Testing revealed preference models with unobserved randomness using random sets

Thomas Demuynck* and Tom Potoms†

December 11, 2023

Abstract

We use random set theory to derive testable implications for revealed preference models in the presence of unobserved randomness. Our test requires that a certain vector belongs to the Aumann expectation of a well defined random set. We present a computationally fast and simple algorithm to obtain a finite sample implementation of our test and we show how to use a subsampling procedure to derive valid inference. We demonstrate the practical relevance of our results by an application to the standard intertemporal consumption model with idiosyncratic income risks and to a model of approximate expected utility maximisation.

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

Over the years, there has been a steady increase of the use of revealed preference (RP) models to test a wide variety of decision models.[1] An RP analysis usually starts from a finite data set on budgets and choices, for example, a dataset on linear

*Corresponding author: 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.

budgets and chosen bundles from these budgets. It then looks for conditions on this dataset such that the observed choices are consistent with some given decision model, like for example utility maximization. Usually the resulting RP tests take the form of a set of combinatorial conditions or a set of (linear) inequalities. Well known examples are the Weak, Strong or Generalized Axiom of Revealed Preference (WARP, SARP or GARP).[2] The most attractive feature of the RP methodology stems from the fact that the method abstains from imposing specific functional forms on the underlying utility (or production) functions, although shape restrictions, like concavity or homogeneity can be imposed if desired. Much of the RP literature has focused on settings without randomness, which excludes realistic and relevant features like uncertainty or misperception regarding prices, income, 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 and econometric literature.

Our paper has two main contributions. First, we use the theory of random sets to present a general framework that is able to analyse RP models with unobserved randomness. We find that testing such RP models boils down to verifying whether a certain vector belongs to the Aumann expectation of a random set. This random set is characterized by the underlying RP model and a collection of moment conditions. Our second contribution relates to the (finite sample) implementation of this test. For this, we develop a fast and simple column generation algorithm. Building upon the existing literature dealing with statistical inference on random sets, we also provide a simple subsampling procedure that can be used for inference. We provide two applications to demonstrate the practical relevance of our results.

Overview We start from a general framework with a given RP model. This RP model consists of an observable component and an unobservable (latent) component. Although the marginal distribution over the observable component is known (or can be identified), the joint distribution over both components is inherently non observable. The support of this joint distribution, however, is restricted by the conditions imposed by the RP model. In addition, we assume that the model imposes some moment conditions on the observable and unobservable variables. Due to its gener-

[2]See the seminal works by Afriat (1967), Diewert (1973) and Varian (1982) for early contributions to RP theory.

ality, this framework is able to capture a large set of applications. Interestingly, 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 object. The theory of random sets has received a lot of attention and has become a central toolkit in the literature on partial identification.[3] In this paper, we show that it also provides a very convenient and flexible framework for the analysis of RP models with unobserved randomness. In particular, we show that, under suitable regularity conditions, the testable restrictions of the model boil down to checking whether or not a particular vector, ψ, which corresponds to the values for the moment conditions, belongs to the Aumann-expectation of a particular random set. The Aumann expectation is a generalization of the usual expectation of random variables to the framework of random sets. This Aumann expectation is identified from the observables, the underlying RP model and a set of moment conditions.

The second focus of this paper is to make this test useful for empirical applications, i.e., when one has only access to a (finite) sample of the observable component. 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 ψ to belong to this set. We do this by verifying whether the distance between the vector ψ and the convex hull of the Minkowski average equals zero. A hurdle towards operationalizing this idea, is that in general, there is no closed form expression for this Minkowski average.

To overcome this problem, we propose an intuitive and fast column generating algorithm. The algorithm is similar in spirit 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). Next, using insights from the statistical literature on random set theory, we also provide a simple subsampling procedure that allows for valid statistical inference in our framework.

The final part of the paper illustrates our algorithm by means of two 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

[3]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.

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 follow a recently suggested model of Echenique, Imai, and Saito (2023) 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. This application also illustrates the usefulness of our subsampling procedure.

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 they introduce measurement error on prices, instead of measurement error on demand. 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.

Recently, Aguiar and Kashaev (2021) analyzed RP problems in the presence of unobserved randomness (measurement error). Their analysis follows the ELVIS approach of Schennach (2014). This approach uses ideas from information theory (i.e., I-projections and maximum entropy) to analyze models characterized via moment conditions. In order to translate the RP model into this setting, one first translate all revealed preference conditions to suitable moment conditions. Next, these conditions are verified by integrating out the unobservable latent variables using a least informative maximum entropy distribution. For the latter Aguiar and Kashaev (2021) use numerical integration based on an underlying hit and run procedure to obtain random samples from the feasible region.

Our approach, on the other hand, is based on the theory of random sets. As the Aumann Expectation of a random set is always convex, this approach has the great

advantage that it enables us to use well known tools from convex analysis, like support functions, to analyse the setting.[4] Concerning the empirical implementation, our approach further has the advantage that it avoids the need for numerical integration. This means that it might be more more precise and requires less computing time. In fact, all results in this paper were obtained by repeatedly solving a collection of linear or quadratic programs, which can be done quite efficiently. All our computations were obtained using a standard laptop in a time-span of less than 2 days (including the subsampling procedure in subsection 5.2).[5] Moreover, the inner loop of the algorithm can easily be parallelized, so further efficiency gains are possible.

In the RP literature, stochastic revealed preference theory takes a related, yet quite different viewpoint on 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.[6] On the other hand, in this paper we have a cross section of individual datasets where each dataset consists of multiple decisions of a single decision maker or household, and one imposes moment conditions across individuals. 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.[7]

Structure of the paper Section 1 presents the random sets framework to test RP models with unobserved heterogeneity and moment restrictions. Section 2 translates this framework to a finite sample setting. Section 3 gives our column generation algorithm and discusses its properties. Section 4 provides statistical inference. Section 5 contains two applications. Finally Section 6 contains a conclusion.

[4]See also, among others, Ekeland, Galichon, and Henry (2010); Beresteanu, Molchanov, and Molinari (2011); Li (2021) who use insights of convex analysis and apply it to random sets. See also Schennach (2014) for a more thorough discussion related to the difference between the random sets and ELVIS approach.

[5]The laptop has an Intel(R) Core(TM) i5-10210U CPU with 2.11 GHz and 16.0 GB RAM.

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

[7]See also Smeulders, Cherchye, and de Rock (2021) who uses a different column generating algorithm to make computational gains for RP analysis in a stochastic revealed preference setting.

1 Framework

Our framework is based on Jovanovic (1989) and inspired by the work of 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, which are collected in the random (finite dimensional) vector $y \in \mathcal{Y} \subseteq \mathbb{R}^{d_Y}$. Next, there is a random vector $u \in \mathcal{U} \subseteq \mathbb{R}^{d_U}$ that provides the unobservable or latent variables (from the point of view of the econometrician).

Example. *In an RP framework, one can think of y as representing a finite (random) dataset $(p_t, q_t, m_t)_{t \leq T}$ (with $T \in \mathbb{N}_0$) on prices $p_t \in \mathbb{R}^k_{++}$, budgets $m_t > 0$ and choices $q_t \in \mathbb{R}^k_+$ from the linear budgets $\{q \in \mathbb{R}^k_+ | \langle p_t, q \rangle \leq m_t\}$. In this case, the stochastic nature of y follows from the assumption that the dataset y is obtained from a random individual in the population. The latent vector u might then, for example, correspond to a collection $(\varepsilon_t)_{t \leq T}$ of vectors in $\mathbb{R}^k$ representing the measurement error of the bundles $(q_t)_{t \leq T}$ in the dataset y.*

We denote by μ the (unknown) joint distribution of (y, u).[8] 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 (μ_Y, Γ, f, ψ) 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 measurable set that gives all combinations $(y, u) \in \mathcal{Y} \times \mathcal{U}$ that are consistent with the RP 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.*
- *$\psi = (\psi^1, \dots, \psi^K) \in \mathbb{R}^K$ is a K-dimensional vector of moment values for these functions.*

[8] This distribution is defined on some underlying measure space $(\mathcal{Y} \times \mathcal{U}, \mathcal{B})$ where $\mathcal{B}$ is the Borel σ-algebra on $\mathcal{Y} \times \mathcal{U} \subseteq \mathbb{R}^{d_U + d_Y}$.

