Testing revealed preference models with unobserved randomness: a column generation approach

Thomas Demuynck
ECARES, Université libre de Bruxelles

Tom Potoms
University of Sussex

December 2022

ECARES working paper 2022-42

ECARES
ULB - CP 114/04
50, F.D. Roosevelt Ave., B-1050 Brussels BELGIUM
www.ecares.org

Testing revealed preference models with unobserved randomness: a column generation approach

Thomas Demuynck* and Tom Potoms†

December 19, 2022

Abstract

We present a computationally fast and simple algorithm to test structural revealed preference models in the presence of unobserved randomness. Towards this end, we operationalize insights from random set theory. We demonstrate the practical relevance of our results by an application to the standard intertemporal consumption model with idiosyncratic income risks and an approximate expected utility model.

Keywords: Partial identification, revealed preferences, column generation approach, random set theory.
JEL codes: C50, C60, D15

Since the seminal works by Afriat (1967), Diewert (1973) and Varian (1982), the use of revealed preference (RP) analysis for testing structural decision models has steadily increased.[1] Its most attractive feature relies on the fact that it is nonparametric, in the sense that it abstains from imposing specific functional forms on underlying utility or production functions, thereby making such tests more robust against violations to such auxiliary parametric assumptions.

Much of the RP literature has focused on settings without randomness, which excludes realistic and relevant features like uncertainty or misperception regarding prices, income,

*Université Libre de Bruxelles, ECARES. Avenue F. d. Roosevelt 50, 1050 Brussels, Belgium. E-mail: thomas.demuynck@ulb.be
†University of Sussex. Jubilee building, Falmer Brighton BN1 9SL, United Kingdom. E-mail: T.Potoms@sussex.ac.uk

[1] See Varian (2006), Crawford and De Rock (2014), Chambers and Echenique (2016) and Demuynck and Hjertstrand (2020) for overviews of the literature.

wages or measurement error. The absence of a general (nonparametric) RP method to incorporate such features has been a severe restriction in terms of its acceptability within the broader economic literature. In this paper, we present a computationally fast method to test for a variety of RP models accounting for such unobserved randomness. Towards this end, we rely on recent insights in the economic literature that uses random set theory to analyze partially identified models and we use ideas from operations research (and computer science) to operationalize these.

Overview We start from a general framework in which we have an economic model, say an RP model, that is represented by an underlying joint distribution over observable and unobservable, i.e., latent, variables. Although the marginal distribution over the observable variables is known (or can be identified), the joint distribution is inherently non observable. The support of this joint distribution, however, is usually restricted by the conditions imposed by the underlying model. In addition, we assume that the model imposes some moment conditions on the observable and unobservable variables. Due to its generality, this framework is able to capture a large set of economic applications including many RP models. As shown in the recent literature, the restrictions imposed by such setting can be expressed in terms of so-called *random sets*. A random set is a generalization of a random variable to sets, or alternatively, a set-valued random variable. The theory of random sets has received a lot of attention and has become a central toolkit in the literature on partial identification.[2] Under suitable regularity conditions, one can show that the testable implications boil down to checking whether or not a particular vector α, which corresponds to the values for the moment conditions, belongs to the Aumann-expectation of a particular random set.[3] This Aumann-expectation is, in principle, identified from the observables and the maintained model's assumptions.

The main focus of this paper is to make this testable implication useful for practical applications, i.e., when one has only access to a (finite) sample of the observable variables. The finite sample equivalent of the Aumann-expectation is given by the convex hull of the Minkowski average, that is: an average of sets. The finite sample analogue of the model's testable implication then requires the same vector α to belong to this set. Equivalently, we can check whether the distance between the vector α and the convex hull of the Minkowski average equals zero. An important caveat towards operationalizing this idea, is that in many cases,

[2]See, among many others, Beresteanu and Molinari (2008); Beresteanu, Molchanov, and Molinari (2012, 2011); Molchanov (2017) and Li (2021). See also Molchanov and Molinari (2014) and Bontemps and Magnac (2017) for an overview.

[3]The Aumann expectation is a generalization of the usual expectation of random variables to the framework of random sets.

including most RP models, there is no closed form expression for this Minkowski average. This makes it far from straightforward to compute this distance.

To solve this problem, we propose an intuitive and fast column generating algorithm. The algorithm is similar in flavor to the popular GJK-algorithm that is widely used to determine the minimal distance between two convex sets in 2 or 3 dimensional space (Gilbert, Johnson, and Keerthi, 1988). Our algorithm consists of a Master program and a set of Auxiliary programs. At the start of the algorithm, we draw a small number of points from the Minkowski average, and consider the convex hull of these points. This forms an inner approximation of the convex hull of the Minkowski average. At every iteration of the algorithm, the Master program finds the projection of the vector α on this inner approximation. If this projection is equal to α, then our algorithm finishes as the minimal distance between α and the convex hull of the Minkowski average equals zero. In cases where the distance is not zero, we expand the inner approximation by adding a new point of the Minkowski average to the current inner approximation. This newly chosen point is found by solving a specific maximization problem based on the support function associated with the normal vector of the most recent projection. At each iteration of the algorithm, we add an additional informative member to the inner approximation until no further improvements can be found (or until convergence is reached). We show that our column generating algorithm is fast and approximates the true distance between α and the convex hull of the Minkowski up to arbitrary precision.

The final part of the paper illustrates our algorithm by provides two illustrative RP applications. The first uses data from the consumer expenditure survey (CEX) to nonparametrically test for consistency of observed consumption behavior with the exponential discounted utility (EDU) model including idiosyncratic income shocks. Interestingly, we find that consumption behavior of singles are more likely to be rationalized by the EDU model with income shocks, compared to married couples (where both spouses work).

In our second application we revisit the problem of choice under risk and uncertainty within the classic expected utility theory. We implement a recently suggested model of Echenique, Imai, and Saito (2021) that allows for the possibility that experimental subjects are prone to misperception error, e.g. misperception in the (objective) probabilities for the different states, prices of the securities or due to the presence of a latent factor in the underlying (Bernoulli) utility function. We use data from Choi, Kariv, Müler, and Silverman (2014) to demonstrate how our method provides an estimate of the lower bound on the variance of the error that is necessary to rationalize a collection of data sets.

Literature overview There have been some earlier approaches to include stochastic features in RP analysis. Most notably Varian (1985) and Epstein and Yatchew (1985), develop a framework that allows for statistical hypothesis testing in RP tests with measurement error on demand. Echenique, Lee, and Shum (2011) follow a similar approach, but instead of measurement error on demand, they introduce measurement error on prices. They further require the marginal utility to be constant at the observed price levels. The strong limitation of these three studies is the assumption that the econometrician needs to have knowledge regarding the distribution of measurement error (e.g. normally distributed with known mean and variance), which seems somewhat at odds with the nonparametric flavor of the RP methodology.

More recently, Aguiar and Kashaev (2020) made an important contribution to the literature. They conduct RP analysis in the presence of (potentially non-classical) measurement error. Their test is based on the Entropic Latent Variable Integration via Simulation (ELVIS) methodology put forward by Shennach (2014). The ELVIS method uses maximum-entropy moments to analyse models with unobserved randomness that are characterized via moment conditions. Though this approach is versatile and has many attractive features, the implementation of the ELVIS-based test relies on numerical integration and might, as a result, be computationally demanding. Our approach, which is based on computing distances to the convex hull of the Minkowski average of a sample of sets might (in many instances) be faster to implement as it avoids the need for numerical integration. All results in this paper only involve repeatedly solving a collection of linear or quadratic programs. They were run on a standard laptop configuration[4] and all of them (except for the subsampling exercise in subsection 4.2) were conducted within a single day.

In the RP literature, stochastic revealed preference theory takes a related, yet quite different approach to include stochastic features within a nonparametric framework (McFadden and Richter, 1971; McFadden, 2005). This theory is based on the Random Utility model and studies whether it is possible to rationalize a given distribution of choices (from a repeated cross section of household consumption behavior) by a (stable) distribution of preferences.[5] In this paper, we are closer to the framework of Aguiar and Kashaev (2020) where we have the availability of a cross section of individual datasets where each dataset consists of multiple decisions of a single decision maker or household. Our framework is therefore more useful to analyze individual decision models that include random features like uncertainty or measurement error rather than randomness originating from individual heterogeneity.

[4] In particular, the laptop has an Intel(R) Core(TM) i5-10210U CPU with 2.11 GHz and 16.0 GB RAM.

[5] See Kawaguchi (2017), Kitamura and Stoye (2018), Cosaert and Demuynck (2018), Cherchye, Demuynck, and De Rock (2019) and Hubner (2022) for recent contributions.

However, see Smeulders, Cherchye, and de Rock (2021) who use a similar style column generating algorithm to make computational gains for RP analysis in a stochastic revealed preference setting.

As an important disclaimer, we would like to mention that the main contribution of our paper deals with the computational and algorithmic aspects. As such, we largely abstain from questions relating to statistical inference. Nevertheless, we refer to appendix B for a succinct discussion of the large sample properties, and possible avenues towards conducting statistical inference in our framework.

Overview Section 1 presents the random sets based framework to test models with unobserved heterogeneity and moment restrictions. Section 2 translates this framework to a finite sample setting. Section 3 gives our algorithm and discusses the convergence results. Section 4 contains two applications. Finally Section 5 contains a conclusion.

1 Framework

We start by presenting the general framework. The setting is based on those presented in Jovanovic (1989) and more recently in Li (2021). Both of these papers study general (structural) economic environments or 'models'. To fix ideas, we first present the basic elements of this framework.

We assume the econometrician has access to a collection of observable variables (e.g. consumer expenditures, prices, income, wages etc.), which are collected in the random (finite dimensional) vector y. The values of y are contained in a set $\mathcal{Y} \subseteq \mathbb{R}^{d_Y}$. Next, u is a finite dimensional random vector of unobservable or latent variables (from the point of view of the econometrician) which takes values in a set $\mathcal{U} \subseteq \mathbb{R}^{d_U}$. We denote by μ the (unknown) joint distribution of (y, u) on some measure space $(\mathcal{Y} \times \mathcal{U}, \mathcal{B})$ where $\mathcal{B}$ is the Borel σ-algebra on $\mathcal{Y} \times \mathcal{U} \subseteq \mathbb{R}^{d_U + d_Y}$. Clearly, this joint distribution contains all the relevant information regarding the interaction between observable variables and latent factors and is therefore a crucial object for any economic model. With these in hand, we can formally define what we mean by a model:

Definition 1 *A* model *consists of a tuple* $(\mu_Y, \Gamma, f, \alpha)$ *where*

- μ_Y *is the marginal probability measure of* μ *with respect to the observable variables* y*,*

- $\Gamma \subseteq \mathcal{Y} \times \mathcal{U}$ *is a* $\mathcal{B}$*-measurable set that gives all combinations* $(y, u) \in \mathcal{Y} \times \mathcal{U}$ *that are consistent with the economic model.*
- $f : \mathcal{Y} \times \mathcal{U} \to \mathbb{R}^K$ *gives a vector of measurable functions* $f = (f^1, \dots f^K)$ *that govern the moment conditions imposed by the economic model.*
- $\alpha = (\alpha^1, \dots, \alpha^K) \in \mathbb{R}^K$ *is a* K*-dimensional vector of moment values for these functions.*

This definition merits some discussion. First of all, a model $(\mu_Y, \Gamma, f, \alpha)$ imposes restrictions on the joint distribution μ on $\mathcal{Y} \times \mathcal{U}$. Given that the random vector y is observed (in principle), any model must satisfy the condition that the marginal distribution of μ with respect to y equals μ_Y, i.e. for all $A \in \mathcal{B}$:

$$\mu_Y(A) = \mu(A \times \mathcal{U}). \tag{1}$$

Next, in terms of interpretation, the set Γ encodes possible values for the variables (y, u) that can arise as an outcome of the model. In other words, Γ puts a restriction on the support for the random vectors (y, u).

$$\mu(\Gamma) = 1. \tag{2}$$

Finally, the functions $f^1, \dots, f^K$ and the values $\alpha^1, \dots, \alpha^K$ encode a (finite) set of moment conditions that also need to be satisfied:

$$\mathbb{E}_\mu f^k = \int_{\mathcal{Y} \times \mathcal{U}} f^k(y, u)\, \mathrm{d}\mu = \int_\Gamma f^k(y, u) \mathrm{d}\mu = \alpha^k, \text{ for all } k = 1, ..., K,$$

Using the notation $f = (f^1, \dots, f^K)$ and $\alpha = (\alpha^1, \dots, \alpha^K)$ we can succinctly write this as:

$$\mathbb{E}_\mu f = \alpha. \tag{3}$$

To summarize, we are interested in the existence of a joint probability distribution μ such that (1), (2) and (3) are satisfied.

