2019

2019-5 Revealed Statistical Consumer Theory

Roy Allen

Pawel Dziewulski

John Rehbeck

Revealed Statistical Consumer Theory

by

Roy Allen, Pawel Dziewulski, and John Rehbeck

Research Report # 2019-5 **October 2019**

Department of Economics

Research Report Series

Department of Economics
Social Science Centre
Western University
London, Ontario, N6A 5C2
Canada

Revealed statistical consumer theory*

Roy Allen

Department of Economics
University of Western Ontario
rallen46@uwo.ca

Paweł Dziewulski

Department of Economics
University of Sussex
P.K.Dziewulski@sussex.ac.uk

John Rehbeck

Department of Economics
Ohio State University
rehbeck.7@osu.edu

October 22, 2019

Abstract

We investigate a model of deterministic stochastic choice for the standard consumer problem. We introduce the framework of *statistical consumer theory* where the individual maximizes their utility with respect to a distribution of bundles that is constrained by a statistic (e.g. mean expenditure). We show that this behavior is observationally equivalent to an individual whose preferences depend only on the statistic of the distribution. Statistical consumer theory neither nests nor is nested in the random utility approach. We provide a formal statistical test of the model accounting for sampling variability and demonstrate it in an illustrative example using data on capuchin monkeys.

*We thank Keith Chen for providing access to the data.

1 Introduction

Stochastic choice occurs when an individual faces a budget set several times and produces a distribution of choices from that budget set. In economics, a large literature has emerged to understand stochastic choice. There are two main approaches. The first one assumes that stochastic choices are generated by random preference shocks. This led to the development of random utility models (RUMs).[1] The second approach assumes that stochastic choices are generated by an individual who chooses a most-preferred distribution of bundles and randomizes according to that distribution. Following the literature, we call this approach *deterministic stochastic choice*, since individuals have a deterministic preference over the distributions of commodity bundles.[2] In this paper, we examine deterministic stochastic choice for the standard consumer problem when an individual maximizes their utility with respect to distributions over consumption bundles that are constrained by a statistic. We call this framework a *statistical choice model*. We show that these restrictions are equivalent to a model where the individual's preferences depend only on the statistic that constrains the distributions.

Within the paper, we focus on restrictions from the mean statistic. We are interested in the case where an individual faces budget restrictions that depend on *mean expenditure*, however the ideas readily extend to other statistics. We consider the standard consumer problem where an individual can choose a distribution of consumption bundles, but faces a budget constraint that mean expenditure is less than income. The main result (Theorem 1) shows that an individual makes choices consistent with the constraint on mean expenditures if and only if the generalized axiom of revealed preference (Afriat, 1967; Diewert, 1973; Varian, 1982) is satisfied on mean consumption bundles. This result follows while only assuming local nonsatiation of the preference over distributions.

The main result also shows that when the budget constraints depend only on mean expenditure, it is equivalent to assume an individual (i) has a nonsatiated preference over *all* distributions, or (ii) has preferences that depend only on mean consumption. More

[1] Random utility models are studied in Thurstone (1927), Luce (1959), Block and Marschak (1960), Falmagne (1978), McFadden and Richter (1990), McFadden (2005), and Kitamura and Stoye (2018).

[2] An early paper to consider this perspective is Machina (1985). For a more recent work in this framework see Swait and Marley (2013), Fudenberg et al. (2015), Freer and Martinelli (2016), Cerreia-Vioglio et al. (2019), and Allen and Rehbeck (2019).

generally, if there are systematic *budgetary* restrictions from a statistic of the distributions, then one can model individual behavior as arising from *preferences* that depend only on the statistic. After stating the main result, we show through examples that this framework neither nests nor is nested by random utility models.

The model of statistical consumer theory we present has intuitive appeal. A natural starting point when considering an individual choosing a distribution of bundles is to consider what constraints are placed on the distributions that can be chosen. Constraints on the distributions by a statistic (e.g. mean) may occur by necessity since an individual cannot overspend their income indefinitely. These constraints may also occur via mental accounting (Thaler, 1980) where an individual allows themselves to purchase bundles beyond their income (e.g. vacation), but does not allow these regularly. Using a statistic may be easier for an individual performing mental accounting since it is a lower dimensional object than the whole distribution.

Statistical consumer theory also provides a new perspective for models of stochastic choice. As mentioned earlier, there are many papers that study stochastic choice. The approach of random utility models (McFadden and Richter, 1990) is often criticized since it is difficult to apply it to welfare analysis because an individual's distribution of preferences is in general not uniquely identified.[3] In contrast, when stochastic choices are generated from a preference over distributions following Machina (1985), a researcher can use the standard welfare framework of consumer theory. This paper provides a bridge that allows a researcher to use the ideas of standard consumer theory to study stochastic choice, since natural statistics such as the mean lead to an analogue of the generalized axiom of revealed preference.

This paper also provides theory that matches the analysis used in experiments on animal decision making following the work of Kagel et al. (1975), Battalio et al. (1981), Kagel et al. (1981), Battalio et al. (1985), and Kagel et al. (1995).[4] Since researchers performing experiments on economic rationality of animals are not able to explain the

[3]This need not be true if one restricts the class of preferences. For example, the single-crossing random utility models in Apesteguia et al. (2017) yield point identification on distributions of preferences.

[4]There are other papers that study rationality of animals. For example, Shafir (1994), Chen et al. (2006), Latty and Beekman (2010), Lea and Ryan (2015), Krasheninnikova et al. (2018), and Natenzon (2019) study rationality properties of choice distributions from bees, capuchin monkeys, amoeboids, frogs, and parrots.

directions to their subjects, these studies are concerned about noise being present from a single choice. For these reasons, rationality of animals is mainly tested using mean choices. While the test of mean rationality is standard in such experiments, it does not follow from any standard revealed preference result. This paper fills this gap.[5]

Following the theory, we provide a simple statistical framework for testing rationality using mean choices. Many experiments examining rationality in animals look at choices from two linear budget sets. These experiments are often designed so that a violation of rationality can only be detected when there is a violation of downward-sloping demand in consumption-compensated budget sets. We present a method to test for violations of mean rationality for *any* two budget sets. In particular, our test can be used for budget sets that are not consumption-compensated. We illustrate how to to perform the test using data on capuchin monkeys from Chen et al. (2006).

This paper also informs work on revealed preference which is discussed in a textbook setting in Chambers and Echenique (2016). In particular, we build off the standard demand setting studied in Afriat (1967), Diewert (1973), and Varian (1982). To the best of our knowledge, the idea of considering how revealed preference restrictions from budget constraints are related to the objective function of the individual is new. However, the results are mathematically similar to those developed in Forges and Minelli (2009). The approach is also related to Richter (1979), Chambers et al. (2019), and Deb et al. (2019), which examine the relation between primal and dual revealed preference relations.

The remainder of the paper is organized as follows. Section 2 defines the model with constraints on mean expenditure and presents the main results. Section 3 contrasts behavior of statistical consumer theory with random utility models. Section 4 presents the results on the statistical methods for common budgets used in tests of mean rationality for animals and provides an illustrative example using data from Chen et al. (2006). Section 5 provides a discussion of the results and some final remarks.

[5] In contrast, experiments on human often use a single choice from each budget, e.g. Harbaugh et al. (2001), Andreoni and Miller (2002), Choi et al. (2007). Nonetheless, a researcher may be concerned that a human does not understand instructions and be interested in examining mean rationality.

2 Model and results

In this section, we present a model of consumer choice where the individual maximizes their utility over a distribution of consumption bundles, but faces a restriction on mean expenditure. We discuss how this extends to general statistics in Appendix A. All proofs not found in the main text are located in Appendix A.

We identify the *consumption space* with the positive orthant of the L-dimensional real space $\mathbb{R}^L_+$. Hence, a consumption bundle is a vector $x \in \mathbb{R}^L_+$, where for $\ell = 1, \ldots, L$ each entry x_ℓ determines the amount of one of the L commodities. We endow $\mathbb{R}^L_+$ with a norm $\| \cdot \|$ and the natural product order $\geq$.[6] Let Δ denote the space of Borel *probability measures* over the consumption space $\mathbb{R}^L_+$. For technical reasons, we restrict our attention to measures that satisfy $\int \|y\| d\mu(y) < \infty$, which is without loss of generality in this framework.[7] Finally, we endow Δ with the weak topology.[8]

We consider a model of consumer choice in which an individual maximizes their utility function with respect to probability distributions over the consumption space. Thus, stochastic choice is generated by an agent who chooses a most-preferred distribution of bundles and randomizes the selection of a particular bundle in $\mathbb{R}^L_+$ according to this distribution. See Sopher and Narramore (2000), Agranov and Ortoleva (2017), and Feldman and Rehbeck (2019) for experimental evidence supporting this view of stochastic choice.

The *primitive dataset* $\mathcal{D} := \left\{ (p^t, m^t, \mu^t) : t \in T \right\}$ consists of a finite number of prices $p^t \in \mathbb{R}^L_{++}$, incomes $m^t \in \mathbb{R}_{++}$, and probability measures $\mu^t \in \Delta$. Here, we interpret the probability measure μ^t as being chosen by a single consumer. We abuse notation and use T both to refer to the number of observations and the set of their labels, where the meaning is clear from the context. Later in the paper, we show that one may assume $m^t = \int (p^t \cdot y) d\mu^t(y)$ without loss of generality for statistical consumer theory.

[6]This is to say that, for any vectors $x = (x_\ell)_{\ell=1}^L$, $y = (y_\ell)_{\ell=1}^L$ in $\mathbb{R}^L$, we have $x \geq y$ if $x_\ell \geq y_\ell$, for all $\ell = 1, \ldots, L$. In addition, the relation is strict and denoted by $x > y$ if $x \geq y$ and $x \neq y$.

[7]Equivalently, the Bochner integral $\int y \, d\mu(y)$ in $\mathbb{R}^L_+$ is well-defined. This follows from Theorem 11.44 in Aliprantis and Border (2006).

[8]A sequence $\{\nu_k\}$ in Δ *weakly converges* to ν if the Lebesgue integral $\int f d\nu_k$ converges to $\int f d\nu$, for any continuous and bounded function $f : \mathbb{R}^L_+ \to \mathbb{R}$. In particular, this space is metrizable.

2.1 Statistical consumer theory

The focus of this paper is on a case of statistical consumer theory where the set of feasible distributions is constrained by mean expenditure, given the observed prices. Specifically, given prices $p \in \mathbb{R}^L_{++}$ and income $m > 0$, a feasible distribution ν must be contained in the *mean expenditure budget* defined by the mapping

$$A(p, m) \;=\; \left\{ \nu \in \Delta : \int (p \cdot y) d\nu(y) \leq m \right\}. \tag{1}$$

The mean expenditure budget allows an individual to purchase bundles that cost more or less than their income. However, on average the expenditure can not exceed the income. Following the interpretation of deterministic stochastic choice, the individual selects a distribution μ to maximize their utility over the budget $A(p, m)$.

