UNIVERSITY OF SUSSEX | BUSINESS SCHOOL

ECONOMICS

Working Paper Series

No. 02-2021

Revealed statistical consumer theory

Roy Allen
University of Western Ontario
rallen46@uwo.ca

Pawel Dziewulski
University of Sussex
p.k.dziewulski@sussex.ac.uk

John Rehbeck
Ohio State University
rehbeck.7@osu.edu

Abstract: We provide a microfoundation for using aggregated data (e.g. mean purchases) when evaluating consumer choice data. We present a model of statistical consumer theory where the individual maximizes their utility with respect to a distribution of bundles that is constrained by a statistic of the distribution (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. This means that despite working with distributions, the empirical content of the model only depends on a finite-dimensional statistic. Statistical consumer theory neither nests nor is nested in the random utility approach. We show this approach generalizes quasilinear utility with random preferences and income, mean-variance preferences, and preferences that depend on arbitrary moments.

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

February, 2021

Abstract

We provide a microfoundation for using aggregated data (e.g. mean purchases) when evaluating consumer choice data. We present a model of *statistical consumer theory* where the individual maximizes their utility with respect to a distribution of bundles that is constrained by a statistic of the distribution (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. This means that despite working with distributions, the empirical content of the model only depends on a finite-dimensional statistic. Statistical consumer theory neither nests nor is nested in the random utility approach. We show this approach generalizes quasilinear utility with random preferences and income, mean-variance preferences, and preferences that depend on arbitrary moments.

* We thank Abi Adams-Prassl, Victor Aguiar, Christopher P. Chambers, Sylvain Chassang, Ian Crawford, Jay Lu, Matthew Polisson, Kota Saito, Gerelt Tserenjigmid, Leeat Yariv, and participants at the conference on "Consumer Behaviour: New Models, New Methods" for helpful comments.

1 Introduction

Outside of experimental settings, data typically do not match the classic consumer problem of utility maximization subject to a budget constraint. A leading example is scanner data. Such data is highly disaggregated, typically at the transaction-level. This raises several concerns when attempting to apply consumer theory. First, it is unclear whether the budget of the consumer is at the transaction or for a longer period of time. A second concern is the "zeros" problem: a typical transaction involves zero quantities of many goods, though when examining many transactions an individual may purchase a wide variety of goods. In practice, researchers typically aggregate (sum up) the disaggregated data across time to address the zeros problem. In this case, budgetary constraints are in terms of aggregated data. For example, Echenique et al. (2011) process transaction-level data to form quantities for four-week periods and analyze this aggregated data without referencing the original primitive data.[1] Several questions arise from this procedure: Precisely, how does the budget enter? What information is lost when working directly with the aggregated data rather than transaction-level quantities?[2]

This paper addresses these questions by providing a demand theory for *distributions*. We hypothesize that given prices, an individual chooses a *distribution* of quantities. An empirical analyst can interpret a transaction as a realization from this distribution. The budgets we consider involve a statistic of the distribution, and we call the general framework *statistcal demand theory*. In our main analysis, we assume the budget restricts the *mean expenditure*. The key insight of this paper is that with this assumption on the budget there is *no loss* in working with aggregated (mean) quantities relative to the entire distribution. Specifically, the testable implications of the framework are exhausted by an acyclicity condition imposed only on the mean quantities.

This equivalence has several important implications for applied work. First, this provides a plausible microfoundation for using mean consumption. A microfoundation for

[1] Historically, aggregating over time has regularly been done when examining demand systems. For example, Barten (1969) looks at yearly data of sixteen commodity groups, Pesaran and Deaton (1978) looks at quarterly data of income and a single consumption index, and Deaton and Muellbauer (1980) looks at yearly data of eight commodities.

[2] A third concern is that prices may not be constant over the period of aggregation; this raises new challenges that are orthogonal to the conceptual contribution of this paper. We are pursuing a flows interpretation in ongoing work that complements the present paper. For a measurement error perspective, see Aguiar and Kashaev (2020).

using aggregate purchases is absent from the literature, even though this is standard practice. Second, when using the model for counterfactual analysis (in settings with similar budgets), the data are informative only about the mean of the distribution of quantities in a new setting. Third, insights from the standard consumer problem apply when one replaces "quantities" in the standard model with "mean quantities" when studying distributional choice. Thus, an applied researcher need only open a standard textbook to find a theory of welfare or counterfactuals in our setting. Fourth, when testing the model of statistical demand theory with a budget on mean expenditure, there is no loss in working with only mean quantities without keeping track of the entire distribution. In particular, to conduct econometric analysis that takes into account sampling variability, it is only necessary to treat sampling variability in the mean quantities rather than the entire distribution of choices.[3]

In addition to providing a justification for what researchers are already doing, we show that statistical demand theory generalizes several different approaches. For example, distributional choice with mean expenditure constraints generalizes behavior from a random quasilinear utility model with random income. This is an important generalization since the income for purchases is often unobserved when studying purchases of a specific group of goods (e.g. groceries). While we focus on the mean expenditure budget, the approach of statistical demand theory applies to other budgets. Building on Forges and Minelli (2009), we characterize statistical demand theory for general statistics other than the mean, and show how this approach covers mean variance preferences and preferences that depend on higher order moments.

We also make a theoretical contribution to the growing literature on stochastic choice. Rather than starting from random utility models as in Block and Marschak (1960) and McFadden and Richter (1990), we start from a preference for randomization following Machina (1985).[4] Machina (1985) studies stochastic choice of a probability distribution

[3] This paper does not directly address sampling variability. However, building on the theoretical insights of this paper, we present a statistical test in Allen et al. (2021) when there are two goods and two prices.

