examples, and outlines an approach to non-asymptotic inference. (He also considers models with more structure.) While we start from a continuous problem and build a (uniform) asymptotic theory, the settings become similar after Proposition 3.1 below. The core difference is that in the most general

[4] BBC's implementation exploits only the Weak Axiom of Revealed Preference (WARP) and therefore a necessary but not sufficient condition for rationalizability. This can be remedied, however. See Blundell, Browning, Cherchye, Crawford, De Rock, and Vermeulen (2015).

case, methods in Manski (2007) will not be practical for a choice universe containing more than a few elements. In a related paper, Manski (2014) uses our computational toolkit for choice extrapolation.

Our setting much simplifies if there are only two goods, an interesting but obviously very specific case. Blundell, Kristensen, and Matzkin (2014) bound counterfactual demand in this setting through bounding quantile demands.[5] They justify this through an invertibility assumption. Hoderlein and Stoye (2015) show that with two goods, this assumption has no observational implications.[6] Hence, Blundell, Kristensen, and Matzkin (2014) use the same assumptions as we do; however, the restriction to two goods is fundamental. Hausman and Newey (2016) nonparametrically bound welfare and income effects under assumptions resembling ours. Beyond a smoothness restriction (that we do not impose in theory, but we do use one in the empirical implementation), most of their results are specific to two goods. Our method conceptually applies to any number of goods and is practically applicable to at least five goods.

With more than two goods, pairwise testing of a stochastic analog of WARP amounts to testing a necessary but not sufficient condition for stochastic rationalizability. This is explored by Hoderlein and Stoye (2014) in a setting that is otherwise ours and also on the same data. Kawaguchi (2016) tests a logically intermediate condition, again on the same data. A different test of necessary conditions was 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 average demand under weak conditions. The resulting test is passed by the same data we use. Dette, Hoderlein, and Neumeyer (2016) propose a similar test using quantiles.

In sum, every paper cited in this section has one of the features (i)-(iii) mentioned in the introduction. We feel that removing aggregation or invertibility conditions is useful because these are usually assumptions of convenience. Testing necessary and sufficient conditions is obviously (at least in principle) sharper than testing necessary ones. And there are many empirical applications with more than two goods.

Section 4 of this paper is (implicitly) about testing multiple inequalities, the subject of a large literature in economics and statistics. See, in particular, Gourieroux, Holly, and Monfort (1982) and Wolak (1991) and also Chernoff (1954), Kudo (1963), Perlman (1969), Shapiro (1988), Takemura and Kuriki (1997), Andrews (1991), Bugni, Canay, and Shi (2015), and Guggenberger, Hahn, and

[5]The α-quantile demand induced by π is the nonstochastic demand system defined by the α-quantiles of π_j across j. It is well defined only if $K = 2$.

[6]A similar point is made, and exploited, by Hausman and Newey (2016).

Kim (2008). For the also related setting of inference on parameters defined by moment inequalities, see furthermore Andrews and Soares (2010), Bugni (2010), Canay (2010), Chernozhukov, Hong, and Tamer (2007), Imbens and Manski (2004), Romano and Shaikh (2010), Rosen (2008), and Stoye (2009). The major difference to these literatures is that moment inequalities, if linear (which most of the papers do not assume), define the faces of a polyhedron. The restrictions generated by our model are more akin to defining the polyhedron's vertices. One cannot in practice switch between these representations in high dimensions, so that we have to develop a new approach. In contrast, the inference theory in Hoderlein and Stoye (2014) only requires comparing two budgets. Because of this simple structure, they can express their hypotheses in terms of faces of polyhedra. While they use the structure of the model to improve on mechanical application of earlier results, no major conceptual advance is needed.

3. Analysis of Population Level Problem

In this section, we show how to verify rationalizability of a known set of cross-sectional demand distributions on J budgets. The main result is a relatively tractable geometric characterization of stochastic rationalizability.

Assume there is a finite sequence of J budget planes

$$\mathcal{B}_j = \{y \in \mathbf{R}_+^K : p_j' y = 1\}, j = 1, ..., J$$

and that the researcher observes the corresponding vector $(P_1, ..., P_J)$ of cross-sectional distributions of demand P_j as defined in 1.3. We will call $(P_1, ..., P_J)$ a stochastic demand system henceforth. Using a "more is better" assumption, we restrict choice to budget planes to simplify notation; this restriction is not essential to the method. We also do not restrict the number of goods K.

Definition 3.1. The stochastic demand system $(P_1, ..., P_J)$ is *stochastically rationalizable* if there exists a distribution P_u over utility functions u so that

$$P_j(Y) = \int 1\{\arg \max_{x \in \mathbf{R}_+^K : p_j' x = 1} u(x) \in Y\} \mathrm{d}P_u, \quad Y \subset \mathcal{B}_j, j = 1, ..., J. \tag{3.1}$$

Remark 3.1. Define

$$D(p, u) := \arg \max_{x \in \mathbf{R}_+^K : p' x = 1} u(x) \tag{3.2}$$

as the (nonstochastic) demand function induced by utility function u. Then 3.1 holds iff there exists a distribution P_u over utility functions that induces P_j as distribution of $D(p_j, u)$, $j = 1, ..., J$.

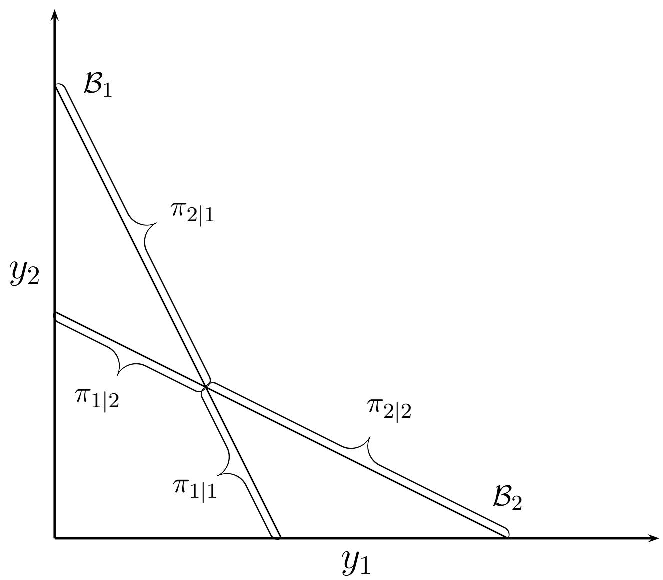

FIGURE 1. Visualization of Example 3.1.

This model is extremely general. While we nominally assume that $\arg\max_{x\in\mathcal{B}_j} u(x)$ is unique, this is really a normalization because a unique demand is observed for each data point. We already pointed out that "more is better" is not essential either. The model is also extremely rich - a parameterization would involve an essentially unrestricted distribution over utility functions. Verifying (3.1) therefore seems formidable. We will dramatically simplify this problem in several steps. Our running example will be the following:

Example 3.1. There are two budgets and they intersect, thus $J = 2$ and there exists $y \in \mathbf{R}^K_{++}$ with $p_1'y = p_2'y$.

The example is illustrated in Figure 1. Certain details of the Figure will be explained later. We will look at much more involved examples at the end of this section.

3.1. **Discretizing the Testing Problem.** In a first step, we replace the nonparametric RUM with a finite dimensional discrete choice model. This transformation involves no loss of information: The stochastic demand system is stochastically rationalizable iff some other demand system defined below for a model with finite universal choice set is. Therefore, while we dramatically simplify the model, we continue to test a necessary and sufficient condition for stochastic rationalizability in the original problem.

Any demand function $D(p,u)$ impacts $(P_1,...,P_J)$ only through $(D(p_1,u),...,D(p_J,u))$. Thus we have:

Remark 3.2. Call a vector $d=(d_1,...,d_J)\in\mathcal{B}_1\times...\times\mathcal{B}_J$ *rationalizable* if $d=(D(p_1,u),...,D(p_J,u))$ for some u. Let the set $\mathcal{D}\subset\mathcal{B}_1\times...\times\mathcal{B}_J$ collect all rationalizable such vectors. Then (3.1) holds iff there exists a distribution P_d on $\mathcal{D}$ such that

$$P_j(Y)=\int 1\{d_j\in Y\}\mathrm{d}P_d,\quad Y\subset\mathcal{B}_j, j=1,...,J. \tag{3.3}$$

Next, a demand vector $(d_1,...,d_J)$ is rationalizable iff it fulfills the Generalized Axiom of Revealed Preference (GARP). (Alternatively, use the strong axiom (SARP) to restrict attention to strictly convex u.) The only information needed to verify this is whether, for different values of j and k, one has $p_j'd_j>p_j'd_k$, $p_j'd_j=p_j'd_k$, or $p_j'd_j<p_j'd_k$. If two distinct demand vectors agree on all of these comparisons, then either both or neither are rationalizable, and no information is lost by treating them as equivalent. This will allow us to restrict attention to a finite subset of $\mathcal{D}$.

Formalizing this requires some definitions.

Definition 3.2. Let $\mathcal{X}:=\{x_1,...,x_I\}$ be the coarsest partition of $\cup_{j=1}^J\mathcal{B}_j$ such that (i) each $\mathcal{B}_j$ equals the union of some subset of the partition; (ii) for any $i\in\{1,...,I\}$ and $j\in\{1,...,J\}$, x_i is either completely on, completely strictly above, or completely strictly below $\mathcal{B}_j$. Elements of $\mathcal{X}$ will be called *patches*. Elements of of $\mathcal{X}$ that are part of more than one budget will also be called *intersection patches*. The number of patches that jointly comprise budget $\mathcal{B}_j$ will be called I_j. Note that $\sum_{j=1}^J I_j\geq I$, strictly so (because of multiple counting of intersection patches) if any two budget planes intersect.

Remark 3.3. $I_j\leq I\leq 3^J$, hence I_j and I are finite.

Definition 3.3. The *discretized equivalent* $\mathcal{Y}^*$ of of $\mathcal{Y}$ is an arbitrary, but henceforth fixed, vector $(y_1^*,...,y_I^*)\in x_1\times...\times x_I$.

Then we have:

Proposition 3.1. *Suppose that two stochastic demand systems $(P_1,...,P_J)$ and $(P_1^*,...,P_J^*)$ agree on the probabilities of all patches, i.e. $P_j(x_i)=P_j^*(x_i)$ for all $j=1,...,J$ and $x_i\in\mathcal{X}$. Then each demand system is rationalizable iff the other one is. In particular, this applies if $(P_1^*,...,P_J^*)$ is concentrated on $\mathcal{Y}^*$.*

Revisiting Example 3.1. The patches generated by Example 3.1 are illustrated in Figure 1. There are five of them. Four are marked $\pi_{1|1}$ and so on; the meaning of this will become clear. These can be described as "on budget 1 but below budget 2" and so on. The fifth patch is the intersection point. Each budget consists of three patches.

3.2. **Geometric Characterization.** Proposition 3.1 allows us to transform the problem into one with finite universal choice set. The next step is a geometric characterization of stochastic rationalizability. This requires a few more definitions.

Definition 3.4. The *vector representation* of $(\mathcal{B}_1, ..., \mathcal{B}_J)$ is a vector of length $\sum_{j=1}^{J} I_j$ whose first I_1 components are the patches comprising $\mathcal{B}_1$, the next I_2 components are the patches comprising $\mathcal{B}_2$, and so forth. The ordering of elements within budgets is arbitrary but henceforth fixed. Note that a patch in the intersection of budgets will appear repeatedly.

Definition 3.5. The *vector representation* of $(P_1, ..., P_J)$ is the vector π of length $\sum_{j=1}^{J} I_j$ whose first I_1 elements are the probabilities assigned by P_1 to patches within $\mathcal{B}_1$ and so forth. The order of elements corresponds to the vector representation of budgets.

In the following, fix $(P_1, ..., P_J)$ and let $(P_1^*, ..., P_J^*)$ be the equivalent (in the sense of Proposition 3.1) demand system that is concentrated on $\mathcal{Y}^*$. Note that $(P_1^*, ..., P_J^*)$ is completely determined by π: The first I_1 elements of π define the probability mass function corresponding to P_1^* and so forth.

Next, if $(P_1^*, ..., P_J^*)$ is rationalizable, then Remark 3.2 must apply, where P_d is furthermore concentrated on nonstochastic demand systems that are, in turn, concentrated on $\mathcal{Y}^*$. Any such nonstochastic demand system has a natural vector representation as well.

Definition 3.6. Consider a nonstochastic demand system d^* that only selects elements of $\mathcal{Y}^*$. The *vector representation* of d^* is the vector representation of the degenerate stochastic demand system that corresponds to d^*. Thus, it it a binary vector a of length $\sum_{j=1}^{J} I_j$ such that exactly one of the first I_1 entries equals 1, exactly one of the next I_2 entries equals 1, and so forth, and where the entries of 1 indicate which element of $\mathcal{Y}^*$ was chosen from the respective budget.

Definition 3.7. The *rational demand matrix* A is the (unique, up to ordering of columns) smallest matrix such that the vector representation a of each rationalizable nonstochastic demand system d^* concentrated on $\mathcal{Y}^*$ is a column of A. The number of columns of A is denoted H.

Remark 3.4. $H \leq \prod_{j=1}^{J} I_j$, hence H is finite.

Then Proposition 3.1 and Remark 3.2 imply the following.

Proposition 3.2. *The stochastic demand system $(P_1, ..., P_J)$ is rationalizable iff its vector representation can be written as*

$$\pi = A\nu \text{ for some } \nu \in \Delta^{H-1}, \tag{3.4}$$

where Δ^{H-1} is the unit simplex in $\mathbf{R}^H$. Furthermore, this representation obtains iff

$$\pi = A\nu \text{ for some } \nu \geq 0. \tag{3.5}$$

The vector ν can be interpreted as distribution over $(1, ..., H)$, hence Proposition 3.2 says that π must be representable as mixture over the columns of A. Intuitively, A is a list of (vector representations of) all rationalizable individual choice types who only choose from $\mathcal{Y}^*$, and ν represents a population distribution over such types that therefore rationalizes the $(P_1^*, ..., P_J^*)$ – and, by implication, any $(P_1, ..., P_J)$ – that corresponds to π. Note that ν is not in general unique. Finally, $A\nu = \pi$ implies $\mathbf{1}_H'\nu = 1$, therefore we do not need to explicitly impose the latter. This detail improves on a similar result in McFadden (2005).

In sum, we reduced the extremely complex (3.1) to the condition that some vector of probabilities must be in the column cone of some matrix A. For known A and π, this is easily checked even if these objects are very large. The main caveat is that the matrix A can be hard to compute.

We conclude this section with a number of remarks.

GARP vs SARP. We reiterate that rationalizability of nonstochastic demand systems can be defined, and our test can therefore be applied, using either GARP or SARP. SARP will define a somewhat smaller matrix A, but nothing else changes. Note that weak revealed preference occurs if demand on some budget $\mathcal{B}_i$ is also on some other budget plane $\mathcal{B}_j$. Therefore, GARP and SARP can differ only in the assessment of demand vectors that contain at least three intersection patches.

Simplification if demand is continuous. Intersection patches are of lower dimension than budget planes. Thus, if the distribution of demand is known to be continuous, their probability is known to be zero, and they can be eliminated from $\mathcal{Y}^*$. In large problems, this will considerably simplify A. Also, each remaining patch is a subset of exactly one budget plane, so that $\sum_{j=1}^J I_j = I$. We impose this simplification henceforth and in our empirical application, but none of our results and algorithms depend on it. Observe finally that SARP agrees with GARP in this case.

Revisiting Example 3.1. Assume that demand is distributed continuously, then the intersection patch can be dropped and $\mathcal{Y}^*$ has $I = 4$ elements. WARP, SARP, and GARP all agree on which choice patrtern to exclude. Index patches such that the excluded demand vector is $d = (1,0,1,0)$. Then

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

The column cone of A can be explicitly written as

$$C = \left\{ \begin{bmatrix} \nu_1 \\ \nu_2 + \nu_3 \\ \nu_2 \\ \nu_1 + \nu_3 \end{bmatrix} : \nu_1, \nu_2, \nu_3 \geq 0 \right\}.$$

The restriction of (P_1, P_2) to patches is a vector $\pi = (\pi_{1|1}, \pi_{2|1}, \pi_{1|2}, \pi_{2|2})$, where the order of probabilities corresponds to the order of patches and where Figure 1 visualizes the interpretation of these patch probabilities. We formalize this notation below. and its components are visualized in Figure 1. The only restriction on π beyond adding-up constraints is that $\pi_{1|1} + \pi_{1|2} \leq 1$ or equivalently $\pi_{2|2} \geq \pi_{1|1}$. This is well known to be the exact implication of a RUM for this example (Matzkin (2006); Hoderlein and Stoye (2015)).

Generality of our characterization. The above characterization applies immediately if the choice universe is finite to begin with or if a discretization trick similar to Proposition 3.1 is available. See Manski (2007) for an example of the former and Deb, Kitamura, Quah, and Stoye (2016) for an example of the latter. Linearity of budgets sets is not required either; see Manski (2014) for an application where budget sets are kinked. Also, in some applications the choice universe is discrete; e.g., a good has to be bought in integer quantities. This is not a problem conceptually. In cases that are covered by Polisson and Quah (2013), e.g. integer constraints superimposed on otherwise standard budgets, it does not even make an operational difference.