With this restriction, we consider *mean expenditure datasets* of the form $\mathcal{D}^A := \left\{(A^t, \mu^t) : t \in T\right\}$ where $A^t := A(p^t, m^t)$ and $\mu^t \in A^t$. Throughout the theoretical analysis, we assume there is no error in measuring the chosen distribution μ^t. However, when applying these results to certain empirical settings, sampling variability may arise since only finitely many realizations of μ^t can be observed. These realizations would be interpreted as "choices" in a standard deterministic or random utility model. In contrast, for our approach, preferences are defined over distributions μ^t and the realizations are drawn from the chosen distributions. We discuss methods to address finite sample issues in Section 4.

We are interested in when the dataset $\mathcal{D}^A$ can be described, or *rationalized*, by utility maximization. Equivalently, we wish to know when there exists a utility function $U : \Delta \to \mathbb{R}$ such that for all $t \in T$, the chosen distribution μ^t satisfies

$$U(\mu^t) \;\geq\; U(\nu), \quad \text{for all } \nu \in A^t. \tag{2}$$

Clearly, with no additional restriction on the utility function, any dataset can be rationalized with a constant function U. For this reason, we restrict our attention to the class of *locally nonsatiated* utility functions U. Local nonsatiation of U requires that, for any probability measure $\nu \in \Delta$ and neighborhood, there is some distribution ν' in the neighborhood that satisfies $U(\nu') > U(\nu)$.

The above framework allows for a natural definition of revealed preference relations. We begin by specifying the *directly revealed preference* relation R defined over the set of observed choices $\{\mu^t\}_{t\in T}$. For any $t, s \in T$, we say that measure μ^t is directly revealed preferred to μ^s, and denote it by $\mu^t \, R \, \mu^s$, when $\int (p^t \cdot y) d\mu^s(y) \leq m^t$; i.e., whenever the measure μ^s was available from the budget A^t. Next, we define the *strictly directly revealed preference* relation P. For any $t, s \in T$, we say that measure μ^t is strictly directly revealed preferred to μ^s, and denote it by $\mu^t \, P \, \mu^s$, when $\int (p^t \cdot y) d\mu^s(y) < m^t$.

It is straightforward to show that both relations are consistent with any locally nonsatiated utility function U that rationalizes the set of observations $\mathcal{D}^A$. Indeed, whenever the consumer selects a measure μ^t at time t, they directly reveal that it is preferable to any other option μ^s that satisfies the mean expenditure constraint $\int (p^t \cdot y) d\mu^s(y) \leq m^t$. Hence, $\mu^t \, R \, \mu^s$ must imply $U(\mu^t) \geq U(\mu^s)$.

One can show that the strict directly revealed preference P is also consistent with a locally nonsatiated utility. First, recall that local nonsatiation of U requires that for any element $\nu \in \Delta$ and neighborhood, there is some ν' in the neighborhood such that $U(\nu') > U(\nu)$. By continuity of function $\nu \to \int (p^t \cdot y) d\nu(y)$, the set $\left\{\nu \in \Delta : \int (p^t \cdot y) d\nu(y) < m^t\right\}$ is open. Therefore, for any element μ^s contained in the set, there must be some distribution ν' such that $U(\nu') > U(\mu^s)$. Since ν' is also available from the A^t budget, the previous claim implies $U(\mu^t) \geq U(\nu') > U(\mu^s)$.

Now we construct the revealed preference relation R^* from the previous direct revealed preference relations. Specifically, for any $t, s \in T$, we say that μ^t is *revealed preferred* to μ^s, denoted by $\mu^t \, R^* \, \mu^s$, if there is a sequence of indices $a, b, c, \ldots, z \in T$ such that

$$\mu^t \, R \, \mu^a, \;\; \mu^a \, R \, \mu^b, \;\; \ldots \;, \;\; \text{and} \;\; \mu^z \, R \, \mu^s. \tag{3}$$

Moreover, we say that μ^t is *strictly revealed preferred* to μ^s, denoted by $\mu^t \, P^* \, \mu^s$, when there is a sequence as in (3) with at least one pair strictly revealed preferred. This immediately implies the testable restriction of *mean acyclicity*.

Definition 1 (Mean acyclicity). *For any cycle $\mathcal{C} = \left\{(a, b), (b, c), \ldots, (z, a)\right\}$ in $T \times T$ such that $\int (p^t \cdot y) d\mu^s(y) \leq m^t$ for $(t, s) \in \mathcal{C}$, we have $\int (p^t \cdot y) d\mu^s(y) = m^t$, for $(t, s) \in \mathcal{C}$.*

The definition of mean acyclicity depends only on the information contained in the primitive dataset, but it is equivalent to restrictions on the revealed preference relations. Indeed, it is satisfied if and only of the revealed strict preference relation P^* is irreflexive.[9]

[9]The relation P^* is irreflexive if *not* $\mu \, P^* \, \mu$, i.e., no element is strictly preferable to itself.

At the same time, the definition coincides with the *generalized axiom of revealed preference* (GARP) on the revealed preference relation R^* so that

$$\mu^t \, R^* \, \mu^s \;\text{ implies }\; \text{not } \mu^s \, P \, \mu^t. \tag{4}$$

We use the definition in equation (4) to construct a statistical test in Section 4.

Mean acyclicity requires further comment. In particular, the condition applies to one element cycles $\mathcal{C} = \{(t,t)\}$. It follows that maximization of a locally nonsatiated utility function requires that $\int (p^t \cdot y) d\mu^t(y) = m^t$, for all $t \in T$. Thus, the budget constraint must be binding for every observed choice. An important practical implication of this fact is that it is not crucial for the econometrician to observe income m^t. Thus, given prices p^t and the distribution μ^t, the average expenditure must satisfy $m^t = \int (p^t \cdot y) d\mu^t(y)$.

Mean acyclicity is a straightforward extension of GARP (as in Afriat, 1967; Diewert, 1973; Varian, 1982) to choices over probability measures, rather than consumption bundles. In fact, if the consumer chooses only degenerate lotteries, GARP coincides with mean acyclicity. To see this, for all $t \in T$ a degenerate lottery satisfies $\mu^t = \delta_{x^t}$, where the latter denotes the Dirac measure concentrated at some $x^t \in \mathbb{R}^L_+$. Therefore, we have $\int (p^t \cdot y) d\mu^s(y) = p^t \cdot x^s$ for all $t, s \in T$, which reduces mean acyclicity to GARP.

Finally, by Lemma 11.45 in Aliprantis and Border (2006), we have

$$\int (p \cdot y) d\nu(y) \;=\; p \cdot \int y \, d\nu(y)$$

for all $\nu \in \Delta$ and $t \in T$. This implies that all relevant information for mean acyclicity is summarized by the mean bundle $\int y \, d\nu(y)$ for the distribution ν. We leverage this observation for the test in Section 4, where we exploit the fact that, when faced with sampling error, estimating mean consumption rather than the whole distribution μ^t is sufficient to test for mean acyclicity.

Since we can represent the budget set as a restriction on mean bundles, it is natural to study utility functions that depend only on the mean bundle. We provide a formal definition of a mean choice rationalization below.

Definition 2. *We say the dataset $\mathcal{D}^A$ is rationalizable with a* mean choice model *if there is a locally nonsatiated function $f : \mathbb{R}^L_+ \to \mathbb{R}$ such that*

$$U(\nu) \;:=\; f\left(\int y \, d\nu(y)\right)$$

rationalizes $\mathcal{D}^A$ as in condition (2).

If the dataset can be rationalized with a mean choice model, then the utility that rationalizes the data only depends on the mean bundle (a vector), rather than the whole set of distributions. In the main theorem below, we argue that locally nonsatiated preferences are equivalent to the mean choice model when the sets of distributions are restricted by the mean expenditure budget defined in equation (1).

Theorem 1. *For any set of observations $\mathcal{D}^A$ with mean expenditure budgets, the following statements are equivalent:*

(i) $\mathcal{D}^A$ satisfies mean acyclicity.

(ii) $\mathcal{D}^A$ is rationalizable with a locally nonsatiated utility function $U : \Delta \to \mathbb{R}$.

(iii) $\mathcal{D}^A$ is rationalizable with a mean choice model*.*

(iv) $\mathcal{D}^A$ is rationalizable with a mean choice model *with a continuous, strictly increasing, and concave function $f : \mathbb{R}^L_+ \to \mathbb{R}$ such that $U(\nu) := f\big(\int y \, d\nu(y)\big)$.*

This result has several implications. First, the revealed preference analysis of statistical consumer theory is simple and parallels classical revealed preference analysis. The main theorem shows this connection when consumer choices are restricted by mean expenditures. However a more general statistical consumer theory is developed in Appendix A.2. Second, in the above framework, the mean choice model is observationally equivalent to locally nonsatiated preferences. Therefore, this property of consumer preference has no additional testable implications. In fact, given statement (iv), there is no loss of generality in restricting attention to a mean choice model with a well-behaved function f. We emphasize that these equivalences depend on the assumptions imposed on the observable budget sets and are not true in general, as shown in Section 2.2.

2.2 Alternative budgets

In the main analysis we focus attention on the mean expenditure budget. However, in some settings it is natural to consider an alternative specification. Given the primitive dataset $\mathcal{D} := \big\{(p^t, m^t, \mu^t) : t \in T\big\}$, we consider *support budgets* defined by the mapping

$$B(p, m) \ := \ \Big\{\nu \in \Delta : \nu\big(\{y \in \mathbb{R}^L_+ : p \cdot y \leq m\}\big) = 1\Big\}, \tag{5}$$

for prices $p \in \mathbb{R}^L_{++}$ and mean income $m > 0$. We call this a support budget since ν is restricted to satisfy affordability for (almost) every realization. When using these support budgets, we denote the corresponding dataset by $\mathcal{D}^B = \{(B^t, \mu^t) : t \in T\}$ where $B^t := B(p^t, m^t)$. Thus, we require for all $t \in T$ that $\mu^t \in B^t$.

This budget is more restrictive than the mean expenditure constraint considered previously. Recall that mean expenditure budget allows violations of $p \cdot y \le m$ with positive probability, as long as the mean expenditure is bounded by m. We study the support budget because it better matches the application of experimental studies on animal rationality. Experiments on animal rationality have each subject (e.g., capuchin monkey) make choices from the same budget set several times. Repeated choices from a budget set in an experiment then give a distribution of choices that can be used to study whether data are consistent with the mean choice model, as in Definition 2.

Analogously to (2), a dataset $\mathcal{D}^B = \{(B^t, \mu^t) : t \in T\}$ is rationalizable by a utility function $U : \Delta \to \mathbb{R}$ if $U(\mu^t) \ge U(\nu)$, for all $\nu \in B^t$ and $t \in T$. Moreover, the set is rationalizable with the mean choice model if U is specified as in Definition 2.

