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 CWP56/17

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

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 restrict 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: December 7, 2017.
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.

Numerous other, recent papers nonparametrically test rationality, bound welfare, or bound demand responses in repeated cross-sections. But all such 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 or low dimensional 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 do not apply. The technique 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 methods also can be (and recently have been) applied to substantively different models that share certain core features with the benchmark model that we now define.

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

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

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.

We initially assume that W and p are nonrandom, while the utility function u is randomly drawn according to probability law P_u:[1]

$$u \sim P_u.$$

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 7, 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 (1.1) is then indexed by the normalized price vector p. Denoting this by $y(p)$, we have a collection of distributions of demand

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

indexed by $p \in \mathbf{R}^K_{++}$. We initially assume that the distribution of utility functions P_u, i.e. the population being observed, is the same across price regimes. 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 initial analysis in Section 5.1 but drop in Section 5.2 and in our empirical application. Throughout the analysis, we do not place substantive restrictions on P_u. Thus we allow for minimally constrained, 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 $1 \leq j \leq J$, suppose N_j random draws of y distributed according to

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

are observed. Define $N = \sum_{j=1}^J N_j$ for later use. Then P_j can be estimated consistently as $N_j \uparrow \infty$ for each j, $1 \leq j \leq J$. The question is whether the estimated distributions may, up to sampling

[1]To keep the presentation simple, here and henceforth we are informal about probability spaces and measurability. See McFadden (2005) for a formally rigorous setup.

[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.

uncertainty, have arisen from a RUM. In the idealized setting in which there are truly J budgets, we show how to test this with minimal further assumptions. Then, to bring the method from this idealized setting to real data, we introduce some smoothness and also control function assumptions.

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 can be further used to obtain counterfactuals and carry out inference about them, once again under minimal assumptions. Third, this paper offers a new statistical test with broad applicability. While motivated by the application that we report, this test has already been used for other models as well, namely to a nonparametric game theoretic model with strategic complementarity by Lazzati, Quah, and Shirai (2017), to a novel model of "price preference" by Deb, Kitamura, Quah, and Stoye (2016), and to a test of the collective household model by Hubner (2017). The method is extended to choice extrapolation (using algorithms from an earlier version of this paper) in Manski (2014) and Adams (2016); the former also 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, as long as the overall set of budgets is finite.

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 and presents algorithms that allow one to compute this characterization in practice. This is a "population level" (all identifiable quantities are known) analysis whose first part builds on classic work by McFadden and Richter (1991) and McFadden (2005). 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 expenditure and adjusting for endogenous expenditure. Section 6 contains a Monte Carlo investigation of the test's finite sample performance, and Section 7 contains our empirical application. Section 8 concludes. All proofs are collected in Appendix A, and pseudocode for some algorithms is in Appendix B, in the Supplemental Materials.

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 the Weak (Samuelson 1938), Strong (Houthakker 1950), or Generalized (Afriat 1967) Axiom of Revealed Preference (WARP, SARP, and GARP henceforth).[3] 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 Section 3.1 clarifies and modestly extends it. 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 a crucial discretization step below. The core difference is that 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.

---

[3]See also especially Richter (1966) and Varian (1982).

[4]BBC's implementation exploits only WARP and therefore a necessary but not sufficient condition for rationalizability. This is remedied in Blundell, Browning, Cherchye, Crawford, De Rock, and Vermeulen (2015).

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 average welfare under assumptions resembling ours, though their approach additionally imposes smoothness restriction to facilitate nonparametric estimation and interpolation. Their approach is based on nonparametric smoothing, hence the curse of dimensionality needs to be addressed in the presence of many goods, though conceptually their main identification results apply to an arbitrary number of goods. Our method also 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 (2017) 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 rationalizable 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 that 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 or three 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 at a time, and Kawaguchi (2017) tests necessary conditions that are directly expressed as faces of a (larger) polyhedron. Therefore, inference in both papers is much closer to the established literature on moment inequalities.

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 results are a tractable geometric characterization of stochastic rationalizability and algorithms for its practical implementation.

3.1. A Geometric Characterization.

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 henceforth call $(P_1, ..., P_J)$ a *stochastic demand system.* Using a "more is better" assumption, we restrict choice to budget planes; this simplification is not essential to the method. We 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

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

This model is extremely general. Uniqueness of nonstochastic demand has no observational implications 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. Testing stochastic rationalizability therefore seems formidable.

However, we next provide a very tractable characterization of stochastic rationalizability. The essential insight was stated in passing by McFadden (2005) and goes as follows: The revealed preference information contained in a given nonstochastic demand system only depends on whether choice from $\mathcal{B}_i$ is on, above, or below $\mathcal{B}_j$ for all different i, j. Thus, for any two demand systems that agree on all of these comparisons, either both or none are rationalizable. This allows one to transform the testing problem into one where the universal choice set consist of sets like "the bit of budget $\mathcal{B}_1$ that is above $\mathcal{B}_2$ and below all other budgets." This set is finite. We then show that, here and in all other settings of this type, the set of stochastic choice functions that are mixtures of rationalizable nonstochastic ones – hence, stochastically rationalizable – can be expressed as a finite cone.

Formalizing this insight requires a few ingredients.

Definition 3.2. Let $\mathcal{X} := \{x_1, ..., x_I\}$ be the coarsest partition of $\cup_{j=1}^{J}\mathcal{B}_j$ such that 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 $\mathcal{X}$ that are part of more than one budget will be called *intersection patches*. Each budget can be uniquely expressed as union of 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 3.1. $I_j \leq I \leq 3^J$, hence I_j and I are finite.

The partition $\mathcal{X}$ is the finite universal choice set alluded to above. The basic idea is that all choices from budget $\mathcal{B}_1$ that are on the same patch induce the same directly revealed preferences, so are equivalent for the purpose of our test. Conversely, stochastic rationalizability does not at all constrain the distribution of demand on any patch. Therefore, stochastic rationalizability of $(P_1, \ldots, P_J)$ can be decided by only considering the probabilities assigned to patches on the respective budgets. We formalize this as follows.

Definition 3.3. The *vector representation* of $(\mathcal{B}_1, ..., \mathcal{B}_J)$ is a $\sum_{j=1}^{J} I_j$-vector

$$(x_{1|1}, \ldots, x_{I_1|1}, x_{1|2}, \ldots, x_{I_J|J})$$

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 patches on budgets is arbitrary but henceforth fixed. Note that intersection patches appear repeatedly.

Definition 3.4. The *vector representation* of $(P_1, ..., P_J)$ is the $\sum_{j=1}^{J} I_j$-vector

$$\pi := (\pi_{1|1}, \dots, \pi_{I_1|1}, \pi_{1|2}, \dots, \pi_{I_J|J}),$$

where $\pi_{i|j} := P_j(x_{i|j})$.

Thus, the vector representation of a stochastic demand system lists the probability masses that it assigns to patches.

Next, a stochastic demand system is (stochastically) rationalizable iff it is a mixture of rationalizable nonstochastic demand systems. To intuit this, one may literally think of the latter as characterizing rational individuals. It follows that the vector representation of a rationalizable stochastic demand system must be the corresponding mixture of vector representations of rationalizable nonstochastic demand systems. The latter vector representations are in $\{0,1\}^I$, so there are only finitely many of them. Thus, define:

Definition 3.5. The *rational demand matrix* A is the (unique, up to ordering of columns) matrix such that the vector representation of each rationalizable nonstochastic demand system is exactly one column of A. The number of columns of A is denoted H.

Remark 3.2. $H \leq \prod_{j=1}^{J} I_j$.

We then have:

Theorem 3.1. *The stochastic demand system* $(P_1, ..., P_J)$ *is rationalizable iff its vector representation can be written as*

$$\pi = A\nu \text{ for some } \nu \in \Delta^{H-1}, \tag{3.2}$$

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

$$\pi = A\nu \text{ for some } \nu \geq 0. \tag{3.3}$$

We conclude this section with a number of remarks.