As we will show in the second part of the paper, the main focus will be in applying this framework to analyze RP models with unobserved randomness. Towards this end, it might be useful to interpret y as a dataset of choices from a (random) consumer. The random latent vector u then captures the randomness that we would like to include in the underlying decision model, e.g. uncertainty or misperception. We can define Γ as consisting of all

vectors (y, u) that satisfy the corresponding RP test. In other words $(y, u) \in \Gamma$ if the particular dataset y satisfies the testable RP conditions given the particular value of u for the unobserved latent variables. In this setting, the moment conditions could then be used to capture additional requirements on the randomness in the model, for example the assumption that the average value of u is is zero, or that particular elements of u are uncorrelated with certain observables in y. We will tackle concrete examples of these in Section 4, when we study some specific RP applications of our formal setting.

Let $\mathcal{H}(\mu_Y, \Gamma)$ collect all the feasible distributions μ such that (1) and (2) are satisfied, i.e., μ_Y is the marginal distribution of μ with respect to y and $\mu(\Gamma) = 1$. Let us define the correspondence $F : \mathcal{Y} \rightrightarrows \mathbb{R}^K$:

$$F(y) = \{f(y, u) : (y, u) \in \Gamma\}. \tag{4}$$

Notice that for a given y, $F(y)$ is a subset of $\mathbb{R}^K$. In fact, as y is random, $F(y)$ is a random set. We impose some regularity conditions on F. In particular, we assume that the following is true throughout the paper:

Assumption 2

(i) The sets $F(y) = \{f(y, u) : (y, u) \in \Gamma\}$ are closed μ_Y-a.s.

(ii) There is a measurable function $g(y)$ that only depends on y such that $\mathbb{E}_\mu g(y) < \infty$ and:

$$g(y) \geq \sup_{(y,u)\in\Gamma} \|f(y, u)\| \qquad \mu_Y - a.s.$$

Where μ_Y-a.s. denotes that something is satisfies almost surely with respect to the marginal distribution μ_Y. Assumption 2 essentially imposes compactness on the sets $F(y)$. Given that the latter are contained in $\mathbb{R}^K$, this is equivalent to assuming they are closed and bounded almost surely. Assumption 2 is very similar to the assumptions made in Li (2021) and Beresteanu, Molchanov, and Molinari (2011).

Given the definition of $\mathcal{H}(\mu_Y, \Gamma)$, we can rewrite the existence of a distribution μ that satisfies

(1), (2) and (3), as follows:[6]

$$\min_{\mu \in \mathcal{H}(\mu_Y, \Gamma)} \mathbb{E}_\mu \| f(y,u) - \alpha \| = 0. \tag{5}$$

where $\|.\|$ is the Euclidean norm on $\mathbb{R}^K$.[7]

Support functions The basis of the further analysis relies on the concept of a *support function*, which was also used in the identification strategies in several earlier papers such as Ekeland, Galichon, and Henry (2010), Beresteanu, Molchanov, and Molinari (2011) and Li (2021).

For a compact set $A \subseteq \mathbb{R}^K$, the support function of A, $h_A : \mathbb{R}^K \to \mathbb{R}$ is given by:

$$h_A(\lambda) = \sup_{x \in A} \langle \lambda, x \rangle,$$

where $\langle ., . \rangle$ denotes the inner product.[8] We will use the notation $\text{co}(A)$ to denote the convex hull of A, that is, the smallest convex set containing A. Furthermore, we will denote by $\overline{\text{co}}(A)$ the convex, closed closure of A (i.e., the smallest convex and closed set that contains A). It is well known that $\overline{\text{co}}(A) = \text{co}(\overline{A})$ where $\overline{A}$ is the closure of A. The closed convex closure can be identified using its support function:

$$\overline{\text{co}}(A) = \{x \in \mathbb{R}^K : \forall \lambda \in \mathbb{S}^K, \langle \lambda, x \rangle \leq h_A(\lambda)\}. \tag{6}$$

where $\mathbb{S}^K = \{\lambda \in \mathbb{R}^K : \|\lambda\| = 1\}$ is the $K-1$ dimensional unit simplex. If condition (3) is satisfied, then by linearity of the expectation, we have that:

$$\mathbb{E}_\mu \langle \lambda, f(y,u) \rangle = \langle \lambda, \alpha \rangle, \qquad \forall \lambda \in \mathbb{S}^K \tag{7}$$

[6] A related, but somewhat weaker condition is that:

$$\inf_{\mu \in \mathcal{H}(\mu_Y, \Gamma)} \mathbb{E}_\mu \| f(y,u) - \alpha \| = 0.$$

This condition is the main focus for the analysis contained in Shennach (2014). The condition is generally weaker than (5), given that the infimum might not be achieved by any distribution $\mu \in \mathcal{H}(\mu_Y, \Gamma)$. However, as shown by Li (2021), the two are indistinguishable given Assumption 2.

[7] Note that we could extend our framework to allow for a set of parameters, say $\theta \in \Theta \subseteq \mathbb{R}^r$ and make Γ conditional on the value of θ. In that case, (5) would define the identified set for the parameters θ. This is the definition employed by, for example, Roehrig (1988), Li (2021) and Ekeland, Galichon, and Henry (2010). Given, however, that our main focus is on testing models, we will not explicitly include such (structural) parameters, so as to not overburden notation.

[8] Support functions are a frequently used tool in convex analysis. See (Rockafellar, 1970) for an in-depth treatment.

Next, by definition of the support function, one obtains from condition (2) that:

$$\mu\left(\left\{(y,u)\in\Gamma:\langle\lambda,f(y,u)\rangle\leq h_{F(y)}\left(\lambda\right)\right\}\right)=1, \qquad \forall\lambda\in\mathbb{S}^K. \tag{8}$$

Combining (7) and (8), we obtain:

$$\mathbb{E}_\mu\left[h_{F(y)}\left(\lambda\right)\right]\geq\mathbb{E}_\mu\langle\lambda,f\left(y,u\right)\rangle=\langle\lambda,\alpha\rangle \qquad \forall\lambda\in\mathbb{S}^K \tag{9}$$

For given $\lambda\in\mathbb{S}^K$, $h_{F(y)}$ is a function solely of y, and μ has support only on Γ. We therefore obtain $\mathbb{E}_\mu\left[h_{F(y)}(\lambda)\right]=\mathbb{E}_{\mu_Y}\left[h_{F(y)}(\lambda)\right]$. Consequently (9) implies that:

$$\mathbb{E}_{\mu_Y}\left[h_F(y)(\lambda)\right]\geq\langle\lambda,\alpha\rangle \qquad \forall\lambda\in\mathbb{S}^K. \tag{10}$$

In particular:

$$\inf_{\lambda\in\mathbb{S}^K}\mathbb{E}_{\mu_Y}\left[h_{F(y)}(\lambda)-\langle\lambda,\alpha\rangle\right]\geq 0.$$

As shown by Li (2021), an even stronger result holds.

Theorem 3 *If Assumption 2 holds, then (5) is satisfied if and only if:*

$$\inf_{\lambda\in\mathbb{S}^K}\mathbb{E}_{\mu_Y}\left[h_{F(y)}(\lambda)-\langle\lambda,\alpha\rangle\right]\geq 0. \tag{11}$$

The condition in (11) is demonstrated and used by Li (2021) to characterize sharp identified regions for a variety of partially identified econometric models. In our case, we will use this condition as a starting point to develop tests for several RP models with unobserved randomness.

There are two issues concerning the practical implementation of (11). First, the expected value in the expression is over the marginal density μ_Y. Although the latter can in principle be identified through data, in practice it needs to be estimated or approximated using a finite sample. Second, to compute the support function $h_{F(y)}$ we would need a succinct description of the random set $F(y)$, which can be challenging in practice. The next two sections will address both these issues and show how to operationalize condition (11).

Aumann expectation As shown by Molchanov (2017, Theorem 2.1.38), if μ_Y is non-atomic (which means that it does not have a point mass) then there exists a unique compact

convex set $\mathbb{E}F \subseteq \mathbb{R}^K$ such that:

$$\mathbb{E}_{\mu_Y} h_{F(y)}(\lambda) = h_{\mathbb{E}F}(\lambda).$$

The set $\mathbb{E}F$ is called the *Aumann expectation* of the random set $F(y)$. It can equivalently be obtained as the set of all expected values of all random variables that take values in $F(y)$ almost surely. The Aumann expectation $\mathbb{E}F$ can be shown to be closed and convex, which means that condition (11) can equivalently be expressed as:

$$\alpha \in \mathbb{E}F. \tag{12}$$

So the testable implication boils down to verifying if the vector α belongs to the Aumann expectation $\mathbb{E}F$.

2 Finite sample approximation

In this section, we assume that the econometrician has access to a (finite) i.i.d. sample from μ_y, say $\mathcal{D} = \{y_1, ..., y_n\}$. Using the standard finite sample analogue principle, we can replace (11) by its finite sample analogue:

$$\inf_{\lambda \in \mathbb{S}^K} \left[\frac{1}{n} \sum_{i=1}^{n} h_{F(y_i)}(\lambda) - \langle \lambda, \alpha \rangle \right] \geq 0, \tag{13}$$

We refer to appendix B for a short discussion on the validity of this finite sample approximation and some other statistical considerations.

Condition (13) replaces the expectation operator by the sample average over the i.i.d. draws. The expression in (13) can be simplified even further. In particular, let us denote by $\overline{F}_n = \frac{1}{n} \sum_i F(y_i)$ the *Minkowski (sample) average* of the sets $\{F(y_1), \ldots, F(y_n)\}$:[9]

$$\overline{F}_n = \left\{ \frac{1}{n} \sum_{i=1}^{n} f_i : f_i \in F(y_i) \right\} = \left\{ \frac{1}{n} \sum_{i=1}^{n} f(y_i, u_i) : (y_i, u_i) \in \Gamma \right\}. \tag{14}$$

Given that the sets $F(y_i)$ are compact (see Assumption 2), the set $\overline{F}_n$ is also compact. Moreover, as the support function is linear in the Minkowski sum, we have that (13) is

[9]Here and in the rest of the paper we will slightly abuse notation and also use f as a generic element of $F(y)$ (or sometimes co $(\overline{F}_n)$).

equivalent to:

$$\inf_{\lambda \in \mathbb{S}^K} \left[h_{\overline{F}_n}(\lambda) - \langle \lambda, \alpha \rangle \right] \geq 0. \tag{15}$$

The condition in (15) can in principle be tested, although a direct test based on this condition is not straightforward. The following gives an easier, equivalent condition (see Luenberger (1969, p.136) for a proof).

Proposition 4 *Condition (15) is satisfied if and only if* $\alpha \in \text{co}(\overline{F}_n)$. *In particular, If* $\alpha \notin \text{co}(\overline{F}_n)$ *then*

$$\inf_{\lambda \in \mathbb{S}^K} \left[h_{\overline{F}_n}(\lambda) - \langle \lambda, \alpha \rangle \right] = -d(\alpha, \text{co}(\overline{F}_n)).$$

where $d(\alpha, \text{co}(\overline{F}_n)) = \inf_{f \in \text{co}(\overline{F}_n)} \|f - \alpha\|$ *is the distance between the vector* α *and the set* $co(\overline{F}_n)$.