The revealed preference analysis of the mean choice model is similar with support budgets. Consider the directly revealed preference relation R defined earlier. Notice that even though the set of distributions available to the consumer differs, such a preference relation is rationalizable with a mean choice model. For example, suppose that $\mu^t \, R \, \mu^s$, or equivalently $p^t \cdot \int y \, d\mu^s(y) \le m^t$, for some $t, s \in T$. The latter condition implies that the mean bundle $\bar{y}_{\mu^s} := \int y \, d\mu^s(y)$ corresponding to the measure μ^s was affordable in observation t. Therefore, it was possible for the individual to select a measure in B^t with the same mean as μ^s, e.g., the Dirac measure $\delta_{\bar{y}_{\mu^s}}$ concentrated at the vector $\bar{y}_{\mu^s}$. Hence, for any mean choice model with the utility function $U(\nu) := f\big(\int y \, d\nu(y)\big)$, it must be that $\mu^t \, R \, \mu^s$ implies $U(\mu^t) \ge U(\delta_{\bar{y}_{\mu^s}}) = U(\mu^s)$. Note that this is true even when the support of measure μ^s is not contained in $\{y \in \mathbb{R}^L_+ : p^t \cdot y \le m^t\}$.

Whenever the corresponding function f is locally nonsatiated, it must be that $\mu^t \, P \, \mu^s$ implies $U(\mu^t) > U(\mu^s)$. Clearly, this suffices for the revealed preference relations R^* and P^* to be consistent with preference induced by any model of mean choice that rationalizes the data $\mathcal{D}^B$ with support budgets. We now arrive at the following proposition.

Proposition 1. *For any set of observations $\mathcal{D}^B$ with support budgets, the following state-*

ments are equivalent:

(i) $\mathcal{D}^B$ *satisfies mean acyclicity.*

(ii) $\mathcal{D}^B$ *is rationalizable with a* mean choice model*.*

(iii) $\mathcal{D}^B$ *is rationalizable with a* mean choice model *with a continuous, strictly increasing, and concave function* $f : \mathbb{R}^L_+ \to \mathbb{R}$ *such that* $U(\nu) := f\big(\int y\, d\nu(y)\big)$*.*

That a mean choice model implies mean acyclicity follows from the argument above. The other implications follow immediately from Theorem 1. Recall that Theorem 1 is stated over mean expenditure budgets rather than support budgets. An important distinction is that with support budgets, the mean choice model is no longer observationally equivalent to the general model with locally nonsatiated utility. Below is an example of a dataset that is inconsistent with the mean choice model but is rationalizable by a locally nonsatiated utility function. We discuss the testable restrictions for other assumptions on preferences with support budgets in Appendix B.

Example 1. Let $\mathcal{D} = \big\{(p^t, m^t, \mu^t) : t = 1, 2\big\}$, with prices $p^1 = (2,1)$, $p^2 = (1,2)$ and incomes $m^1 = m^2 = 1$. Moreover, suppose that measure μ^1 assigns probability $1/4$ to vector $(0,1)$ and $3/4$ to $(1/2, 0)$, while distribution μ^2 assigns probability $1/4$ to point $(1,0)$ and $3/4$ to $(0, 1/2)$. See Figure 1. Notice that the mean vectors corresponding to μ^1, μ^2 are $\bar{y}_{\mu^1} = (3/8, 1/4)$ and $\bar{y}_{\mu^2} = (1/4, 3/8)$, respectively. In particular, since $p^1 \cdot \bar{y}_{\mu^2} = p^2 \cdot \bar{y}_{\mu^1} = 7/8 < 1$, the set $\mathcal{D}$ violates mean acyclicity.

We show that despite this, the dataset can be rationalized by a locally nonsatiated utility $U : \Delta \to \mathbb{R}$ with support budgets. For example, consider the function $U : \Delta \to \mathbb{R}$ given by

$$U(\nu) \;:=\; \min\Big\{\sup\big\{p^t \cdot y : y \in \operatorname{supp}(\nu)\big\} : t = 1, 2\Big\},$$

where $\operatorname{supp}(\nu)$ denotes the support of measure ν. Indeed, it is locally nonsatiated in the weak topology, while $U(\mu^1) = U(\mu^2) = 1$. At the same time, for any measure ν with the support contained in $\big\{y \in \mathbb{R}^2_+ : p^t \cdot y \le 1\big\}$, it must be that $U(\nu) \le 1$ for $t = 1, 2$. Clearly, this suffices for U to rationalize the corresponding set $\mathcal{D}^B$ with support budgets.

To see that the above utility U does not rationalize the data with mean expenditure budgets defined in (1), consider the distribution ν' that assigns probability $1/10$ to $(0,2)$, $1/10$ to $(2,0)$, and $8/10$ to $(1/8, 1/8)$. It follows that $\int (p^t \cdot y) d\nu'(y) = 9/10 < 1$, for

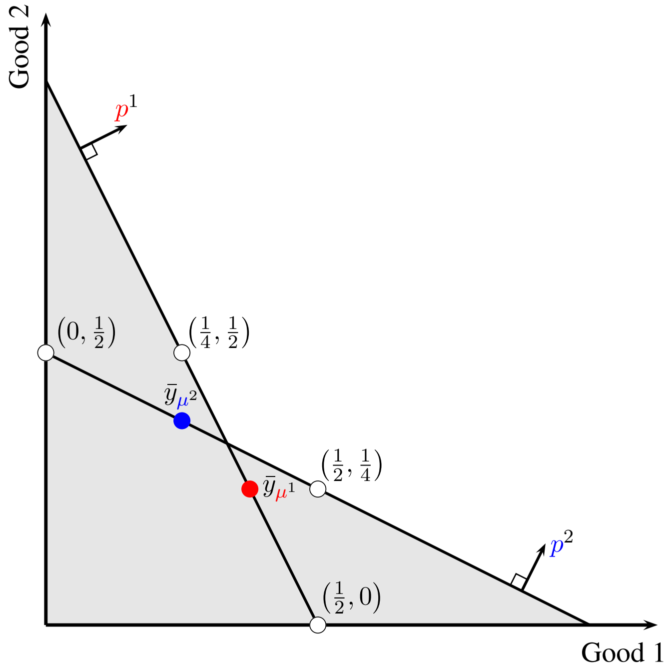

Figure 1: Graphical interpretation of the dataset in Example 1

$t = 1, 2$, so ν' is available for each mean budget. However, $U(\nu') = 4$ which dominates μ^1 and μ^2 and is feasible for the respective budgets.

3 Comparison to random utility models

In this section we compare the mean choice model to random utility models. For a random utility model, randomness arises because an individual's preference is random. Thus, an individual maximizes utility given a random draw of preferences on a standard budget constraint. We follow the definitions of random utility models in McFadden and Richter (1990), McFadden (2005), and Kitamura and Stoye (2018).

Recall that the budget in the standard deterministic consumer problem is given by

$$C(p, m) \;:=\; \Big\{ y \in \mathbb{R}^L_+ : p \cdot y \leq m \Big\}, \tag{6}$$

for prices $p \in \mathbb{R}^L_{++}$ and income $m > 0$. For brevity, we let $C^t := C(p^t, m^t)$. The support budget set described in equation (5) can also be written as

$$B(p, m) \;:=\; \Big\{ \nu \in \Delta : \nu\big(C(p, m)\big) = 1 \Big\}.$$

Distributions of choices generated from random utility models will necessarily be included in the support budget dataset $\mathcal{D}^B = \{(B^t, \mu^t) : t \in T\}$ since the individual maximizes random preferences on each C^t.

We now formally describe random utility models. Let $\mathscr{U}$ be the space of strictly quasiconcave locally nonsatiated utility functions $u : \mathbb{R}^L_+ \to \mathbb{R}$. A set of observations $\mathcal{D}$ is rationalized by a *random utility model* (RUM) if there is a probability measure ρ over the space of functions $\mathscr{U}$ such that, for all $t \in T$:

$$\mu^t(O) \;=\; \rho\Big(\big\{\tilde{u} \in \mathscr{U} : \text{argmax}_{y \in C^t} \tilde{u}(y) \in O\big\}\Big), \tag{7}$$

for any measurable subset $O \subseteq \mathbb{R}^L_+$, where the argmax set is a singleton since $\mathscr{U}$ consists of strictly quasiconcave functions. In other words, the probability of choosing a bundle in the set O is equal to the probability of drawing a utility function that is maximized over C^t at some point in the set O. For a linear programming characterization of RUM see McFadden and Richter (1990), McFadden (2005), and Kitamura and Stoye (2018).[10]

We show through examples that mean choice models neither nest nor are nested by random utility models. In Example 2 we discuss a dataset that can be rationalized only by a mean choice model. Here, there is no distribution over preferences that can generate the observations. Despite this, the mean behavior is consistent with mean acyclicity. In contrast, the dataset in Example 3 is only rationalizable by a RUM. Here, it is critical that the distribution of goods are optimal for a given draw of utility but there are no restrictions on aggregate behavior of the distribution. Thus, the model of mean choice may be more appropriate for an individual whose behavior does not appear systematic on day-to-day purchases, but who has consistent choices when studying, say, weekly purchases from the same budget.

Example 2. Let a primitive dataset be given by $\mathcal{D} = \{(p^1, m^1, \mu^1), (p^2, m^2, \mu^2)\}$, where $p^1 = (2, 1)$, $p^2 = (1, 2)$, and $m^1 = m^2 = 1$. In addition, suppose that measure μ^1 assigns probability $7/12$ to bundle $(1/2, 0)$ and $5/12$ to $(0, 1)$, while μ^2 assigns weight $7/12$ to $(0, 1/2)$ and $5/12$ to $(1, 0)$. Clearly, both measures μ^1, μ^2 belong to the corresponding support budgets B^1, B^2, respectively.

[10] There is also a condition called the *axiom of revealed stochastic preferences* given in McFadden and Richter (1990) and McFadden (2005) that is more similar to GARP.

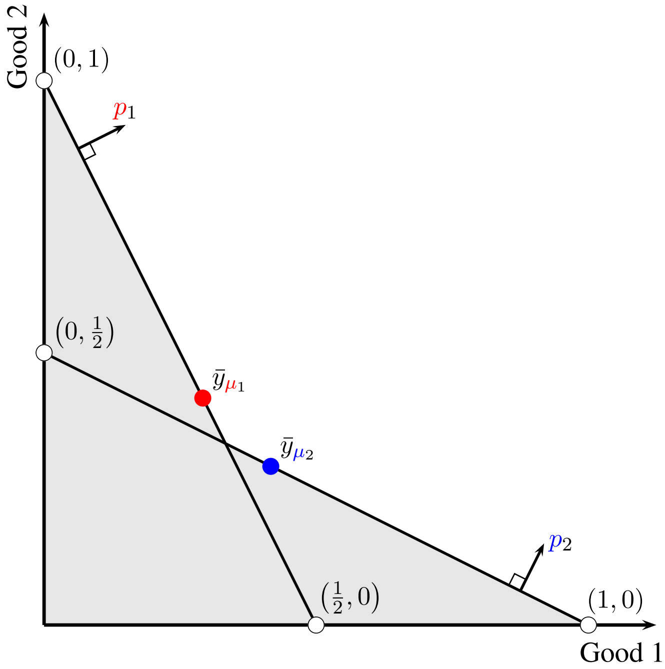

Figure 2: Graphical interpretation of the dataset in Example 2.