GARP vs SARP. 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 disagree only for nonstochastic demand systems that select from at least three intersection patches. The result also applies to the "random utility" or "unobserved heterogeneity" extension of any other revealed preference characterization that allows for discretization of choice space. Indeed, see Deb, Kitamura, Quah, and Stoye (2016) for an application to a different model of individual rationality and Hubner (2017) for an application to the collective household consumption model.

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 $\mathcal{X}$. 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 and GARP agree in this case.

Generality of our characterization. At its very heart, Theorem 3.1 only uses that choice from finitely many budgets can reveal only finitely many distinct revealed preference relations. Thus, it applies to any setting with finitely many budgets, irrespective of budgets' (possibly nonlinear) shape. For example, budget sets are kinked in Manski (2014).

3.2. **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 binary "candidate" vector a is the vector representation of a rationalizable nonstochastic demand system (a is then called *rationalizable* below) and finding all such vectors.

Checking rationalizability of 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 fulfill GARP. The rationalizable demand matrix A collects all complete rationalizable vectors a.

To check rationalizability of a given a, we initially extract a directly revealed preference relation. For example, if (an element of) $x_{i|j}$ is chosen from budget $\mathcal{B}_j$, then it is revealed preferred to all $x_{i|k}$ on or below $\mathcal{B}_j$. This information can be extracted extremely quickly.[7]

[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, $i \leq I_1$, is $(0, -1, 1, 1, 1)$ if $x_{i|1}$ is on budget $\mathcal{B}_1$, below budget $\mathcal{B}_2$, and above the remaining budgets. Each row therefore encodes the revealed preference information implied by the corresponding choice.

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 $\mathcal{X}$ 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 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, we just described how to test 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 potential bottleneck whose cost escalates rapidly (in general more than polynomially) with J. We very briefly mention two approaches that we do not recommend. First, one could check rationalizability of each of the $\prod_{j=1}^{J} I_j$ logically conceivable candidate vectors a. 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 $(x_{1|1}, \ldots, x_{I_1|1})$. Each of these children has I_2 children that correspond to $(x_{1|2}, \ldots, x_{I_2|2})$, 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

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

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 result.

Theorem 3.2. *Suppose that for some $M \geq 1$, $(\mathcal{B}_1, ..., \mathcal{B}_M)$ are either all below or all above $\mathcal{B}_J$. 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.*

If the geometry of budgets allows it – this is particularly likely if budgets move outward over time and even guaranteed if some budget planes are parallel – this theorem can be used to construct columns of A recursively from columns of A-matrices that correspond to a smaller J. The gain can be tremendous because (worst-case) computational cost of finding A increases more than exponentially with A. A caveat is that application of Theorem 3.2 may require to manually reorder budgets so that it applies. Also, while the internal ordering of $(\mathcal{B}_1, \ldots, \mathcal{B}_M)$ and $(\mathcal{B}_{M+1}, \ldots, \mathcal{B}_{J-1})$ does not matter, the Theorem may apply to distinct partitions of the same set of budgets. In that case, any choice of partition will accelerate computations, but we have no general advice on which is best. We tried the refinement in our empirical application, and it improves computation time for some of the largest matrices by orders of magnitude. However, the depth-first search proved so fast that, in order to keep it transparent, our replication code omits this step.

3.3. **Examples.**

We conclude this section with some examples. All examples presume continuous demand and therefore disregard intersection patches.

Example 3.1. The simplest example where RUM is not vacuous consists of two intersecting budgets, thus $J = 2$ and there exists $y \in \mathbf{R}^K_{++}$ with $p_1'y = p_2'y$.

In this example, there are $I = 4$ patches. WARP, SARP, and GARP all agree on which nonstochastic demand system to exclude. Index vector representations as in Figure 1, then the only

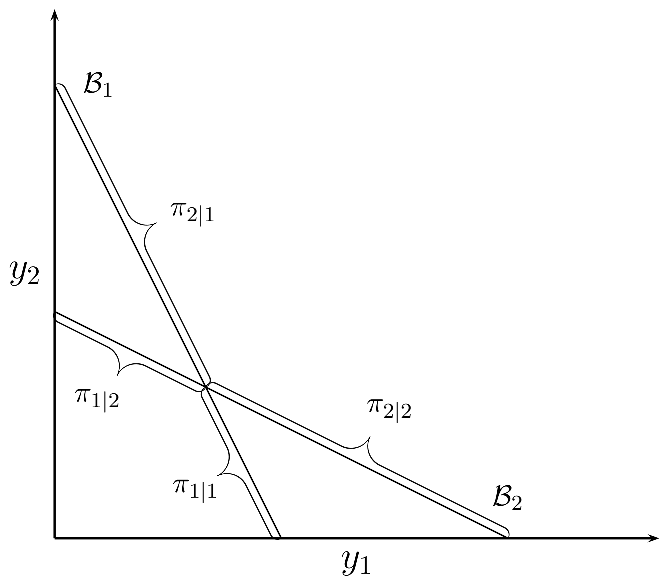

FIGURE 1. Visualization of Example 3.1.

excluded behavior is $(1,0,1,0)'$, thus

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{matrix} x_{1|1} \\ x_{2|1} \\ x_{1|2} \\ x_{2|2} \end{matrix},$$

where the rightmost column in the display is a reminder of which row of A corresponds to which patch. The column cone of A can be explicitly written as

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

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). All of this is illustrated in Figure 1.

Example 3.2. The following is the simplest example in which WARP does not imply SARP, so that applying Example 3.1 to all pairs of budgets will only test a necessary condition. More subtly, it can be shown that the conditions in Kawaguchi (2017) are only necessary as well. Let $K = J = 3$ and assume a maximal pattern of intersection of budgets; for example, prices could be $(p_1, p_2, p_3) = ((1/2, 1/4, 1/4), (1/4, 1/2, 1/4), (1/4, 1/4, 1/2))$. Each budget has 4 patches for a total of $I = 12$ patches, and one can compute

$$A = \left(\begin{array}{ccccccccccccccccccccccccc}
0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&1&0&0&0\\
0&0&0&0&0&1&0&0&0&0&0&0&0&1&0&0&1&0&0&1&0&0&1&0&0\\
0&1&0&0&0&0&1&0&0&0&0&1&0&0&0&0&0&1&0&0&0&0&0&1&0\\
1&0&1&1&0&0&0&1&1&1&1&0&1&0&1&0&0&0&1&0&1&0&0&0&1\\
0&0&0&0&0&0&0&0&1&0&0&0&0&1&1&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&1&1&1&1&0&0&0&0&0&0\\
1&0&0&1&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&1&1&0&0&0&0\\
0&1&1&0&1&1&1&1&0&0&0&1&1&0&0&0&0&0&0&0&0&1&1&1&1\\
1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&1&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&1&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1&1&1&1&1&1&1&1&1
\end{array}\right)
\begin{array}{c}
x_{1|1}\\ x_{2|1}\\ x_{3|1}\\ x_{4|1}\\ x_{1|2}\\ x_{2|2}\\ x_{3|2}\\ x_{4|2}\\ x_{1|3}\\ x_{2|3}\\ x_{3|3}\\ x_{4|3}
\end{array}.$$

Interpreting this matrix requires knowing the geometry of patches. For example, choice of patches $x_{3|1}$, $x_{2|2}$, and $x_{3|3}$ from their respective budgets would violate SARP, thus A does not contain the column $(0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0)'$. We revisit this example in our Monte Carlo study.

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.[9] Figure 2 visualizes one budget in $\mathbf{R}^3$ from our empirical application and its intersection with 6 other budgets. There are total of 10 patches (plus 15 intersection 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, and each cross-section of size N_j is observed over the deterministic (and hence exogenous) budget

[9] In exploratory work using more complex examples, we computed an A matrix with over 2 million columns in a few hours on a standard desktop computer.