Proposition 4 provides an interesting geometric interpretation to condition (15). Checking whether or not the latter is satisfied amounts to verifying if the vector α is contained in the convex closed closure of the Minkowski average $\overline{F}_n$. This directly relates to condition (12) as $\text{co}(\overline{F}_n)$ is in fact the finite sample approximation of the Aumann expectation $\mathbb{E}F$.[10] Furthermore, the criterion in the expression (15) is equal to (minus) the distance between α and $\overline{F}_n$. It therefore suffices to compute the latter and verify whether it is equal to zero, i.e., we can rephrase (15) by:

$$d(\alpha, \text{co}(\overline{F}_n)) = \min_{f \in \text{co}(\overline{F}_n)} \|\alpha - f\| = 0. \tag{16}$$

The solution $\widehat{f}$ to the minimization problem (16) is the projection of α onto the convex set $\text{co}(\overline{F}_n)$. As $\text{co}(\overline{F}_n)$ is convex, this solution is unique. Though $\text{co}(\overline{F}_n)$ is not necessarily a hyperplane, we can use the following (standard) result in convex analysis to characterize this projection (see, for example, Luenberger (1969, p.69) for a proof)

Proposition 5 $\widehat{f}$ *solves (16) if and only if* $\widehat{f} \in \text{co}(\overline{F}_n)$ *and for all* $f \in \overline{F}_n$*:*

$$\langle \alpha - \widehat{f}, f - \widehat{f} \rangle \leq 0. \tag{17}$$

Verifying (17) is still difficult in practice. In particular, in most applications it is very difficult

[10] See also the discussion in Appendix B.

to find a closed form expression for $\text{co}(\overline{F}_n)$ and, as a consequence, we cannot directly find the projection of α on $\text{co}(\overline{F}_n)$. The next section provides a computationally fast algorithm to find the smallest distance between α and $\text{co}(\overline{F}_n)$ using a column generation procedure.

3 The column generating algorithm

Our aim is to solve (16), still operating under the assumption the econometrician only has access of a finite sample of i.i.d. observations for the observables, $\mathcal{D} = \{y_1, ..., y_n\}$.

The general idea of column generation as a method to solve (complex) optimization problems is to start by solving a restricted version of the latter, on a subset of the choice domain. In each iteration, the domain is expanded until there is insufficient improvement in the objective function. This technique has recently been fruitfully applied to nonparametric testing of random utility models in Smeulders, Cherchye, and de Rock (2021).

In the context of the present paper, column generation will take the following form: we start with a finite number of J draws from the Minkowski average $\overline{F}_n$.[11] The idea is that the number of elements J is initially relatively small (e.g., $J = 1$) but will grow as the algorithm proceeds. The convex hull of these points, i.e., the set $\text{co}(\{f_1, \ldots, f_J\})$ can then be considered as an inner approximation of $\text{co}(\overline{F}_n)$. At each iteration, we find the projection of α on this inner approximation and add (if necessary) a new informative element to the inner approximation to start the next iteration.

3.1 The Master and Auxiliary programs

We now formalize the idea behind our column generating algorithm. The algorithm is divided in two main programs, the first one, the Master program, will compute (for a given J) the smallest distance between the vector α and the inner approximation, $\text{co}(\{f_1, \ldots, f_J\})$ of $\text{co}(\overline{F}_n)$:

[11] In practice, we can obtain these by drawing random elements $u_{i,j}, i = 1, \ldots, n; j = 1, \ldots, J$ such that $(y_i, u_{i,j}) \in \Gamma$ for all $j = 1, \ldots, J$. We can then define $f_j = \frac{1}{n} \sum_i f(y_i, u_{i,j}) \in \overline{F}_n$.

$$
\begin{aligned}
\textbf{MASTER:} \qquad v_J = \min_{\substack{\eta \in \mathbb{R}^K \\ w_j \in \mathbb{R}}} \; & \|\eta\|^2, \\
\text{s.t. } & \eta + \sum_{j=1}^{J} w_j f_j = \alpha, \\
& \sum_{j=1}^{J} w_j = 1, \\
& w_j \geq 0, \qquad\qquad \forall j = 1, \ldots, J
\end{aligned}
$$

We can rewrite this master program more succinctly by defining $\widehat{f} = \sum_{j=1}^{J} w_j f_j$ (and therefore $\eta = \alpha - \widehat{f}$, one can think of η as the 'residual term' from projecting α on $\text{co}(\{f_1, \ldots, f_J\})$). This gives the equivalent formulation:

$$
\textbf{MASTER: } v_J = \min_{\widehat{f} \in \text{co}(\{f_1, \ldots, f_J\})} \|\alpha - \widehat{f}\|^2.
$$

The Master program is a quadratic minimization problem, which can be solved quite efficiently using standard software.[12]

There are now two possibilities. First, if the optimal solution v_J equals zero, i.e., $\widehat{f} = \alpha$ or equivalently $\eta = \alpha - \widehat{f} = 0$, then as $\text{co}(\{f_1, \ldots, f_J\}) \subseteq \text{co}(\overline{F}_n)$, we know that (16) also has an optimal solution equal to zero, so $\alpha \in \text{co}(\overline{F}_n)$ which means that (13) is satisfied, and we cannot reject the model's testable restriction. In the other case, we have $v_J > 0$. In this case, one cannot directly know whether one should continue the program using an enlarged inner approximation of $\text{co}(\overline{F}_n)$, or whether one can stop. The following result provides a way to distinguish between these possibilities:

Proposition 6 *Let $v_J > 0$ be the optimal solution of the MASTER program with optimal solutions η and $\widehat{f} = \sum_{j=1}^{J} w_j f_j$. Then $v_J = d(\alpha, \text{co}(\overline{F}_n))^2$ if and only if:*

$$
\frac{1}{n} \sum_{i=1}^{n} h_{F(y_i)}(\eta) \leq \langle \eta, \widehat{f} \rangle. \tag{18}
$$

The result in Proposition 6 easily follows from the characterization of the convex projection in Proposition 5. Proposition 6 is very useful to decide when to stop the algorithm. Indeed, in case the inequality in (18) is satisfied, we know that we have found the distance between

[12] Alternatively, more specialized algorithms are also available (Wolfe, 1976; Sekitani and Yamamoto, 1993).

α and $\text{co}(\overline{F}_n)$, which is what we need in order to test the model, cfr. (16). Now note that (18) can be rewritten as

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n} \sup_{f\in F(y_i)} \langle \eta, f\rangle \leq \langle \eta, \widehat{f}\rangle \\ \iff & \frac{1}{n}\sum_{i=1}^{n} \sup_{f\in F(y_i)} \langle \eta, f-\widehat{f}\rangle \leq 0. \end{aligned} \tag{19}$$

And the condition in (19) suggests a procedure to implement our algorithm, in particular, in case $v_J > 0$, within each iteration of the master program, we can check whether (19) is satisfied by solving the following auxiliary optimization problem:

$$\textbf{AUX:} \sup_{f\in\overline{F}_n} \langle \eta, f-\widehat{f}\rangle. \tag{20}$$

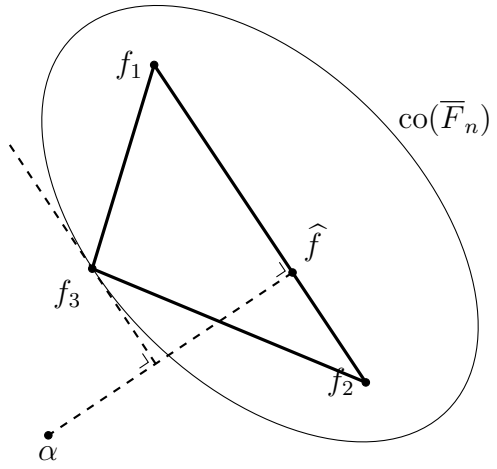
Note that going from (19) to (20), we have made use of the fact that the support function is linear in the Minkowski average. Now, we need to verify whether the value function in problem (20) is non-positive. In case it is, we are done as it means that condition (19) is satisfied and therefore $v_J = d(\alpha, \text{co}(\overline{F}_n))^2$. In this sense, we obtained a certificate of optimality. On the other hand, if the optimal value in (20) is positive, we can return to the Master program and resolve the projection problem, but now with the domain expanded to include an additional element f_{J+1} to the inner approximation $\text{co}(\{f_1, \ldots, f_J\})$. For this additional element, we take $f_{J+1} \in \arg\sup_{f\in\overline{F}_n}\langle \eta, f-\widehat{f}\rangle$, as this is in a sense the most plausible extra element to extend the inner approximation for $\text{co}(\overline{F}_n)$. Given the new set $\{f_1, \ldots, f_{J+1}\}$, we can return to the MASTER program for the next iteration of the algorithm. We should also note that, instead of solving (20), one can choose to solve n auxiliary programs, of the following form:

$$\begin{aligned} \textbf{AUX}(i)\textbf{:}\ h_{F(y_i)}(\eta) &= \sup_{f\in F(y_i)} \langle \eta, f\rangle, \\ &= \sup_{u\in\mathcal{U}} \langle \eta, f(y_i,u)\rangle \text{ s.t. } (y_i,u)\in\Gamma. \end{aligned} \tag{21}$$

And then one needs to check whether the average of all the values of the support functions is below $\langle \eta, \widehat{f}\rangle$. If it is, we are done again as (19) is satisfied. Otherwise, we again expand the domain for the original projection problem in the MASTER program by including he average of the solutions of the Auxiliary programs.

Figure 1 depicts the idea behind the algorithm. There is a convex set $\text{co}(\overline{F}_n)$, whose shape

Figure 1: Illustration algorithm



is only implicitly defined via the set Γ and the functions defining the moment restrictions, as summarized in f. We are interested in measuring the minimal distance between α and $\text{co}(\overline{F}_n)$. Assume the algorithm starts with $J = 2$, with elements $f_1, f_2 \in \overline{F}_n$. The convex hull of these points $\text{co}(\{f_1, f_2\})$ is the line segment between f_1 and f_2. The first step is to solve the MASTER program, which computes the projection of α onto $\text{co}(\{f_1, f_2\})$. This projection is denoted by $\widehat{f}$. The vector defined by the difference between α and $\widehat{f}$ is the vector η. Subsequently, the AUXILIARY programs will find the element $f_3 \in \overline{F}_n$ whose projection onto the vector $\eta = \alpha - \widehat{f}$ is largest. This is given by the vector that solves $\langle \eta, f_3 \rangle = h_{\overline{F}_n}(\eta)$. Next, we update the inner approximation to $\text{co}(\{f_1, f_2, f_3\})$ and repeat the procedure. It is clear to see that at every iteration of the algorithm, this set will expand, and the new distance $\|\eta\| = \|\alpha - \widehat{f}\|$ will decrease.

The figure also makes it clear that the distance between α and $\overline{F}_n$ will equal $\|\alpha - \widehat{f}\|$ if and only if it is impossible to find a vector f_3 for which the projection of $f_3 - \widehat{f}$ onto $\eta = \alpha - \widehat{f}$ is positive, i.e., if the angle between $\eta = \alpha - \widehat{f}$ and $f_3 - \widehat{f}$ is obtuse.

In the next section, we explicitly formulate the algorithm and derive its convergence properties.

3.2 Convergence properties

In this section, we will show that the algorithm presented in the previous section indeed approximates the value of $d(\alpha, \text{co}(\overline{F}_n))$ arbitrarily close, and show its convergence properties.

To start, Algorithm 1 gives the algorithm presented in Section 3.

Algorithm 1 Column generation program

Require: $0 < \delta < 1/2$ and $0 < \varepsilon < 1$

1: $J \leftarrow 0$
2: $D_J = \{f_0\}$ for some $f_0 \in \overline{F}_n$
3: $dist \leftarrow 0$
4: **while** True **do**
5: $\quad \widehat{f}_J \leftarrow \arg\min_{f \in D_J} \|\alpha - f\|^2$ ▷ Master program
6: $\quad v_J \leftarrow \|\alpha - \widehat{f}_J\|^2$
7: $\quad \eta_J = \alpha - \widehat{f}_J$
8: $\quad$ **if** $v_J < \varepsilon$ **then**
9: $\quad\quad$ Return 0
10: $\quad$ **else**
11: $\quad\quad f_{J+1} \leftarrow \arg\max_{f \in \overline{F}_n} \left\langle \eta_J, f - \widehat{f}_J \right\rangle$ ▷ Auxiliary program
12: $\quad\quad$ **if** $\langle \eta_J, f_{J+1} - \widehat{f}_J \rangle > \delta \|\eta_J\|^2$. **then**
13: $\quad\quad\quad D_{J+1} = \text{co}(D_J \cup \{f_{J+1}\})$
14: $\quad\quad\quad J \leftarrow J + 1$
15: $\quad\quad$ **else**
16: $\quad\quad\quad$ Return $\|\eta_J\|$
17: $\quad\quad$ **end if**
18: $\quad$ **end if**
19: **end while**

Lines 1-3 initialize the algorithm. Here we start with a singleton $D_0 = \{f_0\}$, however, in practice, we might use any finite set of elements from $\overline{F}_n$. Then it goes into a loop. Line 5 solves the Master program after which it defines the solution $\widehat{f}_J, \eta_J$ and the optimal value v_J. On line 8 it checks if v_J is smaller than ε if so, the algorithm stops having found a distance below the maintained threshold of ε.