3.3. **Computing A.** We next elaborate how to compute A from a vector of prices $(p_1, ..., p_J)$. For ease of exposition, we assume that intersection patches can be dropped. We add remarks on generalization along the way. We split the problem into two subproblems, namely checking whether a "candidate" vector a is the representation of a rationalizable demand system (a is then called *rationalizable* below) and finding all such vectors.

Checking rationalizability of a vector a.

Consider any binary I-vector a with at most one entry of 1 on each subvector corresponding to one budget. This vector can be thought of as encoding choice behavior on all or some budgets. It is *complete* if it has exactly J entries of 1, i.e. it specifies a choice from each budget. It is called *incomplete* otherwise. It is called *rationalizable* if those choices that are specified jointly fulfil GARP. The rationalizable demand matrix A collects all complete rationalizable vectors a.

To check rationalizability of a given, complete or incomplete, vector a, we initially extract from it a directly revealed preference relation. For example, if y_i^* is chosen from budget $\mathcal{B}_j$, then it is revealed preferred to all y_k^* on or below $\mathcal{B}_j$. This information can be extracted extremely quickly.[7]

We next check whether the transitive completion of directly revealed preference is cyclical. This is done by checking cyclicality of the directed graph in which nodes correspond to elements of $\mathcal{Y}^*$ and in which a directed link indicates directly revealed preference. Operationally, this check can use either the Floyd-Warshall algorithm Floyd (1962), a Depth-first search of the graph, or a Breadth-first search. All methods compute quickly in our application. An important simplification is to notice that revealed preference cycles can only pass through nodes y_i^* that were in fact chosen. Hence, it suffices to look for cycles on the subgraph with the (at most) J corresponding nodes. This dramatically simplifies the problem as I increases rapidly with J. Indeed, increasing K does not increase the size of graphs checked in this step, though it tends to lead to more intricate patterns of overlap between budgets and, therefore, to richer revealed preference relations.

In this heuristic explanation, we interpreted all revealed preferences as strict. This is without loss of generality absent intersection patches because GARP and SARP then agree. If intersection patches are retained, the method just described tests SARP and not GARP. To test GARP, one would have to check rejected vectors a for the possibility that all revealed preference cycles are weak and accept them if this is the case.[8]

Collecting rationalizable vectors.

This step is a bottleneck. We very briefly mention two approaches that we do not recommend. First, one could generate one candidate vector a from each of the $\prod_{j=1}^{J} I_j$ conceivable choice patterns

[7] In a preliminary step of the implementation, we compute a $(I \times J)$-matrix X where, for example, the i-th row of X is $[0, -1, 1, 1, 1]$ if y_i^* is on budget $\mathcal{B}_1$, below budget $\mathcal{B}_2$, and above the remaining budgets. Preferences revealed by choice from budget $\mathcal{B}_j$ can be read off the j-th column of X.

[8] The matrix X from footnote 7 is designed to allow for this as its entries of 0 and 1 differentiate between weak and strict preference. The information is not exploited in our empirical example.

over patches and then check rationalizability of all of these columns. We implemented this approach for debugging purposes, but computational cost escalates rapidly. Similarly, we do not recommend to initially list all possible preference orderings over patches and then generate columns of A from them.

Our benchmark algorithm is based on representing all conceivable vectors a as leaves (terminal nodes) of a rooted tree that is recursively constructed as follows: The root has I_1 children corresponding to elements of $\mathcal{B}_1 \cap \mathcal{Y}^*$. Each of these children has I_2 children that correspond to patches in $\mathcal{B}_2 \cap \mathcal{Y}^*$, and so on for J generations. The leaves of the tree correspond to complete vectors a that specify to choose the leaf and all its ancestors from the respective budgets. Non-terminal nodes can be similarly identified with incomplete vectors.

The algorithm attempts a depth-first search of the tree. At each node, rationalizability of the corresponding vector is checked. If an inconsistency is detected, the node and its entire subtree are deleted. If the node is a leaf and no inconsistency is detected, then a new column of A has been discovered. The algorithm terminates when each node has been either visited or deleted. It discovers each column of A exactly once. Elimination of subtrees means that the vast majority of complete candidate vectors are never visited. Pseudocode for the algorithm is displayed in appendix B.

Finally, a modest amount of problem-specific adjustment can lead to further, dramatic improvement. The key to this is contained in the following proposition.

Proposition 3.3. *Suppose that for some $M \geq 1$, either all of $(\mathcal{B}_1, ..., \mathcal{B}_M)$ are contained in $\mathcal{B}_J$ or they all contain it. Suppose also that choices from $(\mathcal{B}_1, ..., \mathcal{B}_{J-1})$ are jointly rationalizable. Then choices from $(\mathcal{B}_1, ..., \mathcal{B}_J)$ are jointly rationalizable iff choices from $(\mathcal{B}_{M+1}, ..., \mathcal{B}_J)$ are.*

This proposition is helpful whenever not all budgets mutually intersect. In that case, all rationalizable demand systems can be discovered by finding rationalizable demand systems on smaller domains and combining the results. In particular, iterated application of Proposition 3.3 informs the following strategy, which is also provided as pseudocode in appendix B. First, construct a matrix $A_{M+1,J-1}$ only for $(\mathcal{B}_{M+1}, ..., \mathcal{B}_{J-1})$. Next, for each column of $A_{M+1,J-1}$, find all rationalizable completions to $(\mathcal{B}_1, ..., \mathcal{B}_{J-1})$ as well as all rationalizable completions to $(\mathcal{B}_{M+1}, ..., \mathcal{B}_J)$. Each combination of two such completions is rationalizable. No step in this algorithm checks rationalizability on J budgets at once; furthermore, a Cartesian product structure of the columns of A is exploited. In our application, the refinement improves computation time for some of the largest matrices by orders of magnitude, although the depth-first search proved so fast that our replication code omits this step.

3.4. **Additional Examples.** We conclude this section with some more involved examples. For this purpose, we introduce one new notation:

Notation. In this section and the next two, elements of $\mathcal{X}$ defined in Definition 3.2 will be indexed as $\mathcal{X} \cap \mathcal{B}_j = \{x_{1|j}, ..., x_{I_j|j}\}$ for $1 \leq j \leq J$. For example, patch $x_{2|4}$ refers to the second patch on Budget $\mathcal{B}_4$. The ordering of patches within each budget corresponds to the vector representation of budgets.

Example 3.2. The following is the simplest example in which WARP does not imply SARP, so that tests based on applying Example 3.1 to all pairs of budgets would not have full power. Let $K = J = 3$. Let prices be $(p_1, p_2, p_3) = ((1/2, 1/4, 1/4), (1/4, 1/2, 1/4), (1/4, 1/4, 1/2))$.[9] In this example, each budget has 4 patches for a total of $I = 12$ patches. This yields $4^3 = 64$ candidate vectors a, $H = 25$ of whom are rationalizable:

$$A = \begin{array}{c} x_{1|1} \\ x_{2|1} \\ x_{3|1} \\ x_{4|1} \\ x_{1|2} \\ x_{2|2} \\ x_{3|2} \\ x_{4|2} \\ x_{1|3} \\ x_{2|3} \\ x_{3|3} \\ x_{4|3} \end{array} \begin{bmatrix} 0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&1&0&0&0&0&0&0&0&1&0&0&1&0&0&1&0&0&1&0&0 \\ 0&1&0&0&0&0&1&0&0&0&0&1&0&0&0&0&0&1&0&0&0&0&0&1&0 \\ 1&0&1&1&0&0&0&1&1&1&1&0&1&0&1&0&0&0&1&0&1&0&0&0&1 \\ 0&0&0&0&0&0&0&0&1&0&0&0&0&1&1&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&1&1&1&1&0&0&0&0&0&0 \\ 1&0&0&1&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&1&1&0&0&0&0 \\ 0&1&1&0&1&1&1&1&0&0&0&1&1&0&0&0&0&0&0&0&0&1&1&1&1 \\ 1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&1&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&1&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1&1&1&1&1&1&1&1&1 \end{bmatrix}.$$

If one knows the mutual relation of patches, one can make sense of this matrix. For example, a consumer choosing $x_{1|1}$ but also $x_{1|2}$ would violate WARP as both patches lie below the respective other budget. Thus, no column of A equals $[1, 0, 0, 0, 1, ...]$. More subtly, the demand vectors mentioned in footnote 5 lie on patches $x_{3|1}$, $x_{2|2}$, and $x_{3|3}$. Thus, A does not contain the column $[0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0]'$.

[9]To see that WARP does not imply SARP in this specific example, consider demand vectors $(q_1, q_2, q_3) = ((1, 1/2, 3/2), (3/2, 1, 1/2), (1/2, 3/2, 1))$. Then $p_1 q_2 = p_2 q_3 = p_3 q_1 = 9/8$ and $p_1 q_3 = p_2 q_1 = p_3 q_2 = 7/8$, hence choices from pairs of budgets fulfil WARP but violate SARP.

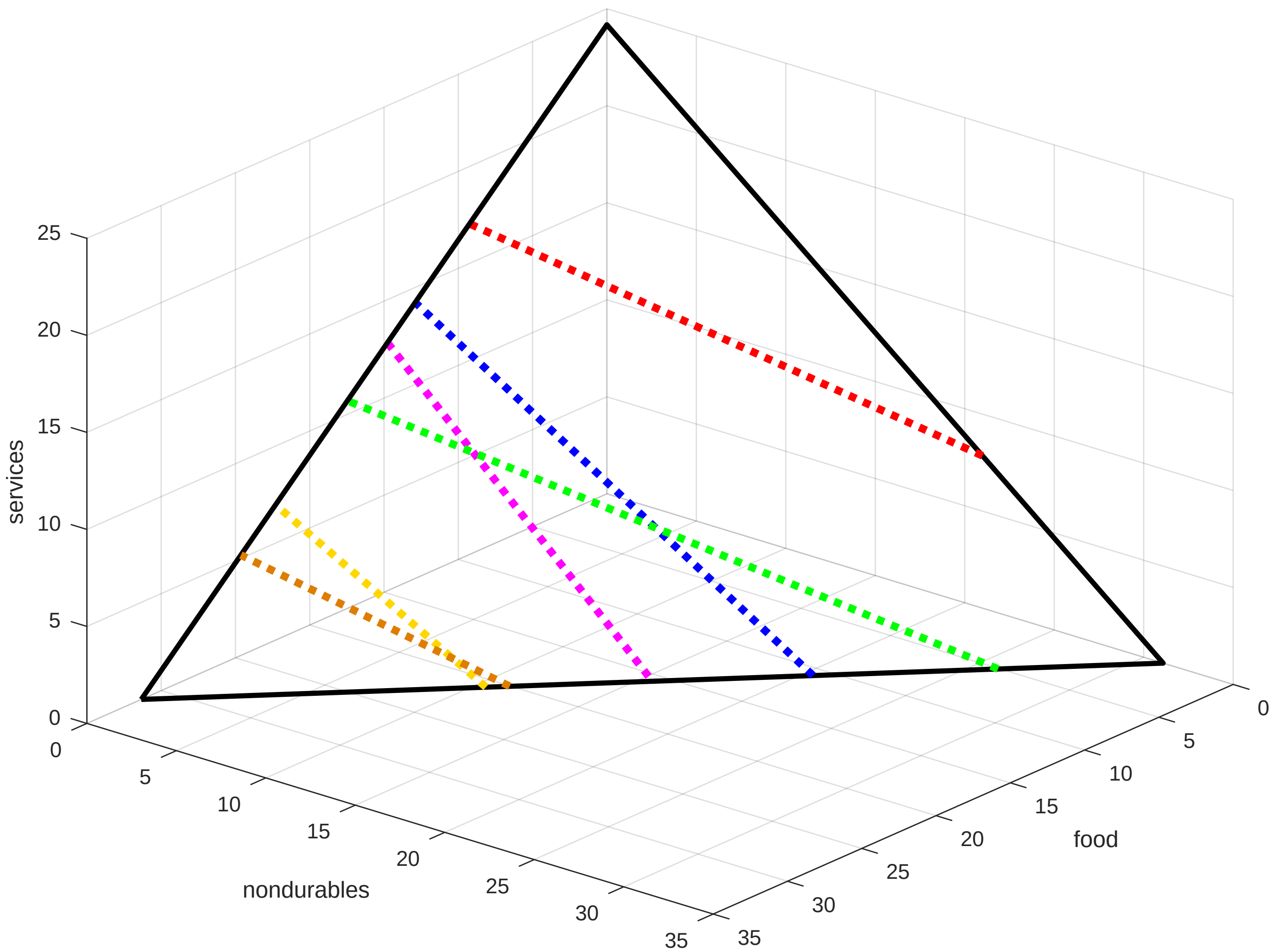

FIGURE 2. Visualization of one budget set in the empirical application.

Example 3.3. Our empirical application has sequences of $J = 7$ budgets in $\mathbf{R}^K$ for $K = 3, 4, 5$ and sequences of $J = 8$ budgets in $\mathbf{R}^3$. The largest A-matrices are of sizes 78×336467 and 79×313440. In exploratory work using longer sequences of budgets, we computed an A matrix with over 2 million columns in a few hours on Cornell's CISER server (12 workers, Intel Xeon CPU E7 4870).

Figure 2 visualizes one budget in $\mathbf{R}^3$ from our empirical application and its intersection with 6 other budgets. There are total of 35 patches, 25 of which are intersection patches, so that in our empirical application, this is treated as a budget with 10 patches.

4. Statistical Testing

This section lays out our statistical testing procedure in the idealized situation where, for finite J, repeated cross-sectional observations of demand over J periods are available to the econometrician, where each cross-section of size N_j is observed over the deterministically determined (and hence exogenous) budget plane for period j. Then the probabilities in π were estimated by corresponding sample frequencies. We define a test statistic and critical value and show that the resulting test is uniformly asymptotically valid over an interesting range of d.g.p.'s.

4.1. **Null Hypothesis and Test Statistic.** Recall from (3.5) that we wish to test:

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

This hypothesis is equivalent to

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

where Ω is a positive definite matrix (restricted to be diagonal in our inference procedure) and $\mathcal{C} := \{A\nu | \nu \geq 0\}$ is a convex cone in $\mathbf{R}^I$. The solution η_0 of ($\mathbf{H}_B$) is the projection of $\pi \in \mathbf{R}^I_+$ onto $\mathcal{C}$ under the weighted norm $\|x\|_\Omega = \sqrt{x'\Omega x}$. The corresponding value of the objective function is the squared 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 ($\mathbf{H}_B$) would be $\min_{\eta \in C}[\hat{\pi} - \eta]'\Omega[\hat{\pi} - \eta]$, where $\hat{\pi}$ estimates π, for example by sample choice frequencies. The usual normalization by sample size yields

$$
\begin{aligned}
J_N \quad : \quad &= N \min_{\eta \in C}[\hat{\pi} - \eta]'\Omega[\hat{\pi} - \eta] \\
&= N \min_{\nu \in \mathbf{R}^H_+}[\hat{\pi} - A\nu]'\Omega[\hat{\pi} - A\nu].
\end{aligned}
\tag{4.1}
$$

Once again, ν is not unique at the optimum, but $\eta = A\nu$ is. Call its optimal value $\hat{\eta}$. Then $\hat{\eta} = \hat{\pi}$, and $J_N = 0$, if the estimated choice probabilities $\hat{\pi}$ are stochastically rationalizable; obviously, our null hypothesis will be accepted in this case.

4.2. **Simulating a Critical Value.** We next explain how to get a valid critical value for J_N under the assumption that $\hat{\pi}$ estimates the probabilities of patches by corresponding sample frequencies and that one has R bootstrap replications $\hat{\pi}^{*(r)}, r = 1, ..., R$. Thus, $\hat{\pi}^{*(r)} - \hat{\pi}$ is a natural bootstrap analog of $\hat{\pi} - \pi$. We will make enough assumption to ensure that its distribution consistently estimates the distribution of $\hat{\pi} - \pi_0$, where π_0 is the true value of π. The main difficulty is that one cannot use $\hat{\pi}$ as bootstrap analog of π_0.

Our bootstrap procedure relies on a tuning parameter τ_N chosen s.t. $\tau_N \downarrow 0$ and $\sqrt{N}\tau_N \uparrow \infty$.[10] Also, let Ω be diagonal and positive definite and let $\mathbf{1}_H$ be a H-vector of ones.[11]. Then our procedure is as follows:

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

$$J_N = \min_{[\nu - \tau_N \mathbf{1}_H / H] \in \mathbf{R}_+^H} N[\hat{\pi} - A\nu]'\Omega[\hat{\pi} - A\nu]$$

(ii) Define the τ_N*-tightened recentered* bootstrap estimators

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

(iii) The bootstrap test statistic is

$$J_N^{*(r)}(\tau_N) = \min_{[\nu - \tau_N \mathbf{1}_H / H] \in \mathbf{R}_+^H} N[\hat{\pi}_{\tau_N}^{*(r)} - A\nu]'\Omega[\hat{\pi}_{\tau_N}^{*(r)} - A\nu],$$

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.