Example. *Going back to our RP setting, μ_Y represents the distribution of the datasets $y = (p_t, q_t, m_t)_{t\leq T}$ over the population of individuals. The set Γ captures all combinations of observable data sets $y = (p_t, q_t, m_t)_{t\leq T}$ and measurement error $u = (\varepsilon_t)_{t\leq T}$ that jointly satisfy the RP conditions. For example we might want to impose that consumption bundles arise from the maximisation of some well-behaved utility function but that bundles are measured with error. In this case, we then should require that the modified dataset $(p_t, q_t + \varepsilon_t, m_t)_{t\leq T}$ satisfies the Generalized Axiom of Revealed Preference (GARP).*[9] *As such, Γ consists of all combinations $(y, u) = ((p_t, q_t, m_t)_{t\leq T}, (\varepsilon_t)_{t\leq T})$ such that $(p_t, q_t + \varepsilon_t, m_t)_{t\leq T}$ satisfies GARP. Next, f represents a collection of moment conditions. Denoting by $\varepsilon_{t,j}$ the measurement error on good $j \leq k$ in at observation t, we could require, for example, that the mean measurement error for all goods j should equal zero: $\mathbb{E}(\varepsilon_{t,j}) = 0$ $(j \leq k, t \leq T)$. Another additional condition could be that we assume that measurement error on good j is uncorrelated with the price of the good $\mathbb{E}(\varepsilon_{t,j}p_{t,j}) = 0$ $(j \leq k, t \leq T)$. This would specify a collection of $K = 2(T + k)$ moment functions $f = (f^1_{t,j}, f^2_{t,j})_{t\leq T, j\leq k}$ with $f^1_{j,t}(y, u) = \varepsilon_{t,j}$ and $f^2_{j,t}(y, u) = p_{t,j}\varepsilon_{t,j}$. Finally, the vector ψ gives the values of the right hand sides of these moment conditions, which in this example is a K dimensional vector of zeros. We will tackle other (more elaborate) settings in Section 5.*

Note that, strictly speaking, the set of moment conditions that one imposes are not part of the underlying RP restrictions. In fact, we see this separation of moment conditions and the underlying RP model as one of the attractive features of the framework because it provides a clear conceptual distinction between the conditions imposed by the RP restrictions (captured by the set Γ) and the additional moment restrictions that the econometrician is willing to impose. This resembles the usual separation in econometrics between functional form restrictions and the additional set of moment conditions required to identify the model.

A model (μ_Y, Γ, f, ψ) imposes restrictions on the joint distribution μ of (y, u). In particular, 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 measurable sets A in $\mathcal{Y}$:

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

[9] See, for example Varian (1982).

Next, the set Γ puts a restriction on the support for the random vectors (y, u).

$$\Pr_\mu\left((y,u)\in\Gamma\right)=1. \tag{2}$$

Finally, the functions $f^1,\ldots,f^K$ and the values $\psi^1,\ldots,\psi^K$ encode a (finite) set of moment conditions:

$$\mathbb{E}_\mu f(y,u)=\psi. \tag{3}$$

To summarize, the aim is to find conditions for the existence of a joint probability distribution μ such that (1), (2) and (3) are jointly satisfied.[10]

A crucial element in our further analysis is the following correspondence $F:\mathcal{Y}\rightrightarrows\mathbb{R}^K$:

$$F(y)=\left\{f(y,u)\in\mathbb{R}^K\middle|(y,u)\in\Gamma\right\}. \tag{4}$$

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.

Example. *For our RP example, if $y=(p_t,q_t,m_t)_{t\leq T}$, $F(y)$ provides the set of all possible vectors $f(y,u)\in\mathbb{R}^K$ with components $f^1_{j,t}(y,u)=\varepsilon_{t,j}$ and $f^2_{j,t}(y,u)=p_{j,t}\varepsilon_{j,t}$ and where the value of $u=(\varepsilon_t)_{t\in T}$ is such that $(p_t,q_t+\varepsilon_t,m_t)_{t\in T}$ satisfies GARP, i.e., $(y,u)\in\Gamma$.*

We impose some regularity conditions on F. In particular, we assume that the following is true throughout the paper:

Assumption 2.

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

(ii) There is a measurable function $v:\mathcal{Y}\to\mathbb{R}$ such that $\mathbb{E}_\mu v(y)<\infty$ and:

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

[10] 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, we would be interested in finding 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.

Assumption 2 essentially imposes compactness on the sets $F(y)$.[11] 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).

Let $\mu_{U|y}$ be the conditional distribution of u given y. From conditions (2), (3), and the definition of $F(y)$, we can write:

$$\psi = \mathbb{E}_\mu f(y,u) = \int_{\mathcal{Y}} \left(\int_{F(y)} f(y,u) d\mu_{U|y} \right) d\mu_Y \tag{5}$$

Now, consider the random vector $s(y)$ that takes the value $\int_{F(y)} f(y,u) d\mu_{U|y}$ for all $y \in \mathcal{Y}$. This random variable is a mean of values in the compact set $F(y)$, so its value is in the closed convex hull of $F(y)$, which we denote by $\operatorname{co}(F(y))$. Note that $\mathbb{E}_{\mu_Y} s(y)$ equals ψ.

Aumann expectation An integrable random vector $s(y) \in \mathbb{R}^K$ that takes values in $\operatorname{co}(F(y))$ with probability one is called a selection of the random set $\operatorname{co}(F(y))$. Above reasoning shows that our model requires the existence of at least one integrable selection of the random set $\operatorname{co}(F(y))$ with expectation ψ.

The set that collects all expecations of integrable selections of a random set is called the Aumann expectation of the random set. We denote the Aumann expectation of $\operatorname{co}(F(y))$ by $\mathbb{E}F$.[12]

$$\mathbb{E}F = \{\mathbb{E}_{\mu_Y} s(y) | s(y) \text{ is an integrable selection of } \operatorname{co}(F(y))\}$$

As shown by Molchanov (2017, Theorem 2.1.26), if μ_Y is non-atomic (which means that it does not have a point mass) then $\mathbb{E}F$ is convex and closed.

Condition (5) implies that:

$$\psi \in \mathbb{E}F. \tag{6}$$

In fact, as we will see, this condition is not only necessary, but also sufficient.

[11] In practice, if the moment functions f are continuous, it is often convenient to enforce compactness of $F(y)$ by imposing that the universal set of latent variables $\mathcal{U}$ is compact.

[12] As shown by Molchanov (2017) under mild conditions, the Aumann expectation of $\operatorname{co}(F(y))$ will be equal to the Aumann expecation of $F(y)$. This is why we use the notation $\mathbb{E}F$ instead of $\mathbb{E}\operatorname{co}(F)$.

Support functions Instead of using Aumann expectations, it is also convenient to reformulate the testable implications in terms of support functions. This strategy is also used as part of the identification strategies in several other papers such as Ekeland, Galichon, and Henry (2010), Beresteanu, Molchanov, and Molinari (2011) and Li (2021).

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

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

where $\langle ., . \rangle$ denotes the inner product.[13] Any convex and compact set can be uniquely recovered from its support functions, so this offers a useful dual framework.

Let $\mathbb{S}^K = \{p \in \mathbb{R}^K | \|p\| = 1\}$ be the K-dimensional sphere. As shown by Li (2021), the following holds.

Theorem 3. *If Assumption 2 holds, then conditions (1), (2) and (3) are satisfied if and only if:*

$$\sup_{p \in \mathbb{S}^K} \mathbb{E}_{\mu_Y} \left[\langle p, \psi \rangle - h_{F(y)}(p) \right] \leq 0. \tag{7}$$

To see how this relates to condition (6) above, notice first that we can replace $h_{F(y)}(p)$ by $h_{\text{co}(F(y))}(p)$ as the support function of a compact set equals the support function of its convex closure. Next, we have the important result that the expectation of the support functions over a random set equals the support function of its Aumann expecations.

$$\mathbb{E}_{\mu_Y} h_{\text{co}(F(y))}(p) = h_{\mathbb{E}F}(p).$$

Given this, (7) can be rewritten as:

$$\sup_{p \in \mathbb{S}^K} \left(\langle p, \psi \rangle - h_{\mathbb{E}F}(p) \right) \leq 0.$$

Now assume that this condition is not satisfied. If μ_Y is non-atomic, then $\mathbb{E}F$ is convex and closed, so a reversal of the inequality would imply that there exists a hyperplane separating $\mathbb{E}F$ and ψ, which effectively demonstrates that $\psi \notin \mathbb{E}F$. On

[13] We refer to Rockafellar (1970) for an in-depth treatment of the use of support functions in convex analysis.

the other hand if this condition is satisfied, there is no separating hyperplane that separates $\mathbb{E}F$ from ψ. As $\mathbb{E}F$ is convex, this effectively demonstrates that $\psi \in \mathbb{E}F$.