If not, it solves the Auxiliary program and saves its solution in f_{J+1}. This can be done by setting:

$$f_{J+1} = \frac{1}{n} \sum_{i=1}^{n} f(y_i, u_i),$$

where for all $i \in \{1, \ldots, n\}$:

$$u_i \in \arg\max_{u \in \mathcal{U}} \langle \eta, f(y_i, u) \rangle \text{ s.t. } (y_i, u) \in \Gamma.$$

Then the algorithm checks if the angle between η_J and $f_{J+1} - \widehat{f}_J$ is close enough to satisfying the condition of Proposition 5. If so, it quits the algorithm with optimal value $\|\eta_J\|$. Else, it updates D_J to $D_{J+1} = \text{co}(D_J \cup \{f_{J+1}\})$ and increments for the next iteration.

Notice that if at some iteration, for some f_{J+1}, we have $\langle \eta_J, f_{J+1} \rangle < \langle \eta_J, \alpha \rangle$, then we can exclude the case $d(\alpha, \text{co}(\overline{F}_n)) = 0$. Indeed, in such case, we have that the hyperplane given by slope η_J through f_{J+1} is a supporting hyperplane of $\text{co}(\overline{F}_n)$ that strictly separates α from $\text{co}(\overline{F}_n)$. So if the goal is to determine whether $\alpha \in \text{co}(\overline{F}_n)$ or not (without computing the actual distance) one can easily add this additional check to the algorithm.

The following Theorem show that if the algorithm stops, then its solution approximates $d(\alpha, \text{co}(\overline{F}_n))$ arbitrarily close for δ sufficiently small.

Theorem 7 *If Algorithm 1 stops (say at iteration J) then, either $d(\alpha, \text{co}(\overline{F}_n)) < \sqrt{\varepsilon}$ (if it stops at line 9) or:*

$$\|\alpha - \widehat{f}_J\|^2 \leq \frac{\left[d(\alpha, \text{co}(\overline{F}_n))\right]^2}{1 - 2\delta},$$

if it stops at line 16.

As such, if δ and ε are chosen sufficiently small, the algorithm will approximate the true distance arbitrarily close. In our implementation, we choose $\varepsilon = 10^{-6}$ and $\delta = 10^{-3}$.[13]

The following Theorem shows that, as long as the algorithm runs, the squared distance $\|\alpha - \hat{f}_J\|^2$ decreases exponentially fast over the iteration. In particular, it demonstrates that the algorithm will stop in finite time.

Theorem 8 *There is a $\theta \in (0, 1)$ such that if the algorithm runs at least J iterations, then:*

$$\|\alpha - \widehat{f}_J\|^2 \leq \theta^J \|\alpha - \widehat{f}_0\|^2.$$

3.3 Solving the Auxiliary programs

The Master program is a simple quadratic minimization program which can be solved efficiently using standard solvers. The auxiliary program however might be somewhat more involved. In practice, the auxiliary program is solved using n maximization programs:

$$\max_{u \in \mathcal{U}} \langle \eta, f(y_i, u) \rangle \text{ s.t. } (y_i, u) \in \Gamma, \qquad \forall i = 1, \ldots, n$$

The simplest setting occurs when $f(y_i, u)$ is a linear function of u and when Γ can be expressed as a set of inequalities, linear in u (conditional on y_i). For such instances, the

[13] Here we are guided by the default tolerances imposed by the Gurobi solver which are 10^{-6}.

problem turns into a simple linear program, which can be solved using standard optimization software. Our first application fits this structure.

Even if the moment functions are quadratic in u, it might still be possible to use quadratic programming to solve the auxiliary problem as we will demonstrate in our second application.

In settings where f is highly nonlinear or if Γ takes on a more complicated structure, other optimization routines must be used. One solution might be to use a nonlinear optimization solver that can also deal with non-linear constraints. Another approach to such instances might be to sample for each observation i a large number of elements u for which $(y_i, u) \in \Gamma$ and to pick the point where the objective $\langle \eta, f(y_i, u) \rangle$ is largest. This gives a lower approximation which will usually improve with the number of sampled points. Such sampling might be performed using a suitable Markov chain Monte Carlo methods that have shown to be quite versatile (and efficient) for sampling points from regions determined by RP constraints (Demuynck, 2021).

4 Applications

The description of a model given in this paper (defined by the correspondence Γ and moment conditions) is relatively general and therefore broad enough to capture several applications in the literature. In this section, we discuss two such applications to revealed preference analysis, which can serve as an illustration for the versatility of the column generation approach.

4.1 Intertemporal consumption with income uncertainty

Our first application revisits the classical setting of intertemporal consumption choice under uncertainty. We consider the standard setting where a household decides between consumption and savings over time and that every household is an exponential discounted utility (EDU) maximizer. The EDU model is clearly a workhorse for many different economic settings. Although intertemporal consumption and savings has received attention in the revealed preference literature, most of this literature focuses on settings without uncertainty (Browning, 1989; Crawford, 2010; Demuynck and Verriest, 2013; Adams, Cherchye, De Rock, and Verriest, 2014).[14]

An important remark is that adding uncertainty to these models without additional struc-

[14] See Aguiar and Kashaev (2020) for an exception.

ture on the shocks yields extremely limited or even vacuous testable restrictions. Indeed, any consumer behavior that is rational in the sense of maximizing the per period utility function, can then also be rationalized as an intertemporal utility maximizer by appropriate incorporation of some random (large enough) unobserved shocks. Therefore, it is necessary to impose some basic structure on latent income shocks. To be specific, we will consider the following environment: consumers have rational expectations, income shocks are i.i.d. over time and there are no aggregate shocks.[15] Furthermore, as standard for nonparametric analysis of consumer behavior, we assume stable (but possibly heterogeneous) preferences for consumers over the relevant time window.

To fix ideas, consider a consumer (household) i with a concave, continuous and monotone instantaneous utility function U^i and discount rate β that chooses consumption amounts according to the standard infinite horizon EDU model:

$$\max_{(c_t^i)_{t=0,1,\dots}} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t U^i\left(c_t^i\right)\middle|\mathcal{I}_0^i\right]$$

$$\text{s.t. } s_{t+1}^i = m_t^i + (1+r_t)s_t^i - \langle p_t^i, c_t^i\rangle, \tag{22}$$

$$s_0^i \text{ given.} \tag{23}$$

Here, c_t^i is the consumption vector of i at period t, $(m_t^i)_{t\in\mathbb{N}}$ is a random income stream for individual i, p_t^i are vectors of time varying (possibly individual specific) strictly positive prices, r_t denotes the interest rate at time t and s_t^i are savings of individual i at time t. Finally, $\mathcal{I}_t^i$ refers to the information set at period t, i.e., all the information available to consumer i, at the start of period t. The Bellman equation for this problem is given by:

$$V^i(s_t^i) = \max_{c_t^i}\left\{U^i(c_t^i) + \beta\mathbb{E}\left[V^i(m_t^i + (1+r_t)s_t^i - \langle p_t^i, c_t^i\rangle)\middle|\mathcal{I}_t^i\right]\right\}$$

Using standard arguments, one can show that this value function is concave. The first order and envelope conditions (for good j) are given by:[16]

$$\frac{\partial U^i(c_t^i)}{\partial c_j} = \beta\mathbb{E}\left[\widehat{\lambda}_{t+1}^i|\mathcal{I}_t^i\right]p_{t,j}^i,$$

$$\widehat{\lambda}_t^i = \beta(1+r_t)\mathbb{E}[\widehat{\lambda}_{t+1}^i|\mathcal{I}_t^i].$$

[15] There is no generally agreed framework in the literature on how to deal with aggregate shocks econometrically, and is beyond the scope of our paper. A recent contribution tackling estimation and testing of models in the presence of aggregate shocks, within a parametric context is Hahn, Kuersteiner, and Mazzocco (2020).

[16] In case u^i is not differentiable, we take corresponding subdifferentials.

where we used the notation $\widehat{\lambda}_t^i = \frac{dV^i(s_t^i)}{ds_t^i}$. If we define discounted prices $\rho_t^i = \frac{p_t^i}{\Pi_{k=1}^t(1+r_k)}$ and $\lambda_t^i = \beta^{t+1}\widehat{\lambda}_t^i \prod_{k=0}^t (1+r_k)$ and substitute the second condition into the first, we obtain the following standard form:

$$\frac{\partial U^i(c_t^i)}{\partial c_j} = \beta^{-t}\lambda_t^i \rho_{t,j}^i, \tag{24}$$

$$\mathbb{E}\left[\lambda_t^i - \lambda_{t+1}^i | \mathcal{I}_t^i\right] = 0. \tag{25}$$

Equation (24) gives the first order condition for consumption. Equation (25) is the equivalent optimality condition for savings. Together they yield a standard set of Euler equations for the intertemporal consumption model. Importantly, we note that (25) is a (conditional) moment condition and implies that any element in $\mathcal{I}_t^i$ must be orthogonal to the growth in marginal utility of income, $\lambda_{t+1}^i - \lambda_t^i$. From this conditional moment condition, we can derive a large set of implied unconditional moment conditions. In particular, for any variable $x_t^i \in \mathcal{I}_t^i$ we need that $\mathbb{E}\left[(\lambda_{t+1}^i - \lambda_t^i)x_t^i\right] = 0$.

We now turn to the rationalizability question. We first assume a large population of households from which we observe a random sample of n households with their (discounted) prices and their (consumption) choices, say $\mathcal{D}^i = \{(\rho_t^i, c_t^i)\}_{t\in\mathcal{T}}$, $i = 1, ..., n$ and where $\mathcal{T} = \{1, ..., T\}$ and $T \in \mathbb{N}_0$.

We first introduce a notion of *rationalizability* conditional on the marginal utility levels. In particular, let $\Lambda^i = (\lambda_t^i)_{t\in\mathcal{T}}$ denote a vector of strictly positive numbers representing the marginal utilities at different time periods.

Definition 9 *The data set $\mathcal{D}^i = \{(\rho_t^i, c_t^i)\}_{t\in\mathcal{T}}$ is said to be rationalizable by the EDU model given $\Lambda^i = (\lambda_t^i)_{t\in\mathcal{T}}$ (and β) if there exists a concave, locally non-satiated and continuous utility function U^i such that (24) is satisfied for the given marginal utility values in Λ^i.*

Given this, we obtain the following well-known result in RP analysis:[17]

Lemma 10 *Let $\beta \in (0,1)$, let $\mathcal{D}^i = (\rho_t^i, c_t^i)_{t\in\mathcal{T}}$ be a dataset and let $\Lambda^i = (\lambda_t^i)_{t\in\mathcal{T}}$ be a set of strictly positive numbers. Then the following are equivalent.*

- *$\mathcal{D}^i$ be rationalizable by EDU given $\Lambda^i = (\lambda_t^i)_{t\in\mathcal{T}}$.*

[17] We omit the proof as it is almost identical to the one in Browning (1989).

- *There are numbers* $(U_t^i)_{t\leq T}$ *such that for all* $t, v \leq T$*:*

$$U_t^i - U_v^i \leq \frac{\lambda_v^i}{\beta^v}\langle \rho_v^i, (c_t^i - c_v^i)\rangle. \tag{26}$$

The system of inequalities in (26) are known as *Afriat inequalities* (Afriat, 1967) and are obtained from combining concavity of the utility function U^i, together with the first order conditions. Obviously, in practice, we do not observe Λ^i, so one could say that Lemma 10 has little practical relevance. One solution would be to study whether there exists at least one instance Λ^i for which (26) holds. However, this amount of flexibility limits the additional bite present in intertemporal models. Indeed, such a loosening of rationalizability would lead to the same testable implications as for static utility maximization (i.e., GARP).[18] In this sense, the dynamic intertemporal utility model with uncertainty imposes no additional restrictions compared to per-period static utility maximization. However, we can leverage the additional set of conditions given by (25) to obtain additional bite.