The object $\hat{\eta}_{\tau_N}$ is the true value of π in the bootstrap population, i.e. it is the bootstrap analog of π_0. It differs from $\hat{\pi}$ through a "double recentering." To disentangle the two recenterings, suppose first that $\tau_N = 0$. Then inspection of step (i) of the algorithm shows that $\hat{\pi}$ would be projected onto the cone $\mathcal{C}$. This is a relatively standard recentering "onto the null" that resembles recentering of the J-statistic in overidentified GMM. However, with $\tau_N > 0$, there is a second recentering because the cone $\mathcal{C}$ itself has been tightened. We next discuss why this recentering is needed.

[10]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 $\underline{N} = \min_j N_j$ and N_j is the number of observations on Budget $\mathcal{B}_j$: see (4.7). This choice corresponds to the "BIC choice" in Andrews and Soares (2010). We will later propose a different τ_N based on how π is in fact estimated.

[11]In principle, $\mathbf{1}_H$ could be any strictly positive H-vector, though a data based choice of such a vector is beyond the scope of the paper.

4.3. **Discussion.** Our testing problem is related to the large literature on inequality testing but adds an important twist. Writing $\{a_1, a_2, ..., a_H\}$ for the column vectors of A, one has

$$\mathcal{C} = \text{cone(A)} := \{\nu_1 a_1 + ... + \nu_H a_H : \nu_h \geq 0\},$$

i.e. the set $\mathcal{C}$ is a finitely generated cone. The following result, known as the Weyl-Minkowski Theorem, provides an alternative representation that is useful for theoretical developments of our statistical testing procedure.

Theorem 4.1. *(Weyl-Minkowski Theorem for Cones) A subset $\mathcal{C}$ of $\mathbf{R}^I$ is a finitely generated cone*

$$\mathcal{C} = \{\nu_1 a_1 + ... + \nu_H a_H : \nu_h \geq 0\} \text{ for some } A = [a_1, ..., a_H] \in \mathbf{R}^{I \times H} \tag{4.2}$$

if and only if it is a finite intersection of closed half spaces

$$\mathcal{C} = \{t \in \mathbf{R}^I | Bt \leq 0\} \text{ for some } B \in \mathbf{R}^{m \times I}. \tag{4.3}$$

The expressions in (4.2) *and* (4.3) *are called a $\mathcal{V}$-representation (as in "vertices") and a $\mathcal{H}$-representation (as in "half spaces") of $\mathcal{C}$, respectively.*

See, for example, Theorem 1.3 in Ziegler (1995).[12] The "only if" part of the theorem (which is Weyl's Theorem) shows that our rationality hypothesis $\pi \in \mathcal{C}, \mathcal{C} = \{A\nu | \nu \geq 0\}$ in terms of a $\mathcal{V}$-representation can be re-formulated in an $\mathcal{H}$-representation using an appropriate matrix B, at least in theory. If such B were computationally feasible, our testing problem would resemble tests of

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

based on test statistics of 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 extensively; see references in Section 2. Its analysis is intricate because the limiting distribution of T_N depends discontinuously on the true value of $B\theta$. One common way to get a critical value is to consider the globally least favorable case, which is $\theta = 0$. A less conservative strategy widely followed in the econometric literature on moment inequalities is "Generalized Moment Selection" (GMS; see Andrews and Soares (2010), Bugni (2010), Canay (2010)). If we had the $\mathcal{H}$-representation of $\mathcal{C}$, we might conceivably use the same technique. However, the duality

[12] See Gruber (2007), Grünbaum, Kaibel, Klee, and Ziegler (2003) and Ziegler (1995) for these results and other materials concerning convex polytopes used in this paper.

between the two representations is purely theoretical: In practice, B cannot be computed from A in high-dimensional cases like ours.

We therefore propose a *tightening* of the cone $\mathcal{C}$ that is computationally feasible and will have a similar effect as GMS. The idea is to tighten the constraint on ν in (4.1). In particular, define $\mathcal{C}_{\tau_N} := \{A\nu | \nu \geq \tau_N \mathbf{1}_H/H\}$ and define $\hat{\eta}_{\tau_N}$ as optimal argument in

$$
\begin{aligned}
J_N(\tau_N) \quad &: = \min_{\eta \in \mathcal{C}_{\tau_N}} N[\hat{\pi} - \eta]'\Omega[\hat{\pi} - \eta] \\
&= \min_{[\nu - \tau_N \mathbf{1}_H/H] \in \mathbf{R}_+^H} N[\hat{\pi} - A\nu]'\Omega[\hat{\pi} - A\nu].
\end{aligned}
\tag{4.4}
$$

Our proof establishes that constraints in the $\mathcal{H}$-representation that are almost binding at the original problem's solution (i.e., their slack is difficult to be distinguished from zero at the sample size) will be binding with zero slack after tightening. Suppose that $\sqrt{N}(\hat{\pi} - \pi) \to_d N(0, S)$ and let $\hat{S}$ consistently estimate S. Let $\tilde{\eta}_{\tau_N} := \hat{\eta}_{\tau_N} + \frac{1}{\sqrt{N}} N(0, \hat{S})$ or a bootstrap random variable and use the distribution of

$$
\begin{aligned}
\tilde{J}_N(\tau_N) \quad &: = \min_{\eta \in \mathcal{C}_{\tau_N}} N[\tilde{\eta}_{\tau_N} - \eta]'\Omega[\tilde{\eta}_{\tau_N} - \eta] \\
&= \min_{[\nu - \tau_N \mathbf{1}_H/H] \in \mathbf{R}_+^H} N[\tilde{\eta}_{\tau_N} - A\nu]'\Omega[\tilde{\eta}_{\tau_N} - A\nu],
\end{aligned}
\tag{4.5}
$$

to approximate the distribution of J_N. This has the same theoretical justification as the inequality selection procedure. Unlike the latter, however, it avoids the use of an $\mathcal{H}$-representation, thus offering a computationally feasible testing procedure.

Revisiting Example 3.1. With two intersecting budget planes, one can verify that the cone $\mathcal{C}$ is represented by

$$
B = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & 0 & 0 & -1 \end{bmatrix}
\tag{4.6}
$$

and then use a standard moment inequalities method, such as GMS, on this. This is the essence of inference in Hoderlein and Stoye (2014). Computation of B is not feasible in high dimensions, in which case our tightening based approach is useful.

4.4. **Theoretical Justification.** We now provide a detailed justification. First, we formalize the notion that choice probabilities are estimated by sample frequencies. Thus, for each budget set $\mathcal{B}_j$, denote the choices of N_j individuals, indexed by $n = 1, ..., N_j$, by

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

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

$$\hat{\pi}_{i|j} = \sum_{n=1}^{N_j} d_{i|j,n}/N_j, i = 1, ..., I_j, j = 1, ..., J. \tag{4.7}$$

The next lemma, among other things, shows that our tightening of the $\mathcal{V}$-representation of $\mathcal{C}$ is equivalent to a tightening its $\mathcal{H}$-representation but leaving B unchanged. For a matrix B, let $\mathrm{col}(B)$ denote its column space.

Lemma 4.1. *For $A \in \mathbf{R}^{I \times H}$, define*

$$\mathcal{C} = \{A\nu | \nu \geq 0\}.$$

and let

$$\mathcal{C} = \{t : Bt \leq 0\}$$

be its $\mathcal{H}$-representation for some $B \in \mathbf{R}^{m \times I}$ such that $B = \begin{bmatrix} B^{\leq} \\ B^{=} \end{bmatrix}$, where the submatrices $B^{\leq} \in \mathbf{R}^{\bar{m} \times I}$ and $B^{=} \in \mathbf{R}^{(m-\bar{m}) \times I}$ correspond to inequality and equality constraints, respectively. For $\tau > 0$ define

$$\mathcal{C}_\tau = \{A\nu | \nu \geq (\tau/H)\mathbf{1}_H\}.$$

Then one also has

$$\mathcal{C}_\tau = \{t : Bt \leq -\tau\phi\}$$

for some $\phi = (\phi_1, ..., \phi_m)' \in \mathrm{col}(B)$ with the properties that (i) $\bar{\phi} := [\phi_1, ..., \phi_{\bar{m}}]' \in \mathbf{R}_{++}^{\bar{m}}$, and (ii) $\phi_k = 0$ for $k > \bar{m}$.

Lemma 4.1 is *not* just a re-statement of the Minkowski-Weyl theorem for polyhedra, which would simply say $\mathcal{C}_\tau = \{A\nu | \nu \geq (\tau/H)\mathbf{1}_H\}$ is alternatively represented as an intersection of closed halfspaces. The lemma instead shows that the inequalities in the $\mathcal{H}$-representation becomes tighter by

$\tau\phi$ after tightening the $\mathcal{V}$-representation by $\tau_N \mathbf{1}_H/H$, with the same matrix of coefficients B appearing both for $\mathcal{C}$ and $\mathcal{C}_\tau$. Note that for notational convenience, we rearrange rows of B so that the genuine inequalities come first and pairs of inequalities that represent equality constraints come last.[13] This is w.l.o.g.; in particular, the researcher does not need to know which rows of B these are. Then as we show in the proof, the elements in ϕ corresponding to the equality constraints are automotically zero when we tighten the space for *all* the elements of ν in the $\mathcal{V}$-representation. This is a useful feature that makes our methodology work in the presence of equality constraints.

The following assumptions are used for our asymptotic theory.

Assumption 4.1. *For all $j = 1, ..., J$, $\frac{N_j}{N} \to \rho_j$ as $N \to \infty$, where $\rho_j > 0$.*

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 $1 \le j \le J$ and $1 \le k \le m$.

Assumption 4.2. *J repeated cross-sections of random samples $\left\{\{d_{i|j,n(j)}\}_{i=1}^{I_j}\right\}_{n(j)=1}^{N_j}$, $j = 1, ..., J$, are observed.*

The econometrician also observes the normalized price vector p_j, which is fixed in this section, for each $1 \le j \le J$.

Next, we impose a mild condition that guarantees stable behavior of the statistic J_N. To this end, we further specify the nature of each row of B. Recall that w.l.o.g. the first $\bar{m}$ rows of B correspond to inequality constraints, whereas the rest of the rows represent equalities. Note that the $\bar{m}$ inequalities include nonnegativity constraints $\pi_{i|j} \ge 0, 1 \le i \le I_j, 1 \le j \le J$, represented by the row of B consisting of a negative constant for the corresponding element and zeros otherwise. Likewise, the identities that $\sum_{i=1}^{I_j} \pi_{i|j}$ is constant across $1 \le j \le J$ are included in the set of equality constraints.[14] We show in the proof that the presence of these "definitional" equalities/inequalities, which always hold by construction of $\hat{\pi}$, do not affect the asymptotic theory even when they are (close to) be binding. Define $\mathcal{K} = \{1, ..., m\}$, and let $\mathcal{K}^D$ be the set of indices for the rows of B corresponding to the above nonnegativity constraints and the constant-sum constraints. Let $\mathcal{K}^R = \mathcal{K} \setminus \mathcal{K}^D$, so that

[13] In the matrix displayed in (4.6), the third and fourth row would then come last.

[14] If we impose the (redundant) restriction $\mathbf{1}_H'\nu = 1$ in the definition of $\mathcal{C}$, then the corresponding equality restrictions would be $\sum_{i=1}^{I_j} \pi_{i|j} = 1$ for every j.

$b_k'\pi \leq 0$ represents an economic restriction if $k \in \mathcal{K}^R$.[15] With these definitions, consider the following requirement:

Condition 4.1. *For each $k \in \mathcal{K}^R$, $\mathrm{var}(b_k(j(k))'d_{j,n}) \geq \epsilon$ holds for at least one $j(k)$, $1 \leq j(k) \leq J$, where ϵ is a positive constant.*

Our uniform size control result stated below relies on a triangular array CLT. It is known that for triangular array $X_{iN}, i = 1, ..., N \sim_{\text{iid}} \text{Bernoulli}(p_N)$, $N = 1, 2, ...$, such a CLT obtains iff the Lyapunov condition $Np_N(1 - p_N) \to \infty$ holds. As long as one of the elements of π appearing in each of the k-th restriction belongs to $[\varepsilon, 1 - \varepsilon]$ for small $\varepsilon > 0$ for each $k \in \mathcal{K}^R$, then $\sum_{j=1}^{J} b_k(j)'d_{j,n}$ will satisfy the Lyapunouv condition. Condition 4.1 guarantees this. Note that this does not require, for example, all of the elements of π to be bounded away from zero. See the proof of Theorem 4.2 in Appendix A for more on this point. Though this condition involves the matrix B, implementing our procedure does not require knowing B. The condition is meant as restriction on the underlying d.g.p. Directly testing Condition 4.1 using data seems to require the setting of the empirical problem to be of a sufficiently small scale so that B can be computed.

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 4.1 for some (common) value of ϵ.

Theorem 4.2. *Choose τ_N so that $\tau_N \downarrow 0$ and $\sqrt{N}\tau_N \uparrow \infty$. Also, let Ω be diagonal, where all the diagonal elements are positive. Then under Assumptions 4.1 and 4.2*

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

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

While it is obvious that our tightening contracts the cone, the result depends on a more delicate feature, namely that we (potentially) turn non-binding inequalities from the $\mathcal{H}$-representation into binding ones but not vice versa. This feature is not universal to cones as they get contracted. Our proof establishes that it generally obtains if Ω is the identity matrix and all corners of the cone are acute. In this paper's application, we can further exploit the cone's geometry to extend the result to any diagonal Ω. Our method immediately applies to other testing problems featuring $\mathcal{V}$-representations if analogous features can be verified.

[15]In (4.6), $\mathcal{K}^R$ contains only the last row of the matrix.

We finally remark that in principle, a critical value for J_N could be computed by regularization. Let $\tilde{\eta}_{\alpha_N} := \hat{\eta} + \sqrt{\frac{\alpha_N}{N}} N(0, \hat{S})$, where α_N is a sequence that goes to infinity slowly. Recall that $\hat{\eta}$ is the projection of the choice frequency vector $\hat{\pi}$ onto the cone $\mathcal{C}$. The distribution of $\tilde{J}_N(\alpha_N) := \frac{N}{\alpha_N} \min_{\nu \in \mathbf{R}_+^H} [\tilde{\eta}_{\alpha_N} - A\nu]' \Omega [\tilde{\eta}_{\alpha_N} - A\nu]$ can be evaluated by simulation. It provides a valid approximation of the distribution of J_N asymptotically, regardless of the position of η_0, the population analog of $\hat{\eta}$, on the cone $\mathcal{C}$. This is basically the idea behind subsampling and the m-out-of-n bootstrap. It is convenient computationally and does not rely on geomtric features of the cone. However, it is subject to practitioners' critiques of subsampling, e.g. sensitivity to choice of α_N; furthermore, Andrews and Guggenberger (2009, 2010) forcefully argue that it can suffer from low power compared to inequality selection methods to which our method is more similar.

5. Extending the Scope of the Test

The methodology outlined in Section 4 requires (i) the observations available to the econometrician are drawn on a finite number of budgets and (ii) the budgets are given exogenously, that is, unobserved heterogeneity and budgets are assumed to be independent. These conditions are naturally satisfied in some applications. The empirical setting in Section 6, however, calls for modifications because Condition (i) is certainly violated in it and imposing Condition (ii) would be very restrictive. We propose to use a series estimator to estimate the conditional choice probability vector π for a specific expenditure W when W is distributed continuously (Section 5.1). Furthermore, a method to test stochastic rationality in the presence of possible endogeneity of income is developed using a control function method (Section 5.2).

The setting in this section is as follows. Let $\tilde{p}_j \in \mathbf{R}_{++}^K$ denote the unnormalized price vector, fixed for each period j. Let $(S, \mathcal{S}, P)$ denote the underlying probability space. Since we have repeated cross-sections over J periods, write $P = \otimes_{j=1}^J P^{(j)}$, a J-fold product measure. Let P_u denote the marginal probability law of u, which we assume does not depend on j. We do not, however, assume that the laws of other random elements, such as income, are time homogeneous. Let $w = \log(W)$ denote log total expenditure, and the researcher chooses a value $\underline{w}_j$ for w for each period j.

5.1. **Test statistic with smoothing.**

This subsection proposes a smoothing procedure based on a series estimator (see, for example, Newey (1997)) for π to deal with a situation where total expenditure W is continuously distributed, yet exogenous. We need some notation and definitions to formally state the asymptotic theory behind our procedure with smoothing.

Let $w_{n(j)}$ be the log total expenditure of consumer $n(j)$, $1 \le n(j) \le N_j$ observed in period j.

Assumption 5.1. *J repeated cross-sections of random samples $\left\{\left(\{d_{i|j,n(j)}\}_{i=1}^{I_j}, w_{n(j)}\right)\right\}_{n(j)=1}^{N_j}, j = 1, ..., J$, are observed.*

The econometrician also observes the unnormalized price vector $\tilde{p}_j$, which is fixed, for each $1 \leq j \leq J$.