To finish the section, we want to make two remarks related to the verification of (7) (or (6)). 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 from a finite sample. Second, to compute the support function $h_{F(y)}(p)$ we would be nice to have a succinct description of the random set $F(y)$, which can be challenging in practice. The next two sections will address both these issues.

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 $\{y_1, ..., y_n\}$. Using the standard finite sample analogue principle, we can replace (7) by its finite sample analogue:

$$\sup_{p \in \mathbb{S}^K} \left[\frac{1}{n} \sum_{i=1}^{n} \langle p, \psi \rangle - h_{F(y_i)}(p) \right] \leq 0, \tag{8}$$

Condition (8) replaces the expectation operator by the sample average over the n i.i.d. draws. The expression in (8) 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 $\{F(y_1), \ldots, F(y_n)\}$:[14]

$$\overline{F}_n = \left\{ \frac{1}{n} \sum_{i=1}^{n} f_i \middle| f_i \in F(y_i) \right\} = \left\{ \frac{1}{n} \sum_{i=1}^{n} f(y_i, u_i) \middle| u_i \in \mathcal{U} \text{ and } (y_i, u_i) \in \Gamma \right\}. \tag{9}$$

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 (8) is equivalent to:

$$\sup_{p \in \mathbb{S}^K} \left[\langle p, \psi \rangle - h_{\overline{F}_n}(p) \right] \leq 0. \tag{10}$$

[14]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 $\text{co}(\overline{F}_n)$).

Condition in (10) can in principle be tested, although a direct implementation is far from straightforward as it is known to be quite hard to compute the Minkowski sum of a collection of sets. The following gives an easier, equivalent condition (see Luenberger (1969, p.136) for a proof).

Proposition 4. *Condition (10) is satisfied if and only if* $\psi \in \text{co}(\overline{F}_n)$. *In particular, If* $\psi \notin \text{co}(\overline{F}_n)$ *then*

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

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

Proposition 4 provides an interesting geometric interpretation to condition (10). Checking whether or not the latter is satisfied amounts to verifying if the vector ψ is contained in the convex closed closure of the Minkowski average $\overline{F}_n$. This directly relates to condition (6) as $\text{co}(\overline{F}_n)$ is the finite sample approximation of the Aumann expectation $\mathbb{E}F$. Furthermore, the criterion in expression (10) equals the distance between ψ and $\overline{F}_n$. It therefore suffices to compute the latter and verify whether it is equal to zero, i.e., condition (10) is equivalent to the condition:

$$d(\psi, \text{co}(\overline{F}_n)) = 0. \tag{11}$$

The solution $\widehat{f}$ to the minimization problem (11) is the unique projection of ψ onto the convex set $\text{co}(\overline{F}_n)$. As this is a convex, closed set, 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. $\widehat{f} = \arg\min_{f \in \text{co}(\overline{F}_n)} \|\psi - f\|$ *if and only if* $\widehat{f} \in \text{co}(\overline{F}_n)$ *and for all* $f \in \overline{F}_n$*:*

$$\langle \psi - \widehat{f}, f - \widehat{f} \rangle \leq 0. \tag{12}$$

The geometric interpretation of lemma 5 is that the vector $\psi - \widehat{f}$ must make an obtuse angle with $f - \widehat{f}$ for all $f \in \overline{F}_n$. Condition (12) provides a straightforward

condition to verify whether a certain element in $\text{co}(\overline{F}_n)$ is the projection of ψ to $\text{co}(\overline{F}_n)$. In particular, it suffices to maximize the linear objective $\langle \psi - \widehat{f}, f \rangle$ with respect to $f \in \overline{F}_n$ and to check whether it is smaller than (or equal to) $\langle \psi - \widehat{f}, \widehat{f} \rangle$.

3 The column generating algorithm

In this section, we provide an algorithm to compute $d(\psi, \text{co}(\overline{F}_n))$. We do this using a column generation procedure. 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, this subset is expanded until insufficient improvement in the objective function can be made.

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 $\overline{F}_n$.[15] 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 ψ 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 an outer and inner optimization problem. The first one, called the Master program, computes (for a given J) the smallest square distance between the vector ψ and the inner approximation, $\text{co}(\{f_1, \ldots, f_J\})$ of $\text{co}(\overline{F}_n)$:

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

[15] 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 $f_j = \frac{1}{n} \sum_i f(y_i, u_{i,j}) \in \overline{F}_n$.

An equivalent, less succinct formulation is the following:[16]

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

The Master program is a quadratic minimization problem, which can be solved quite efficiently using standard software.[17]

Given the solution to this MASTER problem, there are two possibilities. First, if the optimal solution v_J equals zero, i.e., $\widehat{f} = \psi$, then as $\text{co}(\{f_1, \ldots, f_J\}) \subseteq \text{co}(\overline{F}_n)$, we know that (11) also has an optimal solution equal to zero, so $d(\psi, \text{co}(\overline{F}_n)) = 0$ and (8) is satisfied. In the other case, $v_J > 0$. If so, we need to find out whether we should continue the iteration using an enlarged inner approximation of $\text{co}(\overline{F}_n)$, or whether one can stop having effectively found the square distance between ψ and $\text{co}(\overline{F}_n)$. 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 $v_J = d(\psi, \text{co}(\overline{F}_n))^2$ if and only if:*

$$\frac{1}{n}\sum_{i=1}^{n} h_{F(y_i)}(\eta) = h_{\overline{F}_n}(\eta) = \sup_{f \in \overline{F}_n} \langle \underbrace{\psi - \widehat{f}}_{\eta}, f \rangle \leq \langle \underbrace{\psi - \widehat{f}}_{\eta}, \widehat{f} \rangle. \tag{13}$$

Proposition 6 easily follows from the characterization of the convex projection in Proposition 5. This proposition gives us a certificate of optimality that allows us to decide when to stop the algorithm. Indeed, in case the inequality in (13) is satisfied, we know that we have found the distance between ψ and $\text{co}(\overline{F}_n)$. Note that (13) can

[16] To see that the two formulations are equivalent, you can use the transformation $\widehat{f} = \sum_{j=1}^{J} w_j f_j$ and $\eta = \psi - \widehat{f}$

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

be rewritten as

$$\frac{1}{n}\sum_{i=1}^{n}\sup_{f\in F(y_i)}\langle\eta,f\rangle\leq\langle\eta,\widehat{f}\rangle \tag{14}$$

This suggests the following procedure: as long as $v_J > 0$, within each iteration of the master program, solve the following n auxiliary optimization problem:

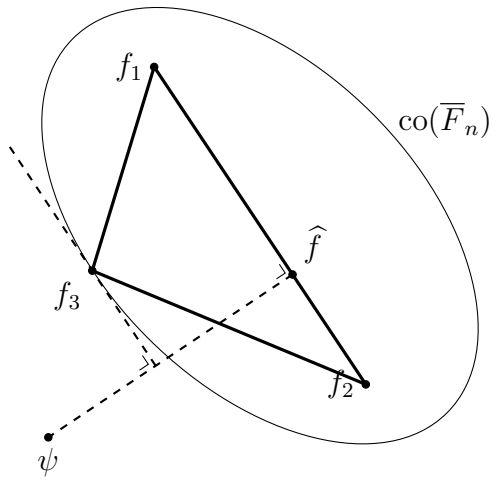
$$\begin{aligned}\textbf{AUX}(i)\textbf{:}\qquad a_i &= \sup_{f\in F(y_i)}\langle\eta,f\rangle \\ &= \sup_{u\in\mathcal{U}}\langle\eta,f(y_i,u)\rangle \text{ s.t. } (y_i,u)\in\Gamma.\end{aligned} \tag{15}$$

Given the optimal values a_i verify whether $\sum_i a_i \leq \langle\eta,\widehat{f}\rangle$. In case it is, (14) is satisfied which means that $v_J = d(\psi,\text{co}(\overline{F}_n))^2$. On the other hand, if $\sum_i a_i > \langle\eta,\widehat{f}\rangle$, we 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 $f_{J+1} = \frac{1}{n}\sum_{i=1}^{n} f_i$ where f_i is a solution to AUX(i). Note that by definition, $f_{J+1}\in\overline{F}_n$ so $\{f_1,\ldots,f_{J+1}\}\subset\overline{F}_n$.