In order to make the connection with our general framework (Section 1), our collection of observable random variables y correspond to all the observed variables in the datasets $\mathcal{D}^i$ where the randomness is obtained from the population (i.e., i is a draw from the population of individuals). Next, the latent (unobserved) variables u correspond to the vector of marginal utilities $\Lambda^i = (\lambda_t^i)_{t\in\mathcal{T}}$. We can again use the Afriat inequalities as given by (26) to define the set Γ of all pairs (u, y) that are consistent with the EDU model, in particular:

$$\Gamma = \left\{ \left(\mathcal{D}^i, \Lambda^i\right) \middle| \exists (U_t^i)_{t\leq T} : U_t^i - U_v^i \leq \frac{\lambda_v^i}{\beta^v}\rho_v^i \left(c_t^i - c_v^i\right), \text{ for all } t, v \in \mathcal{T} \right\}.$$

In order to make the correspondence Γ satisfy Assumption 2, we can limit $\lambda_t^i \in [\lambda_L, \lambda_U]$ for some lower bound λ_L and upper bound λ_U. As is clear from the Afriat inequalities (26), imposing a lower or upper bound is without loss of generality as both variables $(U_t^i)_{t\in\mathcal{T}}$ and $(\lambda_t^i)_{t\in\mathcal{T}}$ can be scaled by a common factor. Imposing both an upper and lower bound, however, might put restrictions on the Afriat inequalities. In our application, we will have $T = 4$ and so setting $\lambda_L = 1$ and $\lambda_U = 10$ seems to be a sufficiently large margin. Finally, we need to specify a collection of moment functions f. These will be derived from the Euler conditions (25). As the information sets $\mathcal{I}_t^i$ might be very big, this condition contains indeed a large (potentially infinite) number of implied moment conditions.

[18] See Varian (1982).

In particular, we impose several conditions of the following form:

$$\mathbb{E}\left[(\lambda_{t+1}^i - \lambda_t^i)x_{k,t}^i\right] = 0. \tag{27}$$

where $x_{k,t}^i$ is some variable in the information set $\mathcal{I}_t^i$ $(k = 1, \ldots, K)$. As $T = 4$, this gives for every k, 3 moment functions: $f_t^k(\mathcal{D}^i, \Lambda^i) = (\lambda_{t+1}^i - \lambda_t^i)x_{k,t}^i$. Our algorithm is quite flexible to increase the number of moment conditions as long as they are linear in the latent variables $\Lambda^i = (\lambda_t^i)_{t\in\mathcal{T}}$ as this means that the AUXILIARY programs (21) takes the form of a linear program (conditional on β). A first, rather obvious choice is $x_{1,t}^i = 1$, which gives:

$$\mathbb{E}\left[\lambda_{t+1}^i - \lambda_t^i\right] = 0, \text{ for all } t \leq T-1, \tag{28}$$

This requires that the mean difference between λ_{t+1}^i and λ_t^i should be zero for all $t \leq T-1$.

To be specific, our MASTER program takes the following form:

$$\begin{aligned}
\textbf{MASTER:}\ & \min_{\eta_{k,t}} \|\eta^2\| \\
& \text{s.t.} \sum_{j=1}^{J} w_j (f_j)_t^k + \eta_{k,t} = 0 && \forall k \leq K, t \leq T \\
& \sum_{j=1}^{J} w_j = 1 \\
& w_j \geq 0, && \forall j = 1, \ldots J.
\end{aligned}$$

Notice that here $\alpha = 0$ as the right hand sides of all moment conditions equal zero. Once we have solved this model, we first check if $\|\eta\|^2 \approx 0$ and if so, we can immediately stop as then $0 \in \text{co}(\overline{F}_n)$. If $\|\eta\|^2 > 0$, then recall that we have to solve a number of auxiliary problems, equal to the size of the cross-section. In particular, for each dataset $\mathcal{D}^i$ we solve the following problem.

$$\begin{aligned}
\textbf{AUXILIARY}(i)\textbf{:}\ & \max_{\lambda_t^i, U_t^i} \sum_{t=1}^{T-1} \sum_{k=1}^{K} \eta_{k,t}(\lambda_{t+1}^i - \lambda_t^i)x_{k,t}^i \\
& \text{s.t. } U_t^i - U_v^i \leq \frac{\lambda_v^i}{\beta^v}\rho_v^i(c_t^i - c_v^i) && \forall t, v \leq T \\
& \lambda_t^i \in [\lambda_L, \lambda_U] && \forall t \leq T
\end{aligned}$$

It is easily to see that AUXILIARY(i) is a linear program (for a given β). In sum, every iteration of the algorithm requires us to solve one quadratic (the MASTER) and n linear

programs (the AUXILIARY programs) and is therefore computationally quite efficient.[19]

4.1.1 Data

We use the Consumer Expenditure Survey (CEX) for the period 1991Q1-2018Q4. This dataset is constructed by the Bureau of Labor Statistics and contains a rich set of measures for expenditures across detailed consumption categories, earnings and demographic characteristics. The CEX has a rotating panel feature, in which about 20 % of the interviewed addresses is new to the sample. All households are selected to be representative for the US population. Importantly for our purposes, the CEX has a (restricted) panel dimension. In particular, respondents are interviewed for a maximum of 5 times. The preliminary interview in which essential CEX recording procedures are explained. They are also asked to keep records of their expenditures. Respondents are traced for a maximum of 4 consecutive quarters. In each of these subsequent interviews, the respondents are asked to provide expenditure details regarding the last 3 months prior to the interview date. In our application, we use data from the Interview part of the CEX, in particular the detailed consumption expenditure (MTBI) files. These files consist of detailed monthly expenditures on a plenitude of consumption categories, and records these expenditures for a span of 3 months before the interview date for the specific wave took place. We follow Blundell et al. (2008) and Cosaert and Demuynck (2018) in grouping the different expenditure categories within three classes: *food*, *nondurables* and *services*. For food, we consider both food away from home and food at home. For nondurables, we aggregate expenditures on clothing, tobacco and alcohol. Finally, for services we aggregate expenditures on fuel, oil, gasoline, other energy services and public transport (fares). These expenditures are on a monthly frequency, whereas most of the other available information regarding households (including salary etc.) is measured at a quarterly frequency. Following Attanasio and Weber (1995) and Mazzocco (2007), we scale these monthly expenditures to a quarterly frequency by considering the monthly expenditures for the month just preceding the interview and multiply these by 3. To obtain prices for these consumption categories, we first gathered associated Consumer Price Indices published by the BLS at the most disaggregated level possible. For each of our 3 consumption classes we then construct a price index which is a geometrically weighted average of the CPI's of its constituent consumption categories. The weights attached to each category is its relative expenditure share within the broader consumption class to which it is categorized. With regards to (gross) interest rates, we used municipal bond rates as published in the economic

[19] The quadratic and linear programs are solved using the commercial GUROBI software, which can be used to solve large scale linear, quadratic and integer programming software. The software is free for academic use.

reports to the president.

In addition to the 3 consumption categories, we also include *leisure* as an additional consumption good. In the CEX there is only information regarding (quarterly) labor supply. To impute leisure, we use a similar procedure as in Mazzocco (2008) and set quarterly available (productive) time to an individual 1,092 hours,[20] and then subtract the individual's labor supply to obtain a measure of quarterly leisure. For wages, we use the detailed member information (MEMI) files from the Interview Survey, which contains salaries at different frequencies, which we first convert in a consistent quarterly measure. Wages are then imputed by salaries divided by quarterly labor supply.

To obtain our final dataset, we impose the following sample selection restrictions: we only keep respondents that are between 19 and 70 years old, who live in urban areas and who have non-missing and non-zero consumption of all our consumption categories. Note that, in contrast to many other papers that use the CEX to study (intertemporal) consumption and savings behavior of individuals, we do not add any further restrictions regarding family composition and/or marital status. There are several reasons to allow for a wide range of (observed) heterogeneity across consumers in our dataset. First, given that our setting is fully nonparametric, we do not require any (parametric) structure on utilities. Consequently, identification arguments are not based on the plausibility of these parametric assumptions which are often dependent on further sample restrictions. Next, our general framework as outlined in this paper (Section 1) relies on a set of moment restrictions. In our specific application these take the form as in (27). Our test based on these restrictions will become stronger with its length, and in that sense having more heterogeneity across observables is useful for our purposes.

4.1.2 Results

In order to be able to apply our RP analysis, we first need to be sure that there is at least one value $\Lambda^i = (\lambda^i_t)_{t\in T}$ for which the Afriat conditions in Lemma 10 hold. As this is not always the case, we first look for the largest $e \leq 1$ for which

$$U^i_t - U^i_v \leq \frac{\lambda^i_v}{\beta^v}\langle \rho^i_v, (c^i_t - e \cdot c^i_v)\rangle \qquad \forall t, v \leq T.$$

[20] This amounts to around 15 productive hours per day.

This value corresponds to the so called Afriat critical cost efficiency index in RP analysis. Then if $e < 1$ we conduct the test conditional on this value, replacing c_v^i by $e \cdot c_v^i$ where necessary.

Concerning the moment conditions, recall that the functions f^k are of the form:

$$(\lambda_{t+1}^i - \lambda_t^i) x_{k,t}^i. \tag{29}$$

In our dataset, we observe households for 4 subsequent periods ($T = 4$), so t ranges from 1 to 3 and for x we choose several covariates, known at period t. First, taking $x_t^i = 1$ gives the usual condition that expected marginal utility should not change. In addition we pick for x_t^i the period t prices and quantities of the 3 good categories, the age of the hh-head, the education level of the hh head (0,1,2 according to low, medium or highly educated) and the expenditure levels of the three categories. This gives us a total of $3 \times 12 = 36$ moment conditions. We standardize all covariates $x_{k,t}^i$ to make sure that all moment conditions get more or less equal weight in the algorithm.

For the implementation, we first divide our sample of households according to their status: single, married and both working and married and only one working. Table 1 gives some summary statistics. Consumption patterns are (on average) very equal across the different household types. Expenditures on each category (food, services and non-durables) are close to 1/3 of the total expenditure. Leisure time of the household head is more or less equal between married (both working) and singles and quite higher for couples where only one of the two spouses is working. The wage is lowest for singles and highest for couples where both are working. For each of the summary statistics, we see that there is considerable variation which is a setting that is particularly tailored for RP analysis.

We conduct our RP test within each of the 3 categories (married and both working, married and one working and singles) and according to the date of the first household observations. This makes it that within each of such subgroup, all time periods are aligned. As we have data from 1991 until 2012, this gives us 84 subsamples per category. For the analysis below, we discard subsamples of size below 40.

For each of these subsamples we run our algorithm. The output gives for each of these subsamples a pass or no pass depending on whether the subsample passes the test or not. Finally, we regress the incidence of rejecting consistency with the model on a dummy indicating the category controlling for the number of observations in the subsample. We do this for values of the discount rate equal to $\beta = 1, 0.99, 0.97$ and 0.95.