Since both $\int(p^1 \cdot y)d\mu^2(y)$ and $\int(p^2 \cdot y)d\mu^1(y)$ are equal to $13/12 > 1 = m^1 = m^2$, this suffices for the set of observations to satisfy mean acyclicity and, thus, be rationalizable by a mean choice model (by Proposition 1). Equivalently, the means of distributions μ^1, μ^2 are given by $\bar{y}_{\mu^1} = (7/24, 5/12)$ and $\bar{y}_{\mu^2} = (5/12, 7/24)$, respectively, where $p_1 \cdot \bar{y}_{\mu^2} = p_2 \cdot \bar{y}_{\mu^1} = 13/12 > 1$. See Figure 2 for a graphical interpretation.

In contrast, the data are inconsistent with the random utility model. Indeed, since $p^1 \cdot (0, 1/2) = 1/2 < m^1$, for a mass of at least $7/12$ of utilities, bundle $(0, 1/2)$ would be strictly inferior to $(1/2, 0)$. Analogously, as $p^2 \cdot (1/2, 0) = 1/2 < m_2$, at least $7/12$ of utilities would have to rank $(0, 1/2)$ strictly over $(1/2, 0)$. However, this would imply that for a mass of at least $1/6$ of all utilities we would have both $u(1/2, 0) > u(0, 1/2)$ and $u(1/2, 0) < u(0, 1/2)$, which yields a contradiction.

Example 3. Consider the primitive dataset $\mathcal{D} = \{(p^1, m^1, \mu^1), (p^2, m^2, \mu^2)\}$ from Example 1, where $p^1 = (2, 1)$, $p^2 = (1, 2)$, and $m^1 = m^2 = 1$; moreover, the measure μ^1 assigns probability $1/2$ to bundles $(1/2, 0)$ and $(1/4, 1/2)$, while μ^2 assigns probability $1/2$ to $(0, 1/2)$ and $(1/2, 1/4)$. Clearly, both measures μ^1, μ^2 belong to the corresponding support budgets B^1, B^2, respectively. Recall Figure 1.

As shown in Example 1, the dataset violates mean acyclicity. At the same time, it is straightforward to show that the set of observations can be rationalized with a random utility model. Clearly, one can always find a function $u_1 : \mathbb{R}^2_+ \to \mathbb{R}$ in $\mathscr{U}$ that is uniquely maximized at $(1/2, 0)$ over $C^1 := \{y \in \mathbb{R}^2_+ : p^1 \cdot y \leq 1\}$ and uniquely maximized at $(1/2, 1/4)$ over $C^2 := \{y \in \mathbb{R}^2_+ : p^2 \cdot y \leq 1\}$. Analogously, there is a function $u_2 : \mathbb{R}^2_+ \to \mathbb{R}$ in $\mathscr{U}$ that is uniquely maximized at $(0, 1/2)$ over C^2 and uniquely maximized at $(1/4, 1/2)$ over C^1. Therefore, a random utility model ρ that assigns probability $1/2$ to each of these utility functions rationalizes these data.

4 Testing mean acyclicity

The previous analysis describes the empirical content of mean choice models when the distributions are known. However in many applications, the distribution μ^t is not known perfectly. Instead, an analyst may see particular selected bundles, each interpreted as a realized draw of consumption from the distribution μ^t. This section describes a method of statistical inference that accommodates sampling variability arising from the fact that the distribution given by μ^t is not known exactly.

4.1 The statistical test

The method of inference we present is designed for application to common datasets used to test rationality of animals. This empirical setting has several features that simplify testing relative to a more general setup. First, prices and income are known exactly, so they do not need to be estimated. Second, there are several realized choices from each budget, which allows us to use the central limit theorem to justify a normal approximation of the means of the sampled distributions. Third, there are only two budgets ($T = 2$).[11] This means there is a single cycle to consider when checking mean acyclicity. After describing the test, we apply the methods to the data from Chen et al. (2006).

We now describe the test in detail. We assume an analyst observes a collection of n^t realizations $\{Y^{it}\}_{i=1}^{n^t}$ corresponding to each observation t. These are treated as random variables and we denote them in upper case. Each realization Y^{it} is a vector of quantities,

[11] The methods of Hsieh et al. (2018) can be used for a statistical test with more than two budgets.

i.e., $Y^{it} \in \mathbb{R}^L_+$. These realizations are draws from the distribution μ^t, which we interpret as the actual choice. We interpret the population mean quantities $\mathbb{E}[Y^{it}]$ as the mean of the distribution μ^t described previously, so that $\mathbb{E}[Y^{it}] = \int y \, d\mu^t(y)$, for all $t \in T$.

When $T = 2$, as in our application, we write the null hypothesis of mean acyclicity as

$$H_0: \; p^s \cdot \mathbb{E}[Y^{it}] \le m^s \;\text{ implies }\; p^t \cdot \mathbb{E}[Y^{is}] \ge m^t, \;\text{ for all } t \neq s \text{ in } \{1,2\}.$$

This is reformulation of condition (4), that uses the fact that $\int (p \cdot y) d\nu(y) = p \cdot \int y \, d\nu(y)$. The alternative hypothesis is

$$H_a: \; p^s \cdot \mathbb{E}[Y^{it}] \le m^s, \text{ for all } t \neq s \text{ in } \{1,2\}, \text{ with at least one inequality strict.}$$

Recall that in the application, prices and income are measured without error, so we use lower case letters to indicate that they are nonrandom.

We reject when sample averages are sufficiently far from the null. Formally, let

$$\overline{Y}^t \;:=\; \frac{1}{n^t} \sum_{i=1}^{n^t} Y^{it}.$$

First note that if $\overline{Y}^t$ were nonrandom and equal to $\mathbb{E}[Y^{it}]$, we could reject when

$$p^1 \cdot \overline{Y}^2 \le m^1 \;\text{ and }\; p^2 \cdot \overline{Y}^1 \le m^2,$$

with at least one inequality strict. If we allow for sampling variability, the empirical means may be different from the theoretical ones. Thus, we propose a test of the form

$$\phi \;:=\; 1\left\{p^1 \cdot \overline{Y}^2 \le m^1 - \hat{\sigma}^2 c_{\sqrt{1-\alpha}};\; p^2 \cdot \overline{Y}^1 \le m^2 - \hat{\sigma}^1 c_{\sqrt{1-\alpha}}\right\},$$

where $1\{\cdot\}$ is the indicator function, $\phi = 1$ denotes rejection, and $\phi = 0$ denotes failure to reject. We denote the nominal size of the test by α (e.g. 0.05), and let $c_{\sqrt{1-\alpha}}$ be the $\sqrt{1-\alpha}$-quantile of the standard normal distribution. In addition,

$$\hat{\sigma}^1 \;:=\; \sqrt{\frac{1}{n^1} \sum_{i=1}^{n^1} \left(p^2 \cdot Y^{i1} - p^2 \cdot \overline{Y}^1\right)^2}$$

$$\hat{\sigma}^2 \;:=\; \sqrt{\frac{1}{n^2} \sum_{i=1}^{n^2} \left(p^1 \cdot Y^{i2} - p^1 \cdot \overline{Y}^2\right)^2},$$

are sample analogue estimators of the standard deviation of $p^2 \cdot \overline{Y}^1$ and $p^1 \cdot \overline{Y}^2$, respectively. The key assumption that justifies this test is that for $t = 1, 2$ there is a large number, n^t, of independent draws Y^{it} from each distribution μ^t.

To provide some intuition behind this test, note that each inequality in the argument of ϕ is motivated by a standard one-sided testing problem. Under the independence assumption and some mild conditions, we have the large sample approximation

$$\text{Prob}\left(p^1 \cdot \mathbb{E}[Y^{i2}] \leq p^1 \cdot \overline{Y}^2 + \hat{\sigma}^2 c_{\sqrt{1-\alpha}}\right) \;\approx\; \sqrt{1-\alpha}.$$

Thus, on the event

$$p_1 \cdot \overline{Y}^2 \;\leq\; m^1 - \hat{\sigma}^2 c_{\sqrt{1-\alpha}},$$

or equivalently

$$p_1 \cdot \overline{Y}^2 + \hat{\sigma}^2 c_{\sqrt{1-\alpha}} \;\leq\; m^1,$$

we have strong evidence against

$$p^1 \cdot \mathbb{E}[Y^{i2}] \;\geq\; m^1.$$

A similar argument holds when testing the other inequality. Thus, when the test function ϕ is 1, we have strong evidence against *both* inequalities

$$p^1 \cdot \mathbb{E}[Y^{i2}] \geq m^1 \;\text{ and }\; p^2 \cdot \mathbb{E}[Y^{i1}] \geq m^2,$$

i.e., strong evidence against H_0. The choice of $c_{\sqrt{1-\alpha}}$ comes from the fact that we have to reject two inequalities. In more detail, this threshold is motivated by

$$\begin{aligned}
\text{Prob}&\left(p^1 \cdot \mathbb{E}[Y^{i2}] \leq p^1 \cdot \overline{Y}^2 + \hat{\sigma}^2 c_{\sqrt{1-\alpha}}\,;\; p^2 \cdot \mathbb{E}[Y^{i1}] \leq p^2 \cdot \overline{Y}^1 + \hat{\sigma}^1 c_{\sqrt{1-\alpha}}\right) \\
&= \text{Prob}\left(p^1 \cdot \mathbb{E}[Y^{i2}] \leq p^1 \cdot \overline{Y}^2 + \hat{\sigma}^2 c_{\sqrt{1-\alpha}}\right) \text{Prob}\left(p^2 \cdot \mathbb{E}[Y^{i1}] \leq p^2 \cdot \overline{Y}^1 + \hat{\sigma}^1 c_{\sqrt{1-\alpha}}\right) \\
&\approx \sqrt{1-\alpha}\sqrt{1-\alpha} \;=\; 1-\alpha,
\end{aligned}$$

where the first equality follows from the independence assumption.

We note that the test we propose is not consistent against alternatives in which one of the inequalities is binding. That is, a configuration such as

$$p^1 \cdot \mathbb{E}[Y^{i2}] = m^1 \;\text{ and }\; p^2 \cdot \mathbb{E}[Y^{i1}] < m^2$$

violates the null, but is not reject with probability 1 as the sample size increases.[12]

There are a few other features of this test to point out. First, note that if a deterministic test (e.g. using $\overline{Y}^t$ in place of $\mathbb{E}[Y^{it}]$) satisfies mean acyclicity, then the statistical test fails to reject for any $\alpha < 0.75$. This is because if $\alpha < 0.75$ then $c_{\sqrt{1-\alpha}} > c_{0.5} = 0$.