This sub-section assumes that the total expenditure is exogenous, in the sense that

$$w \perp\!\!\!\perp u$$

holds under every $P^{(j)}, 1 \leq j \leq J$. This will be relaxed in the next subsection. Define

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

We have

$$\begin{aligned} p_{i|j}(\underline{w}_j) &= \Pr\{d_{i|j,n(j)} = 1 | w_{n(j)} = \underline{w}_j\} \\ &= \Pr\{D(\tilde{p}_j / w_{n(j)}, u) \in x_{i|j} | w_{n(j)} = \underline{w}_j\} \\ &= \Pr\{D(\tilde{p}_j / \underline{w}_j, u) \in x_{i|j}, u \sim P_u\} \end{aligned}$$

where the third equality follows from the exogeneity assumption. Letting

$$\pi_{i|j} = p_{i|j}(\underline{w}_j)$$

and writing $\pi_j := (\pi_{1|j}, ..., \pi_{I_j|j})'$ and $\pi := (\pi'_1, ..., \pi'_J)' = (\pi_{1|1}, \pi_{2|1}, ..., \pi_{I_J|J})'$, the stochastic rationality condition is given by

$$\pi \in \mathcal{C}$$

as before. Define $q^K(w) = (q_{1K}(w), ..., q_{KK}(w))'$, where $q_{jK}(w), j = 1, ..., K$ are basis functions (e.g. power series or splines) of w. Instead of sample frequency estimators, for each $j, 1 \leq j \leq J$ we use

$$\begin{aligned} \hat{\pi}_{i|j} &= q^{K(j)}(\underline{w}_j)' \widehat{Q}^{-}(j) \sum_{n(j)=1}^{N_j} q^{K(j)}(w_{n(j)}) d_{i|j,n(j)} / N_j, \\ \widehat{Q}(j) &= \sum_{n(j)=1}^{N_j} q^{K(j)}(w_{n(j)}) q^{K(j)}(w_{n(j)})' / N_j \\ \hat{\pi}_j &= (\hat{\pi}_{1|j}, ..., \hat{\pi}_{I_j|j})', \\ \hat{\pi} &= (\hat{\pi}'_1, ..., \hat{\pi}'_J)', \end{aligned}$$

to estimate $\pi_{i|j}$, where A^- denotes a symmetric generalized inverse of A and $K(j)$ is the number of basis functions applied to Budget $\mathcal{B}_j$. The estimators $\hat{\pi}_{i|j}$'s may not take their values in $[0, 1]$. This does not seem to cause a problem asymptotically, though as in Imbens and Newey (2009), we may (and do, in the application) instead use

$$\hat{\pi}_{i|j} = G\left(q^{K(j)}(\underline{w}_j)'\widehat{Q}^{-}(j)\sum_{n(j)=1}^{N_j} q^{K(j)}(w_{n(j)})d_{i|j,n(j)}/N_j\right),$$

where G denotes the CDF of Unif$(0,1)$. The smoothed version of J_N is obtained using the above series estimator for $\hat{\pi}$ in (4.1). Then an appropriate choice of τ_N is $\tau_N = \sqrt{\frac{\log \underline{n}}{\underline{n}}}$ with

$$\underline{n} = \min_j N_j I_j/\text{trace}(v_N^{(j)})$$

where $v_N^{(j)}$ is defined below. 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, and we follow this convention. The bootstrapped test statistic $\tilde{J}_N(\tau_N)$ is obtained applying the same replacements to the formula (4.5), 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 $\hat{v}_N^{(j)}$ satisfy $\hat{v}_N^{(j)} v_N^{(j)-1} \to_p \mathbf{I}_{I_j}$, where

$$v_N^{(j)} = [\mathbf{I}_{I_j} \otimes q^{K(j)}(\underline{w}_j)'Q_N(j)^{-1}]\Lambda_N^{(j)}[\mathbf{I}_{I_j} \otimes Q_N^{-1}(j)q^{K(j)}(\underline{w}_j)]$$

with $Q_N(j) := \mathrm{E}[q^{K(j)}(w_{n(j)})q^{K(j)}(w_{n(j)})']$, $\Lambda_N^{(j)} := \mathrm{E}[\Sigma^{(j)}(w_{n(j)}) \otimes q^{K(j)}(w_{n(j)})q^{K(j)}(w_{n(j)})']$, and $\Sigma^{(j)}(w) := \mathrm{Cov}[d_{j,n(j)}|w_{n(j)} = w]$. Note that $\Sigma^{(j)}(w) = \text{diag}\left(p^{(j)}(w)\right) - p^{(j)}(w)p^{(j)}(w)'$ where $p^{(j)}(w) = [p_{1|j}(w), ..., p_{I_j|j}(w)]'$. For example, one may use

$$\hat{v}_N^{(j)} = [\mathbf{I}_{I_j} \otimes q^{K(j)}(\underline{w}_j)'\hat{Q}^{-}(j)]\widehat{\Lambda}(j)[\mathbf{I}_{I_j} \otimes \hat{Q}^{-}(j)q^{K(j)}(\underline{w}_j)]$$

with $\widehat{\Lambda}(j) = \frac{1}{N_j}\sum_{n(j)=1}^{N_j}\left[\widehat{\Sigma}^{(j)}(w_{n(j)}) \otimes q^{K(j)}(w_{n(j)})q^{K(j)}(w_{n(j)})'\right]$, $\widehat{\Sigma}^{(j)}(w) = \text{diag}\left(\widehat{p}^{(j)}(w)\right) - \widehat{p}^{(j)}(w)\widehat{p}^{(j)}(w)'$, $\widehat{p}^{(j)}(w) = [\widehat{p}_{1|j}(w), ..., \widehat{p}_{I_j|j}(w)]'$ and $\widehat{p}_{i|j}(w) = q^{K(j)}(w)'\widehat{Q}^{-}(j)\sum_{n(j)=1}^{N_j} q^{K(j)}(w_{n(j)})d_{i|j,n(j)}/N_j$. 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 $\tilde{\eta}_{\tau_N}(j) := \hat{\eta}_{\tau_N}(j) + \frac{1}{\sqrt{N_j}}N(0, \hat{v}_N^{(j)}), j = 1, ..., J$.

Let $\mathcal{W}_j$ denote the support of $w_{n(j)}$. For a symmetric matrix A, $\lambda_{\min}$ signifies its smallest eigenvalue.

Condition 5.1. *There exist positive constants C, ϵ, δ, and $\zeta(K)$, $K \in \mathbf{N}$ such that the following holds:*

(i) $\pi \in \mathcal{C}$;

(ii) *For each $k \in \mathcal{K}^R$, $\mathrm{var}(b_k(j(k))'d_{j,n}|w_{n(j)}) \geq \epsilon$ holds for at least one $j(k), 1 \leq j(k) \leq J$;*

(iii) $\sup_{w\in\mathcal{W}_j}|p_{i|j}(w) - q^K(w)'\beta_K^{(j)}| \leq CK^{-\delta}$ *holds with some* K*-vector* $\beta_K^{(j)}$ *for every* $K \in \mathbf{N}$, $1 \leq i \leq I_j, 1 \leq j \leq J$*;*

(iv) *Letting* $\widetilde{q}^K := C_{K,j}q^K$, $\lambda_{\min}\mathrm{E}[\widetilde{q}^K(w_{n(j)})\widetilde{q}^K(w_{n(j)})'] \geq C$ *holds for every* K *and* j*, where* $C_{K,j}, K \in \mathbf{N}, 1 \leq j \leq J$ *are constant nonsingular matrices;*

(v) $\max_j \sup_{w\in\mathcal{W}_j}\|\widetilde{q}^K(w)\| \leq C\zeta(K)$ *for every* $K \in \mathbf{N}$.

In what follows, F_j signifies the joint distribution of $(d_{i|j,n(j)}, w_{n(j)})$. Let $\mathcal{F}$ be the set of all $(F_1, ..., F_J)$ that satisfy Condition 5.1 for some $(C, \epsilon, \delta, \zeta(\cdot))$.

Theorem 5.1. *Choose* τ_N *and* $K(j), j = 1, ..., J$ *so that* $\sqrt{N_j}K^{-\delta}(j) \downarrow 0$, $\zeta(K(j))^2K(j)/N_j \downarrow 0$, $j = 1, ..., J$, $\tau_N \downarrow 0$, *and* $\sqrt{\underline{n}}\tau_N \uparrow \infty$*. Also let* Ω *be diagonal where all the diagonal elements are positive. Then under Assumptions 4.1 and 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* $\tilde{J}_N(\tau_N)$, $0 \leq \alpha \leq \frac{1}{2}$.

5.2. **Endogeneity.** We now relax the assumption that consumer's utility functions are realized independently from W. Exogeneity of budget sets is a standard assumption in classical demand analysis based on random utility models; for example, it is assumed, at least implicitly, in **??**. Nonetheless, the assumption can be a concern in applying our testing procedure to a data set such as ours. Recall that the budget sets $\{B_j\}_{j=1}^J$ are based on prices and total expenditure. The latter is likely to be endogenous, and the econometrician should be concerned with its potential effect.

As independence between utility and budgets is fundamental to McFadden-Richter theory, addressing it in our testing procedure might seem difficult. Fortunately, recent advances in nonparametric identification and estimation of models with endogeneity inform a solution. To see this, it is useful to rewrite the model so that we can cast it into a framework of nonseparable models with endogenous covariates. Writing $p_j = \tilde{p}_j/W$, where $\tilde{p}_j$ is the unnormalized price vector, the essence of the problem is as follows: Stochastic rationalizability imposes restrictions on the conditional distributions of $y = D(p, u)$ for different p when u is distributed according to its population marginal distribution P_u, but the observed conditional distribution of y given p does not estimate this when w and u are interrelated.

Well-known results for nonseparable models, such as Chesher (2003), are concerned with (point) identification of the structural function $D(\cdot,\cdot)$ under the assumption that the unobserved heterogeneity (u in our model) is a scalar and that the structural function $D(\cdot,\cdot)$ is monotone in u. In contrast, it is an important part of our contribution that we let the nuisance parameter u be infinite dimensional and leave $D(\cdot,\cdot)$ completely unrestricted. Hence, we do not want to make such assumptions. Fortunately, we do not need to: Imbens and Newey (2009) (see also Blundell and Powell (2003)) note that various "average" counterfactual effects can be identified when u is not even finite dimensional using a control function approach. The patch probabilities that we want to estimate are such average effects.

For each fixed value $\underline{w}_j$ and the unnormalized price vector $\tilde{p}_j$ in period j, $1 \le j \le J$, define the endogeneity corrected conditional probability[16]

$$\begin{aligned}\pi(\tilde{p}_j/e^{\underline{w}_j}, x_{i|j}) &:= \Pr\{D(\tilde{p}_j/e^{\underline{w}_j}, u) \in x_{i|j}, u \text{ distributed according to } P_u\} \\ &= \int \mathbf{1}\{D_j(\underline{w}_j, u) \in x_{i|j}\} dP_u\end{aligned}$$

where $D_j(w,u) := D(\tilde{p}_j/e^{w}, u)$. Then Proposition 3.2 still applies to

$$\pi_{\mathrm{EC}} := [\pi(p_1, x_{1|1}), ..., \pi(p_1, x_{I_1|1}), \pi(p_2, x_{1|2}), ..., \pi(p_2, x_{I_2|2}), ..., \pi(p_J, x_{1|J}), ..., \pi(p_J, x_{I_J|J})]'.$$

If we define $J_{\mathrm{EC}} = \min_{\nu \in \mathbf{R}_+^h} [\pi_{\mathrm{EC}} - A\nu]'\Omega[\pi_{\mathrm{EC}} - A\nu]$, then its value is zero iff stochastic rationality holds. Note that this new definition π_{EC} recovers the definition of π in Sub-section 5.1 when w is exogenous.

Suppose w is endogenous but there exists a control variable ϵ such that

$$w \perp\!\!\!\perp u | \epsilon$$

holds under every $P^{(j)}, 1 \le j \le J$. For example, given a reduced form $w = h_j(z, \epsilon)$ with h_j monotone in ϵ and z is an instrument, one may use $\epsilon = F^{(j)}_{w|z}(w|z)$ where $F^{(j)}_{w|z}$ denotes the conditional CDF of w given z under $P^{(j)}$ when the random vector (w,z) obeys the probability law $P^{(j)}$; see Imbens and Newey (2009) for this type of control variable in the context of cross-sectional data. Note that $\epsilon \sim \mathrm{Uni}(0,1)$ under every $P^{(j)}, 1 \le j \le J$ by construction. Let $P^{(j)}_{y|w,\epsilon}$ denote the conditional probability measure for y given (w,ϵ) corresponding to $P^{(j)}$. Adapting the argument in Imbens and Newey (2009) and Blundell and Powell (2003), under the assumption that $\mathrm{supp}(w) = \mathrm{supp}(w|\epsilon)$ under $P^{(j)}, 1 \le j \le J$

[16]This is the conditional choice probability if p is (counterfactually) assumed to be exogenous. We call it "endogeneity corrected" instead of "counterfactual" because we use the term counterfactual when referring to rationality restricted prediction.

we have

$$\begin{aligned}\pi(p_j, x_{i|j}) &= \int_0^1 \int_u \mathbf{1}\{D_j(\underline{w}_j, u) \in x_{i|j}\} dP^{(j)}_{u|\epsilon} d\epsilon \\ &= \int_0^1 \int_u \mathbf{1}\{D_j(w, u) \in x_{i|j}\} dP^{(j)}_{u|w=\underline{w}_j,\epsilon} d\epsilon \\ &= \int_0^1 P^{(j)}_{y|w,\epsilon}\left\{y \in x_{i|j} | w = \underline{w}_j, \epsilon\right\} d\epsilon, \quad 1 \leq j \leq J\end{aligned}$$

where the second equality follows from $w \perp\!\!\!\perp u|\epsilon$ and the support condition. Note that $P^{(j)}_{y|w,\epsilon}\left\{y \in x_{i|j} | w = \underline{w}_j, \epsilon\right\}$ is observable.

Let $z_{n(j)}$ be the $n(j)$-th observation of the instrumental variable z in period j.

Assumption 5.2. *J repeated cross-sections of random samples $\left\{\left(\{d_{i|j,n(j)}\}_{i=1}^{I_j}, x_{n(j)}, z_{n(j)}\right)\right\}_{n=1}^{N_j}, j = 1, ..., J$, are observed.*

The econometrician also observes the unnormalized price vector $\tilde{p}_j$, which is fixed, for each $1 \leq j \leq J$.

The last result shows that π_{EC} can be estimated nonparametrically. More specifically, we can proceed in two steps as follows. The first step is to obtain control variable estimates $\widehat{\epsilon}_{n(j)}, n(j) = 1, ..., N_j$ for each j. For example, let $\widehat{F}^{(j)}_{w|z}$ be a nonparametric estimator for $F_{w|z}$ for a given instrumental variable z in period j. For concreteness, we consider a series estimator as in Imbens and Newey (2002). Let $r^L(z) = (r_{1L}(z), ..., r_{LL}(z))$, where $r_{\ell L}(z), \ell = 1, ..., L$ are basis functions, then define

$$\widehat{F}^{(j)}_{w|z}(w|z) = r^L(z)' \widehat{R}^{-}(j) \sum_{n(j)=1}^{N_j} r^{L(j)}(z_{n(j)}) \mathbf{1}\{w_{n(j)} \leq w\}/N_j$$

where

$$\widehat{R}(j) = \sum_{n(j)=1}^{N_j} r^{L(j)}(z_{n(j)}) r^{L(j)}(z_{n(j)})'/N_j$$

Let

$$\widetilde{\epsilon}_{n(j)} = \widehat{F}^{(j)}_{w|z}(w_{n(j)}|z_{n(j)}), n(j) = 1, ..., N_j.$$

Choose a sequence $\upsilon_N \to 0, \upsilon_N > 0$ and define $\iota_N(\epsilon) = (\epsilon + \upsilon_N)^2/4\upsilon_N$, then let

$$\gamma_N(\epsilon) = \begin{cases} 1 & \text{if } \epsilon > 1 + \upsilon_N \\ 1 - \iota_N(1-\epsilon) & \text{if } 1 - \upsilon_N < \epsilon \le 1 + \upsilon_N \\ \epsilon & \text{if } \upsilon_N \le \epsilon \le 1 - \upsilon_N \\ \iota_N(\epsilon) & \text{if } -\upsilon_N \le \epsilon \le \upsilon_N \\ 0 & \text{if } \epsilon < -\upsilon_N \end{cases}$$

then our control variable is $\widehat{\epsilon}_{n(j)} = \gamma_N(\widetilde{\epsilon}_{n(j)}), n(j) = 1, ..., N_j$.