Table 2 contains the results. First of all, a larger sample size significantly leads to less

Table 1: Summary Statistics (means)

	married	one working	singles
N	5496	3637	6136
expenditure share food	0.334	0.335	0.337
std	0.075	0.079	0.081
expenditure share nondurables	0.331	0.331	0.330
std	0.072	0.077	0.077
expenditure share services	0.335	0.333	0.333
std	0.074	0.075	0.079
leisure (hours/quarter) of hh head	540.32	883.01	546.39
std	128.27	280.51	121.28
wage of hh head	21.450	18.337	17.546
std	26.00	23.65	16.52

Table 2: Regression results: likelihood to reject model

	$\beta = 1$	$\beta = 0.99$	$\beta = 0.97$	$\beta = 0.95$
constant	0.7145	0.7139	0.5711	0.5507
	(0.1654)	(0.1659)	(0.1610)	(0.1612)
married, both working	0.1403	0.1514	0.1234	0.1189
	(0.0678)	(0.0672)	(0.0650)	(0.0636)
married, one working	-0.1060	-0.0977	-0.0824	-0.0692
	(0.0935)	(0.0933)	(0.08820)	(0.0879)
subsample size	-0.0086	-0.0087	-0.0067	-0.0066
	(0.0025)	(0.0025)	(0.0025)	(0.0025)
R^2	0.0837	0.0906	0.0620	0.0608
N	188	188	188	188

robust standard errors between brackets

rejections of the model. This might indicate that our sample size is somewhat too small and that the model tends to reject a bit too much in small samples. Next, more interestingly, it seems that for married couples where both are working, we reject the model more often. This is in agreement with the literature on collective models of household economics that states that multi-person households do not behave as single individuals. Interestingly, we do not find this when looking at households where only one of the two spouses is working.

4.2 Approximate expected utility maximization

For our second illustration, we study the testable restrictions implied by expected utility theory (EUT), in a nonparametric RP framework. To be more precise, assume consumers make choices over Arrow-Debreu securities over a finite set of states, $\{1, \ldots, S\}$. We assume consumers are risk averse and that they each have a C^1 concave Bernoulli utility function, $U^i : \mathbb{R}_+ \to \mathbb{R}$. We will assume a (lab) experimental setting where the probabilities over states are controlled by the econometrician (hence objective), however, the consumer is allowed to have some misperception of these probabilities, meaning the consumer is maximizing subjective expected utility as in Savage (1954). The choice problem for consumer i over the securities can then be represented as follows:

$$\max_{(x_s^i)_{s\in\mathcal{S}}} \sum_{s=1}^{S} \widetilde{P}_s^i U^i(x_s^i) \text{ s.t. } \sum_{s\in\mathcal{S}} p_s x_s^i \leq m.$$

Here $\widetilde{P}_s^i$ is the subjective probability attached to state s, and p_s is the price of the Arrow-Debreu securities for state s.[21] The values $(x_s^i)_{s\leq S}$ give the amounts of the Arrow-Debreu securities over the states in $\mathcal{S}$. The first order conditions give that:

$$\frac{dU^i(x_s^i)}{dx} = \frac{\lambda^i p_s}{\widetilde{P}_s^i} \qquad \forall s \leq S$$

where λ^i is the Lagrange multiplier for the budget constraint. Taking logs on both sides, we

[21] Although the probabilities $\widetilde{P}_s^i$ must add up to one, we will ignore this, as it is in fact a non-testable restriction. Indeed, notice that we can always normalize the utility function to make sure that probabilities add up to one without changing the optimal choices x_s^i.

obtain:

$$\ln\left(\frac{dU^i(x_s^i)}{dx}\right) = \mu^i + \rho_s - \widetilde{\pi}_s^i \qquad \forall s \leq S$$

where we used the notation $\mu^i = \ln(\lambda^i)$, $\rho_s = \ln(p_s)$ and $\widetilde{\pi}_s^i = \ln\left(\widetilde{P}_s^i\right)$. From the point of view of the econometrician, this framework includes unobserved randomness due to the possibility of misperceptions in the (objective) probabilities by the experimental subjects.

This model is quite similar to Echenique, Imai, and Saito (2021), who studies the minimal error to perceived probabilities required to be able to rationalize choices by EUT. In contrast to our setting, however, they consider a non-stochastic framework. Interestingly though, they show that the model is formally equivalently to a model with minimal (multiplicative) perturbations to prices or to the Bernoulli utility function. In our framework, these different specifications could potentially be distinguished by different sets of moment restrictions.

We will remain in the setting of allowing for errors in perceived probability, and assume that there are (objective) true probabilities P_s that deviate from the subjective ones in the following sense:

$$P_s^i = \widetilde{P}_s^i \exp(\varepsilon_s^i).$$

In other words, we define

$$\varepsilon_s^i = \widetilde{\pi}_s^i - \pi_s^i.$$

where $\pi_s^i = \ln(P_s^i)$. We then get the convenient additive form:

$$\ln\left(\frac{dU^i}{dx}(x^i)\right) = \mu^i + \rho_s - \pi_s - \varepsilon_s^i \qquad \forall s \leq S \tag{30}$$

We assume that the values ε_s^i are unobserved by the econometrician. Now consider a data set $\mathcal{D}^i = \left\{(\pi_{s,t}, \rho_{s,t}, x_{s,t}^i)\right\}_{t\leq T, s\leq S}$, where t denotes an observation. We denote by $\Xi^i = \left\{(\varepsilon_{s,t}^i)\right\}_{s\leq S, t\leq T}$ the set of (potential) errors. We then introduce the following notion of rationalizability:

Definition 11 *We say that the dataset $\mathcal{D}^i$ is expected utility (EU) rationalizable for Ξ^i if there exists a C^1 concave utility function U^i such that condition (30) is satisfied for all $t \leq T$ and $s \leq S$.*

Note that rationalizability is defined conditional on the set of errors. The following result provides a characterization for EU rationalizability:

Lemma 12 *Given* $\Xi^i = \left\{(\varepsilon^i_{s,t})\right\}_{s\leq S, t\leq T}$ *and the dataset* $\mathcal{D}^i = \left\{(\rho_{s,t}, \pi_{s,t}, x^i_{s,t})\right\}_{t\leq T, s\leq T}$. *The following are equivalent:*

- *The dataset* $\mathcal{D}^i$ *is rationalizable for* Ξ^i,
- *There exists a vector* μ_t *such that for all observations* $t, v \leq T$ *and all states* $s, w \leq S$*:*

$$\begin{aligned} &\textit{if } x^i_{s,t} \geq x^i_{w,v} \\ &\textit{then } \mu_t + \rho_{s,t} - \pi_{s,t} - \varepsilon^i_{s,t} \leq \mu_v + \rho_{w,v} - \pi_{w,v} - \varepsilon^i_{w,v}. \end{aligned} \tag{31}$$

A similar result has been derived and proven in Echenique and Saito (2015) and Echenique, Imai, and Saito (2021).[22] The Afriat-style inequalities in (31) are directly related to the diminishing marginal utility of any Bernoulli map that EU-rationalizes a dataset $\mathcal{D}^i$. Indeed, if consumption increases across two particular states and choice instances, then the marginal utility of consumption decreases. It is immediately clear that if we do not observe $\Xi^i = \{(\varepsilon^i_{s,t})\}_{t\leq T, s\leq S}$ the conditions in Lemma 12 have no testable implications as we can always find values that satisfy (31). We therefore need to impose additional restrictions on these errors in order to circumvent making the nonparametric restrictions of EUT void. An attractive possibility in this instance is to augment this model with a set of moment conditions.

To fit this model into our framework of the previous section we identify the observable random variable y with a random dataset $\mathcal{D}^i$ where here i is drawn at random from some population of agents. Next, we identify the latent (unobserved) variables u by the vector $\Xi^i = (\varepsilon^i_{s,t})_{t\leq T, s\in\mathcal{S}}$. Using Lemma 12 we identify Γ as all rationalizable combinations of observables and unobservables.

$$\Gamma = \left\{(\mathcal{D}^i, \Xi^i) \middle| \mathcal{D}^i \text{ is rationalizable for } \Xi^i\right\}$$

In other words, Γ contains all couples $\mathcal{D}^i$ and Ξ^i that satisfy (31). In order to make Γ compact, we can limit $\varepsilon^i_t \in [\varepsilon_L, \varepsilon_H]$ for some lower and upperbounds ε_L and ε_H.[23]

To finalize, we can impose several moment conditions. One natural condition is that ε should have mean 0 over all states and observations:

$$\mathbb{E}\left[\sum_{t\leq T}\sum_{s\leq S}\varepsilon^i_{s,t}\right] = 0.$$

[22] The proof of the result is omitted but almost similar to the proof of the main theorem in Demuynck and Staner (2020).

[23] In our application we set $\varepsilon_L = -10$ and $\varepsilon_U = 10$.

Although this condition fixes the mean, by itself it is not very restrictive, as we can always add a constant to all values $\varepsilon^i_{s,t}$ without influencing the revealed preference conditions (31). On the other hand, we can also assume that these errors are uncorrelated with the prices which imposes that:

$$\mathbb{E}\left[\sum_{t\leq T}\sum_{s\leq S}\varepsilon^i_{s,t}p_{s,t}\right]=0.$$

Finally, we will add a restriction that bounds the error's variance:[24]

$$\mathbb{E}\left[\sum_{t}\sum_{s\leq S}(\varepsilon^i_{s,t})^2\right]\leq TS\sigma^2$$

If we translate this setup into our algorithm, we obtain a Master program, where one of the moment conditions becomes an inequality. This, however is not a issue in the sense that the resulting feasibility set remains convex.

The Master program is similar to before except from the fact that now one of the constraints is an inequality.

$$\begin{aligned}
\textbf{MASTER:}\quad & \min \|\eta^2\| \\
\text{s.t.}\quad & \sum_{j=1}^{J} w_j f_j^k + \eta_k = 0 \qquad k=1,\dots,K-1 \\
& \sum_{j=1}^{J} w_j f_j^K + \eta_K \leq TS\sigma^2 \\
& \sum_{k=1}^{J} w_k = 1 \\
& w_k \geq 0.
\end{aligned}$$

Here $f^K(\mathcal{D}^i,\Xi^i)=\sum_t\sum_{s\leq S}(\varepsilon^i_{s,t})^2$ is the moment function corresponding to the variance bound. Observe that the optimal solution will always have $\eta_K\leq 0$, which is relevant when

[24] Notice that this uses the first mean zero moment condition.

going to the AUXILIARY PROGRAM:

$$\textbf{AUXILIARY(i):} \max_{\varepsilon_t^i} \sum_{k=1}^{K} \eta_k f^k(\mathcal{D}^i, \Lambda^i),$$
$$\text{s.t. } (\mathcal{D}^i, \Lambda^i) \in \Gamma.$$

Conveniently, Γ takes the form of a set of linear inequalities (given by (31)). All moment functions f^k are linear except for the last one that gives a quadratic term:

$$f^K(\mathcal{D}^i, \Lambda^i) = \sum_{t \leq T} \sum_{s \leq S} \left(\varepsilon_{s,t}^i\right)^2.$$

However, given that $\eta_K \leq 0$, the program AUXILIARY(i) takes the form of a quadratic (concave) maximization program, which can still be solved efficiently.

As such, each iteration in the program requires solving $n + 1$ quadratic programs.

4.2.1 Data and Illustration

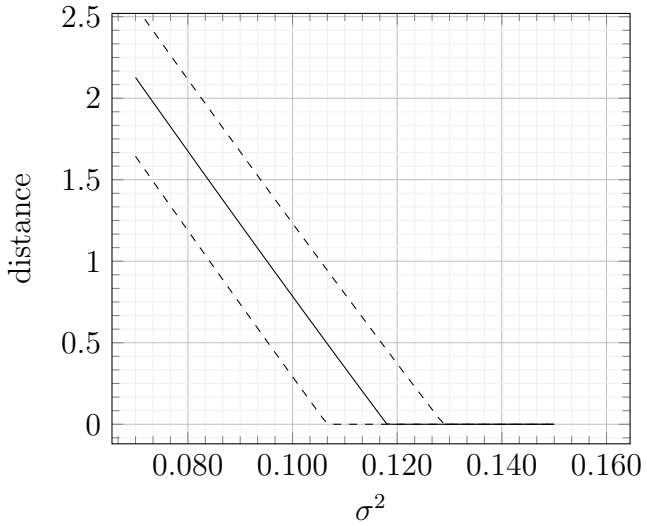
For our illustration, we use experimental dataset of Choi, Kariv, Müler, and Silverman (2014). The dataset is from an experiment on the CentERpanel, which is representative of the Dutch-speaking population in the Netherlands. The data is obtained from an online experiment with 1,182 CentERpanel adults ($n = 1182$).