Figure 1 depicts the idea behind the algorithm. There is a convex set $\text{co}(\overline{F}_n)$, whose shape 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 ψ 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 ψ onto $\text{co}(\{f_1,f_2\})$. This projection is denoted by $\widehat{f}$. The vector defined by the difference between ψ 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 $\eta = \psi - \widehat{f}$ 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 $\|\eta\| = \|\psi - \widehat{f}\|$ will decrease.

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

Figure 1: Illustration algorithm



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 $d(\psi, \mathrm{co}(\overline{F}_n))$ arbitrarily close, and show its convergence properties.

Algorithm 1 gives the full the algorithm presented in the previous subsection. Lines 1-3 initialize the algorithm. Here we start with a singleton $D_0 = \{f_0\}$. In practice, however, we might use any finite set of elements from $\overline{F}_n$. Then we start the loop. Line 5 solves the Master program after which we defines the solution $\widehat{f}_J, \eta_J$ and the optimal value v_J. On line 8 we check 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 programs and saves its solution in f_{J+1}. Then we verify whether the angle between η_J and $f_{J+1} - \widehat{f}_J$ is close enough to satisfying the condition of Proposition 5. If so, we quit the loop with optimal value $\|\eta_J\|$. Else, we update D_J to $D_{J+1} = \mathrm{co}(D_J \cup \{f_{J+1}\})$ and increment for the next iteration.

Remark that the auxilliary optimization problems on line 12 are independent of each other and can therefore be run in parallel, which would speed up the running time.

Algorithm 1 Column generation program

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

$J \leftarrow 0$
$D_J = \{f_0\}$ for some $f_0 \in \overline{F}_n$
$dist \leftarrow 0$
while True **do**
    $\widehat{f}_J \leftarrow \arg\min_{f \in D_J} \|\psi - f\|^2$ ▷ Master program
    $v_J \leftarrow \|\psi - \widehat{f}_J\|^2$
    $\eta_J = \psi - \widehat{f}_J$
    if $v_J < \varepsilon$ **then**
        Return 0
    else
        $f_{J+1} \leftarrow \frac{1}{n} \sum_{i=1}^{n} f_i$ where for all $i \leq n$,
        $f_i \in \arg\max_{f \in F(y_i)} \langle \eta_J, f \rangle$ ▷ Auxiliary programs
        if $\langle \eta_J, f_{J+1} \rangle > \langle \eta_J, \widehat{f}_J \rangle + \delta \|\eta_J\|^2$. **then**
            $D_{J+1} = \mathrm{co}(D_J \cup \{f_{J+1}\})$
            $J \leftarrow J + 1$
        else
            Return $\|\eta_J\|$
        end if
    end if
end while

Further, if at some iteration, for some f_{J+1}, we have $\langle \eta_J, f_{J+1} \rangle < \langle \eta_J, \psi \rangle$, then we can exclude the case $d(\psi, \text{co}(\overline{F}_n)) = 0$. 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 ψ from $\text{co}(\overline{F}_n)$. In this case the true distance $d(\psi, \text{co}(\overline{F}_n))$ will be bounded from below by the difference $\langle \eta_J, \psi \rangle - \langle \eta_J, f_{J+1} \rangle$. So if the goal is to determine whether $\psi \in \text{co}(\overline{F}_n)$ or to obtain a lower bound on the distance (without computing the actual distance) one can easily add this additional check to the algorithm.

Note that the algorithm depends on the two tuning parameters ϵ and δ. In practice, we need these tuning parameters as our implementation will be naturally limited by the tolerances set by the software package (or hardware) used.[18]

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

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

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

if it stops at line 17.

The following Theorem shows that, as long as the algorithm runs, the squared distance $\|\psi - \hat{f}_J\|^2$ decreases exponentially fast over the iterations. In particular, for any value of $\varepsilon > 0$, our algorithm terminates in finite time.

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

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

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 more

[18] This type of limitation is present in most maximisation routines. For example gradient based maximisation need to set a tolerance to decide when to quit the loop and decide that an (approximate) maximum is found. Our algorithm simply make these tolerances explicit. In our implementation, we choose $\varepsilon = 10^{-6}$ and $\delta = 10^{-3}$. Here we are guided by the default tolerances imposed by the Gurobi solver which is 10^{-6}.

involved, depending on the underlying RP model and the moment conditions.

The auxiliary program consists of n maximization problems:

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

The shape of the constraint set Γ is determined by the RP conditions while the objective function is linear in $f(y_i, u)$ and therefore depends on the moment functions. The most straightforward 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 problem turns into a simple linear program, which can be solved using standard optimization software. Our first application fits this structure. In fact, many RP models (based on Afriat inequalities) have the feature that they take the form of a set of linear inequalities (see, for example, Diewert (2012) for an overview). Other (more complicated) models can often be specified as a set of linear inequalities where variables are either continuous or integer values.[19] This transforms above problems to a set of Mixed Integer programming problems, for which there exists standard solvers.

Even if the moment functions $f(y_i, u)$ are quadratic (and concave) in u, it is still possible to use quadratic programming methods to efficiently solve the auxiliary problem as we will demonstrate in our second application.

In settings where Γ takes on a more complicated structure or where f is highly nonlinear, 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 $(y_i, u) \in \Gamma$. Such sampling can be done using MCMC methods like the hit-and-run procedure. Here one starts from a feasible point u for which $(y_i, u) \in \Gamma$ and one iteratively draws a random direction δ and moves u in the direction δ if $(y_i, u + \delta) \in \Gamma$. If not, then depending on the specification of the hit and run procedure, the direction δ might be reduced in length, reversed, or a new draw might be taken. In this way, one moves over the region $\{u \in \mathcal{U} | (y_i, u) \in \Gamma\}$ until one is confident to have obtained a decent coverage of this space.[20] Once such sample is

[19] See for example Cherchye, De Rock, and Vermeulen (2007); Cherchye, Demuynck, De Rock, and Hjertstrand (2015); Demuynck and Rehbeck (2023).

[20] See, for example, Demuynck (2021) for applications of such MCMC methods in RP tests.

obtained one chooses the maximal value of $\langle \eta, f(y_i, u) \rangle$ over values u in this sample. Note that in principle this sampling procedure only has to be conducted once, before the main algorithm is run. Alternatively, one might consider a more directed search based on the values of η and $\widehat{f}$ found in the previous round of the MASTER program, as this might put additional constraints on the subregion of $F(y_i)$ where $\langle \eta, f \rangle$ is maximized.

4 Statistical inference

Recall from section 1 that the test of our model takes the following form:

$$H_0 : d(\psi, \mathbb{E}F) = 0,$$
$$H_1 : d(\psi, \mathbb{E}F) > 0.$$

Inference is then based on the "estimate" $d(\psi, \text{co}(\overline{F}_n))$, determined by the output of Algorithm 1. Using Proposition 4, we obtain:

$$d(\psi, \text{co}(\overline{F}_n)) = \left\{ \sup_{p \in \mathbb{S}^K} \{\langle p, \psi \rangle - h_{\text{co}(\overline{F}_n)}(p)\} \right\}_+ ,$$

where $\{x\}_+ = \max\{x, 0\}$. This distance can be used as the basis for a test of H_0. Consider the following test statistic:

$$T_n = \sqrt{n} d(\psi, \text{co}(\overline{F}_n)).$$

The asymptotic distribution Define the functional $\phi : \mathcal{C}\left(\mathbb{S}^K\right) \to \mathbb{R}$ pointwise by $\phi(g) = \sup_{p \in \mathbb{S}^K} \langle p, \psi \rangle - g(p)$, where $\mathcal{C}\left(\mathbb{S}^K\right)$ denotes the set of continuous functions defined on the unit sphere $\mathbb{S}^K$. We can write

$$T_n = \sqrt{n} \left\{ \phi \left(h_{\text{co}(\overline{F}_n)}(p) \right) \right\}_+$$