The second step is nonparametric estimation of $P^{(j)}_{y|w,\epsilon}\left\{y \in x_{i|j} | w = \underline{w}_j, \epsilon\right\}$. Let $\widehat{\chi}_{n(j)} = (w_{n(j)}, \widehat{\epsilon}_{n(j)})'$, $n(j) = 1, ..., N_j$ for each j. Write $s^{M(j)}(\chi) = (s_{1M(j)}(\chi), ..., s_{M(j)M(j)}(\chi))'$, where $s_{mM(j)}(\chi), \chi \in \mathbf{R}^{K+1}, m = 1, ..., M(j)$ are basis functions, then our estimator for $P^{(j)}_{y|w,\epsilon}\left\{y \in x_{i|j} | w = \cdot, \epsilon = \cdot\right\}$ evaluated at $\chi = (w, \epsilon)$ is

$$\begin{aligned} \widehat{P^{(j)}_{y|w,\epsilon}}\left\{y \in x_{i|j} | w, \epsilon\right\} &= s^{M(j)}(\chi)'\widehat{S}^{-}(j) \sum_{n(j)=1}^{N_j} s^{M(j)}(\widehat{\chi}_{n(j)}) d_{i|j,n(j)}/N_j \\ &= s^{M(j)}(\chi)'\widehat{\alpha}_i^{M(j)} \end{aligned}$$

where

$$\widehat{S}(j) = \sum_{n(j)=1}^{N_j} s^{M(j)}(\widehat{\chi}_{n(j)}) s^{M(j)}(\widehat{\chi}_{n(j)})'/N_j, \quad \widehat{\alpha}_i^{M(j)} := \widehat{S}^{-}(j) \sum_{n(j)=1}^{N_j} s^{M(j)}(\widehat{\chi}_{n(j)}) d_{i|j,n(j)}/N_j.$$

Our endogeneity corrected conditional probability $\pi(p_j, x_{i|j})$ is a linear functional of $P^{(j)}_{y|w,\epsilon}\left\{y \in x_{i|j} | w = \underline{w}_j, \epsilon\right\}$, thus plugging in $\widehat{P^{(j)}_{y|w,\epsilon}}\left\{y \in x_{i|j} | w = \underline{w}_j, \epsilon\right\}$ into the functional, we define

$$\begin{aligned} \pi(\widehat{p_j, x_{i|j}}) &:= \int_0^1 \widehat{P^{(j)}_{y|w,\epsilon}}\left\{y \in x_{i|j} | w = \underline{w}_j, \epsilon\right\} d\epsilon \\ &= D(j)'\widehat{\alpha}_i^{M(j)}, \\ \text{where} \quad D(j) &:= \int_0^1 s^{M(j)}\left(\begin{bmatrix} \underline{w}_j \\ \epsilon \end{bmatrix}\right) d\epsilon \qquad i = 1, ..., I_j, j = 1, ..., J \end{aligned}$$

and

$$\widehat{\pi_{\mathrm{EC}}} = [\pi(\widehat{p_1, x_{1|1}}), ..., \pi(\widehat{p_1, x_{I_1|1}}), \pi(\widehat{p_2, x_{1|2}}), ..., \pi(\widehat{p_2, x_{I_2|2}}), ..., \pi(\widehat{p_J, x_{1|J}}), ..., \pi(\widehat{p_J, x_{I_J|J}})]'.$$

The final form of the test statistic is

$$J_{\mathrm{EC}_N} = N \min_{\nu \in \mathbf{R}_+^h} [\widehat{\pi_{\mathrm{EC}}} - A\nu]' \Lambda [\widehat{\pi_{\mathrm{EC}}} - A\nu].$$

The calculation of critical values can be carried out in the same way as the testing procedure in Section 5.1, though the covariance matrix $v_N^{(j)}$ needs modification. With the nonparametric endogeneity correction, the modified version of $v_N^{(j)}$ is

$$\bar{v}_N^{(j)} = [\mathbf{I}_{I_j} \otimes D(j)' S_N(j)^{-1}] \bar{\Lambda}_N^{(j)} [\mathbf{I}_{I_j} \otimes S_N(j)^{-1} D(j)]$$

where

$$S_N(j) = \mathrm{E}[s^{M(j)}(\chi_{n(j)}) s^{M(j)}(\chi_{n(j)})'], \quad \bar{\Lambda}_N^{(j)} = \bar{\Lambda}_{1_N}^{(j)} + \bar{\Lambda}_{2_N}^{(j)},$$
$$\bar{\Lambda}_{1_N}^{(j)} = \mathrm{E}[\bar{\Sigma}^{(j)}(\chi_{n(j)}) \otimes s^{M(j)}(\chi_{n(j)}) s^{M(j)}(\chi_{n(j)})'], \quad \bar{\Lambda}_{2_N}^{(j)} = \mathrm{E}[m_{n(j)} m_{n(j)}']$$

with

$$\bar{\Sigma}^{(j)}(\chi) := \mathrm{Cov}[d_{j,n(j)} | \chi_{n(j)} = \chi],$$
$$m_{n(j)} := [m_{1,n(j)}', m_{2,n(j)}', \cdots, m_{I_j,n(j)}']',$$

$$m_{i,n(j)} :=$$
$$\mathrm{E}\left[\dot{\gamma}_N(\epsilon_{m(j)}) \frac{\partial}{\partial \epsilon} P_{y|w,\epsilon}^{(j)} \left\{ y \in x_{i|j} | w_{m(j)}, \epsilon_{m(j)} \right\} s^{M(j)}(\chi_{m(j)}) r^{L(j)}(z_{m(j)})' R_N(j)^{-1} r^{L(j)}(z_{n(j)}) u_{mn(j)} \right.$$
$$\left. \Bigg| d_{i|j,n(j)}, w_{n(j)}, z_{n(j)} \right],$$

$$R_N(j) := \mathrm{E}[r^{L(j)}(z_{n(j)}) r^{L(j)}(z_{n(j)})'], \quad u_{mn(j)} := \mathbf{1}\{w_{n(j)} \le w_{m(j)}\} - F_{w|z}^{(j)}(w_{m(j)} | z_{n(j)}).$$

Define

$$\underline{n}_{\mathrm{EC}} = \min_j N_j I_j / \mathrm{trace}(\bar{v}_N^{(j)}),$$

then a possible choice for τ_N is $\tau_N = \sqrt{\frac{\log \underline{n}_{\mathrm{EC}}}{\underline{n}_{\mathrm{EC}}}}$. Proceed as in Section 5.1, replacing $\hat{v}_N^{(j)}$ with a consistent estimator for $\bar{v}_N^{(j)}$ for $j = 1, ..., J$, to define the bootstrap version $\tilde{J}_{\mathrm{EC}}(\tau_N)$.

We impose some conditions to show the validity of the endogeneity-corrected test. Define $\mathcal{X}_j = \mathrm{supp}(\chi_{n(j)})$, and $\mathcal{Z}_j = \mathrm{supp}(z_{n(j)}), 1 \le j \le J$. Following the above discussion, define an $\mathbf{R}^I$-valued functional

$$\pi(P_{y|w,\epsilon}^{(1)}, ..., P_{y|w,\epsilon}^{(J)}) = [\pi_{1|1}(P_{y|w,\epsilon}^{(1)}), ..., \pi_{I_1|1}(P_{y|w,\epsilon}^{(1)}), \pi_{1|2}(P_{y|w,\epsilon}^{(2)}), ..., \pi_{I_2|2}(P_{y|w,\epsilon}^{(2)}), ..., \pi_{1|J}(P_{y|w,\epsilon}^{(J)}), ..., \pi_{I_J|J}(P_{y|w,\epsilon}^{(J)})]'$$

where

$$\pi_{i|j}(P_{y|w,\epsilon}^{(j)}) := \int_0^1 P_{y|w,\epsilon}^{(j)} \left\{ y \in x_{i|j} | w = \underline{w}_j, \epsilon \right\} d\epsilon$$

and $\epsilon_{n(j)} := F^{(j)}_{w|z}(w_{n(j)}|z_{n(j)})$ for every j.

Condition 5.2. *There exist positive constants C, ϵ, δ_1, δ, $\zeta_r(L)$, $\zeta_s(M)$, and $\zeta_1(M)$, $L \in \mathbf{N}$, $M \in \mathbf{N}$ such that the following holds:*

(i) *The distribution of $w_{n(j)}$ conditional on $z_{n(j)} = z$ is continuous for every $z \in \mathcal{Z}_j$, $1 \le j \le J$;*

(ii) $\mathrm{supp}(w_{n(j)}|\epsilon_{n(j)} = \epsilon) = \mathrm{supp}(w_{n(j)})$ *for every $\epsilon \in [0,1]$, $1 \le j \le J$;*

(iii) $\pi(P^{(1)}_{y|w,\epsilon}, ..., P^{(J)}_{y|w,\epsilon}) \in \mathcal{C}$;

(iv) *For each $k \in \mathcal{K}^R$, $\mathrm{var}(b_k(j(k))'d_{j,n}|w_{n(j)}, \epsilon_{n(j)}) \ge \epsilon$ holds for at least one $j(k)$, $1 \le j(k) \le J$;*

(v) *Letting $\widetilde{r}^L := C_{L,j}r^L$, $\lambda_{\min}\mathrm{E}[\widetilde{r}^L(z_{n(j)})\widetilde{r}^L(z_{n(j)})] \ge C$ holds for every L and j, where $C_{L,j}, L \in \mathbf{N}$, $1 \le j \le J$, are constant nonsingular matrices;*

(vi) $\max_j \sup_{z\in\mathcal{Z}_j} \|\widetilde{r}^L(z)\| \le C\zeta_r(L)$ *for every $L \in \mathbf{N}$.*

(vii) $\sup_{w\in\mathcal{W}_j, z\in\mathcal{Z}_j} |F^{(j)}_{w|z}(w,z) - r^L(z)'\alpha^{(j)}_L(w)| \le CL^{-\delta_1}, 1 \le j \le J$ *holds with some L-vector $\alpha^{(j)}_L(\cdot)$ for every $L \in \mathbf{N}$, $1 \le j \le J$;*

(viii) *Letting $\widetilde{s}^M := \bar{C}_{M,j}s^M$, $\lambda_{\min}\mathrm{E}[\widetilde{s}^M(\chi_{n(j)})\widetilde{s}^M(\chi_{n(j)})] \ge C$ holds for every M and j, where $\bar{C}_{M,j}, M \in \mathbf{N}$, $1 \le j \le J$, are constant nonsingular matrices;*

(ix) $\max_j \sup_{\chi\in\mathcal{X}_j} \|\widetilde{s}^M(\chi)\| \le C\zeta_s(M)$ *and* $\max_j \sup_{\chi\in\mathcal{X}_j} \|\partial\widetilde{s}^M(\chi)/\partial\epsilon\| \le C\zeta_1(M)$ *and* $\zeta_s(M) \le C\zeta_1(M)$ *for every $M \in \mathbf{N}$;*

(x) $\sup_{\chi\in\mathcal{X}_j} \left| P^{(j)}_{y|w,\epsilon}\{y \in x_{i|j}|w,\epsilon\} - s^M(\chi)'g^{(i,j)}_M \right| \le CM^{-\delta}$, *holds with some M-vector $g^{(i,j)}_M$ for every $M \in \mathbf{N}$, $1 \le i \le I_j, 1 \le j \le J$;*

(xi) $P^{(j)}_{y|w,\epsilon}\{y \in x_{i|j}|w,\epsilon\}$, $1 \le i \le I_j, 1 \le j \le J$, *are twice continuously differentiable in $\chi = (w,\epsilon)$. Moreover,* $\max_{1\le j\le J}\max_{1\le i\le I_j}\sup_{\chi\in\mathcal{X}_j}\left\|\frac{\partial}{\partial\chi}P^{(j)}_{y|w,\epsilon}\{y \in x_{i|j}|w,\epsilon\}\right\| \le C$ *and* $\max_{1\le j\le J}\max_{1\le i\le I_j}\sup_{\chi\in\mathcal{X}_j}\left\|\frac{\partial^2}{\partial\chi\partial\chi'}P^{(j)}_{y|w,\epsilon}\{y \in x_{i|j}|w,\epsilon\}\right\| \le C$.

In what follows, F_j signifies the joint distribution of $(d_{i|j,n(j)}, w_{n(j)}, z_{n(j)})$. Let $\mathcal{F}_{\mathrm{EC}}$ be the set of all $(F_1, ..., F_J)$ that satisfy Condition 5.2 for some $(C, \epsilon, \delta_1, \delta, \zeta_r(\cdot), \zeta_s(\cdot), \zeta_1(\cdot))$. Then we have:

Theorem 5.2. *Choose τ_N, $M(j)$ and $L(j)$, $j = 1, ..., J$ so that $\tau_N \downarrow 0$, $\sqrt{\underline{n}_{\mathrm{EC}}}\tau_N \uparrow \infty$, $N_jL(j)^{1-2\delta_1} \downarrow 0$, $N_jM(j)^{-2\delta} \downarrow 0$, $M(j)\zeta_1(M(j))^2L^2(j)/N_j \downarrow 0$, $\zeta_s(M(j))^6L^4(j)/N_j \downarrow 0$, and $\zeta_1(M(j))^4\zeta_r(L(j))^4/N_j \downarrow 0$ and also $\underline{C}(L(j)/N_j + L(j)^{1-2\delta_1}) \le \upsilon_N^3 \le \overline{C}(L(j)/N_j + L(j)^{1-2\delta_1})$ holds for some $0 < \underline{C} < \overline{C}$. Also let Ω be diagonal where all the diagonal elements are positive. Then under Assumptions 4.1 and 5.2*

$$\liminf_{N\to\infty} \inf_{(F_1,...,F_J)\in\mathcal{F}_{\mathrm{EC}}} \Pr\{J_{\mathrm{EC}_N} \le \hat{c}_{1-\alpha}\} = 1 - \alpha$$

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

6. Empirical Application

We apply our methods to data from the U.K. Family Expenditure Survey, the same data used by BBC and others. Our testing of a RUM can, therefore, be compared with their revealed preference analysis of a representative consumer. To facilitate comparison of results, we use the exact same selection from these data as they do, namely the time periods from 1975 through 1999 and households with a car and at least one child. The number of data points used varies from 715 (in 1997) to 1509 (in 1975), for a total of 26341. For each year, we extract the budget corresponding to that year's median expenditure and estimate the distribution of demand on that budget by the method outlined in section 5.1 using polynomials of order 3. Like BBC, we assume that all consumers in one year face the same prices, and we use the same price data. While budgets have a tendency to move outward over time, we find that there is substantial overlap of budgets at median expenditure. To account for endogenous expenditure, we use tools from section 5.2 with total household income as instrument. This is also the same instrument used in BBC (2008).

We present results for blocks of eight consecutive periods and the same three composite goods (food, nondurable consumption goods, and services) considered in BBC. For all blocks of seven consecutive years, we analyze the same basket but also increase the dimensionality of commodity space to 4 or even 5. This is done by first splitting nondurables into clothing and other nondurables and then further into clothing, alcoholic beverages, and other nondurables. Thus, the separability assumptions that we (and others) implicitly invoke are successively relaxed. We are able to go further than much of the existing literature in this regard because, while computational expense increases with K,[17] our approach is not subject to a statistical curse of dimensionality,

Regarding the test's statistical power, increasing the dimensionality of commodity space can cut both ways. The number of rationality constraints that we test goes up, which helps if some of the new constraints are violated but adds noise otherwise. Also, the maintained assumptions become weaker: In principle, a rejection of stochastic rationalizability at 3 but not 4 goods might just indicate a failure of separability. In practice, we fail to reject stochastic rationalizability for any combination of time periods and number of goods.

[17]Tables 1 and 2 were computed in a few days on Cornells ECCO cluster (32 nodes). An individual cell of a table can be computed in reasonable time on any desktop computer. Computation of a matrix A took up to one hour and computation of one J_N about five seconds on a laptop.

	3 goods				4 goods				5 goods			
	I	**H**	$\mathbf{J}_N$	**p**	**I**	**H**	$\mathbf{J}_N$	**p**	**I**	**H**	$\mathbf{J}_N$	**p**
75-81	36	6409	3.67	.38	52	39957	5.43	.29	55	53816	4.75	.24
76-82	39	4209	11.6	.14	65	82507	5.75	.39	65	82507	5.34	.31
77-83	41	7137	9.81	.17	65	100728	6.07	.39	68	133746	4.66	.38
78-84	32	3358	7.38	.24	62	85888	2.14	.70	67	116348	1.45	.71
79-85	35	5628	.114	.96	71	202686	.326	.92	79	313440	.219	.94
80-86	38	7104	.0055	.998	58	68738	1.70	.81	66	123462	7.91	.21
81-87	26	713	.0007	.998	42	9621	.640	.89	52	28089	6.33	.27
82-88	15	42	0	1	21	177	.298	.60	31	1283	9.38	.14
83-89	13	14	0	1	15	31	.263	.49	15	31	9.72	.13
84-90	15	42	0	1	15	42	.251	.74	15	42	10.25	.24
85-91	15	63	.062	.77	19	195	3.59	.45	21	331	3.59	.44
86-92	24	413	1.92	.71	33	1859	7.27	.35	35	3739	9.46	.28
87-93	45	17880	1.33	.74	57	52316	6.60	.44	70	153388	6.32	.38
88-94	39	4153	1.44	.70	67	136823	6.95	.38	77	313289	6.91	.38
89-95	26	840	.042	.97	69	134323	4.89	.35	78	336467	5.84	.31
90-96	19	120	.040	.95	56	52036	4.42	.19	76	272233	3.55	.25
91-97	17	84	.039	.93	40	7379	3.32	.26	50	19000	3.27	.24
92-98	13	21	.041	.97	26	897	.060	.93	26	897	.011	.99
93-99	9	3	.037	.66	15	63	0	1	15	63	0	1

TABLE 1. Empirical results with 7 periods. I = number of patches, H = number of rationalizable discrete demand vectors, J_N = test statistic, p = p-value.