[4] This is often called deliberate stochastic choice. For recent work following this framework see Swait and Marley (2013), Fudenberg et al. (2015), Freer and Martinelli (2016), Cerreia-Vioglio et al. (2019), Allen and Rehbeck (2019), and Banerjee et al. (2020). See Sopher and Narramore (2000), Agranov and Ortoleva (2017), Agranov et al. (2020), Agranov and Ortoleva (2020), and Feldman and Rehbeck (2020) for experimental evidence supporting this view.

over *finitely many* alternatives. Following Machina (1985), we take a "demand approach," but study the classical consumer setting with distributions over consumption bundles. We differ from Machina (1985) since distributions over bundles are an infinite dimensional object and we leverage price variation from the standard consumer problem. The main characterization with a mean expenditure constraint demonstrates an important dimension reduction aspect where an infinite dimensional model is observationally equivalent to a finite dimensional one.

Our approach differs from random utility models (RUMs), which are the main paradigm to model stochastic choice and have a rich intellectual history following Thurstone (1927), Luce (1959), Block and Marschak (1960), Falmagne (1978), McFadden and Richter (1990), McFadden (2005), Kitamura and Stoye (2018), among many others. RUMs suppose that each individual has a random distribution of preferences that generate behavior. While RUMs have intuitive appeal, in practice, there are often computational constraints to using these models on consumption data. For example the applications in Kitamura and Stoye (2018) and Deb et al. (2019) consider only a few goods and use data that is aggregated to a commodity index rather than using micro-level data on consumer purchases. There have been some recent advances to compute tests of random utility models with a larger number of goods in Smeulders et al. (2019). In contrast, the approach in this paper could readily be applied to micro-data before aggregating to a commodity index and testing follows by applying methods of Varian (1982) or Aguiar and Kashaev (2020) to mean consumption bundles. Finally, we show through examples that RUMs and statistical consumer theory do not nest each other. Thus, the difference of RUMs and statistical consumer theory could be detected from data.

This paper also informs work on revealed preference, which is discussed in a textbook setting in Chambers and Echenique (2016). In particular, we build on the standard demand setting studied in Afriat (1967), Diewert (1973), and Varian (1982). We show how structure from the budget sets reduces the dimensionality of preferences using techniques 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 characterizes the model with budget constraints on average expenditure and relates this behavior to choice with random incomes. Section 3 contrasts behavior of statistical consumer theory with random utility models. Section 4 describes more general conditions that show how systematic statistical constraints are equivalent to models where the individual only chooses the statistic. We show through examples how this leads to generalizations of mean variance preferences. Section 5 provides a discussion of the results and final remarks.

2 Model and results

In this section, we present a model of consumer choice where an individual maximizes their utility over a distribution of consumption bundles, but faces a restriction on average expenditure. We discuss how this extends to general statistics in Section 4. All proofs not found in the main text are located in Appendix A.

2.1 Preliminaries

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 $\ell = 1, \dots, L$ 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 $\geq$.[5] 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 $\int \|x\| d\mu(x) < \infty$, which is without loss of generality in this framework.[6] Finally, we endow Δ with the topology of weak convergence.[7]

We consider a model of consumer choice in which an individual maximizes a utility function with respect to probability distributions over consumption bundles.[8] We denote the utility function over distributions by $U : \Delta \to \mathbb{R}$. Thus, stochastic choice is generated by an agent who chooses a most-preferred distribution of bundles and randomizes the

[5] That is, 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 $\ell = 1, \dots, L$. In addition, the relation is strict and denoted by $x > y$ if $x \geq y$ and $x \neq y$.

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

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

[8] One could alternatively start with a preference ordering over distributions of consumption bundles. For simplicity of exposition, we start from a utility function.

selection of a particular bundle in $\mathbb{R}^L_+$ according to this distribution.

The *primitive dataset* $\mathcal{D} := \left\{(p^t, m^t, \mu^t) : t \in T\right\}$ consists of a finite number of triplets of prices $p^t \in \mathbb{R}^L_{++}$, expenditure levels $m^t \in \mathbb{R}_{++}$, and probability measures $\mu^t \in \Delta$.[9] Here, we interpret the probability measure μ^t as being chosen by a single consumer. As discussed earlier, a researcher with transaction-level data may treat each transaction as a draw from the probability measure μ^t.[10] 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 $m^t = \int (p^t \cdot x) d\mu^t(x)$ without loss of generality for statistical consumer theory with average expenditure constraints. Thus, only information on (p^t, μ^t) is needed to apply these methods in practice.

2.2 Statistical consumer with average expenditure constraints

Here, we study when the set of feasible distributions is constrained by average expenditure for given observed prices. Specifically, given prices $p \in \mathbb{R}^L_{++}$ and expenditure level $m > 0$, a feasible distribution ν must be contained in the *average expenditure budget*