One may not find the failure to reject an interesting way to differentiate between datasets. Instead, one can compute the least nominal size α^* under which the null is rejected. To do this, one can compute the largest $c^*_{\sqrt{1-\alpha}}$ under which the test rejects the data and then find the corresponding α^*. The construction of α^* is related to a p-value but is distinct because there are configurations consistent with the null hypothesis in which rejection probabilities are not asymptotically equal to nominal size α.

High α^* gives less evidence against the null. To see why this is the case, note that if α^* is high ($\alpha^* > 0.75$), then $c_{\sqrt{1-\alpha^*}}$ is negative. Thus, this requires the data to not only satisfy mean acyclicity when failing to reject, but that the distance of the average bundles must also be far away from rejecting. The closer the α^* is to one, the lower the evidence is against the null since this suggests the mean bundles are statistically far apart.

4.2 Application to Chen et al. (2006)

We apply the above procedure to the data on choices by capuchin monkeys that was previously analyzed in Chen et al. (2006). We examine whether the capuchin monkeys are rational according to standard rationality, random utility models, and mean choice models. Thus, we check whether individuals satisfy the generalized axiom of revealed preference, the axiom of revealed stochastic preference, and mean acyclicity.

The experiment was performed on three different capuchin monkeys. We refer to the three subjects under their abbreviations: AG, FL, and NN. The consumption bundles consisted of two goods: slices of apples and gelatin cubes or grapes (depending on the subject). In the experiment, capuchin monkeys traded tokens for food items under two different exchange rate regimes. In the first regime, one token could be exchanged for

[12] This suggests a potential power issue with the design of Chen et al. (2006), which is constructed to detect violations of the compensated law of demand by pivoting the budget constraint. It is not a deficiency of the test we present. Intuitively, the alternative hypothesis is on the boundary of the null hypothesis, and so it is not possible to statistically distinguish it from nearby points in the null. One may report the minimal nominal size (described below) as a way to describe evidence against the null hypothesis.

	AG	FL	NN
	Subject		
	Budget 1: $p^1 = (1, 1)$		
m^1	12	12	12
n^1	12	11	6
$\bar{Y}^1$	(6.08, 5.92)	(5.64, 6.36)	(5, 7)
	Budget 2: $p^2 = (1/2, 1)$		
m^2	9	10	10
n^2	22	14	10
$\bar{Y}^2$	(9, 4.5)	(13.86, 3.07)	(12.8, 3.6)
GARP	No	No	Yes
RUM	Yes	Yes	Yes
Mean acyclicity	Yes	Yes	Yes
α^*	0.993	0.998	1.00

Table 1: Results of tests

one slice of apple or one gelatin cube/grape. In the second regime, two slices of apple could be exchanged for one gelatin cube/grape. Per each subject, the data generated in the experiment consisted of multiple choices from the two budget sets. For additional information on the experiment, we refer the reader to Chen et al. (2006).

The details and results of the test are contained in Table 1. We note that two of the three subjects from Chen et al. (2006) refute utility maximization, while all three subjects could be described by a random utility model or a mean choice model. We note that all monkeys satisfy mean acyclicity so the statistical test proposed earlier does not reject for any $\alpha < 0.75$. To better discern between the different datasets, we also report the least nominal size α^* that results in a rejection. We note that all α^* computed are greater than 0.99. Thus, there is weak evidence against the hypothesis that subjects satisfy mean acyclicity.

5 Conclusion

This paper shows that models of deterministic stochastic choice can place strong restrictions on data and be taken to applications. In particular, if one constrains the distribution of bundles chosen to have mean expenditure less than income, then maximizing a general utility function over distributions is observationally equivalent to a model of preferences defined over mean consumption only. Thus, intuitive budget constraints greatly reduce the class of deterministic stochastic choice models. In addition, we show how assumptions on the budget constraint crucially affect the interpretation of the revealed preference analysis, and that this approach neither nests or is nested in the random utility model. Finally, we present a simple method of statistical inference for two budgets, which matches the data from Chen et al. (2006) used in the illustrative example.

Overall, this paper highlights that there are often several intuitive mappings of primitive datasets to budget sets that can greatly affect the analysis. For example, even though the datasets with mean expenditure budgets $\mathcal{D}^A$ and support budgets $\mathcal{D}^B$ are derived from the same primitive data, mean acyclicity is not equivalent to a general non-satiated utility function of distributions when studying support budgets. An interesting potential application would be to examine distributions of choices for different periods of time (e.g. weekly, bi-weekly, monthly) using individual panel data such as that found in Cherchye et al. (2017) and compare random utility models to mean choice models to further examine the difference between these models of stochastic choice.

Appendix A Main proofs

Here we present proofs that were not included in the main body of the paper. Unless stated otherwise, we follow the notation introduced in the paper.

A.1 Proof of Theorem 1

It is clear that $(iv) \Rightarrow (iii) \Rightarrow (ii)$. From the main text we have shown $(ii) \Rightarrow (i)$. It remains to show $(i) \Rightarrow (iv)$. Throughout this proof, define the function $g^t : \Delta \to \mathbb{R}$ by

$$g^t(\nu) \ := \ \left[\int (p^t \cdot y) d\nu(y) - m^t \right],$$

for all $t \in T$. Clearly, since $\int \|y\| d\nu(y) < \infty$, for all $\nu \in \Delta$, by 11.45 in Aliprantis and Border (2006), any such function g^t is well-defined. Moreover, it is straightforward to show that it is also continuous (with respect to the weak topology), strictly increasing with respect to the first order stochastic dominance, and concave (in fact, linear).

Mean acyclicity holds if and only if, for any cycle $\mathcal{C} = \{(a,b),(b,c),\ldots,(z,a)\}$ in $T \times T$ such that $g^t(\mu^s) \leq 0$, for all $(t,s) \in \mathcal{C}$, we have $g^t(\mu^s) = 0$, for all $(t,s) \in \mathcal{C}$. By Lemma 2 in Forges and Minelli (2009), there exist numbers $\{\phi^t\}_{t\in T}$ and strictly positive numbers $\{\lambda^t\}_{t\in T}$ such that $\phi^s \leq \phi^t + \lambda^t g^t(\mu^s)$, for all $t,s \in T$.

Take any such numbers $\{\phi^t\}_{t\in T}$, $\{\lambda^t\}_{t\in T}$ and define the function $f : \mathbb{R}^L_+ \to \mathbb{R}$ by

$$f(y) \;:=\; \min\Big\{\phi^s + \lambda^s\big[p^s \cdot y - m^s\big] : s \in T\Big\},$$

which is continuous, strictly increasing, and concave. We claim that function $U : \Delta \to \mathbb{R}$, given by $U(\nu) := f\big(\int y \, d\nu(y)\big)$, rationalizes the set of observations. Indeed, take any $t \in T$ and $\nu \in \Delta$ such that $\int(p^t \cdot y)d\nu(y) \leq m^t$, or equivalently $g^t(\nu) \leq 0$. Then,

$$\begin{aligned}
U(\nu) &= f\left(\int y \, d\nu(y)\right) \\
&= \min\left\{\phi^s + \lambda^s \left[p^s \cdot \int y \, d\nu(y) - m^s\right] : s \in T\right\} \\
&\leq \phi^t + \lambda^t \left[p^t \cdot \int y \, d\nu(y) - m^t\right] \\
&= \phi^t + \lambda^t \left[\int (p^t \cdot y) d\nu(y) - m^t\right] \\
&= \phi^t + \lambda^t g^t(\nu) \\
&\leq \phi^t \\
&\leq \min\left\{\phi^s + \lambda^s g^s(\mu^t) : s \in T\right\} \\
&= \min\left\{\phi^s + \lambda^s \left[\int (p^s \cdot y) d\mu^t(y) - m^s\right] : s \in T\right\} \\
&= \min\left\{\phi^s + \lambda^s \left[p^s \cdot \int y \, d\mu^t(y) - m^s\right] : s \in T\right\} \\
&= f\left(\int y \, d\mu^t(y)\right) \\
&= U(\mu^t),
\end{aligned}$$

where the first inequality follows from the property of the minimum function, the second inequality is implied by $\lambda^t > 0$ and $g^t(\nu) \leq 0$, while the third follows from the construction of the numbers $\{\phi^t\}_{t\in T}$ and $\{\lambda^t\}_{t\in T}$.

It is straightforward to show that function U is strictly increasing in the sense of first order stochastic dominance (see Appendix C). Thus, following Lemma C.4 presented in Appendix C, the function U is locally nonsatiated.

A.2 General statistical consumer theory

Theorem 1 can be extended to more general models of statistical choice. Here, a statistic is a finite-dimensional summary of a distribution. In particular, let $S : \Delta \to \mathbb{R}^L$ be a continuous function from the distribution of bundles to the L-dimensional real vector space.[13] As we require an additional monotonicity condition on S, we assume that it is strictly increasing with respect to first order stochastic dominance.

We assume that the budget constraint for observation $t \in T$ is represented by a continuous function $g^t : \mathbb{R}^k \to \mathbb{R}$ that is strictly increasing on $\mathbb{R}^k$.[14] Distribution $\nu \in \Delta$ is feasible if $g^t\big(S(\nu)\big) \leq 0$. The dataset is given by $\mathcal{D}^G = \big\{(g^t, \mu^t) : t \in T\big\}$. Any variables describing g^t (such as prices) are absorbed into the function, as it is indexed by t.

Definition 3. *The dataset $\mathcal{D}^G$ with statistic S is rationalizable with a* statistical choice model *if there is a locally nonsatiated function $f : \mathbb{R}^L_+ \to \mathbb{R}$ such that the function*

$$U(\nu) \ := \ f\,(S(\nu))$$

rationalizes $\mathcal{D}^G$, i.e. if $g^t\big(S(\nu)\big) \leq 0$ then $U(\mu^t) \geq U(\nu)$, for all $t \in T$.

Similar to the mean expenditure constraint, we can define a direct revealed preference relation R_S by setting $\mu^t \, R_S \, \mu^s$ if $g^t(S(\mu^s)) \leq 0$. Similarly, we can define a strictly directly revealed preference relation P_S so that $\mu^t \, P_S \, \mu^s$ if $g^t(S(\mu^s)) < 0$. Since the statistic S and the budget constraint functions g^t are assumed continuous, we can repeat the arguments in the main text to show that these relations are consistent with a rationalization of the choices by a locally nonsatiated function.