As a reminder, Figure 2 illustrates the application. The budget is the 1993 one as embedded in the 1986-1993 block of periods, i.e. the Figure corresponds to a row of Table 2. All six budgets from 1987 through 1992 intersect the 1993 one.

Tables 1 and 2 summarize our empirical findings. They display test statistics, p-values, and the numbers I of patches and H of rationalizable demand vectors; thus, matrices A are of size $(I \times H)$. All entries that show $J_N = 0$ and a corresponding p-value of 1 were verified to be true zeroes, i.e. $\hat{\pi}_{EC}$ is rationalizable. All in all, it turns out that estimated choice probabilities are typically not stochastically rationalizable, but also that this rejection is not statistically significant.

	3 goods			
	I	**H**	**J$_N$**	**p**
75-82	51	71853	11.4	.17
76-83	64	114550	9.66	.24
77-84	52	57666	9.85	.20
78-85	49	76746	7.52	.26
79-86	55	112449	.114	.998
80-87	41	13206	3.58	.58
81-88	27	713	0	1
82-89	16	42	0	1
83-90	16	42	0	1
84-91	20	294	.072	.89
85-92	27	1239	2.24	.68
86-93	46	17880	1.54	.75
87-94	48	39913	1.55	.75
88-95	42	12459	1.68	.70
89-96	27	840	.047	.97
90-96	24	441	.389	.83
91-98	22	258	1.27	.52
92-99	14	21	.047	.96

TABLE 2. Empirical results with 8 periods. I = number of patches, H = number of rationalizable discrete demand vectors, J_N = test statistic, p = p-value.

Among many other validation exercises, we manually inspected data underlying the 84-91 entry in Table 2 because J_N is rather low. The entry is not in error and illustrates an interesting phenomenon. The matrix X (see footnote 7) corresponding to this entry contains the following three rows:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 0 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 & -1 & -1 & 0 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & 0 \end{bmatrix}$$

The corresponding entries of $\hat{\pi}$ are $(0.107, 0, 0.899)$. This violates the stochastic implications of WARP (see Example 3.1) between budgets $\mathcal{B}_5$ and $\mathcal{B}_8$ because the estimated probabilities for the displayed patches add to more than 1. However, the violation is very small and not significant. This phenomenon occurs frequently and may partly cause the many positive but insignificant values of J_N. If two budgets are slight rotations of each other and demand distributions change continuously in response, then population probabilities for the relevant patches of these two budgets will sum to just less than 1. If these probabilities are estimated independently across budgets, the sample analog of the sum will frequently (just) exceed 1. With 7 or 8 mutually intersecting budgets, there are many opportunities for such reversals. The test statistic is then likely to be positive, but our bootstrap procedure accounts for this.[18]

The phenomenon of estimated choice frequencies typically not being rationalizable means that there is need for a statistical testing theory and also a theory of rationality constrained estimation. The former is this paper's main contribution. We leave the latter for future research.

7. Further Applications and Extensions

7.1. **Partial Identification of Counterfactual Choices.** The toolkit developed in this paper is also useful for counterfactual analysis. For exposition, abstract from estimation issues and take a rationalizable π to be known. Then to bound the value of any function $f(\nu)$ subject to the constraint that ν rationalizes π, solve the program

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

Empirically, one will have to replave π with $\hat{\eta}$ to ensure feasibility of the program; of course, $\hat{\eta} = \hat{\pi}$ whenever $\hat{\pi}$ is rationalizable.

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. 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 BBC 2008, Adams (2016), Manski (2007, 2014), and others.

[18]We also checked whether small but positive test statistics are caused by adding-up constraints, i.e. by the fact that all components of $\hat{\pi}$ that correspond to one budget must add to the same sum across budgets. The estimator $\hat{\pi}$ can slightly violate this because we force it to be inside $[0,1]^I$. Adding-up failures occur but are at least one order of magnitude smaller than the distance from a typical $\hat{\pi}$ to the corresponding projection $\hat{\eta}$.

In the remainder of this this subsection only, assume that demand on budget $\mathcal{B}_J$ is not observed but is to be bounded. Write

$$A = \begin{bmatrix} A_{-J} \\ A_J \end{bmatrix}, \pi = \begin{bmatrix} \pi_{-J} \\ \pi_J \end{bmatrix}, \pi_J = \begin{bmatrix} \pi_{1|J} \\ \vdots \\ \pi_{I_J|J} \end{bmatrix}$$

and let e_i be the i-th unit vector. We begin by bounding components of π_J. This is not of immediate interest in this paper's empirical application, but will be if the true choice universe is discrete.

Corollary 7.1. *$\pi_{i|J}$ is bounded by*

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

where

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

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

$$\begin{aligned} \underline{d}_k(J) &: = [\underline{d}_k(1|J), ..., \underline{d}_k(I_J|J)] \\ \overline{d}_k(J) &: = [\overline{d}_k(1|J), ..., \overline{d}_k(I_J|J)] \end{aligned}$$

with components

$$\begin{aligned} \underline{d}_k(i|J) &: = \min\{y_k : y \in x_{i|J}\}, \quad 1 \leq i \leq I_J \\ \overline{d}_k(i|J) &: = \max\{y_k : y \in x_{i|J}\}, \quad 1 \leq i \leq I_J, \end{aligned}$$

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

Corollary 7.2. *Expected demand for good k on budget $\mathcal{B}_J$ is bounded by*

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

where

$$\begin{aligned} \underline{\delta}_k(J) &: = \min\{\underline{d}_k(J) A_J \nu\} \quad s.t. \quad A_{-J}\nu = \pi_{-J}, \quad \nu \geq 0 \\ \overline{\delta}_k(J) &: = \max\{\overline{d}_k(J) A_J \nu\} \quad s.t. \quad A_{-J}\nu = \pi_{-J}, \quad \nu \geq 0. \end{aligned}$$

Finally, consider bounding the c.d.f. $F_k(z) = \Pr(y_k \leq z)$. This quantity must be bounded in two steps. The event $(y_k \leq z)$ will in general not correspond to a precise set of patches, that is, it is not measurable with respect to (the algebra generated by) $\mathcal{X}$. 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. $y_k \leq z$ 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. $y_k \leq z$ holds for *all* $y \in x_{i|J}$.[19] We thus have:

Corollary 7.3. *For $k = 1, ..., K$ and $z \geq 0$, $F_k(z)$ is bounded from below by*

$$\min_{\nu \in \mathbf{R}_+^H} \{ \sum_{\substack{i \in \{1,\ldots,I_J\}: \\ \overline{d}_k(i|J) \leq z}} e_i' A_J \nu \} \quad s.t. \quad A_{-J}\nu \;=\; \pi_{-J}$$

and from above by

$$\max_{\nu \in \mathbf{R}_+^H} \{ \sum_{\substack{i \in \{1,\ldots,I_J\}: \\ \underline{d}_k(i|J) \leq z}} e_i' A_J \nu \} \quad s.t. \quad A_{-J}\nu \;=\; \pi_{-J},$$

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 will 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.[20] In his recent analysis of optimal taxation of labor, Manski (2014) solves programs like this to find informative bounds.

7.2. **More General Choice Problems.** The methods developed in this paper, including the extension to counterfactual choice, immediately apply whenever the universal choice space is finite or can be made finite by an argument similar to Proposition 3.1. We already mentioned Deb, Kitamura, Quah, and Stoye (2016) as an example of the latter. To further illustrate the point, we briefly elaborate on a setting that has received some attention by theorists, namely random choice from binary sets. Thus, choice probabilities for pairs of options,

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

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

[20] An adaptation of this paper's inference method is in progress. Adaptation of inference procedures in Bugni, Canay, and Shi (2016) and Kaido, Molinari, and Stoye (2016) could be interesting as well, but both would have to overcome the absence of a $\mathcal{H}$-representation of $\mathcal{C}$.

are observed for 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 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

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

If choice probabilities for all pairs are observed, then 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 our toolkit allows to numerically resolve the question in complicated cases and also to perform a statistical test that applies whenever probabilities are estimated. To see this, let $J \leq (\#\mathcal{A})(\#\mathcal{A}-1)/2$ "budgets" be the number of distinct pairs $a,b \in \mathcal{A}$ for whom π_{ab} is observed, 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 $\mathcal{A}$. Computation of A and all other steps then work as before.

To illustrate, let $\mathcal{A} = \{a,b,c\}$ and assume that choice probabilities for all three pairs are observable, then

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

and it is readily verified that $A\nu = \pi$ for some $\nu \in \Delta^5$ iff both $\pi_{ab}+\pi_{bc}+\pi_{ca} \leq 2$ and $\pi_{cb}+\pi_{ba}+\pi_{ac} \leq 2$, confirming sufficiency of the triangle condition.

8. Conclusion

This paper presented asymptotic theory and computational tools for nonparametric testing of Random Utility Models. Again, the null to be tested was that data was 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 testing the model is nonetheless possible. 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, and some of these have already been explored. We close by mentioning further salient issues.

(1) We provide algorithms (and code) that work for reasonably sized problem, but it would be extremely useful to make further improvements in this dimension.

(2) The extension to infinitely many budgets would be of obvious interest. Theoretically, it 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 $J_N^{\max} := \max_{1 \le m \le M} J_N^m$ or a weighted average of them. 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. This requires incorporating nonparametric smoothing in estimating choice probabilities as in Section 5.1, 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. Existing proposals for specification testing in moment inequality models (Andrews and Guggenberger (2009), Andrews and Soares (2010), Bugni, Canay, and Shi (2015), Romano and Shaikh (2010)) use a similar test statistic but work with

$\mathcal{H}$-representations. In settings in which theoretical restrictions inform a $\mathcal{V}$-representation of a cone, the $\mathcal{H}$-representation will typically not be available in practice. We expect that our method can be used in many such cases.

9. Appendix A: Proofs

Proof of Proposition 3.1. Call two demand vectors $(d_1, ..., d_J)$ and $(d'_1, ..., d'_J)$ *equivalent* if they select from the same patches, i.e. $d_j \in x_i \Leftrightarrow d'_j \in x_i$. Then for all j and k, $p'_j d_j \geq p'_j d_k \Leftrightarrow p'_j d'_j \geq p'_j d'_k$. Thus, either both or none of any two equivalent vectors are rationalizable. Next, if $(P_1, ..., P_J)$ is rationalizable, then there exists a distribution P_d that rationalizes it; see Remark 3.2. Let P_d^* be the distribution induced by drawing d from P_d and replacing it with the equivalent d^* that only contains elements of $\mathcal{Y}^*$. (By construction of $\mathcal{Y}^*$, this d^* exists and is unique.) Then P_d^* stochastically rationalizes $(P_1^*, ..., P_J^*)$. The converse reasoning holds true as well. This establishes the special case; the general case follows by applying the special case twice. □

Proof of Proposition 3.2. By Proposition 3.1, $(P_1, ..., P_J)$ is rationalizable iff $(P_1^*, ..., P_J^*)$ is. By remark 3.2, $(P_1^*, ..., P_J^*)$ is rationalizable iff it is induced by a distribution P_d. Because $(P_1^*, ..., P_J^*)$ is concentrated on $\mathcal{Y}^*$, any d in the support of P_d has to be concentrated on $\mathcal{Y}^*$ as well.

By Remark 3.4, the number H of distinct nonstochastic demand systems d^* concentrated on $\mathcal{Y}^*$ is finite. Endow them with an arbitrary but henceforth fixed ordering. Then P_d can be identified with a vector $\nu \in \Delta^{H-1}$.

Consider now $P_j^*(D(p_j, u) = y_i^*)$. Any given d^* either picks y_i^* from $\mathcal{B}_j$ or not; the corresponding component of its vector representation is an indicator of this event. Therefore, $P_j^*(D(p_j, u) = y_i^*)$ is the inner product of ν with the vector that lists all these components in corresponding order. Applying this reasoning to each component of the vector representation of $(\mathcal{B}_1, ..., \mathcal{B}_J)$, one sees that $\pi = A\nu$, where A collects the vector representations of rationalizable demand systems d^* in that same order. □

Proof of Proposition 3.3. We begin with some preliminary observations. Throughout this proof, $c(\mathcal{B}_i)$ denotes the object actually chosen from budget $\mathcal{B}_i$.

(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 $c(\mathcal{B}_i) \succ c(\mathcal{B}_j) \succ c(\mathcal{B}_k) \succ ... \succ c(\mathcal{B}_i)$. If $c(\mathcal{B}_k) \succ c(\mathcal{B}_i)$, then a length 3 cycle has been discovered. Else, there exists a length $N-1$ choice cycle $c(\mathcal{B}_i) \succ c(\mathcal{B}_k) \succ ... \succ c(\mathcal{B}_i)$. 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 $\mathcal{B}_1$ is below $\mathcal{B}_3$, thus $c(\mathcal{B}_3) \succ c(\mathcal{B}_1)$ by monotonicity. If there is a choice cycle, we must have $c(\mathcal{B}_1) \succ c(\mathcal{B}_2) \succ c(\mathcal{B}_3)$. $c(\mathcal{B}_1) \succ c(\mathcal{B}_2)$ implies that $c(\mathcal{B}_2)$ is below $\mathcal{B}_1$, thus it is below $\mathcal{B}_3$. $c(\mathcal{B}_2) \succ c(\mathcal{B}_3)$ implies that $c(\mathcal{B}_3)$ is below $\mathcal{B}_2$. Thus,choice from $(\mathcal{B}_2, \mathcal{B}_3)$ violates WARP.

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 $(\mathcal{B}_1, ..., \mathcal{B}_{J-1})$ is rationalizable but choice from $(\mathcal{B}_1, ..., \mathcal{B}_J)$ is not, then choice from $(\mathcal{B}_{M+1}, ..., \mathcal{B}_J)$ cannot be rationalizable. By observation (ii), if $(\mathcal{B}_1, ..., \mathcal{B}_J)$ is not rationalizable, it contains either a 2-cycle or an irreducible 3-cycle. Because choice from all triplets within $(\mathcal{B}_1, ..., \mathcal{B}_{J-1})$ is rationalizable by assumption, it is either the case that some $(\mathcal{B}_i, \mathcal{B}_J)$ constitutes a 2-cycle or that some triplet $(\mathcal{B}_i, \mathcal{B}_k, \mathcal{B}_J)$, where $i < k$ w.l.o.g., reveals an irreducible choice cycle. In the former case, $\mathcal{B}_i$ must intersect $\mathcal{B}_J$, hence $i > M$, hence the conclusion. In the latter case, if $k \leq M$, the choice cycle must be a 2-cycle in $(\mathcal{B}_i, \mathcal{B}_k)$, contradicting rationalizability of $(\mathcal{B}_1, ..., \mathcal{B}_{J-1})$. If $i \leq M$, the choice cycle is reducible by (iii). Thus, $i > M$, hence the conclusion. $\square$

Proof of Lemma 4.1. Letting $\nu_\tau = \nu - (\tau/H)\mathbf{1}_H$ in $\mathcal{C}_\tau = \{A\nu | \nu \geq (\tau/H)\mathbf{1}_H\}$ we have

$$\begin{aligned} \mathcal{C}_\tau &= \{A[\nu_\tau + (\tau/H)\mathbf{1}_H] | \nu_\tau \geq 0\} \\ &= \mathcal{C} \oplus (\tau/H)A\mathbf{1}_H \\ &= \{t : t - (\tau/H)A\mathbf{1}_H \in \mathcal{C}\} \end{aligned}$$

where $\oplus$ signifies Minkowski sum. Define

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

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

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

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 $\mathcal{C}$ implies that $Ae_h \in \mathcal{C}$, and thus

$$BAe_h \leq 0$$

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

$$\Phi_{kh} = -e_k' BAe_h \geq 0, \quad 1 \leq k \leq m, 1 \leq h \leq H. \tag{9.1}$$

But if $k \leq \bar{m}$, it cannot be that

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

whereas

$$b_k' a_h = 0$$

holds for $\bar{m} + 1 \leq k \leq m, 1 \leq h \leq H$. Therefore if $k \leq \bar{m}$, $\Phi_{kh} = b_k' a_h$ is nonzero at least for one $h, 1 \leq h \leq H$, whereas if $k > \bar{m}$, $\Phi_{kh} = 0$ for every h. Since (9.1) implies that all of $\{\Phi_{kh}\}_{h=1}^H$ are non-negative, we conclude that

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

for every $k \leq \bar{m}$ and $\phi_k = 0$ for every $k > \bar{m}$. We now have

$$\mathcal{C}_\tau = \{t : Bt \leq -\tau\phi\}$$

where ϕ satisfies the stated properties *(i)* and *(ii)*. □

Proof of Theorem 4.2. By applying the Minkowski-Weyl theorem and Lemma 4.1 to J_N and $\tilde{J}_N(\tau_N)$, we see that our procedure is equivalent to comparing

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

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

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

with $\phi = [\bar{\phi}', (0, ..., 0)']'$, $\bar{\phi} \in \mathbf{R}^{\bar{m}}_{++}$, where

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

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

Suppose B has m rows and $\text{rank}(B) = \ell$. Define an $\ell \times m$ 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

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

Let

$$\mathcal{U}_1 = \left\{ \begin{pmatrix} K\gamma \\ h \end{pmatrix} : \gamma = Bt, h = Mt, B^{\leq} t \leq 0, B^{=} t = 0, t \in \mathbf{R}^I \right\}$$