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

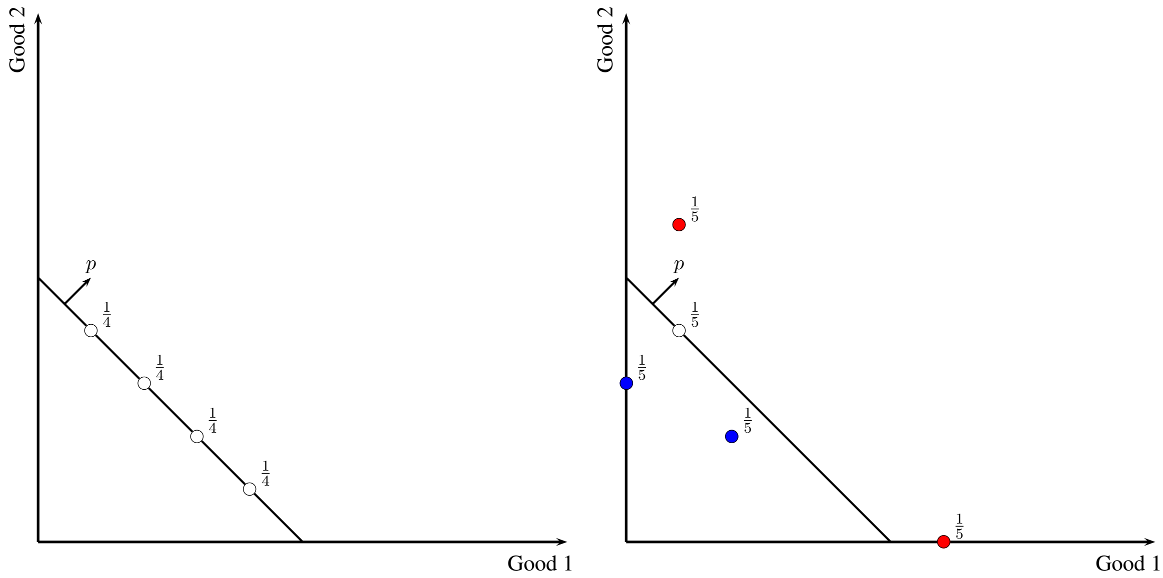
The average expenditure budget allows an individual to choose consumption bundles that cost more or less than the expenditure level. However, only the average expenditure on purchases is constrained. One could interpret the average expenditure constraint as a type of mental accounting (Thaler, 1980). An individual interacts with the average expenditure budget by selecting a distribution μ to maximize their utility U over the budget $A(p, m)$.

Here, it is useful to consider the types of feasible distributions the average expenditure constraint allows. Note the two examples in Figure 1. The points represent the support of a feasible distribution over consumption bundles with the corresponding probabilities next to each point. The line is the average expenditure constraint at given prices. In

[9] Here we assume some time period has been chosen over which the researcher aggregates. We abstracts from what is the "right" time period to estimate a distribution. This same problem is abstracted from in standard consumer theory where the "right" time period to aggregate purchases is not well studied. In practice, applications use a time period of several weeks.

[10] For this paper, we assume there is no error in measuring the chosen distribution μ^t, and later discuss how our results speak to empirical settings in which μ^t is measured with stochastic error.

Figure 1a, all bundles chosen with positive probability have expenditure equal to the average expenditure. In contrast, Figure 1b has bundles chosen with positive probability that are below (blue) or above (red) the average expenditure constraint. One could interpret purchases below average expenditure as "saving" behavior whereas a realization of purchases above average expenditure could be interpreted as "splurging" behavior. Moreover, this distribution has several realizations with zero purchases of different goods that are common in data. The distinction between saving and splurging purchases could be carried forward in a dynamic model and interact with future average expenditure constraints, but we do not pursue this here.



(a) Feasible budget A

(b) Feasible budget B

Figure 1: Example feasible distributions for average expenditure constraint

Notes: The points represent the support of a feasible distribution over consumption bundles with the corresponding probabilities next to each point. The line is the average expenditure constraint at given prices p. Red and blue points denote a "splurge" and "savings," respectively.

With this restriction, we consider *average expenditure datasets* $\mathcal{D}^A := \{(A^t, \mu^t) : t \in T\}$, where $A^t := A(p^t, m^t)$ and $\mu^t \in A^t$. We are interested in when the dataset $\mathcal{D}^A$ can be described, or *rationalized*, by utility maximization. Equivalently, when there is 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. Local nonsatiation of U requires that, for any probability measure $\nu \in \Delta$ and any neighborhood of ν, there is some distribution ν' in its neighborhood (i.e, a nonempty open set containing ν') 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 $\int (p^t \cdot x) d\mu^s(x) \leq m^t$ since the measure μ^s was available from the budget A^t and μ^t was chosen. 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 $\int (p^t \cdot x) d\mu^s(x) < m^t$.

It is straightforward to show that the relation R is 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 average expenditure constraint $\int (p^t \cdot x) d\mu^s(x) \leq m^t$. Hence, $\mu^t \; R \; \mu^s$ must imply $U(\mu^t) \geq U(\mu^s)$. One can also show that the strict directly revealed preference P is consistent with a locally nonsatiated utility.[11]

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 $\mu^t \; R^* \; \mu^s$, when 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 is ordered with respect to P. This immediately implies the testable restriction of *mean acyclicity.*

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

[11] First, recall that local nonsatiation of U requires that, for any element $\nu \in \Delta$ and its neighborhood, there is some ν' in the neighborhood such that $U(\nu') > U(\nu)$. By continuity of the function $\nu \to \int (p^t \cdot x) d\nu(x)$, the set $\big\{\nu \in \Delta : \int (p^t \cdot x) d\nu(x) < m^t\big\}$ 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 $U(\mu^t) \geq U(\nu') > U(\mu^s)$.

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 $\int (p^t \cdot x) d\mu^t(x) = m^t$, 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 analyst to observe the expenditure level m^t. Thus, one only needs a dataset $\{(p^t, \mu^t)\}_{t\in T}$ while setting the expenditure level $m^t = \int (p^t \cdot x) d\mu^t(x)$ to check mean acyclicity.

The definition of mean acyclicity is equivalent to restrictions on the revealed preference relations. In particular, mean acyclicity coincides with the *generalized axiom of revealed preference* (GARP) on the revealed preference relation R^* so that

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

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, then 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 $\int (p^t \cdot x) d\mu^s(x) = p^t \cdot x^s$ for all $t, s \in T$, which reduces mean acyclicity to GARP.

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

$$\int (p \cdot x) d\nu(x) \ = \ p \cdot \int x \, d\nu(x)$$

for all $\nu \in \Delta$ and $t \in T$. This implies that all relevant information for mean acyclicity is summarized by the L-dimensional mean bundle $\int x \, d\nu(x)$ for the distribution ν. In practice, estimating mean bundles rather than the whole distribution μ^t is sufficient to check the mean acyclicity condition.

Since we can represent mean acyclicity on mean bundles, it is natural to study utility functions that depend only on mean bundles.