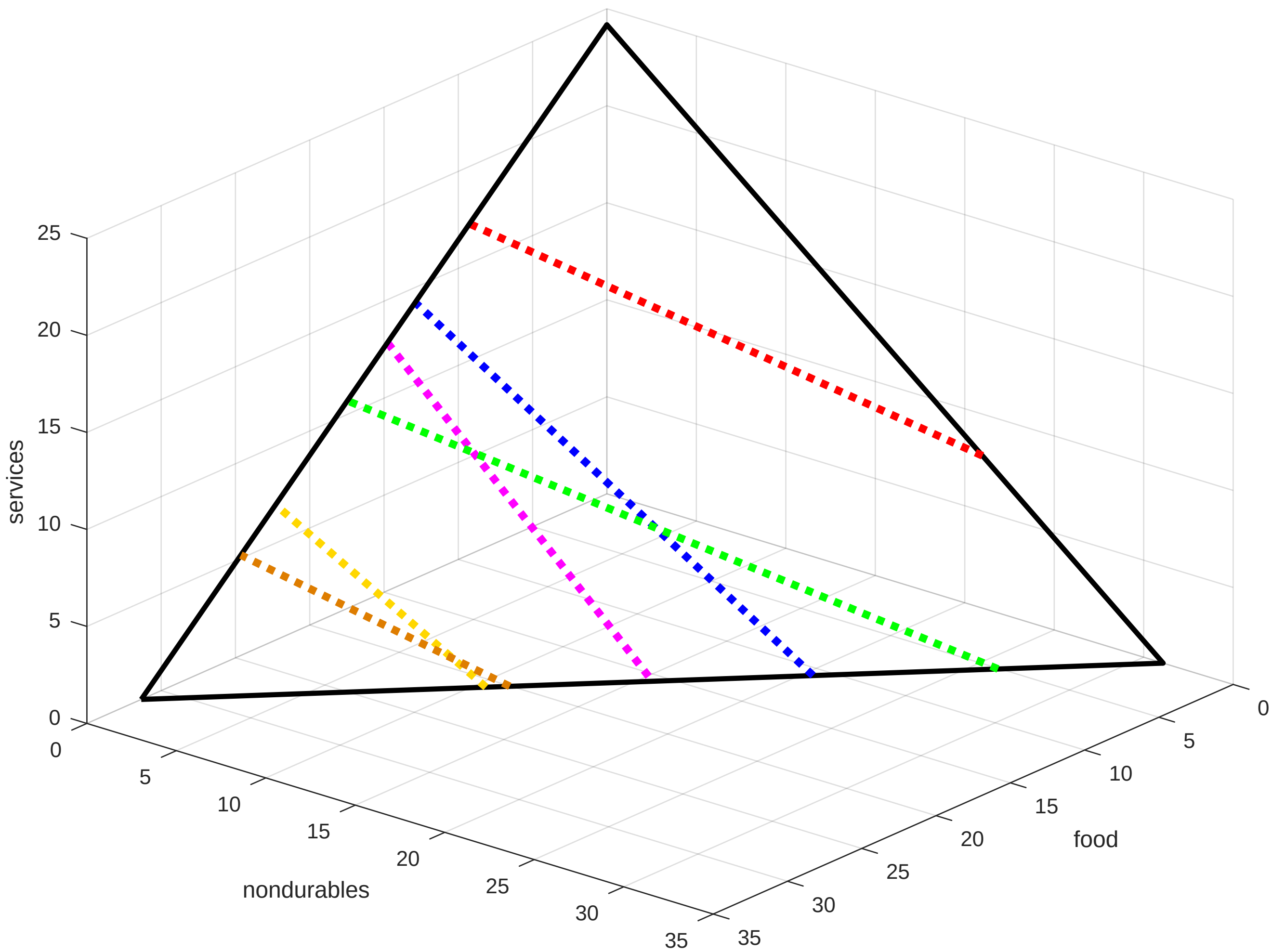

Figure 2. Visualization of one budget set in the empirical application.

plane for period j. 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 (3.3) that we wish to test:

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

This hypothesis is equivalent to

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

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 $\min_{\eta\in\mathcal{C}}[\hat\pi-\eta]'\Omega[\hat\pi-\eta]$, where $\hat\pi$ estimates π, for example by sample choice frequencies. The usual normalization by sample size yields

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

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 $\mathcal{J}_N=0$, 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 $\mathcal{J}_N$ 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, we restrict Ω to be diagonal and positive definite and let $\mathbf{1}_H$ be a H-vector of ones.[11]. The restriction on Ω is important: Together with a geometric feature of the column vectors of the matrix

[10] In this section's simplified setting and if $\hat\pi$ collects sample frequencies, a 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 (4.8). 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.

A, it ensures that among constraints implied by (3.3), those which are fulfilled but with small slack become binding through the *Cone Tightening* algorithm we are about to describe. A non-diagonal weighting matrix can disrupt this property. For further details on this point and its proof, the reader is referred to Appendix A. Our procedure is as follows:

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

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

(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

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

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

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

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.

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.[12]

[12]See Gruber (2007), Grünbaum, Kaibel, Klee, and Ziegler (2003), and Ziegler (1995), especially Theorem 1.3, for these results and other materials concerning convex polytopes used in this paper.

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.*

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 available, 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 a quadratic form of the empirical discrepancy between $B\theta$ and η minimized over $\eta \in \mathbf{R}^q_+$. This type of problem has been studied extensively; see references in Section 2. Its analysis is intricate because the limiting distribution of such a statistic 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 between the two representations is purely theoretical: In practice, B cannot be computed from A in high-dimensional cases like our empirical application.

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

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

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

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

to approximate the distribution of $\mathcal{J}_N$. 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.

To further illustrate the duality between $\mathcal{H}$- and $\mathcal{V}$-representations, we revisit the first two examples. It is not possible to compute B-matrices in our empirical application.

Example 3.1 continued. With two intersecting budget planes, the cone $\mathcal{C}$ is represented by

$$\tag{4.6} B = \begin{array}{c} \begin{matrix} x_{1|1} & x_{2|1} & x_{1|2} & x_{2|2} \end{matrix} \\ \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & 0 & 0 & -1 \end{pmatrix} \end{array}.$$

The first two rows of B are nonnegativity constraints (the other two such constraints are redundant), the next two rows are an equality constraint forcing the sum of probabilities to be constant across budgets, and only the last constraint is a substantive economic constraint. If the estimator $\hat{\pi}$ fulfills the first four constraints by construction, then the testing problem simplifies to a test of $(1,0,0,-1)\pi \leq 0$.

Example 3.2 continued. Eliminating nonnegativity and adding-up constraints for brevity, numerical evaluation reveals

$$\tag{4.7} B = \begin{array}{c} \begin{matrix} x_{1|1} & x_{2|1} & x_{3|1} & x_{4|1} & x_{1|1} & x_{1|2} & x_{3|1} & x_{4|1} & x_{1|1} & x_{2|1} & x_{3|1} & x_{4|1} \end{matrix} \\ \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & -1 & 0 & -1 \\ 0 & -1 & -1 & -1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 \end{pmatrix} \end{array}.$$

The first three rows are constraints on pairs of budgets that mirror the last row of 4.6. The next two constraints are not implied by these, nor by the SARP-like constraints in Kawaguchi (2017), but they imply the latter.

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

$$\hat{\pi}_{i|j} = \sum_{n=1}^{N_j} d_{i|j,n}/N_j, i = 1, ..., I_j, j = 1, ..., J. \tag{4.8}$$

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 col(B) denote its column space.

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

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

Also 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 $\phi = (\phi_1, ..., \phi_m)' \in \text{col}(B)$ *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 automatically 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* $\rho_j > 0$*.*

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 $1 \leq j \leq J$. Let $\mathcal{P}$ denote the set of all π's that satisfy Condition S.1 in Appendix A for some (common) value of (c_1, c_2).

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*

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

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

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

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

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.

5. Extending the Scope of the Test

The methodology outlined in Section 4 requires that (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 7, 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 rationalizability 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 suppose the researcher chooses a value $\underline{w}_j$ for w for each period j. Note that our algorithm and asymptotic theory remain valid if multiple values of w are chosen for each period.