One can show that ϕ is Hadamard (directionally) differentiable,[21] with directional

[21] A map ϕ between two normed vector spaces, $\mathcal{A}$ and $\mathcal{B}$ is said to be Hadamard directionally differentiable at g in the direction k if, for all sequences $(t_n)_{n \in \mathbb{N}_0}$ of positive numbers converging to zero, and all sequences $(k_n)_{n \in \mathbb{N}_0}$ converging to $k \in \mathcal{A}$, the limit $\phi'_g(k) = \lim_n \frac{\phi(g + t_n k_n) - \phi(g)}{t_n}$ exists.

derivative $\phi'_g(k) = \sup_{p\in\Lambda_0} -k(p)$, where $\Lambda_0 = \arg\max_{p\in\mathbb{S}^K} \langle p,\psi\rangle - g(p)$.[22] In addition, the Central Limit Theorem for random variables that are $\mathcal{C}\left(\mathbb{S}^K\right)$-valued of Jain and Marcus (1975) implies:

$$\sqrt{n}\left(h_{\text{co}(\overline{F}_n)}(p) - h_{\mathbb{E}F}(p)\right) \xrightarrow{d} Z(p),$$

where Z is a Gaussian random process with $\mathbb{E}Z(p) = 0$, and covariance $\mathbb{E}Z(p)Z(p') = \mathbb{E}h_{F(y)}(p)h_{F(y)}(p') - \mathbb{E}h_{F(y)}(p)\mathbb{E}h_{F(y)}(p')$. Using this result and the fact that ϕ is Hadamard directionally differentiable allows us to apply Lemma D.4 in Kaido (2016), which yields that for all ψ on the boundary of $\mathbb{E}F$

$$T_n \xrightarrow{d} \sup_{p\in\Lambda_0} \{-Z(p)\}_+ \tag{16}$$

where now,

$$\Lambda_0 = \arg\max_{p\in\mathbb{S}^K} \langle p,\psi\rangle - h_{\mathbb{E}F}(p).$$

In theory, asymptotic inference can be based on (16) by finding the critical value $c_{1-\alpha}$ corresponding to the $1-\alpha$ quantile of $\sup_{p\in\Lambda_0}\{-Z(p)\}_+$. Doing this, however, is complicated by the fact that Λ_0 is not known. As such, we will consider an alternative subsampling procedure.

Algorithm 2 Subsampling

Require: $0 < \alpha < 1$, $0 < b < n$. Let $N_{n,b} = \binom{n}{b}$. A sequence $(\zeta_n)_{n\in\mathbb{N}_0}$ such that $\lim_n \zeta_n = +\infty$, and $\lim_n \frac{\zeta_n}{\sqrt{n}} = 0$.

1: **for** $k = 1,\ldots,N_{n,b}$ **do**
2: Collect the kth subsample of size b (without replacement) and let $\text{co}(\overline{F}_{n,b,k})$
3: be the convex hull of the Minkowski average for this k'th subsample.
4: Compute $d(\psi,\text{co}(\overline{F}_{n,b,k})) = \sup_{p\in\mathbb{S}^K}\langle p,\psi\rangle - h_{\text{co}(\overline{F}_{n,b,k})}(p)$.
5: **end for**
6: Compute $c_{n,b,1-\alpha}$ as the $(1-\alpha)$'th quantile of the subsampling distribution

$$H_{n,b}(z) = \frac{1}{N_{n,b}}\sum_{k=1}^{N_{n,b}} 1\left\{\sqrt{b}\left\{d(\psi,\text{co}(\overline{F}_{n,b,k})) - d(\psi,\text{co}(\overline{F}_n)) + \frac{\zeta_n}{\sqrt{n}}\right\}_+ \le z\right\}$$

[22] A derivation of this is available from the authors upon request.

Subsampling Algorithm 2 gives our subsampling procedure.[23] For each subsample, it determines the distance $d(\psi, \text{co}(\overline{F}_{n,b,k}))$ where $\text{co}(\overline{F}_{n,b,k})$ is the convex hull of the Minkowski average over the sets $F(y_i)$ in the subsample. This distance can readily be computed using Algorithm 1. The critical value for the null hypothesis is obtained as the $1-\alpha$ percentile of the subsampling distribution of $\sqrt{b}\left\{d(\psi, \text{co}(\overline{F}_{n,b,k})) - d(\psi, \text{co}(\overline{F}_n)) + \frac{\zeta_n}{\sqrt{n}}\right\}_+$.
The difference in distances accounts for the difference $d(\psi, \text{co}\,\overline{F}_n)) - d(\psi, \mathbb{E}F)$ between the finite sample test statistic and the actual distance. The additional small correction term $\frac{\zeta_n}{\sqrt{n}}$ accounts for the fact that Λ_0 is unknown and somehow needs to be approximated. A test of $H_0 : d(\psi, \mathbb{E}F) = 0$ based on the empirical distribution of the test statistic, computed through this subsampling approach can be shown to be consistent, in the following sense:

Theorem 9. *Let $(b_n)_{n\in\mathbb{N}}$ be a sequence such that $b_n \to +\infty$ and $(b_n/n) \to 0$ and let $c_{n,b_n,1-\alpha}$ be the $1-\alpha$ quantile of the empirical distribution $H_{n,b_n}(z)$ as determined by Algorithm 2. Consider $\hat{c}_{n,b_n,1-\alpha} = c_{n,b_n,1-\alpha} + \epsilon$, where ϵ is an arbitrarily small positive (real) number. Then, if $\psi \in \mathbb{E}F$, and $\alpha \in (0, 0.5)$, it follows that:*

$$\limsup_n \ \Pr\left(T_n > \tilde{c}_{n,b,1-\alpha}\right) \leq \alpha,$$

and for $\psi \notin \mathbb{E}F$ and $\alpha \in (0,1)$,

$$\lim_n \ \Pr\left(T_n > \tilde{c}_{n,b,1-\alpha}\right) = 1.$$

The proof of this Theorem follows from Corollary 3.2 in Kaido (2016) together with a modification of his bootstrap procedure (we refer to the proof in the Appendix for more details).

5 Applications

In this section, we provide two RP applications of our results.

[23] We refer to Politis, Romano, and Wolf (1999) for a full treatment on subsampling.

5.1 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[24] (Browning, 1989; Crawford, 2010; Demuynck and Verriest, 2013; Adams, Cherchye, De Rock, and Verriest, 2014).

An important remark is that adding uncertainty to these models without additional structure 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.[25] 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 random consumer (household) with a concave, continuous and monotone instantaneous utility function U and discount rate β that chooses consumption amounts according to the standard infinite horizon EDU model:

$$\begin{aligned} &\max_{(c_t)_{t=0,1,\ldots}} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t U\left(c_t\right) \middle| \mathcal{I}_0\right] \\ &\text{s.t. } s_{t+1} = m_t + (1+r_t)s_t - \langle p_t, c_t \rangle, \\ &s_0 \text{ given.} \end{aligned}$$

[24] Aguiar and Kashaev (2021) is an exception.

[25] 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).

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

The Bellman equation for this problem is given by:

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

Using standard arguments, one can show that this value function is concave. The first order and envelope conditions (for good j) are given by:[26]

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

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

$$\frac{\partial U(c_t)}{\partial c_j} = \beta^{-t}\lambda_t \rho_{t,j}, \tag{17}$$
$$\mathbb{E}\left[\lambda_t - \lambda_{t+1}|\mathcal{I}_t\right] = 0. \tag{18}$$

Equation (17) gives the first order condition for consumption. Equation (18) 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 (18) is a (conditional) moment condition and implies that any element in $\mathcal{I}_t$ must be orthogonal to the growth in marginal utility of income, $\lambda_{t+1} - \lambda_t$. From this conditional moment condition, we can derive a large set of implied unconditional moment conditions. For example, for any variable $x_t \in \mathcal{I}_t$ we need that $\mathbb{E}\left[(\lambda_{t+1} - \lambda_t)x_t\right] = 0$.

To make the connection with the framework outlined in Section 1, we correspond the observable component y to the (random) data set: $y = (\rho_t, c_t, \mathcal{I}_t)_{t\leq T}$. We specify the

[26] In case U^i is not differentiable, we take corresponding subdifferentials.

unobserved latent component of the model to be the collection of marginal utility of income levels: $u = (\lambda_t)_{t\leq T}$. Next, we need to define the set Γ of all combinations (y, u) that satisfy the model. In order to do this, we need to establish the relevant RP conditions. First we define what it means to be rationalizable in our setting.