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

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

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

If the dataset can be rationalized with a mean choice model, then any utility function that rationalizes the data only depends on the mean bundle (a vector), rather than all information in the distribution. In the main theorem below, we show that locally nonsatiated preferences are observationally equivalent to the mean choice model when the sets of distributions are restricted by the average 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) $\mathcal{D}^A$ *is rationalizable with a locally nonsatiated utility function* $U : \Delta \to \mathbb{R}$.

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

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

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

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. This connection means we can use the framework of Varian (1982) for counterfactual and welfare analysis. In addition, like Afriat (1967), we see there is no loss of generality restricting attention to a mean choice model with a well-behaved function f, i.e., a continuous, strictly increasing, and concave. Thus, the utility function U itself can be treated as continuous, strictly increasing, and concave, without loss of generality. A more general statistical consumer theory is developed in Section 4.

Second, a model with a locally nonsatiated preference over distributions is observationally equivalent to a mean choice model. Roughly, this means that a researcher cannot gain more flexibility by modelling a preference over distributions than modelling a preference that only depends on mean consumption for the average expenditure budget.

Third, Theorem 1(ii) has important implications when applying the mean choice model in certain empirical settings. For example, sampling variability may arise when 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 applied work. In contrast, for our approach, preferences are defined over distributions μ^t and the realizations are drawn from the chosen distributions. Theorem 1 shows that estimation error in

μ^t only matters insofar as it leads to error in estimating the *mean* of consumption. This is because checking mean acyclicity is possible by checking restrictions on *means*. We leverage the dimension reduction aspect to incorporate variability in empirically relevant datasets in Allen et al. (2021).

2.3 Relation to a random income model

The model of distributional preferences above may seem stylized, but we show it generalizes a class of random utility models where income is also random and unobserved. In applied analysis, it is often assumed that the income of an individual is equal to expenditure on goods. However, since income is often unobserved, it might make more sense to treat it as a random variable. If preferences and income are random, then we are essentially in the case of the Sonnenschein-Mantel-Debreu (Debreu, 1974) anything goes result for mean demands.[12] However, we show that a random quasilinear utility maximizer with random income is rationalizable with a mean choice model.

To formalize this, let (η, ε) be random variables that govern random utility functions and income. We assume these variables are independent of prices, but allow preferences to be potentially correlated with income. We give a precise definition of this model below.

Definition 3. *A* random quasilinear utility and income model *with random variables* (η, ε) *has individuals make choices according to*

$$\begin{aligned} \max_{(x,y)\in\mathbb{R}^L_+\times\mathbb{R}} \quad & u(x;\eta) + y \\ s.t. \quad & p \cdot x + y \leq m(\varepsilon). \end{aligned}$$

Thus, the realization of the random variable η gives a random draw of preferences, while the random variable ε governs the realization of income. To show the relation to a mean choice model and distributional choice, we will make some assumptions. For technical simplicity, we suppose that (η, ε) takes values in the finite set $N \times E$ and for each realization of η the utility function $u : \mathbb{R}^L_+ \times N \to \mathbb{R}$ yields a unique maximizer for all prices $\{p^t\}_{t\in T}$. We let $x^{*,t}(\eta, \varepsilon)$ denote the unique choice given price p^t and unobservables (η, ε). To map to our previous analysis, given a price p^t, a distribution of choices arises

[12] To see why this is the case, a realization of a random draw of preference and incomes is equivalent to the existence of some household in the Sonnenschein-Mantel-Debreu theorem.

because $x^{*,t}(\eta,\varepsilon)$ is random due to (η,ε). The quasilinear model has recently been studied in Brown and Calsamiglia (2007) and Allen and Rehbeck (2020a,b).

The paper by Allen and Rehbeck (2020b) shows that the solutions to the maximization problem satisfy cyclic monotonicity (cf. Rockafellar, 1970) on the expectation of the choices so that for any sequence $\{t_m\}_{m=1}^{M}$ with $t_m \in T$ the inequality

$$\sum_{m=1}^{M} p^{t_m} \cdot \Big(\mathbb{E}[x^{*,t_m}(\eta,\varepsilon)] - \mathbb{E}[x^{*,t_{m+1}}(\eta,\varepsilon)] \Big) \;\leq\; 0$$

holds, where $t_{M+1} = t_1$. It can be shown that this condition implies that mean acyclicity is satisfied. From Theorem 1, this means that these choices are observationally equivalent to a locally nonsatiated utility function that chooses the distribution of choices or a mean choice model. We record this in the proposition below.

Proposition 1. *Suppose that for prices $\{p^t\}_{t\in T}$ that data is generated by a random quasi-linear utility and income model with random variables (η,ε) that take values in the finite set $N \times E$ and for every realization of η the utility $u(x;\eta)$ yields a unique maximizer, then the distribution of choices satisfies mean acyclicity.*

3 Comparison to random utility models

We have just shown that when quasilinear preferences and income are random, the distribution of choices is rationalizable with a mean choice model. In this section we compare the mean choice model to random utility models that have random preferences (not necessarily quasilinear), but a fixed budget. For the random utility model, an individual maximizes utility for every 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). Here, random utility models are generally defined over the budget set from the standard deterministic consumer problem. Recall the standard budget set is given by

$$B(p,m) \;:=\; \Big\{ x \in \mathbb{R}^L_+ : p \cdot x \leq m \Big\}, \tag{5}$$

for prices $p \in \mathbb{R}^L_{++}$ and $m > 0$. For brevity, we let $B^t := B(p^t, m^t)$.