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 $1 \le j \le J$.

This subsection assumes that the total expenditure is exogenous, in the sense that

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

holds under every $P^{(j)}, 1 \le j \le J$. This will be relaxed in the next subsection. Let $p_{i|j}(w) := \Pr\{d_{i|j,n(j)} = 1 | w_{n(j)} = w\}$ then we have

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

where the second 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. Note that π can be estimated by standard nonparametric procedures. For concreteness, we use a series estimator, as defined and analyzed in Appendix A. The smoothed version of $\mathcal{J}_N$ (also denoted $\mathcal{J}_N$ for simplicity) is obtained using the series estimator for $\hat{\pi}$ in (4.1). In Appendix A we also present an algorithm for obtaining the bootstrapped version $\tilde{\mathcal{J}}_N$ of the smoothed statistic.

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 S.2 in Appendix A for some $(c_1, c_2, \delta, \zeta(\cdot))$.

Theorem 5.1. *Let Condition S.3 hold. Also let Ω be diagonal where all the diagonal elements are positive. Then under Assumptions 4.1 and 5.1*

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

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

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 McFadden and Richter (1991). Nonetheless, the assumption can be a concern in applying our testing procedure to a data set such as ours. Recall that the budget sets $\{\mathcal{B}_j\}_{j=1}^J$ are based on prices and total expenditure. The latter is likely to be endogenous, which should be a concern to the econometrician.

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 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. More formally, for each fixed value $\underline{w}_j$ and the unnormalized price vector $\tilde{p}_j$ in period j, $1 \leq j \leq J$, define the endogeneity corrected conditional probability[14]

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

where $D_j(w,u) := D(\tilde{p}_j/e^w, u)$. Then Theorem 3.1 still applies to

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

If we define $\mathcal{J}_{\text{EC}} = \min_{\nu \in \mathbf{R}_+^h} [\pi_{\text{EC}} - A\nu]'\Omega[\pi_{\text{EC}} - A\nu]$, then $\mathcal{J}_{\text{EC}} = 0$ iff stochastic rationalizability holds. Note that this new definition π_{EC} recovers the definition of π in Section 5.1 when w is exogenous.

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

$$w \perp\!\!\!\perp u | \varepsilon$$

holds under every $P^{(j)}, 1 \leq j \leq J$. For example, given a reduced form $w = h_j(z, \varepsilon)$ with h_j monotone in ε and z is an instrument, one may use $\varepsilon = F^{(j)}_{w|z}(w|z)$ 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 $\varepsilon \sim \text{Uni}(0,1)$ under every $P^{(j)}, 1 \leq j \leq J$ by construction. Let $P^{(j)}_{y|w,\varepsilon}$ 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 $\text{supp}(w) = \text{supp}(w|\varepsilon)$ under $P^{(j)}, 1 \leq j \leq J$

[14] This is the conditional choice probability if p is (counterfactually) assumed to be exogenous. We call it "endogeneity corrected" instead of "counterfactual" to avoid confusion with rationality constrained, counterfactual prediction.

we have

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

This means that π_{EC} can be estimated nonparametrically. We propose to use a fully nonparametric two-step estimator, denoted by $\widehat{\pi_{\mathrm{EC}}}$, to define our endogeneity-corrected test statistic $\mathcal{J}_{\mathrm{EC}_N}$; see Appendix A for details. For this, the bootstrap procedure needs to be adjusted appropriately to obtain the bootstrapped statistic $\tilde{\mathcal{J}}_{\mathrm{EC}_N}$: once again, the reader is referred to Appendix A.

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 $1 \le j \le J$.

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 S.4 in Appendix A for some $(c_1, c_2, \delta_1, \delta, \zeta_r(\cdot), \zeta_s(\cdot), \zeta_1(\cdot))$. Then we have:

Theorem 5.2. *Let Condition S.5 hold. Also let Ω be diagonal where all the diagonal elements are positive. Then under Assumptions 4.1 and 5.2,*

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

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

6. Monte Carlo Simulations

We next analyze the performance of Cone Tightening in a small Monte Carlo study. To keep examples transparent and to focus on the core novelty, we model the idealized setting of Section 4, i.e. sampling distributions are multinomial over patches. In addition, we focus on Example 3.2, for which an $\mathcal{H}$-representation in the sense of Weyl-Minkowski duality is available; indeed, see displays (3.2)

	π_0	π_1	π_2	π_3	π_4	π_5	π_6
$x_{1\|1}$	.181	.190	.2	.240	.3	.167	.152
$x_{2\|1}$	.226	.238	.25	.213	.2	.167	.107
$x_{3\|1}$	.226	.238	.25	.213	.2	.333	.441
$x_{4\|1}$	.367	.333	.3	.333	.3	.333	.3
$x_{1\|2}$	.181	.190	.2	.240	.3	.333	.486
$x_{2\|2}$	.226	.238	.25	.213	.2	.167	.107
$x_{3\|2}$	.226	.238	.25	.213	.2	.167	.107
$x_{4\|2}$	.367	.333	.3	.333	.3	.333	.3
$x_{1\|3}$	.181	.190	.2	.240	.3	.333	.486
$x_{2\|3}$	.226	.238	.25	.213	.2	.167	.107
$x_{3\|3}$	.226	.238	.25	.213	.2	.167	.107
$x_{4\|3}$	.367	.333	.3	.333	.3	.333	.3

TABLE 1. The π-vectors used for the main simulations. Entries reported to three digits are rounded.

and (4.7) for the relevant matrices.[15] This allows us to alternatively test rationalizability through a moment inequalities test that ensures uniform validity through GMS.[16]

Data were generated from a total of 31 d.g.p.'s described below and for sample sizes of $N_j \in \{100, 200, 500, 1000\}$; recall that these are per budget, i.e. each simulated data set is based on 3 such samples. The d.g.p.'s are parameterized by the π-vectors reported in Table 1. They are related as follows: π_0 is in the interior of $\mathcal{C}$; π_2, π_4, and π_6 are outside it; and π_2, π_3, and π_5 are on its boundary. Furthermore, $\pi_1 = (\pi_0 + \pi_2)/2$, $\pi_3 = (\pi_0 + \pi_4)/2$, and $\pi_5 = (\pi_0 + \pi_6)/2$. Thus, the line segment connecting π_0 and π_2 intersects the boundary of $\mathcal{C}$ precisely at π_1 and similarly for the next two pairs of vectors. We compute "power curves" along those 3 line segments at 11 equally spaced points, i.e. changing mixture weights in increments of .1, for a total of 31 distinct d.g.p.'s. This is replicated 500

[15]This is also true of Example 3.1, but that example is too simple because the test reduces to a one-sided test about the sum of two probabilities, and the issues that motivate Cone Tightening or GMS go away. We verified that all testing methods successfully recover this and achieve excellent size control, including if tuning parameters are set to 0.

[16]The implementation uses a "Modified Method of Moments" criterion function, i.e. S_1 in the terminology of Andrews and Soares (2010), and the hard thresholding GMS function, i.e. studentized intercepts above $-\kappa_N$ were set to 0 and all others to $-\infty$. The tuning parameter is set to $\kappa_N = \sqrt{ln(N_j)}$.