Similarly, we can take the transitive closure of R_S, which must be consistent with the utility function. For any $t, s \in T$, we say that μ^t is *revealed preferred* to μ^s, denoted by $\mu^t \, R^*_S \, \mu^s$, if there is a sequence of indices $a, b, c, \ldots, z \in T$ such that

$$\mu^t \, R_S \, \mu^a, \ \ \mu^a \, R_S \, \mu^b, \ \ \ldots \ \ , \ \text{and} \ \ \mu^z \, R_S \, \mu^s. \tag{A1}$$

Moreover, we say that μ^t is *strictly revealed preferred* to μ^s, denoted by $\mu^t \, P^*_S \, \mu^s$, when there is a sequence as in condition (A1) with at least one pair is ordered with P^*_S. From this discussion, we consider the following acyclicity condition.

[13] More generally the statistic can map from Δ into a space with dimension that is not equal to L.

[14] Function g^t is strictly increasing when $y'_\ell \geq y_\ell$, for all $\ell = 1, \ldots, K$, implies $g(y') \geq g(y)$, for any $y, y' \in \mathbb{R}^L$, where the latter inequality is strict if $y'_\ell > y_\ell$, for some $\ell = 1, \ldots, L$.

Definition A.1 (Statistical acyclicity)**.** *For any cycle $\mathcal{C} = \{(a,b),(b,c),\ldots,(z,a)\}$ in $T \times T$ such that $g^t\big(S(\mu^s)\big) \leq 0$, for $(t,s) \in \mathcal{C}$, we have $g^t\big(S(\mu^s)\big) = 0$, for all $(t,s) \in \mathcal{C}$.*

The above definition is equivalent to the revealed strict preference relation P_S^* being irreflexive. Equivalently, following the proof of Theorem 1, we see that statistical acyclicity is necessary and sufficient for a statistical choice model.

Theorem A.1. *For any set of observations $\mathcal{D}^G$, the following statements are equivalent:*

(i) $\mathcal{D}^G$ satisfies statistical acyclicity.

(ii) $\mathcal{D}^G$ is rationalizable with a locally nonsatiated utility function $U : \Delta \to \mathbb{R}$.

(iii) $\mathcal{D}^G$ is rationalizable with a statistical choice *model.*

(iv) $\mathcal{D}^G$ is rationalizable with a statistical choice *model with a continuous and strictly increasing function $f : \mathbb{R}_+^L \to \mathbb{R}$ such that $U(\nu) := f\big(S(\nu)\big)$.*

Thus, while we presented the result for mean choice in Theorem 1 of the main body of the paper, analogous result holds for any continuous statistic of the distribution of bundles for restrictions enforced by the functions g^t, for all $t \in T$.

Appendix B Alternative models

In this subsection we discuss the testable implications of three alternative models of deliberately stochastic choice when the budget set is specified as

$$B(p,m) \; := \; \Big\{\nu \in \Delta : \nu\big(\{y \in \mathbb{R}_+^L : p \cdot y \leq m\}\big) = 1\Big\}, \tag{A2}$$

for prices $p \in \mathbb{R}_{++}^L$ and income $m \in \mathbb{R}_{++}$. In particular, we look at the implications of local nonsatiation and first order stochastic dominance.

We begin with the most general question concerning when the set of observations $\mathcal{D}^B := \big\{(B^t, \mu^t) : t \in T\big\}$ can be rationalized by a locally nonsatiated utility function $U : \Delta \to \mathbb{R}$. Perhaps surprisingly, this model places no restrictions on choice behavior. In contrast, when the utility U is strictly increasing with respect to first order stochastic dominance (defined below), there are restrictions on behavior for the dataset $\mathcal{D}^B$.

We again use the notation $C^t := \{y \in \mathbb{R}^L_+ : p^t \cdot y \leq m^t\}$ for the standard deterministic budget constraint, with $\partial C^t := \{y \in \mathbb{R}^L_+ : p^t \cdot y = m^t\}$ corresponding to the budget line, for all $t \in T$. Therefore, we have $B^t = \{\nu \in \Delta : \nu(C^t) = 1\}$, for all $t \in T$. Since each set B^t consists of measures with their supports included in the compact set C^t, we may restrict our attention to the space of probability measures with compact supports, without loss of generality. Recall that the *support* of a probability measure ν is the smallest (by set inclusion) closed set K such that $\nu(K) = 1$.

B.1 Locally nonsatiated models

Here we determine the necessary and sufficient conditions under which a set of observations $\mathcal{D}^B = \{(B^t, \mu^t) : t \in T\}$ is rationalizable as in Section 2.2 with a *locally nonsatiated* utility $U : \Delta \to \mathbb{R}$. Before we proceed, notice that each set B^t can be re-defined as

$$B^t = \Big\{\nu \in \Delta : \operatorname{supp}(\nu) \subseteq C^t\Big\},$$

where by $\operatorname{supp}(\nu)$ we denote the support of ν.

Proposition B.1. *Any set of observations $\mathcal{D}^B$ is rationalizable with a locally nonsatiated utility function $U : \Delta \to \mathbb{R}$.*

Unlike the main analysis in Section 2 with the dataset $\mathcal{D}^A$, whenever choices of individuals are limited to probability distributions over budget sets B^t, the locally nonsatiated model has no testable implications. To see this define $U : \Delta \to \mathbb{R}$ by

$$U(\nu) := \begin{cases} 0 & \text{if } \operatorname{supp}(\nu) \subseteq \bigcup_{t \in T} C^t, \\ \sup\big\{\mathbf{1} \cdot y : y \in \operatorname{supp}(\nu)\big\} & \text{otherwise;} \end{cases}$$

where $\mathbf{1}$ denotes the unit vector. Since we focus on measures with compact supports, the function is well-defined. Moreover, it trivially rationalizes the set of observations since $U(\mu^t) = 0$ for each t. To show that it is locally nonsatiated, take any measure ν and a vector $y \notin \bigcup_{t \in T} C^t$ such that $y > z$, for all $z \in \operatorname{supp}(\nu)$. Given the compact support of the measure, such a vector always exists. Consider now a sequence of measures $\{\nu_k\}$, where $\nu_k := (1 - 1/k)\nu + (1/k)\delta_y$, for all k, and where δ_y denotes the Dirac measure concentrated at y. We see that the sequence $\{\nu_k\}$ converges to ν in the weak topology. Since $\operatorname{supp}(\nu_k) = \operatorname{supp}(\nu) \cup \{y\}$, we have $U(\nu_k) = 1 \cdot y > U(\nu)$, for all k.

The above result follows from the fact that weak convergence of probability measures does not imply convergence in their supports. In particular, for any measure ν with a support contained in a budget set C^t, one can find a measure arbitrarily close to ν with its support not contained in C_t. Because of this, choices reveal no information about the strict preference over probability measures and the model has no testable implications.

B.2 Strictly monotone models

We now investigate the conditions under which a support budget dataset $\mathcal{D}^B$ is rationalizable as in (2) with a utility function $U : \Delta \to \mathbb{R}$ that is strictly increasing with respect to the first order stochastic dominance $\succeq$. More formally, for any $\mu, \nu \in \Delta$, if $\mu \succ \nu$ then $U(\mu) > U(\nu)$. See Appendix C for a discussion on first order stochastic dominance.

The above framework allows for construction of a *directly revealed preference* relation over observed choices $\{\mu^t\}_{t\in T}$, similar to the one in Section 2. We say that the measure μ^t is *directly revealed preferred* to μ^s, and denote it by $\mu^t\ R_F\ \mu^s$, whenever $\mu^s(C^t) = 1$. Equivalently, $\text{supp}(\mu^s) \subseteq C^t$. Moreover, we define the *strictly directly revealed preferred* relation as $\mu^t\ P_F\ \mu^s$ if $\mu^s(C^t) = 1$ and $\mu^s(\partial C^t) < 1$. Therefore, we have $\mu^t\ P_F\ \mu_s$ when $\mu^t\ R_F\ \mu^s$ and the support of μ^s is not contained in the corresponding budget line.

The above relation is weaker than the one defined in Section 2. Indeed, it is true that $\mu^t R_F \mu^s$ $(\mu^t P_F \mu^s)$ implies $\int (p^t \cdot y) d\mu^s(y) \leq (<)\ m^t$. However, since the above relation requires the support of the measure μ^s to be contained in C^t, the converse is not true. Moreover, it is clear that the weak relation R_F is consistent with any utility function that rationalizes the data. In the following lemma, we claim that P_F is also consistent if U is strictly increasing in the sense of first order stochastic dominance.

Lemma B.1. *For any strictly increasing utility $U : \Delta \to \mathbb{R}$ that rationalizes a set of observations $\mathcal{D}^B$, it follows that for any $t \in T$, if $\nu\big(\partial C^t\big) < \nu\big(C^t\big) = 1$ then $U(\mu^t) > U(\nu)$.*

Proof. Take any probability measure $\nu \in \Delta$ with $\nu\big(\partial C^t\big) < \nu\big(C^t\big) = 1$. From Lemma C.1, there is a probability space $(\Omega, \mathcal{F}, \tau)$ and random variable $X : \Omega \to \mathbb{R}^L_+$ such that

$$\nu(O) \;=\; \tau\Big(\big\{\omega \in \Omega : X(\omega) \in O\big\}\Big),$$

for any measurable set $O \subseteq \mathbb{R}^L_+$. Given the assumption imposed on the probability measure ν, there is a set $\Omega' \subseteq \Omega$ that satisfies

$$\tau\Big(\big\{\omega \in \Omega' : p^t \cdot X(\omega) < m^t\big\}\Big) \;>\; 0.$$

Take any random variable $X' : \Omega \to \mathbb{R}^L_+$ such that $p^t \cdot X'(\omega) \leq m^t$ and $X'(\omega) \geq X(\omega)$, for all $\omega \in \Omega$, and $X'(\omega) > X(\omega)$, for some $\omega \in \Omega'$ which clearly exists.[15] Define measure

$$\nu'(O) \;:=\; \tau\Big(\big\{\omega \in \Omega : X'(\omega) \in O\big\}\Big),$$

for all measurable sets $O \subseteq \mathbb{R}^L_+$. By Lemma C.1 in Appendix C, we have $\nu' \succ \nu$. Given that the function U is strictly increasing in the first order stochastic sense, while $\nu'(C^t) = 1$, we obtain $U(\mu^t) \geq U(\nu') > U(\nu)$ which concludes the proof. □

As in the main body of the paper, we denote the transitive closure of R_F by R^*_F. Therefore, we have $\mu^t \, R^*_F \, \mu^s$ if there is a sequence of indices $a, b, c, \ldots, z \in T$ such that

$$\mu^t \, R_F \, \mu^a, \;\; \mu^a \, R_F \, \mu^b, \;\; \ldots \;, \;\; \text{and} \;\; \mu^z \, R_F \, \mu^s.$$

Moreover, the relation is *strict*, or $\mu^t \, P^*_F \, \mu^s$, if at least one comparison is strictly directly revealed preferred. Lemma B.1 leads to the following acyclicity condition.