Definition 10. *The pair $(y, u) = ((\rho_t, c_t, \mathcal{I}_t)_{t\leq T}, (\lambda_t)_{t\leq T})$ is said to be rationalizable by the EDU model if there exists a concave, locally non-satiated and continuous utility function U such that (17) is satisfied given the values of ρ_t, c_t and λ_t.*

We define Γ as the collection of all (y, u) that are EDU rationalizable. The following revealed preference characterization of the set will be helpful.[27]

Lemma 11. *Let $\beta \in (0, 1)$, the pair $(y, u) = ((\rho_t, c_t, \mathcal{I}_t)_{t\leq T}, (\lambda_t)_{t\leq T})$ is EDU rationalizable if and only if there are numbers $(U_t)_{t\leq T}$ such that for all $t, v \leq T$:*

$$U_t - U_v \leq \frac{\lambda_v}{\beta^v}\langle \rho_v, (c_t - c_v)\rangle. \tag{19}$$

The Lemma provides us with the following definition of the set Γ:

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

In order to make the correspondence Γ satisfy Assumption 2, we can further limit $\lambda_t \in [\lambda_L, \lambda_U]$ for some lower bound λ_L and upper bound λ_U. As is clear from the Afriat inequalities (19), imposing a lower or upper bound is without loss of generality as both variables $(U_t)_{t\in\mathcal{T}}$ and $(\lambda_t)_{t\in\mathcal{T}}$ 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.

The system of inequalities in (19) are known as *Afriat inequalities* and are obtained from combining concavity of the utility function U, together with the first order conditions. Obviously, in practice, we do not observe the realizations of $u = (\lambda_t)_{t\in T}$. One solution would be to study whether there exists for each realization y_i of y at least one one instance u_i for which (19) holds. However, this amount of flexibility limits the additional bite present in intertemporal models. Indeed, such a loosening

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

of rationalizability would lead to the same testable implications as for static utility maximization (i.e., GARP).[28] 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 using the Euler conditions (18).

As the information sets $\mathcal{I}_t$ might be large, this condition contains indeed a large (potentially infinite) number of implied moment conditions.

For the application, we impose several conditions of the following form:

$$\mathbb{E}\left[(\lambda_{t+1} - \lambda_t)x_{k,t}\right] = 0. \tag{20}$$

where $x_{k,t}$ is some variable in the information set $\mathcal{I}_t$ $(k = 1, \ldots, m)$. As $T = 4$, this gives for every k, 3 moment functions: $f_{k,t}(y, u) = (\lambda_{t+1} - \lambda_t)x_{k,t}$, so there are in total $K = 3m$ moment functions. Our algorithm is quite flexible to increase the number of moment conditions as long as they are linear in the latent variables $u = (\lambda_t)_{t \leq T}$ as this means that the AUXILIARY programs (15) takes the form of a linear program (conditional on β). A first, rather obvious choice is $x_{1,t} = 1$, which gives:

$$\mathbb{E}\left[\lambda_{t+1} - \lambda_t\right] = 0, \text{ for all } t \leq T - 1, \tag{21}$$

This requires that the mean difference between λ_{t+1} and λ_t (over the population of individuals) to be zero for all $t \leq T - 1$.

To be specific, given the set $\{f_1, \ldots, f_j, \ldots, f_J\}$ of elements in $\text{co}(\overline{F}_n) \subset \mathbb{R}^K$, our MASTER program takes the following form:

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

[28] See Varian (1982).

Notice that here $\psi = 0$ 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 $y_i = (\rho_t^i, c_t^i)_{t \leq T}$ from the sample $(y_1, \ldots, y_i, \ldots, y_n)$, we solve the following problem.

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

Here we indexed all variables by i to convey that these differ for different observations y_i. It is easily to see that AUX(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.[29]

5.1.1 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, including a 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

[29] 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.

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, Browning, and Crawford (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 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,[30] 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.

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

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. Moreover, our test allows each household to have distinct preferences. 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 (20). 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.

5.1.2 Results

In order to be able to apply our RP analysis, we first need to be sure that there is at least one value $u = (\lambda_t)_{t\in T}$ for which the Afriat conditions in Lemma 11 hold. In other words the sets $F(y_i)$ are non-empty. As this is not always the case, we first look for the largest $e \leq 1$ for Such that for some U^i_t and $\lambda^i_t > 0$ $(t = 1, \ldots, T)$:

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

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^i_v by $e \cdot c^i_v$ where necessary.

Concerning the moment conditions, recall that the moment functions $f_{k,t}(y, u)$ are of the form:

$$f_{k,t}(y, u) = (\lambda_{t+1} - \lambda_t)x_{k,t}. \tag{22}$$

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

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

$x_{k,t} = 1$ gives the usual condition that expected marginal utility should not change. In addition we pick for $x_{k,t}$ also 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 $x_{k,t}$ 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.

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 the RP tests all time periods are aligned. As we have data from 1991 until 2012, this gives us 84 samples per category. For the analysis below, we discard samples with less than 40 observations.

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

For each of these samples 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 (i.e. whether $d(\psi, co(\overline{F}_n)) = 0$ or not).[31] 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 $\beta \in \{1, 0.99, 0.97, 0.95\}$.

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

5.2 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

[31] In practice we consider the test not rejected iff $d(\psi, \text{co}(\overline{F}_n)) \leq 0.01$. Here we do not use the bootstrap inference as the number of observations in every sample is already quite low. Note, however, that in the regression, we do control for the number of observations in each sample.

a (radom) decision maker making choices over Arrow-Debreu securities over a finite set of states, $\{1, \ldots, S\}$. We assume that decision makers are risk averse and that each one has a a C^1 concave and strictly increasing Bernoulli utility function, $U : \mathbb{R}_+ \to \mathbb{R}$. We will assume a (lab) experimental setting where the probabilities over states are controlled by the econometrician (hence objective), however, the decision maker is allowed to have some misperception of these probabilities, meaning that she is maximizing subjective expected utility as in Savage (1954). The choice problem for a certain (random) decision maker over the securities can then be represented as follows:

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

Here $\widetilde{P}_s$ is the subjective probability attached to state s, and p_s is the price of the Arrow-Debreu securities for state s.[32] The values $(x_s)_{s\leq S}$ give the amounts of the Arrow-Debreu securities over the states in $\mathcal{S}$. The first order conditions require:

$$\begin{aligned} &\frac{dU(x_s)}{dx} = \frac{\lambda p_s}{\widetilde{P}_s} && \forall s \leq S, \\ &\sum_{s\leq S} p_s x_s = m. \end{aligned}$$

where λ is the Lagrange multiplier for the budget constraint. Taking logs on both sides of the first equation, we obtain:

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

where we used the notation $\mu = \ln(\lambda)$, $\rho_s = \ln(p_s)$ and $\widetilde{\pi}_s = \ln\left(\widetilde{P}_s\right)$. 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.

[32] Although the probabilities $\widetilde{P}_s$ 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.

This model is quite similar to Echenique, Imai, and Saito (2023), 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:

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

In other words:

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

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

$$\ln\left(\frac{dU}{dx}(x)\right) = \mu + \rho_s - \pi_s - \varepsilon_s \qquad \forall s \leq S \tag{23}$$

We assume that the values ε_s are unobserved by the econometrician.

To make the connection to the setting of Section 1, we will relate the observed (random) component y to the finite dataset $y = (\pi_{s,t}, \rho_{s,t}, x_{s,t})_{t\leq T, s\leq S}$, where $t \leq T$ denotes an observation. The unobserved component is given by the set of errors $u = (\varepsilon_{s,t})_{s\leq S, t\leq T}$, where $\varepsilon_{s,t}$ gives the error at observation t for state s..

For defining the set Γ, we use the following notion of rationalizability:

Definition 12. *We say that the pair $(y,u) = ((\pi_{s,t}, \rho_{s,t}, x_{s,t})_{s\leq S, t\leq T}, (\varepsilon_{s,t})_{s\leq S, t\leq T})$ is expected utility (EU) rationalizable if there exists a C^1, strictly increasing and concave utility function U such that condition (23) is satisfied for all $t \leq T$ and $s \leq S$.*

The set Γ then equals the collection of all pairs (y,u) that are EU rationalizable. The following result provides its revealed preference characterization.