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

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

Also define

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

where $\text{col}(B)$ 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 $\gamma^* = Bt^*$. Define

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

then $Bt^{**} = Bt^* = \gamma^*$, therefore $B^{\leq} t^{**} \leq 0$ and $B^{=} t^{**} = 0$. Also, $Mt^{**} = Mt^* + MM'h^* - MM'Mt^* = h^*$, therefore $\binom{K\gamma^*}{h^*}$ is an element of $\mathcal{U}_1$ as well. Consequently,

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

We now have

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

Define

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

and

$$t(x):=\min_{y\in\mathbf{R}^{I-\ell}} T(x,y), \quad s(g):=\min_{\gamma=[\gamma^{\leq\prime},\gamma^{=\prime}]',\gamma^{\leq}\leq 0,\gamma^{=}=0,\gamma\in\mathrm{col}(B)}.$$

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

$$\begin{aligned} J_N &= N\min_{\gamma=[\gamma^{\leq\prime},\gamma^{=\prime}]',\gamma^{\leq}\leq 0,\gamma^{=}=0,\gamma\in\mathrm{col}(B)} t(K[\hat{g}-\gamma]) \\ &= Ns(\hat{g}) \\ &= s(\sqrt{N}\hat{g}). \end{aligned}$$

We now show that tightening can turn non-binding inequality constraints into binding ones but not vice versa. Note that, as will be seen below, this observation uses diagonality of Ω and the specific geometry of the cone $\mathcal{C}$. Let $\hat{\gamma}^k_{\tau_N}$, $\hat{g}^k$ and ϕ^k denote the k-th elements of $\hat{\gamma}_{\tau_N}=B\hat{\eta}_{\tau_N}$, $\hat{g}$ and ϕ. Moreover, define $\gamma_\tau(g)=[\gamma^1(g),...,\gamma^m(g)]'=\mathrm{argmin}_{\gamma=[\gamma^{\leq\prime},\gamma^{=\prime}]',\gamma^{\leq}\leq-\tau\bar{\phi},\gamma^{=}=0,\gamma\in\mathrm{col}(B)} t(K[g-\gamma])$ for $g\in\mathrm{col}(B)$, and let $\gamma^k_\tau(g)$ be its k-th element. Then $\hat{\gamma}_{\tau_N}=\gamma_{\tau_N}(\hat{g})$. Finally, define $\beta_\tau(g)=\gamma_\tau(g)+\tau\phi$ for $\tau>0$ and let $\beta^k_\tau(g)$ denote its k-th element. Note $\gamma^k_\tau(g)=\phi^k=\beta^k_\tau(g)=0$ for every $k>\bar{m}$ and g. Now we show that for each $k\leq\bar{m}$ and for some $\delta>0$,

$$\beta^k_\tau(g)=0$$

if $|g^k|\leq\tau\delta$ and $g^j\leq\tau\delta, 1\leq j\leq\bar{m}$. In what follows we first show this for the case with $\Omega=\mathbf{I}_I$, where $\mathbf{I}_I$ denotes the I-dimensional identity matrix, then generalize the result to the case where Ω can have arbitrary positive diagonal elements.

For $\tau>0$ and $\delta>0$ define hyperplanes

$$H^\tau_k=\{x:b'_k x=-\tau\phi^k\},$$

$$H_k=\{x:b'_k x=0\},$$

half spaces

$$H^\tau_{\angle k}(\delta)=\{x:b'_k x\leq\tau\delta\},$$

and also

$$S_k(\delta) = \{x \in \mathcal{C} : |b_k' x| \le \tau\delta\}$$

for $1 \le k \le m$. Define

$$L = \cap_{k=\bar{m}+1}^{m} H_k,$$

a linear subspace of $\mathbf{R}^I$. In what follows we show that for small enough $\delta > 0$, every element $x^* \in \mathbf{R}^I$ such that

$$x^* \in S_1(\delta) \cap \cdots \cap S_q(\delta) \cap H^\tau_{\angle q+1}(\delta) \cap \cdots H^\tau_{\angle m}(\delta) \text{ for some } q \in \{1, ..., \bar{m}\} \tag{9.2}$$

satisfies

$$x^* | \mathcal{C}_\tau \in H_1^\tau \cap \cdots \cap H_q^\tau \cap L \tag{9.3}$$

where $x^*|\mathcal{C}_\tau$ denotes the orthogonal projection of x^* on $\mathcal{C}_\tau$. Let $g^{*k} = b_k' x^*, k = 1, ..., m$. Note that an element x^* fulfils (9.2) iff $|g^{*k}| \le \tau\delta, 1 \le k \le q$ and $g^{*j} \le \tau\delta, q+1 \le j \le \bar{m}$. Likewise, (9.3) holds iff $\beta_k^\tau(g^*) = 0, 1 \le k \le q$ (recall $\beta_k^\tau(g^*) = 0$ always holds for $k > \bar{m}$). Thus in order to establish the desired property of the function $\beta_\tau(\cdot)$, we show that (9.2) implies (9.3). Suppose it does not hold; then without loss of generality, for an element x^* that satisfies (9.2) for an arbitrary small $\delta > 0$, we have

$$x^*|\mathcal{C}_\tau \in H_1^\tau \cap \cdots \cap H_r^\tau \cap L \quad \text{and} \quad x^*|\mathcal{C}_\tau \notin H_j^\tau, r+1 \le j \le q \tag{9.4}$$

for some $1 \le r \le q-1$. Define halfspaces

$$H^\tau_{\angle k} = \{x : b_k' x \le -\tau\phi^k\},$$

$$H_{\angle k} = \{x : b_k' x \le 0\}$$

for $1 \le k \le m, \tau > 0$ and also let

$$F = H_1 \cap \cdots \cap H_r \cap L \cap \mathcal{C},$$

then for (9.4) to hold for some $x^* \in \mathbf{R}^I$ satisfying (9.2) for an arbitrary small $\delta > 0$ we must have

$$F|\,(H_1^\tau \cap \cdots \cap H_r^\tau \cap L) \subset \mathrm{int}(H^\tau_{\angle r+1} \cap \cdots \cap H^\tau_{\angle q})$$

(Recall the notation | signifies orthogonal projection. Also note that if $\dim(F) = 1$, then (9.4) does not occur under (9.2).) Therefore if we let

$$\Delta(J) = \{x \in \mathbf{R}^I : \mathbf{1}_I' x = J, x \ge 0\},$$

i.e. the simplex with vertices $(J, 0, \cdots, 0), \cdots, (0, \cdots, 0, J)$, we have

$$(F \cap \Delta(J)) \,|\, (H_1^\tau \cap \cdots \cap H_r^\tau \cap L) \subset \mathrm{int}(H_{\angle r+1}^\tau \cap \cdots \cap H_{\angle q}^\tau). \tag{9.5}$$

Let $\{a_1, ..., a_H\} = \mathcal{A}$ denote the collection of the column vectors of A. Then $\{$the vertices of $F \cap \Delta(J)\} \in \mathcal{A}$. Let $\bar{a}, \bar{\bar{a}} \in F \cap \Delta(J)$. Let $B(\varepsilon, x)$ denote the ε-(open) ball with center $x \in \mathbf{R}^I$. By (9.5),

$$B\left(\varepsilon, \left(\bar{a} | \cap_{j=1}^r H_j^\tau \cap L\right)\right) \subset \mathrm{int}(H_{\angle r+1}^\tau \cap \cdots \cap H_{\angle q}^\tau) \cap H_{\angle 1} \cap \cdots \cap H_{\angle r}$$

holds for small enough $\varepsilon > 0$. Let $\bar{a}^\tau := \bar{a} + \tau$, $\bar{\bar{a}}^\tau := \bar{\bar{a}} + \tau$, then

$$\begin{aligned} \left(\left(\bar{a}|(\cap_{j=1}^r H_j^\tau) \cap L\right) - \bar{a}\right)' (\bar{\bar{a}} - \bar{a}) &= \left(\left(\bar{a}|(\cap_{j=1}^r H_j^\tau) \cap L\right) - \bar{a}\right)' (\bar{\bar{a}}^\tau - \bar{a}^\tau) \\ &= 0 \end{aligned}$$

since $\bar{a}^\tau, \bar{\bar{a}}^\tau \in (\cap_{j=1}^r H_j^\tau) \cap L$. We can then take $z \in B\left(\varepsilon, \left(\bar{a}|(\cap_{j=1}^r H_j^\tau) \cap L\right)\right)$ such that $(z - \bar{a})'(\bar{\bar{a}} - \bar{a}) < 0$. By construction $z \in \mathcal{C}$, which implies the existence of a triplet $(a, \bar{a}, \bar{\bar{a}})$ of distinct elements in $\mathcal{A}$ such that $(a - \bar{a})'(\bar{\bar{a}} - \bar{a}) < 0$. In what follows we show that this cannot happen, then the desired property of β_τ is established.

So let us now show that

$$(a_1 - a_0)'(a_2 - a_0) \geq 0 \text{ for every triplet } (a_0, a_1, a_2) \text{ of distinct elements in } \mathcal{A}. \tag{9.6}$$

Noting that $a_i' a_j$ just counts the number of budgets on which i and j agree, define

$$\phi(a_i, a_j) = J - a_i' a_j,$$

the number of disagreements. Importantly, note that $\phi(a_i, a_j) = \phi(a_j, a_i)$ and that ϕ is a distance (it is the taxicab distance between elements in $\mathcal{A}$, which are all 0-1 vectors). Now

$$\begin{aligned} &(a_1 - a_0)'(a_2 - a_0) \\ &= a_1' a_2 - a_0' a_2 - a_1' a_0 + a_0' a_0 \\ &= J - \phi(a_1, a_2) - (J - \phi(a_0, a_2)) - (J - \phi(a_0, a_1)) + J \\ &= \phi(a_0, a_2) + \phi(a_0, a_1) - \phi(a_1, a_2) \geq 0 \end{aligned}$$

by the triangle inequality.

Next we treat the case where Ω is not necessarily $\mathbf{I}_I$. Write

$$\Omega = \begin{bmatrix} \omega_1^2 & 0 & \dots & 0 \\ 0 & \omega_2^2 & \dots & 0 \\ & & \ddots & \\ 0 & \dots & 0 & \omega_I^2 \end{bmatrix}.$$

The statistic J_N in (4.1) can be rewritten, using the square-root matrix $\Omega^{1/2}$,

$$J_N = \min_{\eta^*=\Omega^{1/2}\eta:\eta\in C} [\hat{\pi}^* - \eta^*]'[\hat{\pi}^* - \eta^*]$$

or

$$J_N = \min_{\eta^*\in C^*} [\hat{\pi}^* - \eta^*]'[\hat{\pi}^* - \eta^*]$$

where

$$\begin{aligned} \mathcal{C}^* &= \{\Omega^{1/2} A\nu | \nu \geq 0\} \\ &= \{A^*\nu | \nu \geq 0\} \end{aligned}$$

with

$$A^* = [a_1^*, ..., a_H^*], a_h^* = \Omega^{1/2} a_h, 1 \leq h \leq H.$$

Then we can follow our previous argument replacing a's with a^*'s, and using

$$\Delta^*(J) = \text{conv}([0, ..., \omega_i, ...0]' \in \mathbf{R}^I, i = 1, ..., I).$$

instead of the simplex $\Delta(J)$. Finally, we need to verify that the acuteness condition (9.6) holds for $\mathcal{A}^* = \{a_1^*, ..., a_H^*\}$.

For two I-vectors a and b, define a weighted taxicab metric

$$\phi_\Omega(a, b) := \sum_{i=1}^{I} \omega_i |a_i - b_i|,$$

then the standard taxicab metric ϕ used above is ϕ_Ω with $\Omega = \mathbf{I}_I$. Moreover, letting $a^* = \Omega^{1/2} a$ and $b^* = \Omega^{1/2} b$, where each of a and b is an I-dimensional 0-1 vector, we have

$$a^{*\prime} b^* = \sum_{i=1}^{I} \omega_i [1 - |a_i - b_i|] = \bar{\omega} - \phi_\Omega(a, b)$$

with $\bar{\omega} = \sum_{i=1}^{I} \omega_i$. Then for every triplet (a_0^*, a_1^*, a_2^*) of distinct elements in $\mathcal{A}^*$

$$\begin{aligned}(a_1^* - a_0^*)'(a_2^* - a_0^*) &= \bar{\omega} - \phi_\Omega(a_1, a_2) - \bar{\omega} + \phi_\Omega(a_0, a_2) - \bar{\omega} + \phi_\Omega(a_0, a_1) + \bar{\omega} - \phi_\Omega(a_0, a_0) \\ &= \phi_\Omega(a_1, a_2) - \phi_\Omega(a_0, a_2) - \phi_\Omega(a_0, a_1) \\ &\geq 0,\end{aligned}$$

which is the desired acuteness condition. Since J_N can be written as the minimum of the quadratic form with identity-matrix weighting subject to the cone generated by a^*'s, all the previous arguments developed for the case with $\Omega = \mathbf{I}_I$ remain valid.

Defining $\xi \sim \mathrm{N}(0, \hat{S})$ and $\zeta = B\xi$,

$$\begin{aligned}\tilde{J}_N(\tau_N) &\sim \min_{Bt \leq -\tau_N \phi} N \left[\begin{pmatrix} KB \\ M \end{pmatrix} (\hat{\eta}_{\tau_N} + N^{-1/2}\xi - t) \right]' P^{-1\prime} \Omega P^{-1} \left[\begin{pmatrix} KB \\ M \end{pmatrix} (\hat{\eta}_{\tau_N} + N^{-1/2}\xi - t) \right] \\ &= N \min_{\gamma = [\gamma^{\leq\prime}, \gamma^{=\prime}]', \gamma^{\leq} \leq -\tau_N \bar{\phi}, \gamma^{=} = 0, \gamma \in \mathrm{col}(B)} t\left(K \left[\hat{\gamma}_{\tau_N} + N^{-1/2}\zeta - \gamma \right] \right).\end{aligned}$$

Moreover, defining $\gamma^\tau = \gamma + \tau_N \phi$ in the above, and using the definitions of $\beta_\tau(\cdot)$ and $s(\cdot)$

$$\begin{aligned}\tilde{J}_N(\tau_N) &\sim N \min_{\gamma^\tau = [\gamma^{\tau \leq\prime}, \gamma^{\tau =\prime}]', \gamma^{\tau \leq} \leq 0, \gamma^{\tau =} = 0, \gamma^\tau \in \mathrm{col}(B)} t\left(K \left[\hat{\gamma}_{\tau_N} + \tau_N \phi + N^{-1/2}\zeta - \gamma^\tau \right] \right) \\ &= N \min_{\gamma^\tau = [\gamma^{\tau \leq\prime}, \gamma^{\tau =\prime}]', \gamma^{\tau \leq} \leq 0, \gamma^{\tau =} = 0, \gamma^\tau \in \mathrm{col}(B)} t\left(K \left[\gamma_{\tau_N}(\hat{g}) + \tau_N \phi + N^{-1/2}\zeta - \gamma^\tau \right] \right) \\ &= N \min_{\gamma^\tau = [\gamma^{\tau \leq\prime}, \gamma^{\tau =\prime}]', \gamma^{\tau \leq} \leq 0, \gamma^{\tau =} = 0, \gamma^\tau \in \mathrm{col}(B)} t\left(K \left[\beta_{\tau_N}(\hat{g}) + N^{-1/2}\zeta - \gamma^\tau \right] \right) \\ &= s\left(N^{1/2} \beta_{\tau_N}(\hat{g}) + \zeta \right)\end{aligned}$$

Let $\varphi_N(\xi) := N^{1/2}\beta_{\tau_N}(\tau_N \xi)$ for $\xi = (\xi_1, ..., \xi_m)' \in \mathrm{col}(B)$, then from the property of β_τ shown above, its k-th element φ_N^k for $k \leq \bar{m}$ satisfies

$$\varphi_N^k(\xi) = 0$$

if $|\xi^k| \leq \delta$ and $\xi^j \leq \delta, 1 \leq j \leq m$ for large enough N. Note $\varphi_N^k(\xi) = N^{1/2}\beta_N^k(\tau_N \xi) = 0$ for $k > \bar{m}$. Define $\hat{\xi} := \hat{g}/\tau_N$ and using the definition of φ_N, we write

$$\tilde{J}_N(\tau_N) \sim s\left(\varphi_{\tau_N}(\hat{\xi}) + \zeta \right). \tag{9.7}$$

Now we invoke Theorem 1 of Andrews and Soares (2010, AS henceforth). 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 their Assumptions 1-3 hold for the function s defined above if signs are adjusted appropriately as our formulae deal with negativity constraints, whereas AS work with positivity constraints. (Note that Assumption 1(b) does not apply here since we use a fixed weighting matrix.) The function

φ_N in (9.7) satisfies the properties of φ in AS used in their proof of Theorem 1. AS imposes a set of restrictions on the parameter space (see their Equation (2.2) on page 124). Their condition (2.2) (vii) is a Lyapounov condition for a triangular array CLT. Following AS, consider a sequence of distributions $\pi_N = [\pi'_{1N}, ..., \pi'_{JN}]', N = 1, 2, ...$ in $\mathcal{P} \cap \mathcal{C}$ 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. The Lyapounov condition holds for $b_k(j)' d_{j,n}, n = 1, ..., N(j)$ under π_N for $k \in \mathcal{K}^R$ for at least one j as Condition 4.1 is imposed for $\pi_N \in \mathcal{P}$. We do not impose Condition 4.1 for $k \in \mathcal{K}^D$, therefore it is possible that $\lim_{N\to\infty} \mathrm{var}_{\pi_{jN}}(b_k(j)' d_{j,n}) = 0$, even for every j, when $k \in \mathcal{K}^D$. Note, however, that: (i) The equality $b'_k \hat{\pi} \le 0$ holds by construction for every $k \in \mathcal{K}^D$ and therefore its behavior does not affect J_N; in particular whether $\mathrm{var}_{\pi_{jN}}(b_k(j)' d_{j,n})$ converges to zero or not does not matter; (ii) If $\mathrm{var}_{\pi_N}(b'_k d_n)$, $d_n := [d'_{1,n}, ..., d'_{J,N}]'$, converges to zero for some $k \in \mathcal{K}^D$, then $\sqrt{N} b'_k[\tilde{\eta}_{\tau_N} - \hat{\eta}_{\tau_N}] = o_p(1)$ and therefore its contribution to $\tilde{J}_N(\tau_N)$ is asymptotically negligible in the size calculation. The other conditions in AS10, namely (2.2)(i)-(vi), hold trivially. Finally, Assumptions GMS 2 and GMS 4 of AS10 are concerned with their thresholding parameter κ_N 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

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

□

Proof of Theorem 5.1. We begin by introducing some notation.