Method	N_j	π_0	·	·	·	·	π_1	·	·	·	·	π_2
Cone	100	.002	.004	.008	.008	.018	.024	.060	.110	.178	.238	.334
Tightening	200	0	0	.004	.008	.012	.040	.088	.164	.286	.410	.544
	500	0	0	.004	.004	.026	.066	.166	.310	.500	.690	.856
	1000	0	0	0	0	.010	.058	.206	.466	.764	.924	.984
GMS	100	0	0	0	.002	.004	.004	.010	.024	.032	.066	.100
	200	0	0	0	0	.004	.010	.018	.048	.082	.174	.262
	500	0	0	.002	.002	.026	.050	.150	.266	.442	.640	.814
	1000	0	0	0	0	.008	.050	.194	.460	.758	.924	.976

Method	N_j	π_0	·	·	·	·	π_3	·	·	·	·	π_4
Cone	100	.002	.002	.008	.006	.012	.016	.036	.070	.098	.148	.200
Tightening	200	0	0	.004	.008	.010	.036	.068	.112	.194	.296	.404
	500	0	0	.004	.004	.026	.064	.156	.296	.456	.664	.786
	1000	0	0	0	0	.010	.058	.200	.460	.756	.916	.974
GMS	100	0	0	0	.002	.004	.004	.006	.018	.028	.066	.120
	200	0	0	0	0	.004	.004	.012	.024	.056	.110	.216
	500	0	0	.002	0	.004	.012	.034	.084	.196	.348	.544
	1000	0	0	0	0	.002	.028	.080	.180	.406	.694	.892

Method	N_j	π_0	·	·	·	·	π_5	·	·	·	·	π_6
Cone	100	.002	.002	.006	.004	.020	.052	.126	.326	.548	.766	.934
Tightening	200	0	0	.002	.004	.008	.044	.202	.490	.836	.962	.996
	500	0	0	.002	.004	.012	.072	.374	.880	.992	1	1
	1000	0	0	0	0	.006	.052	.606	.992	1	1	1
GMS	100	0	0	0	.004	.008	.022	.072	.170	.388	.604	.808
	200	0	0	0	0	.006	.020	.112	.300	.640	.880	.974
	500	0	0	0	.002	.004	.072	.292	.716	.952	1	1
	1000	0	0	0	0	0	.042	.498	.956	1	1	1

Table 2. Monte Carlo results. See Table 1 for definition of π-vectors. Recall that $\{\pi_1, \pi_3, \pi_5\}$ are on the boundary of $\mathcal{C}$ and π_0 is interior to it. All entries computed from 500 simulations and 499 replications per bootstrap.

times at a bootstrap size of $R = 499$. Nominal size of the test is $\alpha = .05$ throughout. Ideally, it should be exactly attained at the vectors $\{\pi_1, \pi_3, \pi_5\}$.

Results are displayed in Table 2. Noting that the vectors are not too different, we would argue that the simulations indicate reasonable power. Adjustments that ensure uniform validity of tests do tend to cause conservatism for both GMS and cone tightening, but size control markedly improves with sample size.[17] While Cone Tightening appears less conservative than GMS in these simulations, we caution that the tuning parameters and the distance metrics underlying the test statistics are not directly comparable.

The differential performance across the three families of d.g.p.'s is expected because the d.g.p.'s were designed to pose different challenges. For both π_1 and π_3, one constraint is binding and three more are close enough to binding that, at the relevant sample sizes, they cannot be completely ignored. This is more the case for π_3 compared to π_1. It means that GMS or Cone Tightening will be necessary, but also that they are expected to be conservative. The vector π_5 has three constraints binding, with two more somewhat close. This is a worst case for naive (not using Cone Tightening or GMS) inference, which will rarely pick up all binding constraints. Indeed, we verified that inference with $\tau_N = 0$ or $\kappa_N = 0$ leads to overrejection. Finally, we note that π_2 and π_4 fulfill the necessary conditions identified by Kawaguchi (2017), so that his test will have no asymptotic power at a parameter value in the first two panels of Table 2.

7. Empirical Application

We apply our methods to data from the U.K. Family Expenditure Survey, the same data used by BBC. Our testing of a RUM can, therefore, be compared with their revealed preference analysis of a representative consumer. To facilitate this comparison, we use the same selection from these data, 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, following Section 5.1, estimate the distribution of demand on that budget with 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, there is substantial overlap of budgets at median expenditure. To account for endogenous expenditure, we follow Section 5.2, using total household income as instrument. This is also the same instrument used in BBC (2008).

[17] We attribute some very slight nonmonotonicities in the "power curves" to simulation noise.

We present results for blocks of eight consecutive periods and the same three composite goods (food, nondurable consumption goods, and services) considered in BBC.[18] 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, our approach is not subject to a statistical curse of dimensionality.[19]

Regarding the test's statistical power, increasing the dimensionality of commodity space can in principle cut both ways. The number of rationality constraints increases, and this 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.

Tables 3 and 4 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 $\mathcal{J}_N = 0$ and a corresponding p-value of 1 were verified to be true zeros, 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.[20]

We identified a mechanism that may explain this phenomenon. Consider the 84-91 entry in Table 4, where $\mathcal{J}_N$ is especially low. It turns out that one patch on budget $\mathcal{B}_5$ is below $\mathcal{B}_8$ and two patches on $\mathcal{B}_8$ are below $\mathcal{B}_5$. By the reasoning of Example 3.1, probabilities of these patches must

[18]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 4.

[19]Tables 3 and 4 were computed in a few days on Cornell's 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 $\mathcal{J}_N$ about five seconds on a laptop.

[20]In additional analyses not presented here, we replicated these tables using polynomials of degree 2, as well as setting $\tau_N = 0$. The qualitative finding of many positive but insignificant test statistics remains. In isolation, this finding may raise questions about the test's power. However, the test exhibits reasonable power in our Monte Carlo exercise and also rejects rationalizability in an empirical application elsewhere (Hubner 2017).

We also checked whether small but positive test statistics are caused by adding-up constraints, i.e. by the fact that all components of π that correspond to one budget must jointly be on some unit simplex. The estimator $\hat{\pi}$ can slightly violate this. 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 $\hat{\eta}$.

	3 goods				**4 goods**				**5 goods**			
	I	H	$\mathcal{J}_N$	p	I	H	$\mathcal{J}_N$	p	I	H	$\mathcal{J}_N$	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 3. Empirical results with 7 periods. I = number of patches, H = number of rationalizable discrete demand vectors, $\mathcal{J}_N$ = test statistic, p = p-value.

add to less than 1. The estimated sum equals 1.006, leading to a tiny and statistically insignificant violation. This phenomenon occurs frequently and seems to cause the many positive but insignificant values of $\mathcal{J}_N$. The frequency of its occurrence, in turn, has a simple cause that may also appear in other data: If two budgets are slight rotations of each other and demand distributions change continuously in response, then rationalizable 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 estimates will frequently add to slightly more than 1. With 7 or 8 mutually intersecting

3 goods

	I	H	$\mathcal{J}_N$	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 4. Empirical results with 8 periods. I = number of patches, H = number of rationalizable discrete demand vectors, $\mathcal{J}_N$ = test statistic, p = p-value.

budgets, there are many opportunities for such reversals, and positive but insignificant test statistics may become ubiquitous.

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.

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. 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 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 $\mathcal{J}_N$-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 $\mathcal{J}_N$-statistic for each of these partitions. Given the resulting M-statistics, say $\mathcal{J}_N^1, \cdots, \mathcal{J}_N^M$, we can consider $\mathcal{J}_N^{\max} := \max_{1 \le m \le M} \mathcal{J}_N^m$ 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 $\mathcal{J}_N$-statistics over the covariates. This extension will be pursued.

(4) Natural next steps after rationality testing are extrapolation to (bounds on) counterfactual demand distributions and welfare analysis, i.e. along the lines of BBC (2008) or, closer to our own setting, Adams (2016) and Deb, Kitamura, Quah, and Stoye (2016). This extension is being pursued.

(5) The econometric techniques outlined here can be potentially useful in much broader contexts. Existing proposals for 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 or, more generally, a polyhedron, the $\mathcal{H}$-representation will typically not be available in practice. We expect that our method can be used in many such cases.

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 (2000): "Endogeneity in nonparametric and semiparametric regression models," Manuscript.

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

Bugni, F. A. (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.

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.

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.

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.

Hubner, S. (2017): "Its complicated: A Nonparametric Test of Preference Stability between Singles and Couples," Preprint, University of Oxford.

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.

Kawaguchi, K. (2017): "Testing rationality without restricting heterogeneity," *Journal of Econometrics*, 197(1), 153 – 171.