We now formally describe random utility models. Let $\mathscr{U}$ be the space of strictly quasiconcave locally nonsatiated utility functions $\tilde{u} : \mathbb{R}^L_+ \to \mathbb{R}$. 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 $t \in T$:

$$\mu^t(O) \;=\; \rho\Big(\big\{\tilde{u} \in \mathscr{U} : \operatorname{argmax}_{y \in B^t} \tilde{u}(x) \in O\big\}\Big), \tag{6}$$

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 B^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).[13]

Distributions of choices generated by a random utility model are in the set

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

A key difference between the mean choice model and random utility model is that budget sets $A(p,m)$ and $C(p,m)$ are different. The mean choice model allows choice of consumption bundles that exceed the expenditure level m^t, which is not allowed in random utility models. Recall that as emphasized above, for the mean choice model only the average expenditure need be measured. The random utility model has a fixed budget, in which case the average expenditure is the same as the expenditure for each realization of the random utility. To further compare the models, we restrict attention to distributions when the support of μ^t is a subset of B^t,[14] for all $t \in T$. We show that mean choice models neither nest nor are nested in random utility models for such distributions.

In Example 1, we discuss a dataset that can be rationalized only by a mean choice model. Here, there is no RUM that can generate the observations. Despite this, the mean behavior is consistent with mean acyclicity. In contrast, the dataset in Example 2 is only rationalizable by a RUM. Since the models describe different behavior, one can discriminate between mean choice models and RUMs using field data or experiments.

Example 1. Let a primitive dataset be given by $\mathcal{D} = \big\{(p^1, m^1, \mu^1), (p^2, m^2, \mu^2)\big\}$, 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 probability $7/12$ to $(0, 1/2)$ and $5/12$ to $(1,0)$.

[13] Alternatively, random utility models are characterized by the *axiom of revealed stochastic preferences* in McFadden and Richter (1990) and McFadden (2005) which is conceptually more similar to GARP.

[14] The support of μ^t is the smallest closed set K such that $\mu^t(K) = 1$.

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

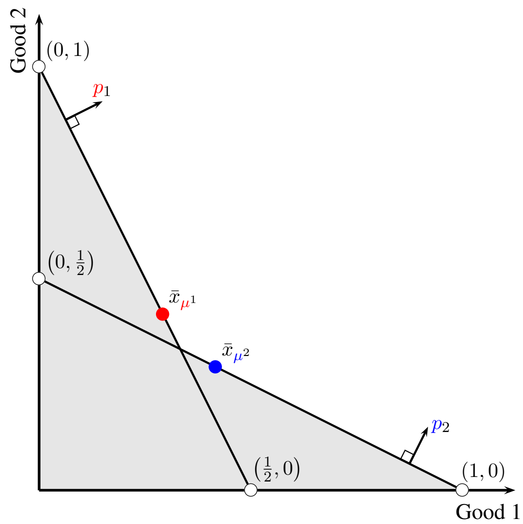


Figure 2: Graphical interpretation of the dataset in Example 1.

In contrast, the data are inconsistent with the random utility model. Indeed, since $p^1 \cdot (0, 1/2) = 1/2 < m^1$, there must be probability of at least $7/12$ on utilities where bundle $(0, 1/2)$ is strictly inferior to $(1/2, 0)$. Analogously, as $p^2 \cdot (1/2, 0) = 1/2 < m_2$, at least a probability of $7/12$ on utilities must rank $(0, 1/2)$ strictly over $(1/2, 0)$. However, this would imply that for a probability 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 2. Let the 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 $m^1 = m^2 = 1$; 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 $(1/2, 1/4)$. See Figure 3 for a graphical representation.

One can easily show that 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 $B^1 := \{x \in \mathbb{R}^2_+ : p^1 \cdot x \leq 1\}$ and uniquely maximized at

$(1/2, 1/4)$ over $B^2 := \{x \in \mathbb{R}^2_+ : p^2 \cdot x \leq 1\}$. 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 B^2 and uniquely maximized at $(1/4, 1/2)$ over B^1. Therefore, a random utility model ρ that assigns probability $1/2$ to each of these utility functions rationalizes these data.

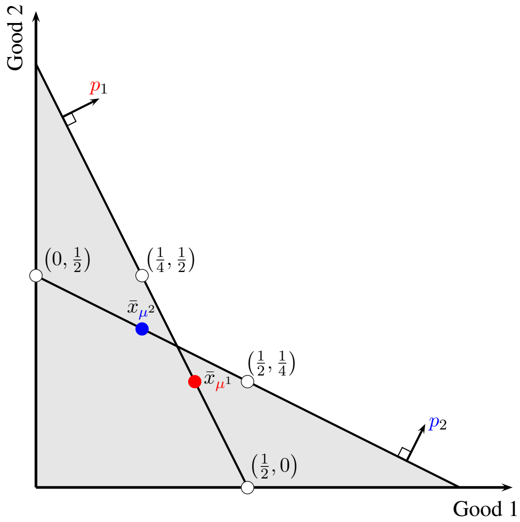


Figure 3: Graphical interpretation of the dataset in Example 2

4 General statistical consumer theory

We have focused on the mean choice model due to the empirical relevance and intuitive appeal of a budget on average expenditure. This section presents a more general *statistical choice model*, where the distribution of choices is restricted through a more general statistic than the mean. Here, we say a statistic is a finite-dimensional summary of a distribution. In particular, let $S : \Delta \to \mathbb{R}^K$ be a continuous function from the distribution of bundles to the K-dimensional real vector space.[15] In addition, we assume that the function S is locally nonsatiated, i.e., for every neighborhood of $\nu \in \Delta$ there is a measure ν' in the neighborhood such that $S(\nu') > S(\nu)$.[16]

Following Forges and Minelli (2009), we assume that the budget constraint for observation $t \in T$ is represented by a continuous function $g^t : \mathbb{R}^K \to \mathbb{R}$. A distribution $\nu \in \Delta$

[15] Note that the statistic does not need to map to the same dimension as the consumption bundle.