Definition B.1 (F-acyclicity). *For any cycle $\mathcal{C} = \big\{(a, b), (b, c), \ldots, (y, z)\big\}$ in $T \times T$ such that $\mu^s(C^t) = 1$, for all $(t, s) \in \mathcal{C}$, we have $\mu^s(\partial C^t) = 1$, for all $(t, s) \in \mathcal{C}$.*

Equivalently, the revealed strict preference relation P^*_F is irreflexive. We now prove that F-acyclicity is also sufficient for a dataset $\mathcal{D}^B$ to be rationalizable with a strictly increasing utility function U. We summarize this result in the following proposition.

Proposition B.2. *For any set of observations $\mathcal{D}^B$ with support budgets, the following statements are equivalent:*

(i) $\mathcal{D}^B$ satisfies F-acyclicity.

(ii) $\mathcal{D}^B$ rationalizable with a utility function $U : \Delta \to \mathbb{R}$ that is strictly increasing in the sense of first order stochastic dominance.

(iii) $\mathcal{D}^B$ rationalizable with a utility function $U : \Delta \to \mathbb{R}$ that is strictly increasing in the sense of first order stochastic dominance, continuous, and concave.

[15] For example, take any $\ell = 1, \ldots, L$ and define $X'(\omega) := X(\omega) + \frac{1}{p^t_\ell}\big[m^t - p^t \cdot X(\omega)\big]e_\ell$, where e_ℓ denotes the ℓth basis vector, i.e., an element of $\mathbb{R}^L$ with the ℓ-th entry equal to one and all others equal to zero.

Proof. Clearly $(iii) \Rightarrow (ii) \Rightarrow (i)$ from the arguments in the text. We show $(i) \Rightarrow (iii)$ constructively. First, take any number β such that

$$\beta \;>\; \max\left\{-\frac{\int_{C^t}(p^t\cdot y-m^t)d\mu^s(y)}{\int_{\mathbb{R}^L_+\setminus C^t}(p^t\cdot y-m^t)d\mu^s(y)} : \mu^s\left(\mathbb{R}^L_+\setminus C^t\right)>0 \text{ and } t,s\in T\right\}.$$

For every $t\in T$ define a function $f_t:\mathbb{R}^L_+\to\mathbb{R}$ by

$$f^t(y) \;:=\; \begin{cases} (p^t\cdot y-m^t) & \text{if } \; y\in C^t, \\ \beta(p^t\cdot y-m^t) & \text{otherwise;} \end{cases}$$

which is continuous and strictly increasing.

Define the function $g^t:\Delta\to\mathbb{R}$ by $g^t(\nu):=\int f^t(y)d\nu(y)$, which is continuous, concave, and strictly increasing in the sense of first order stochastic dominance. Moreover for all $s\in T$, we have $g^t(\mu^s)\le 0$ if and only if $\mu^s(C^t)=1$. Indeed, if $\mu^s(C^t)=1$ then

$$\begin{aligned} g^t(\mu^s) \;&=\; \int f^t(y)d\mu^s(y) \;=\; \int_{C^t} f^t(y)d\mu^s(y)+\int_{\mathbb{R}^L_+\setminus C^t} f^t(y)d\mu^s(y) \\ &=\; \int_{C^t} f^t(y)d\mu^s(y) \;=\; \int_{C^t}(p^t\cdot y-m^t)d\mu^s(y) \;\le\; 0. \end{aligned}$$

To show the converse, suppose that $\mu^s\big(\mathbb{R}^L_+\setminus C^t\big)>0$. Then:

$$\begin{aligned} g^t(\mu^s) \;&=\; \int f^t(y)d\mu^s(y) \\ &=\; \int_{C^t} f^t(y)d\mu^s(y)+\int_{\mathbb{R}^L_+\setminus C^t} f^t(y)d\mu^s(y) \\ &=\; \int_{C^t}(p^t\cdot y-m^t)d\mu^s(y)+\beta\int_{\mathbb{R}^L_+\setminus C^t}(p^t\cdot y-m^t)d\mu^s(y) \\ &>\; \int_{C^t}(p^t\cdot y-m^t)d\mu^s(y)-\int_{C^t}(p^t\cdot y-m^t)d\mu^s(y) \\ &=\; 0, \end{aligned}$$

where the strict inequality follows from the choice of β.

The above observation implies that $\mathcal{D}^B$ satisfies F-acyclicity if and only if, for any cycle $\mathcal{C}=\big\{(a,b),(b,c),\ldots,(z,a)\big\}$ in $T\times T$ such that $g^t(\mu^s)\le 0$, for all $(t,s)\in\mathcal{C}$, it must be that $g^t(\mu^s)=0$, for all $(t,s)\in\mathcal{C}$. Following Lemma 2 in Forges and Minelli (2009), there exist numbers $\{\phi^t\}_{t\in T}$, $\{\lambda^t\}_{t\in T}$ such that $\phi^s\le\phi^t+\lambda^t g^t(\mu^s)$, for all $t,s\in T$, where $\lambda^t>0$, for all $t\in T$. Given that $\nu(C^t)=1$ implies $g^t(\nu)\le 0$, for any measure $\nu\in\Delta$, one can show that the function $U:\Delta\to\mathbb{R}$ defined as $U(\nu):=\min\big\{\phi_s+\lambda_s g_s(\nu):s\in T\big\}$

rationalizes the set of observations by arguments that mimic the proof of Theorem 1. Moreover, by the properties of functions g^t, it is also continuous, strictly increasing in the sense of first order stochastic dominance, and concave. $\square$

Clearly, the model of consumer choice discussed in Section 2.2 is a special case of the model that satisfies strict monotonicity in the sense of first order stochastic dominance. Indeed, given a dataset $\mathcal{D}^B$ it is easy to show that mean acyclicity implies F-acyclicity. However, unlike the model introduced in Section 2.2 (which neither nests nor is nested by random utility models), the strictly monotone model is a generalization of random utility models (McFadden and Richter, 1990). However, there are strictly monotone models that are not rationalized by random utility models (recall Example 2).

Proposition B.3. *If a dataset $\mathcal{D}^B$ is rationalized by a random utility model as described in (7), then there is a utility function $U : \Delta \to \mathbb{R}$ that is strictly monotone in the sense of first order stochastic dominance and rationalizes $\mathcal{D}^B$.*

Proof. We prove the result by showing that whenever a dataset $\mathcal{D}^B$ violates F-acyclicity then it can not be rationalized with a random utility model. Suppose there is some cycle $\mathcal{C} = \big\{(a,b),(b,c),\ldots,(z,a)\big\}$ such that $\mu^s(C^t) = 1$, for all $(t,s) \in \mathcal{C}$, and $\mu^s(\partial C^t) < 1$, for some $(t,s) \in \mathcal{C}$. For any probability measure ρ as in (7), this implies that the measure of utilities $\tilde{u}$ such that $\max_{y \in C^t} \tilde{u}(y) \geq \max_{y \in C^s} \tilde{u}(y)$ is equal to one, for all $(t,s) \in \mathcal{C}$. Moreover, there is a set of non-zero measure of utilities for which the inequality is strict, for some $(t,s) \in \mathcal{C}$. However, this implies that there is a non-zero measure set of utilities $\tilde{u}$ that induce a strict cycle, which yields a contradiction. $\square$

Appendix C First order stochastic dominance

In this section we discuss properties of the first order stochastic dominance. Let Δ_X denote a Borel space of probability distributions over some $X \subseteq \mathbb{R}^L$. We consider the usual partial order over $\mathbb{R}^L$, i.e., for $x, y \in X \subseteq \mathbb{R}^L$, $x \geq y$ if and only if $x_i \geq y_i$ for each $\ell = 1, \ldots, L$. Distribution μ first order stochastically dominates ν, or $\mu \succeq \nu$, whenever $\int f d\mu \geq \int f d\nu$, for any measurable, bounded, and nondecreasing function $f : X \to \mathbb{R}$.

One can show that $\succeq$ is a partial order over Δ_X. This follows from Theorem 2 in Kamae and Krengel (1978) and the fact that $\mathbb{R}^L$ is a Polish space.

Lemma C.1. *Suppose that $\mu \succeq \nu$, for some $\mu, \nu \in \Delta_X$. There is a probability space $(\Omega, \mathcal{F}, \tau)$ and random variables $X_\mu, X_\nu : \Omega \to X$ such that*

(i) X_μ and X_ν are distributed according to μ and ν respectively, i.e., for any Borel measurable set $O \subseteq X$ we have

$$\mu(O) = \tau\Big(\big\{\omega \in \Omega : X_\mu(\omega) \in O\big\}\Big) \quad \textit{and} \quad \nu(O) = \tau\Big(\big\{\omega \in \Omega : X_\nu(\omega) \in O\big\}\Big);$$

(ii) $X_\mu(\omega) \geq X_\nu(\omega)$, for all $\omega \in \Omega$.

See Lemma 4 in Kamae and Krengel (1978) for the proof. Using Lemma C.1, we say that $\mu \succ \nu$ when $\mu \succeq \nu$, $\mu \neq \nu$, and there exists a measurable set $F \subseteq \Omega$ such that for all $\omega \in F$ we have $X_\mu(\omega) > X_\nu(\omega)$. We now prove a series of lemmas.

Lemma C.2. *The distribution μ first order stochastically dominates ν, or $\mu \succeq \nu$, if and only if $\mu(D) \geq \nu(D)$, for any measurable and upward comprehensive set D.*[16]

Proof. We prove the implication ($\Rightarrow$) by contradiction. Suppose that $\mu \succeq \nu$, but there is some measurable, upward comprehensive set D such that $\mu(D) < \nu(D)$. Let χ_D be the indicator function, taking values $\chi_D(x) = 0$, for $x \notin D$, and $\chi_D(x) = 1$ otherwise. The function is obviously bounded. Since D is upward comprehensive, the above function is increasing. Since the simple function is defined on a measurable set, it is measurable. However, it must be that $\int \chi_D d\mu = \mu(D) < \nu(D) = \int \chi_D d\nu$, which contradicts that μ first order stochastic dominates ν.

The converse follows directly from the definition of Lebesgue integration. Suppose that, for any upward comprehensive and measurable set D, we have $\mu(D) \geq \nu(D)$. Clearly, D is upward comprehensive if and only if its complement $\mathbb{R}^L \setminus D$ is downward comprehensive. Thus, for any such set E, we have $\mu(E) \leq \nu(E)$.

Take any bounded, measurable, and increasing function $f : \mathbb{R}^L \to \mathbb{R}$. Clearly, for all $r \in \mathbb{R}$ any sets of the form $\big\{y \in \mathbb{R}^L : f(y) > r\big\}$ and $\big\{y \in \mathbb{R}^L : f(y) < r\big\}$ are upward and downward comprehensive, respectively. Moreover, they are both measurable, by

[16] Set $D \subset \mathbb{R}^L$ is *upward (downward) comprehensive* if $y \in D$ and $x \geq (\leq)\ y$ implies $x \in D$.

measurability of f. This implies that

$$\begin{aligned}\int f d\mu &= \int_0^\infty \mu\Big(\{y \in \mathbb{R}^L : f(y) > r\}\Big)dr - \int_0^\infty \mu\Big(\{y \in \mathbb{R}^L : f(y) < r\}\Big)dr \\ &\geq \int_0^\infty \nu\Big(\{y \in \mathbb{R}^L : f(y) > r\}\Big)dr - \int_0^\infty \nu\Big(\{y \in \mathbb{R}^L : f(y) < r\}\Big)dr \\ &= \int f d\nu.\end{aligned}$$

Since this is true for any increasing function f, the proof is complete. □

Before we state the next result, a function $f : X \to \mathbb{R}$ is strictly increasing if $y'_\ell \geq y_\ell$, for all $\ell = 1, \ldots, L$, and $y'_\ell > y_\ell$, for some ℓ, implies $f(y') > f(y)$, for any $y, y' \in X$.