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

Lazzati, N., J. Quah, and K. Shirai (2017): "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, Uncertainty 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.

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.

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.

SUPPLEMENTAL MATERIALS

Appendix A: Proofs and Further Details of Inferential Procedures

Proof of Theorem 3.1. The proof uses nonstochastic demand systems, which can be identified with vectors $(d_1, \ldots, d_J) \in \mathcal{B}_1 \times \cdots \times \mathcal{B}_J$. Such a system is rationalizable if $d_j \in \arg\max_{y \in \mathcal{B}_j} u(y), j = 1, \ldots, J$ for some utility function u.

Rationalizability of nonstochastic demand systems is well understood. In particular, and irrespective of whether we define rationalizability by GARP or SARP, it is decidable from knowing the preferences directly revealed by choices, hence from knowing patches containing $(d_1, \ldots, d_J)$. It follows that for all nonstochastic demand systems that select from the same patches, either all or none are rationalizable.

Fix $(P_1, ..., P_J)$. Let the set $\mathcal{Y}^*$ collect one "representative" element (e.g., the geometric center point) of each patch. Let $(P_1^*, ..., P_J^*)$ be the unique stochastic demand system concentrated on $\mathcal{Y}^*$ and having the same vector representation as $(P_1, ..., P_J)$. The previous paragraph established that demand systems can be arbitrarily perturbed within patches, so $(P_1, ..., P_J)$ is rationalizable iff $(P_1^*, ..., P_J^*)$ is. It follows that rationalizability of $(P_1, ..., P_J)$ can be decided from its vector representation π, and that it suffices to analyze stochastic demand systems supported on $\mathcal{Y}^*$.

Now, any stochastic demand system is rationalizable iff it is a mixture of rationalizable nonstochastic ones. Since $\mathcal{Y}^*$ is finite, there are finitely many nonstochastic demand systems supported on it; of these, a subset will be rationalizable. Noting that these demand systems are characterized by binary vector representations corresponding to columns of A, the statement of the Theorem is immediate for the restricted class of stochastic demand systems supported on $\mathcal{Y}^*$. □

Proof of Theorem 3.2. 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 $k \le M$, 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 $i \le M$, the choice cycle is reducible by (iii). Thus, $i > M$, hence the conclusion. □

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

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

where $\oplus$ signifies Minkowski sum. Define

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

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

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

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

$$BAe_h \le 0$$

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

$$\Phi_{kh} = -e_k' BAe_h \ge 0, \quad 1 \le k \le m, 1 \le h \le H. \tag{S.1}$$

But if $k \le \bar{m}$, 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 $\bar{m}+1 \le k \le m, 1 \le h \le H$. Therefore if $k \le \bar{m}$, $\Phi_{kh} = b_k' a_h$ is nonzero at least for one $h, 1 \le h \le H$, whereas if $k > \bar{m}$, $\Phi_{kh} = 0$ for every h. Since (S.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 $k \le \bar{m}$ and $\phi_k = 0$ for every $k > \bar{m}$. We now have

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

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

Before we present the proof of Theorem 4.2, it is necessary to specify a class of distributions, to which we impose a mild condition that guarantees stable behavior of the statistic $\mathcal{J}_N$. 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 $1 \leq j \leq J$ are included in the set of equality constraints.[21] 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 $b_k' \pi \leq 0$ represents an economic restriction if $k \in \mathcal{K}^R$.[22] Recalling the choice vectors $(d_{j|1}, ..., d_{j|N_j})$ are IID-distributed within each time period $j, 1 \leq j \leq J$, let d_j denote the choice vector of a consumer facing budget j (therefore w.l.o.g we can let $d_j = d_{j|1}$). Define $d = [d_1', ..., d_J']'$, a random I-vector of binary variables. Note $\mathrm{E}[d] = \pi$. Let

$$\begin{aligned} g &= Bd \\ &= [g_1, ..., g_m]'. \end{aligned}$$

With these definitions, consider the following requirement:

Condition S.1. *For each* $k \in \mathcal{K}^R$, $\mathrm{var}(g_k) > 0$ *and* $\mathrm{E}[|g_k/\sqrt{\mathrm{var}(g_k)}|^{2+c_1}] < c_2$ *hold, where* c_1 *and* c_2 *are positive constants.*

This type of condition is standard in the literature; see, for example, Andrews and Soares (2010).

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

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

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

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

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

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

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

[21] 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.

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

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

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

Let

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

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

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

Also define

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

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 $B^{\le} t^{**} \le 0$ 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

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

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

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

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

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

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

$$S_k(\delta) = \{x \in \mathcal{C} : |b'_k x| \leq \tau\delta\}$$

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

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

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

$$x^* \in S_1(\delta) \cap \cdots \cap S_q(\delta) \cap H^\tau_{\angle q+1}(\delta) \cap \cdots H^\tau_{\angle m}(\delta) \text{ for some } q \in \{1, ..., \bar{m}\} \tag{S.2}$$

satisfies

$$x^*|\mathcal{C}_\tau \in H_1^\tau \cap \cdots \cap H_q^\tau \cap L \tag{S.3}$$

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^* fulfills (S.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, (S.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 (S.2) implies (S.3). Suppose it does not hold; then without loss of generality, for an element x^* that satisfies (S.2) for an arbitrary small $\delta > 0$, we have

$$x^*|\mathcal{C}_\tau \in H_1^\tau \cap \cdots \cap H_r^\tau \cap L \quad \text{and} \quad x^*|\mathcal{C}_\tau \notin H_j^\tau, r+1 \le j \le q \tag{S.4}$$

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

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

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

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

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

then for (S.4) to hold for some $x^* \in \mathbf{R}^I$ satisfying (S.2) for an arbitrary small $\delta > 0$ we must have

$$F| \left(H_1^\tau \cap \cdots \cap H_r^\tau \cap L\right) \subset \text{int}(H_{\angle r+1}^\tau \cap \cdots \cap H_{\angle q}^\tau)$$

(Recall the notation | signifies orthogonal projection. Also note that if $\dim(F) = 1$, then (S.4) does not occur under (S.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

$$(F \cap \Delta(J)) \,|\, (H_1^\tau \cap \cdots \cap H_r^\tau \cap L) \subset \text{int}(H_{\angle r+1}^\tau \cap \cdots \cap H_{\angle q}^\tau). \tag{S.5}$$

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 (S.5),

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

holds for small enough $\varepsilon > 0$. Let $\bar{a}^\tau := \bar{a} + \frac{\tau}{H} A\mathbf{1}_H$, $\bar{\bar{a}}^\tau := \bar{\bar{a}} + \frac{\tau}{H} A\mathbf{1}_H$, 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

$$(a_1 - a_0)'(a_2 - a_0) \geq 0 \text{ for every triplet } (a_0, a_1, a_2) \text{ of distinct elements in } \mathcal{A}. \tag{S.6}$$

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 $\mathcal{J}_N$ in (4.1) can be rewritten, using the square-root matrix $\Omega^{1/2}$,

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

or

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

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 (S.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 $\mathcal{J}_N$ 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 $\xi \sim \text{N}(0, \hat{S})$ and $\zeta = B\xi$,

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

conditional on data $\{\{d_{i|j,n}\}_{i=1}^{I_j}\}_{n=1}^{N_j}$, $j = 1, 2, ..., J$. Moreover, defining $\gamma^\tau = \gamma + \tau_N\phi$ in the above, and using the definitions of $\beta_\tau(\cdot)$ and $s(\cdot)$

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

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

$$\tilde{\mathcal{J}}_N \sim s\left(\varphi_N(\hat{\xi}) + \zeta\right). \tag{S.7}$$

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 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 (S.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 $b'_k\hat{\pi}$ under π_N for $k \in \mathcal{K}^R$ as Condition S.1 is imposed for $\pi_N \in \mathcal{P}$. We do not impose Condition S.1 for $k \in \mathcal{K}^D$. Note, however, that: (i) The equality $b'_k\hat{\pi} \leq 0$ holds by construction for every $k \in \mathcal{K}^D$ and therefore its behavior does not affect $\mathcal{J}_N$; (ii) If $\mathrm{var}_{\pi_N}(g_k)$ 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 $\tilde{\mathcal{J}}_N$ 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