[16] That is, we have $S_i(\nu') \geq S_i(\nu)$, for all dimensions $i = 1, \ldots, K$, and $S_i(\nu') > S_i(\nu)$, for some i.

is feasible if $g^t\big(S(\nu)\big) \leq 0$. Here a general dataset is given by $\mathcal{D}^G = \big\{(g^t, \mu^t) : t \in T\big\}$. Here we make the high level assumption that $g^t\big(S(\nu)\big)$ is continuous and, for any ν and its neighborhood, there is some ν' in the neighborhood such that $g^t\big(S(\nu')\big) > g^t\big(S(\nu)\big)$, for all $t \in T$.[17] Any variables describing g^t (such as prices) are absorbed into the function, since it is indexed by t. For example, the average expenditure constraint is represented by $g^t\big(S(\nu)\big) = p^t \cdot \int x \, d\nu(x) - m^t$, where $S(\nu) = \int x \, d\nu(x)$ and $g^t(y) = p^t \cdot y - m^t$.

Definition 4. *The dataset $\mathcal{D}^G$ with statistic S is rationalizable with a* statistical choice model *if there is a function $f : \mathbb{R}^K \to \mathbb{R}$ such that the function*

$$U(\nu) \;:=\; f\big(S(\nu)\big)$$

is locally nonsatiated and rationalizes $\mathcal{D}^G$, i.e., if $g^t\big(S(\nu)\big) \leq 0$ then $U(\mu^t) \geq U(\nu)$.

Similar to the mean expenditure constraint, we can define a direct revealed preference relation R_S by setting $\mu^t \, R_S \, \mu^s$ when $g^t\big(S(\mu^s)\big) \leq 0$. Similarly, we can define a strictly directly revealed preference relation P_S so that $\mu^t \, P_S \, \mu^s$ when $g^t\big(S(\mu^s)\big) < 0$. The continuity and locally nonsatiation properties on the budget constraint and statistic allow us to mirror arguments in Section 2.2 to show that these relations are consistent with a rationalization of the choices by a locally nonsatiated function $U : \Delta \to \mathbb{R}$.

Similarly, for any $t, s \in T$, we say that μ^t is *revealed preferred* to μ^s, denoted by $\mu^t \, R_S^* \, \mu^s$, when there is a sequence of indices $a, b, c, \ldots, z \in T$ such that

$$\mu^t \, R_S \, \mu^a, \;\; \mu^a \, R_S \, \mu^b, \;\; \ldots \;, \;\; \text{and} \;\; \mu^z \, R_S \, \mu^s. \tag{7}$$

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 (7) with at least one pair ordered with respect to P_S. This motivates the following acyclity condition.

Definition 5 (Statistical acyclicity)**.** *For any cycle $\mathcal{C} = \big\{(a,b), (b,c), \ldots, (z,a)\big\}$ 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 has the same link to GARP as in Section 2.2 but with a general budget constraint and statistic. We show that statistical acyclicity and a statistical choice model are observationally equivalent for data generated from a locally nonsatiated utility function $U : \Delta \to \mathbb{R}$ with a general budgets and a statistic S.

[17] One sufficient condition for this is when each g^t is strictly increasing and S is locally nonsatiated.

Theorem 2. *For any set of observations $\mathcal{D}^G$, the following statements are equivalent:*

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

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

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

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

Thus, the result for the mean choice model in Theorem 1 holds more generally. In particular, a general model of preferences over distributions with systematic restrictions via a statistic in the budget constraint is observationally equivalent to a model where the statistic enters preferences. We note that this is not simply a mathematical curiosity, but can be used to study some empirically relevant types of preferences. In particular, one can study a generalization of mean-variance preferences or preferences that depend on arbitrary moments of consumption.

Example 3. (Mean-variance preferences) Using the results for general consumer theory, we can obtain restrictions that generalize mean-variance preferences. Suppose a consumer maximises a utility $U : \Delta \to \mathbb{R}$ subject to constraints on the mean consumption and variance. For each observation $t \in T$ and each good $\ell = 1, \dots, L$, there are positive numbers $\bar{x}^t_\ell$ and $\bar{v}^t_\ell$ such that a feasible distribution $\nu \in \Delta$ must satisfy

$$\mathbb{E}_\ell(\nu) \;:=\; \int x_\ell \, d\nu(x) \;\leq\; \bar{x}^t_\ell \quad \text{and} \quad \mathrm{Var}_\ell(\nu) \;:=\; \int \big(x_\ell - \mathbb{E}_\ell(\nu)\big)^2 \, d\nu(x) \;\leq\; \bar{v}^t_\ell.$$

One can represent the above set of constraints with a single function $g^t : \mathbb{R}^{2L}_+ \to \mathbb{R}$,

$$g^t\big(S(\nu)\big) \;:=\; \max\Big\{ \max_\ell \big\{\mathbb{E}_\ell(\nu) - \bar{x}^t_\ell\big\}, \max_\ell \big\{\mathrm{Var}_\ell(\nu) - \bar{v}^t_\ell\big\}\Big\},$$

where $S(\nu) = \big(\mathbb{E}_1(\nu), \dots \mathbb{E}_L(\nu), \mathrm{Var}_1(\nu), \dots, \mathrm{Var}_L(\nu)\big)$. The composition of $g^t\big(S(\nu)\big)$ is continuous and satisfies the additional high level nonsatiation condition.[18] Following Theorem 2, choices made with this budget constraint can be rationalized with a utility function

$$U(\nu) \;:=\; f\big(\mathbb{E}(\nu), \mathrm{Var}(\nu)\big)$$

[18] This is because, for any $\nu \in \Delta$ and its neighborhood, one can find a distribution ν' in the neighborhood that has a strictly greater mean and variance than ν, for each $\ell = 1, \dots, L$. Therefore, we have $g^t\big(S(\nu')\big) > g^t\big(S(\nu)\big)$, for all $t \in T$.

where $\mathbb{E}(\nu) = \big(\mathbb{E}_\ell(\nu)\big)_{\ell=1}^L$ and $\text{Var}(\nu) = \big(\text{Var}_\ell(\nu)\big)_{\ell=1}^L$.