Lemma 13. *The couple $(y,u) = ((\pi_{s,t}, \rho_{s,t}, x_{s,t})_{s\leq S, t\leq T}, (\varepsilon_{s,t})_{s\leq S, t\leq T})$ is EU rationalizable if and only if there exist values $(\mu_t)_{t\leq T}$ such that for all observations $t, v \leq T$*

and all states $s, w \leq S$*:*

$$\begin{aligned} &\textit{if } x_{s,t} \geq x_{w,v} \\ &\textit{then } \mu_t + \rho_{s,t} - \pi_{s,t} - \varepsilon_{s,t} \leq \mu_v + \rho_{w,v} - \pi_{w,v} - \varepsilon_{w,v}. \end{aligned} \tag{24}$$

A similar result has been derived and proven in Echenique and Saito (2015) and Echenique, Imai, and Saito (2023).[33] The Afriat-style inequalities in (24) are directly related to the diminishing marginal utility of any Bernoulli map that EU-rationalizes (y, u). Indeed, if consumption increases across two particular states and choice instances, then the marginal utility of consumption decreases.

Using Lemma 13, we define Γ by all couples (y, u) for which we can find $(\mu_t)_{t \leq T}$ such that condition (24) is satisfied.

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

$$\mathbb{E}\left[\sum_{t \leq T} \sum_{s \leq S} \varepsilon_{s,t}\right] = 0. \tag{25}$$

Although this condition fixes the mean, by itself it is not very restrictive, as we can always add a constant to all values $\varepsilon_{s,t}$ without influencing the revealed preference conditions (24). As an additional condition, we assume that these errors are uncorrelated with the prices which imposes that:

$$\mathbb{E}\left[\sum_{t \leq T} \sum_{s \leq S} \varepsilon_{s,t} p_{s,t}\right] = 0. \tag{26}$$

Finally, we will add a restriction that bounds the error's variance:[34]

$$\mathbb{E}\left[\sum_{t} \sum_{s \leq S} (\varepsilon_{s,t})^2\right] \leq TS\sigma^2 \tag{27}$$

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

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

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

sense that the resulting feasibility set remains convex.

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

The first $K-1$ components of f_j correspond to the moment functions related to (25) and (26). The last component $(f_j)_K$ relates to the variance condition (27). In particular, $f_K(y,u) = \sum_t \sum_{s \leq S} (\varepsilon_{s,t})^2$. Observe that the optimal solution will always have $\eta_K \leq 0$, which is relevant when going to the auxiliary programs:

$$
\begin{aligned}
\textbf{AUX(i):} \max_{\varepsilon^i_{s,t}} & \sum_{k=1}^{K} \eta_k f_k(y_i, u_i), \\
& \text{s.t. } (y_i, u_i) \in \Gamma.
\end{aligned}
$$

Conveniently, Γ takes the form of a set of linear inequalities (given by (24)). All moment functions $f_k(y,u)$ are linear in u except for the last one that gives a quadratic term:

$$
f_K(y_i, u_i) = \sum_{t \leq T} \sum_{s \leq S} \left(\varepsilon^i_{s,t}\right)^2.
$$

However, given that $\eta_K \leq 0$, the program AUXI(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.

5.2.1 Data and Illustration

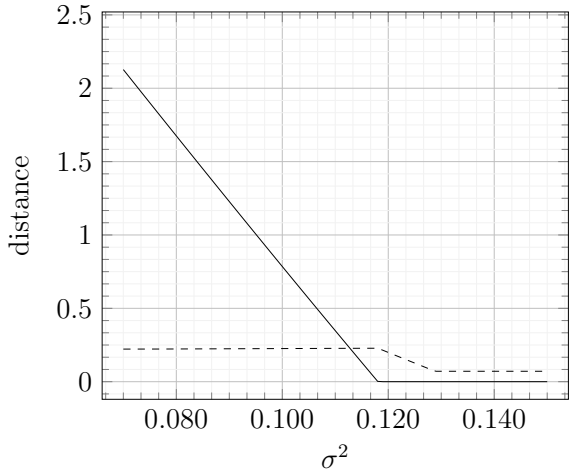
For our illustration, we use experimental dataset of Choi, Kariv, Müler, and Silverman (2014). The dataset is from an experiment on the CentERpanel, which is representa-

tive of the Dutch-speaking population in the Netherlands. The data is obtained from an online experiment with 1182 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 $d(\psi, \text{co}(\overline{F}_n))$ computed by the algorithm as a function of σ^2. We also perform the subsampling procedure outlined in section 4. In particular, for each value of σ^2 in a grid we draw 500 subsamples (without replacement) of size $141 \approx n^{0.7}$ and we specify $\zeta_n = \ln(n)$. The dashed line in the figure plots the value of $d(\psi, \text{co}(\overline{F}_n))$ as a solid line and the critical value $\frac{c_{n,b,0.95}}{\sqrt{n}}$ as a dashed line. We see that the distance linearly decreases until σ^2 is around 0.12 after which it is indistinguishable from zero. We cannot reject the null hypothesis that $d(\psi, \mathbb{E}F) = 0$ at the 5% significance level for values of σ^2 above 0.113. This gives a minimal standard deviation of around 0.34. This gives a value of about 1.4 when exponentiated, which is quite sizable.

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



Note: the solid line gives the value of $d(\psi, \text{co}(\overline{F}_n))$ for various values of σ^2. The dashed curves gives the critical value for the null that $d(\psi, \mathbb{E}F) = 0$.

6 Conclusion

We presented a conceptual framework based on random sets to test revealed preference models with unobserved randomness. To be more precise, we have shown that a test of an RP model with randomness boils down to testing whether or not a vector, ψ, 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 the Minkowski average. As a final step in our theoretical contribution, we showed that checking whether ψ lies in the Minkowski average can be done efficiently using a column generating method. We further provide a subsampling procedure to conduct inference.

The second half of the paper 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 (2021). 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., B. De Rock, and F. Vermeulen (2007). The collective model of household consumption: a nonparametric characterization. *Econometrica 75*, 553–574.

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

Cherchye, L., T. Demuynck, B. De Rock, and P. Hjertstrand (2015). Revealed preference tests for weak separability: an integer programming approach. *Journal of Econometrics 186*, 129–141.

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 preference consistent 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 J. Rehbeck (2023). Computing revealed preference goodness of fit measures with integer programming. *Economic Theory 76*, 1175–1195.

Demuynck, T. and C. Staner (Forthcoming). An efficient revealed preference test for the maxmin expected utility model. *American Economic Journal: Microeconomics*.

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, E. (2012). Afriat's theorem and some extensions to choice under uncertainty. *Economic Journal 122*, 305–331.

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

Echenique, F., T. Imai, and K. Saito (2023). Approximate expected utility rationalization. *Journal of the European Economic Association 21*(5).

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*.

Jain, N. C. and M. B. Marcus (1975). Central limit theorems for $c(s)$-valued random variables. *Journal of Functional Analysis 19*(3), 216–231.

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

Kaido, H. (2016). A dual approach to inference for partially identified econometric models. *Journal of Econometrics 192*, 269–290.

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.

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

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.

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 Theorem 7

If the Algorithm stops at line 9, then $d(\psi, \text{co}(\overline{F}_n)) \leq \|\psi - \widehat{f}_J\|^2 \leq \sqrt{\varepsilon}$, which shows the first claim. Next assume that the algorithm stops at line 17 at iteration J. Let $\widehat{f} = \arg\min_{f \in \text{co}(\overline{F}_n)} \|\psi - f\|^2$ be the projection of ψ onto $\text{co}(\overline{F}_n)$. We have that:

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

As such:

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

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 17) then:

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

This gives:

$$\|\psi - \widehat{f}_J\|^2 \leq \frac{\|\psi - \widehat{f}\|^2}{1 - 2\delta} = \frac{d(\psi, \text{co}(\overline{F}_n))^2}{1 - 2\delta}$$

A.2 Proof of Theorem 8

Assume that the algorithm does not terminate before iteration J, which means that for all $j < J$, $\langle \eta_k, f_{j+1} - \widehat{f}_j\rangle > \delta\|\psi - \widehat{f}_j\|^2$ and $\|\psi - \widehat{f}_j\|^2 \geq \varepsilon$.
Notice that for all $\lambda \in [0, 1]$, $\lambda f_{j+1} + (1-\lambda)\widehat{f}_j \in \text{co}(\{f_1, \ldots, f_{j+1}\}$ (as $\widehat{f}_j \in \text{co}(\{f_1, \ldots, f_j\})$).