Lemma C.3. *Suppose that $\mu \succ \nu$, for some μ, $\nu \in \Delta_X$. For any strictly increasing function $f : X \to \mathbb{R}$, we have $\int f d\nu > \int f d\mu$.*

Proof. Given that $\nu \succeq \mu$, Lemma C.1 implies that there is a probability space $(\Omega, \mathcal{F}, \tau)$ and random variables X_μ, $X_\nu : \Omega \to X$ that are distributed according to μ, ν respectively, and $X_\mu(\omega) \geq X_\nu(\omega)$, for all $\omega \in \Omega$. Since $\mu \succ \nu$, let $\Omega' \subseteq \Omega$ be defined so that

$$\Omega' = \big\{\omega \in \Omega : X_\mu(\omega) > X_\nu(\omega)\big\}.$$

where $\tau(\Omega') > 0$ (recall Lemma C.1). For any strictly increasing $f : X \to \mathbb{R}$, we have

$$\begin{aligned}\int f d\nu - \int f d\mu &= \int_\Omega \Big[f\big(X_\nu(\omega)\big) - f\big(X_\mu(\omega)\big)\Big] d\tau(\omega) \\ &= \int_{\Omega'} \Big[f\big(X_\nu(\omega)\big) - f\big(X_\mu(\omega)\big)\Big] d\tau(\omega) \;>\; 0.\end{aligned}$$

This competes the proof. □

Lemma C.4. *Suppose that $X + \mathbb{R}^L_+ \subseteq X$. For any measure $\mu \in \Delta_X$ and neighborhood, we have $\nu \succ \mu$, for some ν in the neighborhood.*

Proof. We show that for any $\mu \in \Delta$ there is a sequence $\{\mu^k\}$ in Δ that weakly converges to μ and $\mu^k \succ \mu$, for all k. Take any probability space $(\Omega, \mathcal{F}, \tau)$ and the random variable $X_\mu : \Omega \to X$ that is distributed according to μ, i.e., for any measurable $O \subseteq X$ we have

$$\mu(O) \;=\; \tau\Big(\big\{\omega \in \Omega : X_\mu(\omega) \in O\big\}\Big).$$

Take any sequence $\{X^k\}$ of random variables $X^k : \Omega \to \mathbb{R}$ that pointwise converge to X_μ and satisfy $X^k(\omega) > X_\mu(\omega)$, for all $\omega \in \Omega$.

For each k, define a probability measure μ^k so that for any measurable $O \subseteq X$

$$\mu^k(O) \;:=\; \tau\Big(\big\{\omega \in \Omega : X^k(\omega) \in O\big\}\Big).$$

Since $X + \mathbb{R}^L_+ \subseteq X$, we have $\mu^k \in \Delta_X$. Moreover, for any measurable, upward comprehensive set D, it must be that

$$\mu^k(D) \;=\; \tau\Big(\big\{\omega \in \Omega : X^k(\omega) \in D\big\}\Big) \;\geq\; \tau\Big(\big\{\omega \in \Omega : X_\mu(\omega) \in D\big\}\Big) \;=\; \mu(D),$$

since for all $\omega \in \Omega$ if $X_\mu(\omega) \in A$ then $X_k(\omega) \in D$. Therefore, by Lemma C.2 and since $X_k(\omega) > X_\mu(\omega)$, for all $\omega \in \Omega$, it follows for all k that $\mu^k \succ \mu$. Finally, take any continuous, bounded function $f : X \to \mathbb{R}$ and notice that

$$\lim_{k\to\infty}\left|\int f d\mu^k - \int f d\mu\right| \;=\; \lim_{k\to\infty}\left|\int \Big[f\big(X^k(\omega)\big) - f\big(X_\mu(\omega)\big)\Big] d\tau(\omega)\right| \;=\; 0,$$

since $X^k(\omega) \to X_\mu(\omega)$, for all $\omega \in \Omega$. Thus $\{\mu^k\}$ weakly converges to μ. Clearly, this implies that, for any neighborhood around μ, there is some μ^k in the neighborhood such that $\mu^k \succ \mu$. □

References

Afriat, S. N. (1967). The construction of utility functions from expenditure data. *International economic review 8*(1), 67–77.

Agranov, M. and P. Ortoleva (2017). Stochastic choice and preferences for randomization. *Journal of Political Economy 125*(1), 40–68.

Aliprantis, C. D. and K. C. Border (2006). *Infinite dimensional analysis: A hitchhiker's guide* (3rd ed.). Springer.

Allen, R. and J. Rehbeck (2019). Revealed stochastic choice with attributes. *Available at SSRN 2818041*.

Andreoni, J. and J. Miller (2002). Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica 70*(2), 737–753.

Apesteguia, J., M. A. Ballester, and J. Lu (2017). Single-crossing random utility models. *Econometrica 85*(2), 661–674.

Battalio, R. C., J. H. Kagel, and D. N. MacDonald (1985). Animals' choices over uncertain outcomes: Some initial experimental results. *American Economic Review 75*(4), 597–613.

Battalio, R. C., J. H. Kagel, H. Rachlin, and L. Green (1981). Commodity-choice behavior with pigeons as subjects. *Journal of Political Economy 89*(1), 67–91.

Block, H. D. and J. Marschak (1960). Random orderings and stochastic theories of responses. *Contributions to probability and statistics*, 97–132.

Cerreia-Vioglio, S., D. Dillenberger, P. Ortoleva, and G. Riella (2019). Deliberately stochastic. *American Economic Review*. Forthcoming.

Chambers, C., C. Liu, and J. Rehbeck (2019). Costly information acquisition. *Available at SSRN 2951887*.

Chambers, C. P. and F. Echenique (2016). *Revealed preference theory*, Volume 56. Cambridge University Press.

Chen, M. K., V. Lakshminarayanan, and L. R. Santos (2006). How basic are behavioral biases? Evidence from capuchin monkey trading behavior. *Journal of Political Economy 114*(3), 517–537.

Cherchye, L., B. De Rock, R. Griffith, M. O'Connell, K. Smith, and F. Vermeulen (2017). A new year, a new you? Heterogeneity and self-control in food purchases. CEPR Discussion Paper No. DP12499.

Choi, S., R. Fisman, D. Gale, and S. Kariv (2007). Consistency and heterogeneity of individual behavior under uncertainty. *American economic review 97*(5), 1921–1938.

Deb, R., Y. Kitamura, J. K.-H. Quah, and J. Stoye (2019). Revealed price preference: Theory and empirical analysis. *arXiv preprint arXiv:1801.02702*.

Diewert, W. E. (1973). Afriat and revealed preference theory. *Review of Economic Studies 40*(3), 419–425.

Falmagne, J. C. (1978). A representation theorem for finite random scale systems. *Journal of Mathematical Psychology 18*(1), 52–72.

Feldman, P. and J. Rehbeck (2019). Revealing a preference for mixing: An experimental study of risk. Working Paper.

Forges, F. and E. Minelli (2009). Afriat's theorem for general budget sets. *Journal of Economic Theory 144*(1), 135–145.

Freer, M. and C. Martinelli (2016). A representation theorem for general revealed preference. GMU Working Paper in Economics.

Fudenberg, D., R. Iijima, and T. Strzalecki (2015). Stochastic choice and revealed perturbed utility. *Econometrica 83*(6), 2371–2409.

Harbaugh, W. T., K. Krause, and T. R. Berry (2001). GARP for kids: On the development of rational choice behavior. *American Economic Review 91*(5), 1539–1545.

Hsieh, Y.-W., X. Shi, and M. Shum (2018). Inference on estimators defined by mathematical programming. *Available at SSRN 3041040*.

Kagel, J. H., L. Battalio, R. C. Battalio, L. Green, et al. (1995). *Economic choice theory: An experimental analysis of animal behavior*. Cambridge University Press.

Kagel, J. H., R. C. Battalio, H. Rachlin, and L. Green (1981). Demand curves for animal consumers. *Quarterly Journal of Economics 96*(1), 1–15.

Kagel, J. H., R. C. Battalio, H. Rachlin, L. Green, R. L. Basmann, and W. R. Klemm (1975). Experimental studies of consumer demand behavior using laboratory animals. *Economic Inquiry 13*(1), 22–38.

Kamae, T. and U. Krengel (1978). Stochastic partial ordering. *Annals of Probability*, 1044–1049.

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

Krasheninnikova, A., F. Höner, L. O'Neill, E. Penna, and A. M. von Bayern (2018). Economic decision-making in parrots. *Scientific reports 8*(1), 12537.

Latty, T. and M. Beekman (2010). Irrational decision-making in an amoeboid organism: Transitivity and context-dependent preferences. *Proceedings of the Royal Society B: Biological Sciences 278*(1703), 307–312.

Lea, A. M. and M. J. Ryan (2015). Irrationality in mate choice revealed by túngara frogs. *Science 349*(6251), 964–966.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. New York: Wiley.

Machina, M. J. (1985). Stochastic choice functions generated from deterministic preferences over lotteries. *Economic Journal*, 575–594.

McFadden, D. and M. K. Richter (1990). Stochastic rationality and revealed stochastic preference. *Preferences, Uncertainty, and Optimality, Essays in Honor of Leo Hurwicz, Westview Press: Boulder, CO*, 161–186.

McFadden, D. L. (2005). Revealed stochastic preference: A synthesis. *Economic Theory 26*(2), 245–264.

Natenzon, P. (2019). Random choice and learning. *Journal of Political Economy 127*(1), 419–457.

Richter, M. K. (1979). Duality and rationality. *Journal of Economic Theory 20*(2), 131–181.

Shafir, S. (1994). Intransitivity of preferences in honey bees: Support for 'comparative' evaluation of foraging options. *Animal Behaviour 48*(1), 55–67.

Sopher, B. and J. M. Narramore (2000). Stochastic choice and consistency in decision making under risk: An experimental study. *Theory and Decision 48*(4), 323–350.

Swait, J. and A. Marley (2013). Probabilistic choice (models) as a result of balancing multiple goals. *Journal of Mathematical Psychology 57*(1), 1–14.

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization 1*(1), 39–60.

Thurstone, L. (1927). A law of comparative judgment. *Psychological review 34*(4), 273.

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