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

$\square$

Further details of the procedure in Section 5.1 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$(0,1)$. 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 $\Sigma^{(j)}(w) = \mathrm{diag}\left(p^{(j)}(w)\right) - p^{(j)}(w)p^{(j)}(w)'$ 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 $\widehat{\Lambda}(j) = \frac{1}{N_j}\sum_{n(j)=1}^{N_j}\left[\widehat{\Sigma}^{(j)}(w_{n(j)}) \otimes q^{K(j)}(w_{n(j)})q^{K(j)}(w_{n(j)})'\right]$, $\widehat{\Sigma}^{(j)}(w) = \mathrm{diag}\left(\widehat{p}^{(j)}(w)\right) - \widehat{p}^{(j)}(w)\widehat{p}^{(j)}(w)'$, $\widehat{p}^{(j)}(w) = [\widehat{p}_{1|j}(w), ..., \widehat{p}_{I_j|j}(w)]'$ 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 $\tilde{\eta}_{\tau_N}(j) := \hat{\eta}_{\tau_N}(j) + \frac{1}{\sqrt{N_j}}N(0, \hat{v}_N^{(j)}), j = 1, ..., J$.

Noting $\{\{d_{i|j,n(j)}\}_{i=1}^{I_j}, w_{n(j)}\}_{n(j)=1}^{N_j}$ are IID-distributed within each time period $j, 1 \le j \le J$, let (d_j, w_j) denote the choice-log-expenditure pair of a consumer facing budget j. Let $d = [d_1', ..., d_J']'$ and $\mathbf{w} = [w_1, ..., w_J]'$, and define $g = Bd = [g_1, ..., g_m]'$ as before. Let $\mathcal{W}_j$ denote the support of $w_{n(j)}$. For a symmetric matrix A, $\lambda_{\min}$ signifies its smallest eigenvalue.

Condition S.2. *There exist positive constants c_1, c_2, δ, and $\zeta(K)$, $K \in \mathbf{N}$ such that the following holds:*

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

(ii) *For each $k \in \mathcal{K}^R$, $\mathrm{var}(g_k|\mathbf{w} = (\underline{w_1}, ..., \underline{w_J})') \ge s^2(F_1, ..., F_J)$ and $\mathrm{E}[(g_k/s(F_1, ..., F_J))^4|\mathbf{w} = (\underline{w_1}, ..., \underline{w_J})'] < c_1$ hold for every $(\underline{w_1}, ..., \underline{w_J}) \in \mathcal{W}_1 \times \cdots \mathcal{W}_J$;*

(iii) $\sup_{w \in \mathcal{W}_j} |p_{i|j}(w) - q^K(w)'\beta_K^{(j)}| \le c_1 K^{-\delta}$ *holds with some K-vector $\beta_K^{(j)}$ for every $K \in \mathbf{N}$, $1 \le i \le I_j, 1 \le j \le J$;*

(iv) *Letting $\widetilde{q}^K := C_{K,j}q^K$, $\lambda_{\min}\mathrm{E}[\widetilde{q}^K(w_{n(j)})\widetilde{q}^K(w_{n(j)})'] \ge c_2$ holds for every K and j, where $C_{K,j}, K \in \mathbf{N}, 1 \le j \le J$ are constant nonsingular matrices;*

(v) $\max_j \sup_{w \in \mathcal{W}_j} \|\widetilde{q}^K(w)\| \le c_2\zeta(K)$ *for every $K \in \mathbf{N}$.*

□

Condition S.2(ii) is a version of Condition S.1 that accommodates the conditioning by w and series estimation. Conditions S.2(iii)-(v) are standard regularity commonly used in the series regression literature: (iii) imposes a uniform approximation error bound, (iv) avoids singular design (note the existence of the matrices $C_{K,j}$ suffices) and (v) controls the lengths of the series terms used.

The next condition imposes restrictions on tuning parameters.

Condition S.3. *τ_N and $K(j), j = 1, ..., J$ satisfy $\sqrt{N_j}K^{-\delta}(j) \downarrow 0$, $\zeta(K(j))^2K(j)/N_j \downarrow 0$, $j = 1, ..., J$, $\tau_N \downarrow 0$, and $\sqrt{\underline{n}}\tau_N \uparrow \infty$.*

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

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

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

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

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

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

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 $\sqrt{N_j/K(j)}B^{(j)}\pi_{F_N}^{(j)} \to h_j, h_j \leq 0, 1 \leq j \leq J$ as $N \to \infty$. Define $Q_{F_N}^{(j)} = \text{E}_{F_N}[q^{K(j)}(w_{n(j)})q^{K(j)}(w_{n(j)})']$ and $\Xi_{F_N}^{(j)} = \text{E}_{F_N}[B^{(j)}\Sigma_{F_N}^{(j)}(w_{n(j)})B^{(j)'} \otimes q^{K(j)}(w_{n(j)})q^{K(j)}(w_{n(j)})']$, 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. □

Further details of the procedure in Section 5.2 To estimate $\widehat{\pi_{\text{EC}}}$, 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 $\widehat{F}_{w|z}^{(j)}$ 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 $\iota_N(\varepsilon) = (\varepsilon + \upsilon_N)^2/4\upsilon_N$, then let

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

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

The second step is nonparametric estimation of $P^{(j)}_{y|w,\varepsilon}\left\{y \in x_{i|j}|w = \underline{w}_j, \varepsilon\right\}$. Let $\widehat{\chi}_{n(j)} = (w_{n(j)}, \widehat{\varepsilon}_{n(j)})'$, $n(j) = 1, ..., N_j$ 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 $P^{(j)}_{y|w,\varepsilon}\left\{y \in x_{i|j}|w = \cdot, \varepsilon = \cdot\right\}$ evaluated at $\chi = (w, \varepsilon)$ is

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

where

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

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

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

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

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

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 $v^{(j)}_N$ needs modification. With the nonparametric endogeneity correction, the modified version of $v^{(j)}_N$ is

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

where

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

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

with

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

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

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

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

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 $\tilde{J}_{\mathrm{EC}}$.

We impose some conditions to show the validity of the endogeneity-corrected test.

Let $\varepsilon_{n(j)}$ be the value of the control variable ε for the $n(j)$-th consumer facing budget j. Noting $\{\{d_{i|j,n(j)}\}_{i=1}^{I_j}, w_{n(j)}, \varepsilon_{n(j)}\}_{n(j)=1}^{N_j}$ are IID-distributed within each time period $j, 1 \leq j \leq J$, let $(d_j, w_j, \varepsilon_j)$ denote the choice-log expenditure-control variable triplet of a consumer facing budget j. Let $d = [d_1', ..., d_J']'$, $\mathbf{w} = [w_1, ..., w_J]'$, $\mathbf{e} = [\varepsilon_1, ..., \varepsilon_J]'$ and define $g = Bd = [g_1, ..., g_m]'$ as before. Let $\mathcal{X}_j = \mathrm{supp}(\chi_{n(j)})$, $\mathcal{Z}_j = \mathrm{supp}(z_{n(j)})$, and $\mathcal{E}_j = \mathrm{supp}(\varepsilon_{n(j)})$, $1 \leq j \leq J$. Following the above discussion, define an $\mathbf{R}^I$-valued functional

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

where

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

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

Condition S.4. *There exist positive constants c_1, c_2, 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 $z \in \mathcal{Z}_j$, $1 \leq j \leq J$;*