This shows that it is observationally equivalent to behavior of preferences that depend only on the mean and variance of the distribution of choices. We note that the preferences on mean and variance are separable under this formulation. Thus, this rationalization generalizes the standard mean-variance preferences that are additively separable. In practice, one would set $\bar{x}_\ell^t = \mathbb{E}_\ell(\mu^t)$ and $\bar{v}_\ell^t = \text{Var}_\ell(\mu^t)$ to check statistical acyclicity.

We note that the above restrictions do not depend on prices. However, prices could easily be incorporated into the above formulation. For example, one could add the average expenditure constraint so that

$$g^t\big(S(\nu)\big) \;:=\; \max\left\{\max_\ell\left\{\mathbb{E}_\ell(\nu) - \bar{x}_\ell^t\right\}, \max_\ell\left\{\text{Var}_\ell(\nu) - \bar{v}_\ell^t\right\}, \int (p^t \cdot x)\, d\nu(x) - m^t\right\}$$

and the resulting preferences are still separable in mean and variance expenditure. One could also have considered constraints on average expenditure on specific goods or constraints on the variance of expenditure on goods.

Example 4 (Moment preferences)**.** Let a consumer maximize a utility $U : \Delta \to \mathbb{R}$ subject to constraints imposed on moments of consumption. Let $m_{\ell,j}(\nu) := \int x_\ell^j \, d\mu(x)$ denote the j-th moment of the consumption for the ℓ-th good according to the distribution $\nu \in \Delta$. Suppose that, for each observation $t \in T$, a constraint is imposed on the moments of order $j = 1, \dots, J$ for each good $\ell = 1, \dots, L$. Therefore, for each $\ell = 1, \dots, L$ and $j = 1, \dots, J$, we require that feasible distributions satisfy the inequality $m_{\ell,j}(\nu) \le \bar{\kappa}_{\ell,j}^t$ for some positive number $\bar{\kappa}_{\ell,j}^t$. Note that, in such a case, the budget constraint can be described by the function $g^t : \mathbb{R}^{LJ} \to \mathbb{R}$ of

$$g^t\big(S(\nu)\big) \;:=\; \max_{\ell,j}\left\{m_{\ell,j}(\nu) - \bar{\kappa}_{\ell,j}\right\}$$

where $S(\nu) = \big(m_{1,1}(\nu), \dots, m_{1,J}(\nu), m_{2,1}(\nu), \dots, m_{L,1}(\nu), \dots, m_{L,J}(\nu)\big)$. It is straightforward to show that $g^t\big(S(\nu)\big)$ is continuous and satisfies the additional high level nonsatiation condition.[19]

Therefore, following Theorem 2 a utility function given by

$$U(\nu) \;:=\; f\Big(\big(m_{\ell,j}(\nu)\big)_{(\ell,j)\in L\times J}\Big),$$

[19] Given that each function $m_{\ell,j}(\nu)$ is increasing with respect to first order stochastic dominance, for any distribution $\nu \in \Delta$ and its neighborhood, one can find some ν' in the neighborhood such that $m_{\ell,j}(\nu') > m_{\ell,j}(\nu)$, for all $(\ell, j) \in L \times J$. This suffices to show that $g^t\big(S(\nu')\big) > g^t\big(S(\nu)\big)$, for all $t \in T$.

rationalizes the data where $L = \{1, \ldots, L\}$ and $J = \{1, \ldots, J\}$. Thus, one can rationalize the choice of distributional choice that only depend on the moments of the distribution. Straightforward extensions of this could place restrictions on moments across goods. As in the previous example, one could also add the restriction of average expenditure constraints to the function that described the budget constraint and still obtain a preference that only depends on moments of the distribution.

We note that as additional moments of a distribution are modeled in the constraints, there are fewer revealed preference comparisons. To see this, note that if one adds an additional moment restriction to the constraints, then a distribution must satisfy an additional inequality to be revealed preferred. Thus, introducing additional moments to the constraints will necessarily describe more datasets. One potentially interesting exercise would be to look for the least moment restrictions that rationalize a dataset.

5 Conclusion

This is the first paper to provide a microfoundation for using aggregated data to examine consumer preferences. Even though many papers empirically analyze models using aggregate choices, until this paper there was no formal microfoundation that justified this practice. We show that if individuals have a preference for randomization, then it is without loss of generality to use data on aggregate choices.

More broadly, this paper relates to the growing literature on stochastic choice. For example, we show how a random quasilinear utility model with random incomes is nested in the approach. We also show that in practice statistical choice models can be differentiated from random utility models that have a fixed budget. While the main results study the average expenditure constraint, we show that the results also apply to general constraints that depend on a *statistic* of the distribution. We show how this can be used in practice to characterize a generalization of mean-variance preferences and preferences that depend on arbitrary moments.

One implication of the results for the mean choice model is that welfare analysis is possible by building on existing results from the standard consumer problem. Moreover, since it is without loss of generality to study models that depend on mean consumption, an applied researcher can use their favorite functional forms from consumer theory. For example, one can use a Cobb-Douglas model and replace the choice of consumption bundles with means. Lastly, in ongoing work (Allen et al., 2021) we present a statistical test of the deterministic stochastic choice model with mean expenditure constraint and apply it to data from capuchin monkeys. That paper builds on the insight in this paper that testability of the mean choice model depends on the mean quantities, and not the entire distribution of quantities.

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. We start with the proof of Theorem 2. Theorem 1 follows immediately after verifying certain properties when $g^t\big(S(\nu)\big) = p^t \cdot \int x\, d\nu(x) - m^t \leq 0$, where $S(\nu) = \int x\, d\nu(x)$.