As such, for all $\lambda \in [0,1]$:

$$\begin{aligned}\|\psi - \widehat{f}_{j+1}\|^2 &\leq \|\psi - (\lambda f_{j+1} + (1-\lambda)\widehat{f}_j)\|^2, \\ &= \|\psi - \widehat{f}_j\|^2 + \lambda^2\|\widehat{f}_j - f_{j+1}\|^2 + 2\lambda\langle \psi - \widehat{f}_j, \widehat{f}_j - f_{j+1}\rangle, \\ &= \|\psi - \widehat{f}_j\|^2 + \lambda^2\|\widehat{f}_j - f_{j+1}\|^2 - 2\lambda\langle \eta_j, f_{j+1} - \widehat{f}_j\rangle,\end{aligned}$$

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_{j+1}. As such, the optimal value is either interior or equal to 1. If it is interior, the the first order condition gives:

$$\begin{aligned}2\lambda\|f_{j+1} - \widehat{f}_j\|^2 - 2\langle \eta_j, f_{j+1} - \widehat{f}_j\rangle = 0, \\ \rightarrow \lambda = \frac{\langle \eta_j, f_{j+1} - \widehat{f}_j\rangle}{\|f_{j+1} - \widehat{f}_j\|^2} \geq \frac{\delta\|\psi - \widehat{f}_j\|^2}{\|f_{j+1} - \widehat{f}_j\|^2} \geq \frac{\delta\varepsilon}{A}.\end{aligned}$$

Here we used the fact that $\|\psi - \widehat{f}_k\|^2 \geq \varepsilon$ 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, if necessary further increasing A, we can guarantee that, $0 < \frac{\delta\varepsilon}{A} < 1$ Next, if $\lambda = 1$ is optimal, then:

$$\begin{aligned}2\|f_{j+1} - \widehat{f}_j\|^2 - 2\langle \eta_j, f_{j+1} - \widehat{f}_j\rangle \leq 0, \\ \rightarrow \|f_{j+1} - \widehat{f}_j\|^2 \leq \langle \eta_j, f_{j+1} - \widehat{f}_j\rangle.\end{aligned}$$

Now, if λ is interior, then:

$$\begin{aligned}\|\psi - \widehat{f}_{j+1}\|^2 &\leq \|\psi - \widehat{f}_j\|^2 + \lambda^2\|\widehat{f}_j - f_{j+1}\|^2 + 2\lambda\langle\eta_j, \widehat{f}_j - f_{j+1}\rangle, \\ &= \|\psi - \widehat{f}_j\|^2 - \lambda\langle\eta_j, f_{j+1} - \widehat{f}_j\rangle, \\ &\leq \|\psi - \widehat{f}_j\|^2 - \lambda\delta\|\psi - \widehat{f}_j\|^2, \\ &= (1 - \lambda\delta)\|\psi - \widehat{f}_j\|^2, \\ &\leq (1 - \frac{\delta^2\varepsilon}{A})\|\psi - \widehat{f}_j\|^2.\end{aligned}$$

If $\lambda = 1$ then:

$$\begin{aligned}\|\psi - \widehat{f}_{j+1}\|^2 &\leq \|\psi - \widehat{f}_j\|^2 + \|\widehat{f}_j - f_{j+1}\|^2 - 2\langle\eta, f_{j+1} - \widehat{f}_j\rangle, \\ &\leq \|\psi - \widehat{f}_j\|^2 - \langle\eta_j, f_{j+1} - \widehat{f}_j\rangle, \\ &\leq (1 - \delta)\|\psi - \widehat{f}_j\|^2.\end{aligned}$$

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

$$\|\psi - \widehat{f}_{j+1}\|^2 \leq \theta\|\psi - \widehat{f}_j\|^2.$$

As this holds for all $j < J$ we get:

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

which proves the result.

A.3 Proof of Theorem 9

Let

$$\hat{\Lambda}_n = \left\{p \in \mathbb{S}^K : \langle p, \psi\rangle - h_{\mathrm{co}(\overline{F}_n)}(p) \geq \sup_{p' \in \mathbb{S}^K}\left(\langle p', \psi\rangle - h_{\mathrm{co}(\overline{F}_n)}(p')\right) - \frac{\zeta_n}{\sqrt{n}}\right\},$$

Set,

$$\widetilde{Z}_{n,b,k}(p) = \sqrt{b}\left(h_{\mathrm{co}(\overline{F}_{n,b,k})}(p) - h_{\mathrm{co}(\overline{F}_n)}(p)\right)$$

and define,

$$\widetilde{H}_{n,b}(z) = \frac{1}{N_{n,b}} \sum_{k=1}^{N_{n,b}} 1\left\{ \sup_{p \in \hat{\Lambda}_n} \{-Z_{n,b,k}(p)\}_+ \leq z \right\}$$

Let $\widetilde{c}_{n,b,1-\alpha}$ be the $1-\alpha$ quantile of the empirical distribution $\widetilde{H}_{n,b}(z)$. From Corollary 3.2 and the subsampling procedure proposed in Kaido (2016), we have that Theorem 9 holds when we replace for the statement of the theorem the critical value $c_{n,b,1-\alpha}$ by the critical value $\widetilde{c}_{n,b,1-\alpha}$ of $\widetilde{H}_{n,b}(z)$. For the proof to hold, this means that we only need to show that $\widetilde{c}_{n,b,1-\alpha}$ is no larger than $c_{n,b,1-\alpha}$.

First, from the construction of $\hat{\Lambda}_n$ we have that for all $p \in \hat{\Lambda}_n$:

$$d(\psi, \text{co}(\overline{F}_n)) - \frac{\zeta_n}{\sqrt{n}} \leq \langle \psi, p \rangle - h_{\text{co}(\bar{F}_n)}(p).$$

Then

$$\widetilde{Z}_{n,b,k}(p) \geq \sqrt{b}\left(d(\psi, \text{co}(\bar{F}_n)) - \langle \psi, p \rangle + h_{\text{co}(\bar{F}_{n,b,k})}(p) - \frac{\zeta_n}{\sqrt{n}} \right)$$

So

$$\begin{aligned}
\sup_{p \in \hat{\Lambda}_n} \left\{ -\tilde{Z}_{n,b,k}(p) \right\}_+ &\leq \sqrt{b} \sup_{p \in \hat{\Lambda}_n} \left\{ \langle \psi, p \rangle - h_{\text{co}(\bar{F}_{n,b,k})}(p) - d(\psi, \text{co}(\bar{F}_n)) + \frac{\zeta_n}{\sqrt{n}} \right\}_+, \\
&\leq \sqrt{b} \sup_{p \in \mathbb{S}^K} \left\{ \langle \psi, p \rangle - h_{\text{co}(\bar{F}_{n,b,k})}(p) - d(\psi, \text{co}(\bar{F}_n)) + \frac{\zeta_n}{\sqrt{n}} \right\}_+, \\
&= \sqrt{b} \left\{ d(\psi, \text{co}(\overline{F}_{n,b,k})) - d(\psi, \text{co}(\overline{F}_n)) + \frac{\zeta_n}{\sqrt{n}} \right\}
\end{aligned}$$

As such,

$$\begin{aligned}
\widetilde{H}_{n,b}(z) &\geq \frac{1}{N_{n,b}} \sum_{k=1}^{N_{n,b}} 1\left\{ \sqrt{b} \left\{ d(\psi, \text{co}(\overline{F}_{n,b,k})) - d(\psi, \text{co}(\overline{F}_n)) + \frac{\zeta_n}{\sqrt{n}} \right\}_+ \leq z \right\}, \\
&= H_{n,b}(z).
\end{aligned}$$

By definition $H_{n,b}(c_{n,b,1-\alpha}) = (1-\alpha) = \widetilde{H}_{n,b}(\widetilde{c}_{n,b,1-\alpha}) \geq H_{n,b}(\widetilde{c}_{n,b,1-\alpha})$. As $H_{n,b}$ is a weakly increasing function, we have that $c_{n,b,1-\alpha} \geq \widetilde{c}_{n,b,1-\alpha}$.