Subjects were presented with a sequence of 25 decision problems under risk ($T = 25$). Each decision problem consisted of making a choice from a two-dimensional budget line representing the possible allocations over two Arrow-Debreu securities relating to two states with equal probability ($S = 2$ and $P_1^i = P_2^i = 0.5$). The budget lines selected for each subject and in each decision problems were independent of each other and of the sets selected for any of the other subjects in their decision problems. We refer to Choi, Kariv, Müler, and Silverman (2014) for more information on the data set.

We run the algorithm for each value of σ^2 in a grid. Figure 2 shows a plot of the distance $d(\alpha, \text{co}(\overline{F}_n))$ computed by the algorithm as a function of σ^2. We also compute confidence intervals based on a subsampling procedure (Politis, Romano, and Wolf, 1999). In particular, we draw 500 subsamples (without replacement) of size $141 \approx n^{0.7}$. The figure plots the 5th and 95th quantile over these subsamples. We clearly see that the distance linearly decreases until σ^2 is around 0.12 after which it is indistinguishable from zero. This gives a clear set-identification of σ^2 in the sense that any value of σ^2 above 0.12 provides a rationalization

but anything below does not. This gives a minimal standard deviation of around 0.35, which gives 1.42 when exponentiated.

Figure 2: $d(\alpha, \text{co}(\overline{F}_n)$ as a function of σ^2



Note: the solid line gives the estimate of σ^2 on the full dataset. The dashed curves give the 5th and 95th percentile of the 500 subsamples.

5 Conclusion

In this paper we have presented a computationally fast and simple algorithm to test revealed preference models including randomness. We have thereby made use of insights from the partial identification literature, in particular the concept of a random set. To be more precise, we have shown that a nonparametric test of a model boils down to testing whether or not a vector, α, summarizing the restrictions, is contained in the Aumann expectation of a particular random set. To operationalize this test towards realistic empirical settings, we replaced the Aumann expectation by a Minkowski average. As a final step in our theoretical contribution, we have shown that checking whether α lies in the Minkowski average can be done efficiently using a column generating method.

In the second half of the paper, we illustrate how our algorithm can be applied to test models nonparametrically using both survey datasets and experimental data. In particular, we conducted a nonparametric test of a standard intertemporal consumption model with uncertain income using the consumer expenditure survey (CEX). We find that inconsistency with the model is more often observed for married couples where both are working, which

is in line with the collective model of household behavior. Interestingly, this pattern is not found for married couples where only one spouse is working. Finally, we also tested rationalizability of choices by approximate expected utility using experimental data.

Though we have focused on testing models using revealed preference analysis, it should be noted that it might be possible for our algorithm to be fruitfully applied to more general inference of set-identified (structural) parameters. We already illustrated this in our application studying approximate expected utility, where we provided a set-identification for the maximum variance of errors, σ^2. Similarly, in the application with the EDU intertemporal consumption model, we conditioned on β, so this could also provide a manner to set-identify the discount factor. We leave a fuller treatment of such extensions as a promising avenue for future research.

References

Adams, A., L. Cherchye, B. De Rock, and E. Verriest (2014). Consume now or later? Time inconsistency, collective choice, and revealed preference. *American Economic Review 104*, 4147–4183.

Afriat, S. N. (1967). The construction of utility functions from expenditure data. *International Economic Review 8*, 67–77.

Aguiar, V. and N. Kashaev (2020). Stochastic revealed preferences with measurement error. *The Review of Economic Studies 88*, 2042–2093.

Attanasio, O. P. and G. Weber (1995). Is consumption growth consistent with intertemporal optimization? evidence from the consumer expenditure survey. *Journal of Political Economy 103*, 1121–1157.

Beresteanu, A., I. Molchanov, and F. Molinari (2011). Sharp identification regions in models with convex moment predictions. *Econometrica 79*, 1785–1821.

Beresteanu, A., I. Molchanov, and F. Molinari (2012). Partial identification using random set theory. *Journal of Econometrics 166*, 17–32.

Beresteanu, A. and F. Molinari (2008). Asymptotic properties for a class of partially identified models. *Econometrica 76*, 763–814.

Blundell, R., M. Browning, and I. Crawford (2008). Best nonparametric bounds on demand responses. *Econometrica 76*, 1227–1262.

Bontemps, C. and T. Magnac (2017). Set identification, moment restrictions and inference. *Annual Review of Economics 9*, 103–129.

Browning, M. (1989). A nonparametric test of the life-cycle rational expections hypothesis. *International Economic Review 30*, 979–992.

Chambers, C. P. and F. Echenique (2016). *Revealed Preference Theory*. Cambridge University Press.

Cherchye, L., T. Demuynck, and B. De Rock (2019). Bounding counterfactual demand with unobserved heterogeneity and endogenous expenditures. *Journal of Econometrics 211*, 483–506.

Choi, S., S. Kariv, W. Müler, and D. Silverman (2014). Who is (more) rational? *American Economic Review 104*, 1518–1550.

Cosaert, S. and T. Demuynck (2018). Nonparametric welfare and demand analysis with unobserved individual heterogeneity. *Review of Economic and Statistics 100*, 349–361.

Crawford, I. (2010). Habits revealed. *The Review of Economic Studies 77*, 1382–1402.

Crawford, I. and B. De Rock (2014). Empirical revealed preference. *Annual Review of Economics*, 503–524.

Demuynck, T. (2021). A markov chain monte carlo procedure to generate revealed preferenceconsistent datasets,. *Journal of Mathematical Economics 97*(102523).

Demuynck, T. and P. Hjertstrand (2020). *Paul Samuelson, Master of Modern Economics*, Chapter Samuelson' s Approach to Revealed Preference Theory: Some Recent Advances. Palgrave MacMillan.

Demuynck, T. and C. Staner (2020). An efficient revealed preference test for the maxmin expected utility model. Ecares Working Paper 2020-31.

Demuynck, T. and E. Verriest (2013). I'll never forget my first cigarette: A revealed preference analysis of the habits as durables model. *International Economic Review 54*, 717–738.

Diewert, W. E. (1973). Afriat and revealed preference theory. *Review of Economic Studies 40*, 419–425.

Echenique, F., T. Imai, and K. Saito (2021). Approximate expected utility rationalization. Technical report.

Echenique, F., S. Lee, and M. Shum (2011). The money pump as a measure of revealed preference violations. *Journal of Political Economy 119*, 1201–1223.

Echenique, F. and K. Saito (2015). Savage in the market. *Econometrica 83*, 1467–1495.

Ekeland, I., A. Galichon, and M. Henry (2010). Optimal transportation and the falsifiability of incompletely specified economic models. *Economic Theory 42*, 355–374.

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

Gilbert, E. G., D. W. Johnson, and S. S. Keerthi (1988). A fast procedure for computing the distance beteen comple objects in three-dimensional space. *IEEE Journal of Robotics and Automation 4*, 193–203.

Hahn, J., G. Kuersteiner, and M. Mazzocco (2020). Estimation with aggregate shocks. *The Review of Economic Studies 87*, 1365–1398.

Hubner, S. (2022). Identification of unobserved distribution factors and preferences in the collective household model. *Journal of Econometrics*.

Jovanovic, B. (1989). Observable implications of models with multiple equilibria. *Econometrica 57*, 1431–1437.

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

Kitamura, Y. and J. Stoye (2018). Nonparametric analysis of random utility models. *Econometrica 86*(6), 1883–1909.

Li, L. (2021). Identification of structural and counterfactual parameters in a large class of structural econometric models. Technical report.

Luenberger, D. G. (1969). *Optimization by Vector Space Methods*. Wiley.

Mazzocco, M. (2007). Household intertemporal behaviour: a collective characterization and a test of commitment. *The Review of Economic Studies 74*, 857–895.

Mazzocco, M. (2008). Individual rather than household euler equations: identification and estimation of individual preferences using household data. Technical report.

McFadden, D. and M. K. Richter (1971). On the extension of a set function on a set of events to a probability on the generated boolean σ-algebra. Technical report, University of California, Berkeley.

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

Molchanov, I. (2017). *Theory of random sets*. Springer.

Molchanov, I. and F. Molinari (2014). Applications of random set theory in econometrics. *Annual Review of Economics 6*, 229–251.

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

Rockafellar, R. T. (1970). *Convex analysis*. Princeton University Press.

Roehrig, C. S. (1988). Conditions for identification in nonparametric and parametric models. *Econometrica 56*, 433–447.

Savage, L. J. (1954). *Foundations of Statistics*. Wiley.

Sekitani, K. and Y. Yamamoto (1993). A recursive algorithm for fining the minimum norm point in a polytope and a pair of closest points in two polytopes. *Mathematical programming 61*, 233–249.

Seri, R. and C. Choirat (2004). *Computational Science and its applications*. Springer.

Shennach, S. M. (2014). Entropic latent variable integration via simulation. *Econometrica 82*, 345–385.

Smeulders, B., L. Cherchye, and B. de Rock (2021). Non-parametric analysis of random utility models: computational tools for statistical testing. *Econometrica 89*, 437–455.

Varian, H. (1985). Non-parametric analysis of optimizing behavior with measurement error. *Journal of Econometrics 30*, 445–485.

Varian, H. R. (1982). The nonparametric approach to demand analysis. *Econometrica 50*, 945–973.

Varian, H. R. (2006). *Samuelson economics and the 21st century*, Chapter Revealed Preference. Oxford University Press.

Wolfe, P. (1976). Finding the nearest point in a polytope. *Mathematical Programming 11*, 128–149.

A Proofs

A.1 Proof of Proposition 6

Proof. From uniqueness of the projection and by (17) we know that $v_M = \|\alpha - \widehat{f}\| = d(\alpha, co(\overline{F}_n))$ if and only if for all $f \in \overline{F}_n$:

$$\begin{aligned} &\langle \alpha - \widehat{f}, f - \widehat{f} \rangle \leq 0, \\ \iff &\langle \eta, f \rangle \leq \langle \eta, \widehat{f} \rangle. \end{aligned}$$

In other words:

$$\sup_{f \in \overline{F}_n} \langle \eta, f \rangle = h_{\mathrm{co}(\overline{F})}(\eta) \leq \langle \eta, \widehat{f} \rangle.$$

As the support function is linear in the Minkowski sum, the left hand side can also be written as:

$$\frac{1}{n} \sum_{i=1}^{n} h_{F(y_i)}(\eta) \leq \langle \eta, \widehat{f} \rangle$$

as was to be shown. ■

A.2 Proof of Theorem 7

Proof. Assume that $d(\alpha, \mathrm{co}(\overline{F}_n)) > \sqrt{\varepsilon}$. Then the algorithm will never stop at line 9 as $\|\eta_k\| \geq d(\alpha, \mathrm{co}\, \overline{F}_n)$. Let $\widehat{f} = \arg\min_{f \in \mathrm{co}(\overline{F}_n)} \|\alpha - f\|^2$ be the projection of α onto $\mathrm{co}(\overline{F}_n)$. We have that:

$$\begin{aligned} \|\alpha - \widehat{f}\|^2 &= \|\alpha - \widehat{f}_J + \widehat{f}_J - \widehat{f}\|^2, \\ &= \|\alpha - \widehat{f}_J\|^2 + \|\widehat{f}_J - \widehat{f}\|^2 + 2\langle \eta_J, \widehat{f}_J - \widehat{f} \rangle. \end{aligned}$$

As such:

$$\begin{aligned} \|\alpha - \widehat{f}_J\|^2 &= \|\alpha - \widehat{f}\|^2 - \|\widehat{f}_J - \widehat{f}\|^2 + 2\langle \eta_J, \widehat{f} - \widehat{f}_J \rangle, \\ &\leq \|\alpha - \widehat{f}\|^2 - \|\widehat{f}_J - \widehat{f}\|^2 + 2\langle \eta_J, f_{J+1} - \widehat{f}_J \rangle, \\ &\leq \|\alpha - \widehat{f}\|^2 + 2\delta \|\alpha - \widehat{f}_J\|^2 \end{aligned}$$

The first inequality follows from the fact that:

$$\max_{f \in \text{co}(\overline{F}_n)} \langle \eta_J, f - \widehat{f}_J \rangle = \max_{f \in \overline{F}_n} \langle \eta_J, f - \widehat{f}_J \rangle = \langle \eta, f_{J+1} - \widehat{f}_J \rangle.$$

as all extreme points of $\text{co}(\overline{F}_n)$ are in $\overline{F}_n$. The second inequality follows from the fact that if the algorithm stops (at line 16) then:

$$\langle \eta, f_{J+1} - \widehat{f}_J \rangle \leq \delta \|\alpha - \widehat{f}_J\|^2.$$

This gives:

$$\|\alpha - \widehat{f}_J\|^2 \leq \frac{\|\alpha - \widehat{f}\|^2}{1 - 2\delta}$$

■

A.3 Proof of Theorem 8

Proof. Assume that the algorithm does not terminate before iteration K, which means that for all $k < K$, $\langle \eta_k, f_{k+1} - \widehat{f}_k \rangle > \delta \|\alpha - \widehat{f}_k\|^2$ and $\|\alpha - \widehat{f}_k\|^2 \geq \varepsilon$.