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

$$p_F^{(j)}(w) := E_F[d_{j,n(j)} | w_{n(j)} = w], \quad \pi_F^{(j)} = p_F^{(j)}(\underline{w}_j), \qquad \pi_F = [\pi_F^{(1)'}, ..., \pi_F^{(J)'}]'$$

and

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

Note that $\Sigma_F^{(j)}(w) = \mathrm{diag}\left(p_F^{(j)}(w)\right) - p_F^{(j)}(w) p_F^{(j)}(w)'$.

The proof mimics the proof of Theorem 4.2, except for the treatment of $\hat{\pi}$. Instead of the sequence $\pi_N, N = 1, 2, ...$ in $\mathcal{P} \cap \mathcal{C}$, consider a sequence of distributions $F_N = [F_{1N}, ..., F_{JN}], N = 1, 2, ...$ in $\mathcal{F}$ such that $\sqrt{N_j / K(j)} B^{(j)} \pi_{F_N}^{(j)} \to h_j, h_j \le 0, 1 \le j \le J$ as $N \to \infty$. Define $Q_{F_N}^{(j)} = \mathrm{E}_{F_N}[q^{K(j)}(w_{n(j)}) q^{K(j)}(w_{n(j)})']$ and $\Xi_{F_N}^{(j)} = \mathrm{E}_{F_N}[B^{(j)} \Sigma_{F_N}^{(j)}(w_{n(j)}) B^{(j)'} \otimes q^{K(j)}(w_{n(j)}) q^{K(j)}(w_{n(j)})']$, and let

$$V_{F_N}^{(j)} := [\mathbf{I}_m \otimes q^{K(j)}(\underline{w}_j)' {Q_{F_N}^{(j)}}^{-1}] \Xi_{F_N}^{(j)} [\mathbf{I}_m \otimes {Q_{F_N}^{(j)}}^{-1} q^{K(j)}(\underline{w}_j)]$$

and

$$V_{F_N} := \sum_{j=1}^{J} V_{F_N}^{(j)}.$$

Then by adapting the proof of Theorem 2 in Newey (1997) to the triangle array for the repeated crosssection setting, we obtain

$$\sqrt{N} V_{F_N}{}^{-\frac{1}{2}} B[\hat{\pi} - \pi_{F_N}] \overset{F_N}{\rightsquigarrow} N(0, \mathbf{I}_m).$$

The rest is the same as the proof of Theorem 4.2. □

Proof of Theorem 5.2. The proof follows the same steps as those in the proof of Theorem 4.2, except for the treatment of the estimator for π. Therefore, instead of the sequence $\pi_N, N = 1, 2, ...$ in $\mathcal{P} \cap \mathcal{C}$, consider a sequence of distributions $F_N = [F_{1N}, ..., F_{JN}], N = 1, 2, ...$ in $\mathcal{F}_{\mathrm{EC}}$ and the corresponding conditional distributions $P^{(j)}_{y|w,\epsilon;F_N}\{y \in x_{i|j}|w, \epsilon\}$ and $F^{(j)}_{w|z_N}$, $1 \leq i \leq I_j, 1 \leq j \leq J$, $N = 1, 2, ...$ such that $\sqrt{N_j/(M(j) \vee L(j)))} B^{(j)} \pi^{(j)}_{F_N} \to h_j, h_j \leq 0, 1 \leq j \leq J$ as $N \to \infty$, where $\pi_{F_N} = \pi(P^{(1)}_{y|w,\epsilon_N}, ..., P^{(J)}_{y|w,\epsilon_N})$ whereas the definitions of $\overline{V}^{(j)}_{F_N}, 1 \leq j \leq J$ are given shortly. Define $S^{(j)}_{F_N} = \mathrm{E}_{F_N}[s^{M(j)}(\chi_{n(j)}) s^{M(j)}(\chi_{n(j)})']$ as well as

$$\bar{\Xi}^{(j)}_{1_{F_N}} = \mathrm{E}_{F_N}[B^{(j)} \bar{\Sigma}^{(j)}_{F_N}(\chi_{n(j)}) B^{(j)\prime} \otimes s^{M(j)}(\chi_{n(j)}) s^{M(j)}(\chi_{n(j)})']$$

and

$$\bar{\Xi}^{(j)}_{2_{F_N}} = [B^{(j)} \otimes \mathbf{I}_{M(j)}] \mathrm{E}_{F_N}[m_{n(j);F_N} m'_{n(j);F_N}][B^{(j)\prime} \otimes \mathbf{I}_{M(j)}]$$

where

$$\Sigma^{(j)}_{F_N}(\chi) := \mathrm{Cov}_{F_N}[d_{j,n(j)}|\chi_{n(j)} = \chi],$$

$$m_{n(j);F_N} := [m'_{1,n(j);F_N}, m'_{2,n(j);F_N}, \cdots, m'_{I_j,n(j);F_N}]',$$

$$\begin{aligned} & m_{i,n(j);F_N} := \\ & \mathrm{E}_{F_N}\Bigg[\dot{\gamma}_N(\epsilon_{m(j)}) \frac{\partial}{\partial \epsilon} P^{(j)}_{y|w,\epsilon;F_N}\left\{y \in x_{i|j}|w_{m(j)}, \epsilon_{m(j)}\right\} s^{M(j)}(\chi_{m(j)}) r^{L(j)}(z_{m(j)})' R_{F_N}(j)^{-1} r^{L(j)}(z_{n(j)}) u_{mn(j);F_N} \\ & \qquad \Bigg| d_{i|j,n(j)}, w_{n(j)}, z_{n(j)} \Bigg], \end{aligned}$$

$$R_{F_N}(j) := \mathrm{E}_{F_N}[r^{L(j)}(z_{n(j)}) r^{L(j)}(z_{n(j)})'], \quad u_{mn(j);F_N} := \mathbf{1}\{w_{n(j)} \leq w_{m(j)}\} - F^{(j)}_{w|z_N}(w_{m(j)}|z_{n(j)}).$$

With these definitions, let

$$\overline{V}^{(j)}_{F_N} := \left[\mathbf{I}_m \otimes D(j)' S^{(j)}_{F_N}{}^{-1}\right] \bar{\Xi}^{(j)}_{F_N} \left[\mathbf{I}_m \otimes S^{(j)}_{F_N}{}^{-1} D(j)\right]$$

with $\bar{\Xi}^{(j)}_{F_N} = \bar{\Xi}^{(j)}_{1_{F_N}} + \bar{\Xi}^{(j)}_{2_{F_N}}$. Define

$$\overline{V}_{F_N} := \sum_{j=1}^{J} \overline{V}^{(j)}_{F_N}.$$

Then by adapting the proof of Theorem 7 in Imbens and Newey (2002) to the triangle array for the repeated crosssection setting, for the j's that satisfy Condition (iv) we obtain

$$\sqrt{N}\overline{V}_{F_N}{}^{-\frac{1}{2}} B[\hat{\pi} - \pi_{F_N}] \overset{F_N}{\rightsquigarrow} N(0, \mathbf{I}_m).$$

The rest is the same as the proof of Theorem 4.2. $\square$

10. Appendix B: Algorithms for Computing A

This appendix details 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. All implementations are in MATLAB using CVX 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. This works best in our implementation, but could also be a depth-first search.

Algorithms use notation introduced in the proof of Proposition 3.3.

Computing A as in Proposition 3.2.

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

2. `Initialize` $l = 2$.

3. `Set` $c(\mathcal{B}_1) = m_1, ..., c(\mathcal{B}_l) = m_l$. `FW-test` $(c(\mathcal{B}_1), ..., c(\mathcal{B}_l))$.

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.

Refinement using Proposition 3.3

Let budgets be arranged s.t. $(\mathcal{B}_1, ..., \mathcal{B}_M)$ do not intersect $\mathcal{B}_J$; for exposition of the algorithm, assume $\mathcal{B}_J$ 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 $(\mathcal{B}_{M+1}, ..., \mathcal{B}_J)$, though using the full X corresponding to budgets $(\mathcal{B}_1, ..., \mathcal{B}_J)$.[21]

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.

[21]This matrix has more rows than an A matrix that is only intended to apply to choice problems $(\mathcal{B}_{M+1}, ..., \mathcal{B}_J)$.

References

Adams, A. (2016): "Mutually consistent revealed preference bounds," Manuscript, University of Oxford.

Afriat, S. N. (1967): "The Construction of Utility Functions from Expenditure Data," *International Economic Review*, 8(1), 67–77.

Andrews, D. W. (1991): "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, 59, 817–858.

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

——— (2010): "Asymptotic size and a problem with subsampling and with the m out of n bootstrap," *Econometric Theory*, 26(02), 426–468.

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

Blundell, R., M. Browning, L. Cherchye, I. Crawford, B. De Rock, and F. Vermeulen (2015): "Sharp for SARP: Nonparametric Bounds on Counterfactual Demands," *American Economic Journal: Microeconomics*, 7(1), 43–60.

Blundell, R., M. Browning, and I. Crawford (2003): "Nonparametric Engel Curves and Revealed Preference," *Econometrica*, 71(1), 205–240.

——— (2007): "Improving Revealed Preference Bounds On Demand Responses," *International Economic Review*, 48(4), 1227–1244.

——— (2008): "Best Nonparametric Bounds on Demand Responses," *Econometrica*, 76(6), 1227–1262.

Blundell, R., D. Kristensen, and R. Matzkin (2014): "Bounding quantile demand functions using revealed preference inequalities," *Journal of Econometrics*, 179(2), 112 – 127.

Blundell, R., and J. L. Powell (2003): "Endogeneity in nonparametric and semiparametric regression models," *Econometric Society Monographs*, 36, 312–357.

Bugni, F. (2010): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, 78(2), 735–753.

Bugni, F. A., I. A. Canay, and X. Shi (2015): "Specification tests for partially identified models defined by moment inequalities," *Journal of Econometrics*, 185(1), 259 – 282.

——— (2016): "Inference for Subvectors and Other Functions of Partially Identified Parameters in Moment Inequality Models," *Quantitative Economics*, forthcoming.

Canay, I. A. (2010): "{EL} inference for partially identified models: Large deviations optimality and bootstrap validity," *Journal of Econometrics*, 156(2), 408 – 425.

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(5), 1243–1284.

Chesher, A. (2003): "Identification in nonseparable models," *Econometrica*, 71(5), 1405–1441.

DEB, R., Y. KITAMURA, J. QUAH, AND J. STOYE (2016): "Revealed Price Preference: Theory and Stochastic Testing," Manuscript, University of Toronto.

DETTE, H., S. HODERLEIN, AND N. NEUMEYER (2016): "Testing multivariate economic restrictions using quantiles: The example of Slutsky negative semidefiniteness," *Journal of Econometrics*, 191(1), 129–144.

FISHBURN, P. C. (1992): "Induced binary probabilities and the linear ordering polytope: a status report," *Mathematical Social Sciences*, 23(1), 67 – 80.

FLOYD, R. W. (1962): "Algorithm 97: Shortest Path," *Commun. ACM*, 5(6), 345–.

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(1), 63–80.

GRUBER, P. (2007): *Convex and Discrete Geometry*, Grundlehren der mathematischen Wissenschaften. New York: Springer.

GRÜNBAUM, B., V. KAIBEL, V. KLEE, AND G. M. ZIEGLER (2003): *Convex polytopes.* New York: Springer.

GUGGENBERGER, P., J. HAHN, AND K. KIM (2008): "Specification Testing under Moment Inequalities," *Econometrics Letters*, 99, 375–378.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50(4), 1029–54.

HAUSMAN, J. A., AND W. K. NEWEY (2016): "Individual Heterogeneity and Average Welfare," *Econometrica*, 84(3), 1225–1248.

HODERLEIN, S. (2011): "How many consumers are rational?," *Journal of Econometrics*, 164(2), 294–309.

HODERLEIN, S., AND J. STOYE (2014): "Revealed Preferences in a Heterogeneous Population," *The Review of Economics and Statistics*, 96(2), 197–213.

——— (2015): "Testing stochastic rationality and predicting stochastic demand: the case of two goods," *Economic Theory Bulletin*, 3(2), 313–328.

HOUTHAKKER, H. S. (1950): "Revealed Preference and the Utility Function," *Economica*, 17(66), 159–174.

IMBENS, G. W., AND C. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857.

IMBENS, G. W., AND W. K. NEWEY (2002): "Identification and Estimation of Triangular Simultaneous Equation Models without Additivity," NBER Technical Working Paper Series.

——— (2009): "Identification and estimation of triangular simultaneous equations models without additivity," *Econometrica*, 77(5), 1481–1512.

KAIDO, H., F. MOLINARI, AND J. STOYE (2016): "Confidence Intervals for Projections of Partially Identified Parameters," Manuscript. Department of Economics, Cornell University.

KAWAGUCHI, K. (2016): "Testing Rationality without Restricting Heterogeneity," Manuscript, Department of Economics, Hitotsubashi U.

KUDO, A. (1963): "A Multivariate Analogue of the One-Sided Test," *Biometrika*, 50(3/4), 403–418.

LAZZATI, N., J. QUAH, AND K. SHIRAI (2015): "A revealed preference theory of monotone choice and strategic complementarity," Manuscript, University of California Santa Cruz.

LEWBEL, A. (2001): "Demand Systems with and without Errors," *American Economic Review*, 91(3), 611–618.

Manski, C. F. (2007): "Partial Identification Of Counterfactual Choice Probabilities," *International Economic Review*, 48(4), 1393–1410.

——— (2014): "Identification of incomeleisure preferences and evaluation of income tax policy," *Quantitative Economics*, 5(1), 145–174.

Matzkin, R. (2006): "Heterogeneous Choice," in *Advances in Econometrics: Proceedings of the 9th World Congress*. Budapest: Cambridge University Press.

McFadden, D., and K. Richter (1991): "Stochastic rationality and revealed stochastic preference," in *Preferences, Uncettainty and Rationality*, ed. by J. Chipman, D. McFadden, and K. Richter. Boulder: Westview Press, pp. 161–186.

McFadden, D. L. (2005): "Revealed Stochastic Preference: A Synthesis," *Economic Theory*, 26(2), 245–264.

Newey, W. K. (1997): "Convergence rates and asymptotic normality for series estimators," *Journal of Econometrics*, 79(1), 147–168.

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

Polisson, M., and J. K.-H. Quah (2013): "Revealed Preference in a Discrete Consumption Space," *American Economic Journal: Microeconomics*, 5(1), 28–34.

Richter, M. K. (1966): "Revealed Preference Theory," *Econometrica*, 34(3), 635–645.

Romano, J., and A. Shaikh (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78(1), 169–211.

Rosen, A. M. (2008): "Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities," *Journal of Econometrics*, 146(1), 107–117.

Samuelson, P. A. (1938): "A Note on the Pure Theory of Consumer's Behaviour," *Economica*, 5(17), 61–71.

Shapiro, A. (1988): "Towards a Unified Theory of Inequality Constrained Testing in Multivariate Analysis," *International Statistical Review*, 56(1), 49–62.

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

——— (2010): "Partial identification of spread parameters," *Quantitative Economics*, 1(2), 323–357.

Takemura, A., and S. Kuriki (1997): "Weights of $\overline{\chi}^2$ distribution for smooth or piecewise smooth cone alternatives," *The Annals of Statistics*, 25(6), 2368–2387.

Varian, H. R. (1982): "The Nonparametric Approach to Demand Analysis," *Econometrica*, 50(4), 945–973.

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

Ziegler, G. M. (1995): *Lectures on polytopes*, Graduate texts in mathematics. New York: Springer.

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`