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

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

(iv) *For each $k \in \mathcal{K}^R$, $\mathrm{var}(g_k|\mathbf{w} = (\underline{w_1}, ..., \underline{w_J})', \mathbf{e} = (\underline{\varepsilon_1}, ..., \underline{\varepsilon_J}))' \geq s^2(F_1, ..., F_J)$ and $\mathrm{E}[(g_k/s(F_1, ..., F_J))^4 |\mathbf{w} = (\underline{w_1}, ..., \underline{w_J})', \mathbf{e} = (\underline{\varepsilon_1}, ..., \underline{\varepsilon_J})'] < c_1$ hold for every $(\underline{w_1}, ..., \underline{w_J}, \underline{\varepsilon_1}, ..., \underline{\varepsilon_J}) \in \mathcal{W}_1 \times \cdots \mathcal{W}_J \times \mathcal{E}_1 \times \cdots \mathcal{E}_J$;*

(v) *Letting $\widetilde{r}^L := C_{L,j} r^L$, $\lambda_{\min} \mathrm{E}[\widetilde{r}^L(z_{n(j)}) \widetilde{r}^L(z_{n(j)})] \geq c_2$ holds for every L and j, where $C_{L,j}, L \in \mathbf{N}$, $1 \leq j \leq J$, are constant nonsingular matrices;*

(vi) $\max_j \sup_{z \in \mathcal{Z}_j} \|\widetilde{r}^L(z)\| \leq c_1 \zeta_r(L)$ *for every $L \in \mathbf{N}$.*

(vii) $\sup_{w \in \mathcal{W}_j, z \in \mathcal{Z}_j} |F_{w|z}^{(j)}(w, z) - r^L(z)' \alpha_L^{(j)}(w)| \leq c_1 L^{-\delta_1}, 1 \leq j \leq J$ *holds with some L-vector $\alpha_L^{(j)}(\cdot)$ for every $L \in \mathbf{N}$, $1 \leq j \leq J$;*

(viii) *Letting $\widetilde{s}^M := \bar{C}_{M,j} s^M$, $\lambda_{\min} \mathrm{E}[\widetilde{s}^M(\chi_{n(j)}) \widetilde{s}^M(\chi_{n(j)})] \geq c_2$ holds for every M and j, where $\bar{C}_{M,j}, M \in \mathbf{N}$, $1 \leq j \leq J$, are constant nonsingular matrices;*

(ix) $\max_j \sup_{\chi \in \mathcal{X}_j} \|\widetilde{s}^M(\chi)\| \leq C\zeta_s(M)$ *and* $\max_j \sup_{\chi \in \mathcal{X}_j} \|\partial \widetilde{s}^M(\chi)/\partial \varepsilon\| \leq c_1 \zeta_1(M)$ *and* $\zeta_s(M) \leq C\zeta_1(M)$ *for every* $M \in \mathbf{N}$*;*

(x) $\sup_{\chi \in \mathcal{X}_j} \left| P^{(j)}_{y|w,\varepsilon}\{y \in x_{i|j}|w, \varepsilon\} - s^M(\chi)' g^{(i,j)}_M \right| \leq c_1 M^{-\delta}$*, holds with some* M*-vector* $g^{(i,j)}_M$ *for every* $M \in \mathbf{N}$*,* $1 \leq i \leq I_j, 1 \leq j \leq J$*;*

(xi) $P^{(j)}_{y|w,\varepsilon}\{y \in x_{i|j}|w, \varepsilon\}$*,* $1 \leq i \leq I_j, 1 \leq j \leq J$*, are twice continuously differentiable in* $\chi = (w, \varepsilon)$*. Moreover,* $\max_{1 \leq j \leq J} \max_{1 \leq i \leq I_j} \sup_{\chi \in \chi_j} \left\| \frac{\partial}{\partial \chi} P^{(j)}_{y|w,\varepsilon}\left\{y \in x_{i|j}|w, \varepsilon\right\}\right\| \leq c_1$ *and* $\max_{1 \leq j \leq J} \max_{1 \leq i \leq I_j} \sup_{\chi \in \chi_j} \left\| \frac{\partial^2}{\partial \chi \partial \chi'} P^{(j)}_{y|w,\varepsilon}\left\{y \in x_{i|j}|w, \varepsilon\right\}\right\| \leq c_1$*.*

□

Since we use the control function approach to deal with potential endogeneity in w (income), Conditions S.4(i)-(ii) are essential. See Blundell and Powell (2000) and Imbens and Newey (2009) for further discussion on these types of restrictions. Just like Condition S.2(ii), Condition S.4(iv) is a version of Condition S.1 that accommodates the two-step series estimation. Conditions S.4(iv)-(xi) corresponds to standard regularity conditions stated in the context of the two-step approach adopted in this section: (iv) and (x) imposes uniform approximation error bounds, (v) and (viii) avoid singular designs (note the existence of the matrices $C_{L,j}$ and $\bar{C}_{M,j}$ suffices), and (vi) and (ix) control the lengths of (the derivatives of) the series terms used. Condition S.4(xi) imposes reasonable smoothness restrictions on the (observable) conditional probabilities $P^{(j)}_{y|w,\varepsilon}\{y \in x_{i|j}|w, \varepsilon\}$, $1 \leq i \leq I_j, 1 \leq j \leq J$.

The next condition impose restrictions on tuning parameters.

Condition S.5. *Let* τ_N*,* $M(j)$ *and* $L(j)$*,* $j = 1, ..., J$ *satisfy* $\tau_N \downarrow 0$*,* $\sqrt{\underline{n}_{\mathrm{EC}}}\tau_N \uparrow \infty$*,* $N_j L(j)^{1-2\delta_1} \downarrow 0$*,* $N_j M(j)^{-2\delta} \downarrow 0$*,* $M(j)\zeta_1(M(j))^2 L^2(j)/N_j \downarrow 0$*,* $\zeta_s(M(j))^6 L^4(j)/N_j \downarrow 0$*, and* $\zeta_1(M(j))^4 \zeta_r(L(j))^4/N_j \downarrow 0$ *and also* $\underline{C}(L(j)/N_j + L(j)^{1-2\delta_1}) \leq \upsilon_N^3 \leq \overline{C}(L(j)/N_j + L(j)^{1-2\delta_1})$*, for some* $0 < \underline{C} < \overline{C}$*.*

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 $P^{(j)}_{y|w,\varepsilon;F_N}\{y \in x_{i|j}|w, \varepsilon\}$ 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 $\pi_{F_N} = \pi(P^{(1)}_{y|w,\varepsilon_N}, ..., P^{(J)}_{y|w,\varepsilon_N})$ 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

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

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

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

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

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

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

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 triangular array for the repeated cross-section 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. □

Appendix B: Algorithms for Computing A

This appendix details algorithms for computation of A. The first algorithm is the depth-first algorithm for generic computation of A. The second algorithm is a refinement using Theorem 3.2. All implementations are in MATLAB using CVX and are available from the authors. Algorithms use notation introduced in the proof of Theorem 3.2.

Computing A as in Theorem 3.1.

1. Initialize $m_1 = ... = m_J = 1$.
2. Initialize $l = 2$.
3. Set $c(\mathcal{B}_1) = x_{m_1|1}, \ldots, c(\mathcal{B}_l) = x_{m_l|l}$. Check for revealed preference cycles.

4. If a cycle is detected, move to step 7. Else:

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

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

7a. If $m_l < I_l$, set $m_l = m_l + 1$ and return to step 3.

7b. 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.

7c. 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.

(...)

7z. Terminate.

Refinement using Theorem 3.2

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.

1. Use preceding algorithm 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 $(\mathcal{B}_1, ..., \mathcal{B}_J)$.[23]

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 (using preceding algorithm) all vectors $a_{1\to M}$ s.t.$(a_{1\to M}, a_{M+1\to J-1})$ is rationalizable.

2.2 Compute (using preceding algorithm) all vectors a_J s.t.$(a_{M+1\to J-1}, a_J)$ is rationalizable.

2.3 All stacked vectors $(a'_{1\to M}, a'_{M+1\to J-1}, a'_J)'$ are valid columns of A.

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

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

Departments of Economics, Cornell University, Ithaca, NY 14853, and University of Bonn, Germany; Hausdorff Center for Mathematics, Bonn, Germany.

E-mail address: stoye@cornell.edu

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