Notice that for all $\lambda \in [0, 1]$, $\lambda f_{k+1} + (1 - \lambda)\widehat{f}_k \in \text{co}(\{f_1, \ldots, f_{k+1}\}$ (as $\widehat{f}_k \in \text{co}(\{f_1, \ldots, f_k\})$). As such, for all $\lambda \in [0, 1]$:

$$\begin{aligned}
\|\alpha - \widehat{f}_{k+1}\|^2 &\leq \|\alpha - (\lambda f_{k+1} + (1 - \lambda)\widehat{f}_k)\|^2, \\
&= \|\alpha - \widehat{f}_k\|^2 + \lambda^2 \|\widehat{f}_k - f_{k+1}\|^2 + 2\lambda \langle \alpha - \widehat{f}_k, \widehat{f}_k - f_{k+1} \rangle, \\
&= \|\alpha - \widehat{f}_k\|^2 + \lambda^2 \|\widehat{f}_k - f_{k+1}\|^2 - 2\lambda \langle \eta_k, f_{k+1} - \widehat{f}_k \rangle,
\end{aligned}$$

The idea is to take the minimum of the right hand side over $\lambda \in [0, 1]$.

It is clear to see that the right hand side is convex in λ and the derivative is strictly negative for $\lambda = 0$ as the last term is strictly negative by construction of f_{k+1}. As such, the optimal value is either interior or equal to 1. If it is interior, the the first order condition gives:

$$\begin{aligned}
2\lambda \|f_{k+1} - \widehat{f}_k\|^2 - 2\langle \eta_k, f_{k+1} - \widehat{f}_k \rangle = 0, \\
\rightarrow \lambda = \frac{\langle \eta_k, f_{k+1} - \widehat{f}_k \rangle}{\|f_{k+1} - \widehat{f}_k\|^2} \geq \frac{\delta \|\alpha - \widehat{f}_k\|^2}{\|f_{k+1} - \widehat{f}_k\|^2} \geq \frac{\delta \varepsilon}{A}.
\end{aligned}$$

Here we used the fact that $\|\alpha - \widehat{f}_k\|^2 \geq \varepsilon$ and we set A to be the maximum of the diameter

of the set $\overline{F}_n$:

$$A = \max\left\{ \sup_{f,f' \in \overline{F}_n} \|f - f'\|^2 \right\}$$

Notice that A is well defined as every set $F(y_i)$ is compact (and n is finite). Also notice that $0 < \dfrac{\delta\varepsilon}{A} < 1$ Next, if $\lambda = 1$ is optimal, then:

$$\begin{aligned} 2\|f_{k+1} - \widehat{f}_k\|^2 - 2\langle \eta_k, f_{k+1} - \widehat{f}_k \rangle \leq 0, \\ \rightarrow \|f_{k+1} - \widehat{f}_k\|^2 \leq \langle \eta_k, f_{k+1} - \widehat{f}_k \rangle. \end{aligned}$$

Now, if λ is interior, then:

$$\begin{aligned} \|\alpha - \widehat{f}_{k+1}\|^2 &\leq \|\alpha - \widehat{f}_k\|^2 + \lambda^2 \|\widehat{f}_k - f_{k+1}\|^2 + 2\lambda \langle \eta_k, \widehat{f}_k - f_{k+1} \rangle, \\ &= \|\alpha - \widehat{f}_k\|^2 - \lambda \langle \eta_k, f_{k+1} - \widehat{f}_k \rangle, \\ &\leq \|\alpha - \widehat{f}_k\|^2 - \lambda\delta \|\alpha - \widehat{f}_k\|^2, \\ &= (1 - \lambda\delta)\|\alpha - \widehat{f}_k\|^2, \\ &\leq (1 - \frac{\delta^2\varepsilon}{A})\|\alpha - \widehat{f}_k\|^2. \end{aligned}$$

If $\lambda_k = 1$ then:

$$\begin{aligned} \|\alpha - \widehat{f}_{k+1}\|^2 &\leq \|\alpha - \widehat{f}_k\|^2 + \|\widehat{f}_k - f_{k+1}\|^2 + 2\langle \eta, \widehat{f}_k - f_{k+1} \rangle, \\ &\leq \|\alpha - \widehat{f}_k\|^2 - \langle \eta_k, f_{k+1} - \widehat{f}_k \rangle, \\ &\leq (1 - \delta)\|\alpha - \widehat{f}_k\|^2. \end{aligned}$$

Setting $\theta = \min\{1 - \delta, 1 - \delta^2\varepsilon/A\} < 1$, we have that:

$$\|\alpha - \widehat{f}_{k+1}\|^2 \leq \theta \|\alpha - \widehat{f}_k\|^2.$$

As this holds for all $k < K$ we get:

$$\|\alpha - \widehat{f}_K\|^2 \leq \theta^K \|\alpha - \widehat{f}_0\|^2.$$

which proves the result. ■

B Statistical inference

Although statistical inference is not the main focus of the paper, we still want to highlight some well-known results from the literature that could be used to derive inference based on the value of $d(\alpha, \text{co}(\overline{F}_n))$. More details concerning inference on random sets can be found in, among others, Molchanov (2017), Beresteanu and Molinari (2008) and Seri and Choirat (2004).

Recall from section 1, by introducing the Aumann expectation $\mathbb{E}F$ of the random set $F(y)$, our testable implication (11) can be written as:

$$\alpha \in \mathbb{E}F \leftrightarrow d(\alpha, \mathbb{E}F) = 0.$$

This is the hypothesis that we would like to test given the "estimate" $d(\alpha, \text{co}(\overline{F}_n))$, which as we recall is the (smallest) distance between α and $\text{co}(\overline{F}_n)$. This means that we are interested in the approximation properties of the finite sample estimator $\text{co}(\overline{F}_n)$ of the Aumann expectation $\mathbb{E}F$. To that end, one usually uses the notation of the Hausdorff distance between two sets A and $B \subseteq \mathbb{R}^K$:

$$\rho(A, B) = \max \left\{ \sup_{x \in A} d(x, B), \sup_{y \in B} d(y, A) \right\}.$$

One of the reasons why much of the literature on random sets and its applications has focused on the support function is due to the following insight: if A and B are both convex and compact then:

$$\rho(A, B) = \sup_{\|\lambda\|=1} |h_A(\lambda) - h_B(\lambda)|.$$

Given that $\mathbb{E}F$ is convex and closed, we have that:

$$\rho(\mathbb{E}F, \text{co}(\overline{F}_n)) = \sup_{\|\lambda\|=1} \left| h_{\mathbb{E}F}(\lambda) - h_{\text{co}(\overline{F}_n)}(\lambda) \right|.$$

The following law of large numbers is known (see Molchanov (2017, p.322) for the exact conditions).

$$\rho(\text{co}(\overline{F}_n), \mathbb{E}F) \xrightarrow{n} 0 \qquad \text{w.p. } 1.$$

If $d(\alpha, \text{co}(\overline{F}_n)) > 0$, then using the result of Proposition 4, we obtain:

$$\begin{aligned}
d(\alpha, \text{co}(\overline{F}_n)) &= \sup_{\|\lambda\|=1} \{\langle \lambda, \alpha \rangle - h_{\text{co}(\overline{F}_n)}(\lambda)\}, \\
&= \sup_{\|\lambda\|=1} \{\langle \lambda, \alpha \rangle - h_{\mathbb{E}F(y)}(\lambda) + h_{\mathbb{E}F(y)}(\lambda) - h_{\text{co}(\overline{F}_n)}(\lambda)\}, \\
&\leq \sup_{\|\lambda\|=1} \{\langle \lambda, \alpha \rangle - h_{\mathbb{E}F(y)}(\lambda)\} + \sup_{\|\lambda\|=1} \{h_{\mathbb{E}F(y)}(\lambda) - h_{\text{co}(\overline{F}_n)}(\lambda)\}, \\
&\leq d(\alpha, \mathbb{E}F(y)) + \sup_{\|\lambda\|=1} \{h_{\mathbb{E}F(y)}(\lambda) - h_{\text{co}(\overline{F}_n)}(\lambda)\}, \\
&\leq d(\alpha, \mathbb{E}F(y)) + \sup_{\|\lambda\|=1} |h_{\mathbb{E}F(y)}(\lambda) - h_{\text{co}(\overline{F}_n)}(\lambda)|.
\end{aligned}$$

reversing the roles of $\text{co}(F_n)$ and $\mathbb{E}F(y)$ we also get that if $\alpha \notin \mathbb{E}F(y)$ then:

$$d(\alpha, \mathbb{E}F(y)) \leq d(\alpha, \text{co}(\overline{F}_n)) + \sup_{\|\lambda\|=1} |h_{\text{co}(\overline{F}_n)}(\lambda) - h_{\mathbb{E}F(y)}(\lambda)|,$$

Notice that these inequalities also hold if $\alpha \in \text{co}(F_n)$ or $\alpha \in \mathbb{E}F(y)$. This gives:

$$\left| d(\alpha, \text{co}(\overline{F}_n) - d(\alpha, \mathbb{E}F(\mathbf{y})) \right| \leq \rho(\mathbb{E}F(y), \text{co}(\overline{F}_n)) \xrightarrow{n} 0 \qquad \text{w.p. 1.}$$

This shows that $d(\alpha, \text{co}(\overline{F}_n))$ is a consistent estimator of $d(\alpha, \mathbb{E}F)$.

For inference, we can rely on the following central limit theorem (see Molchanov (2017, p345) for the exact conditions):

$$\sqrt{n}\rho(\mathbb{E}F, \text{co}(\overline{F}_n)) \xrightarrow{d} \sup_{\|\lambda\|=1} Z(\lambda),$$

where $\{Z(\lambda) : \|\lambda\| \leq 1\}$ is a zero mean Gaussian random field on $\mathbb{S}^K$ with covariance:

$$\mathbb{E}Z(t)Z(w) = \mathbb{E}h_{F(y)}(t)h_{F(y)}(w) - \mathbb{E}h_{F(y)}(t)\mathbb{E}h_{F(y)}(w).$$

Under the null that $d(\alpha, \mathbb{E}F) = 0$ we therefore get that:

$$\begin{aligned}
\sqrt{n}d(\alpha, \text{co}(\overline{F}_n)) &\leq \sqrt{n}d(\alpha, \mathbb{E}F(y)) + \sqrt{n}\rho(\mathbb{E}F, \text{co}(\overline{F}_n)), \\
&= \sqrt{n}\rho(\mathbb{E}F, \text{co}(\overline{F}_n)) \xrightarrow{d} \sup_{\|\lambda\|=1} Z(\lambda).
\end{aligned}$$

To obtain a statistical test for the hypothesis that $d(\alpha, \mathbb{E}F) = 0$, define a critical value c_α

such that:

$$(1-\alpha) = \mathbb{P}\left(\sup_{\|\lambda\|=1} Z(\lambda) \leq c_\alpha\right).$$

Now notice that the probability of not rejecting, under the null is given by:

$$\mathbb{P}(\sqrt{n}d(\alpha, \text{co}(\overline{F}_n)) \leq c_\alpha) \geq \mathbb{P}\left(\sqrt{n}\rho(\mathbb{E}F, \text{co}(\overline{F}_n)) \leq c_\alpha\right) \approx 1-\alpha.$$

So, for large enough samples, we reject the null hypothesis with probability less than α. See Seri and Choirat (2004) for an algorithm to estimate the value of c_α.