We assume throughout that the function $g^t\big(S(\nu)\big)$ is continuous, for all $t \in T$, and for any $\nu \in \Delta$ and its neighborhood, there is some ν' in the neighborhood such that $g^t\big(S(\nu')\big) > g^t\big(S(\nu)\big)$, for all $t \in T$.

Proof of Theorem 2. It is clear that (iii) $\Rightarrow$ (iv) $\Rightarrow$ (i). We show that (i) $\Rightarrow$ (ii). Clearly, if $g^t\big(S(\nu)\big) \leq 0$ then $U(\mu^t) \geq U(\nu)$, for any $\nu \in \Delta$ and $t \in T$. Moreover, given local nonsatiation of U, it is also true that $g^t\big(S(\nu)\big) < 0$ implies $U(\mu^t) > U(\nu)$. Indeed, since $g^t\big(S(\cdot)\big)$ is continuous, the set $\big\{\nu \in \Delta : g^t(S(\nu)) < 0\big\}$ is open and contains ν. In particular, there must be some ν' such that $g^t\big(S(\nu')\big) < 0$ and $U(\nu') > U(\nu)$. Thus, by our previous observation, we obtain $U(\mu^t) \geq U(\nu') > U(\nu)$.

Now, towards contradiction, suppose there is a dataset that is rationalizable with the statistical choice model, but violates statistical acyclicity. Thus, there is a cycle $\mathcal{C} = \big\{(a, b), (b, c), \ldots, (z, a)\big\}$ in $T \times T$ such that $g^t\big(S(\mu^s)\big) \leq 0$, for all $(t, s) \in \mathcal{C}$, and $g^a\big(S(\mu^b)\big) < 0$, without loss. However, by our initial claim, this implies that

$$U(\mu^a) \;>\; U(\mu^b) \;\geq\; \ldots \;\geq\; U(\mu^z) \;\geq\; U(\mu^a),$$

which contradicts U being a well-defined function.

Next, we now show that (ii) $\Rightarrow$ (iii). Statistical acyclicity holds if and only if, for any cycle $\mathcal{C} = \big\{(a, b), (b, c), \ldots, (z, a)\big\}$ in $T \times T$ such that $g^t\big(S(\mu^s)\big) \leq 0$, 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 are numbers $\{\phi^t\}_{t\in T}$ and strictly positive numbers $\{\lambda^t\}_{t\in T}$ such that $\phi^s \leq \phi^t + \lambda^t g^t\big(S(\mu^s)\big)$, for all $t, s \in T$. Take any such numbers $\{\phi^t\}_{t\in T}$, $\{\lambda^t\}_{t\in T}$ and define the function $f : \mathbb{R}^K_+ \to \mathbb{R}$,

$$f(y) \ := \ \min\Big\{\phi^s + \lambda^s g^s(y) : s \in T\Big\},$$

which is continuous since each function g^t is continuous. We claim that the function $U : \Delta \to \mathbb{R}$, given by $U(\nu) := f\big(S(\nu)\big)$, rationalizes the set of observations and is locally nonsatiated. First, we show U rationalizes the dataset. Take any $t \in T$ and $\nu \in \Delta$ such that $g^t\big(S(\nu)\big) \leq 0$. Then,

$$\begin{aligned}
U(\nu) &= f\big(S(\nu)\big) \\
&= \min\Big\{\phi^s + \lambda^s g^s\big(S(\nu)\big) : s \in T\Big\} \\
&\leq \phi^t + \lambda^t g^t\big(S(\nu)\big) \\
&\leq \phi^t \\
&\leq \min\Big\{\phi^s + \lambda^s g^s\left(S(\mu^t)\right) : s \in T\Big\} \\
&= f\big(S(\mu^t)\big) \\
&= U(\mu^t),
\end{aligned}$$

where the first inequality follows from the property of the minimum function, the second inequality is implied by $\lambda^t > 0$ and $g^t\big(S(\nu)\big) \leq 0$, while the third follows from the construction of the numbers $\{\phi^t\}_{t\in T}$ and $\{\lambda^t\}_{t\in T}$. Finally, given the high-level nonsatiation assumption on $g^t\big(S(\nu)\big)$ and the fact that $\lambda^t > 0$, for all $t \in T$, it is straightforward to show that the function U is also locally nonsatiated. $\square$

Proof of Theorem 1. For the primitive dataset $\mathcal{D}$, define $g^t(x) := p^t \cdot x - m^t$ and let $S(\nu) = \int x \, d\nu(x)$. Clearly, each g^t is strictly increasing, since $p^t \in \mathbb{R}^L_{++}$. Moreover, $S(\nu) = \int x \, d\nu(x)$ is strictly increasing with respect to first order stochastic dominance by Lemma B.3, i.e., if ν' strictly first order stochastically dominates ν, then $S(\nu') > S(\nu)$. Thus, by Lemma B.4, it is locally nonsatiated. In particular, the functions $g^t\big(S(\nu)\big)$, for all $t \in T$, satisfy the high-level assumption (recall footnote 17). Following Theorem 2, this suffices for existence of a continuous and strictly increasing function f such that the function $U(\nu) := f\big(S(\nu)\big)$ is locally nonsatiated and rationalizes the data. Finally, the fact that f is concave and strictly increasing follows directly from the construction of the function in the proof of Theorem 2 and the fact that for all $t \in T$ the function g^t is concave (in fact, linear) and strictly increasing. $\square$

Proof of Proposition 1. Suppose that the data is generated by a random quasilinear utility and income model as in the statement of the proposition. For each p^t, let

$$\left(x^{*,t}(\eta,\varepsilon), y^{*,t}(\eta,\varepsilon)\right) \;:=\; \underset{(x,y)\in\mathbb{R}^L_+\times\mathbb{R}}{\operatorname{argmax}} \left\{ U(x,\eta) + y \,\middle|\, p^t\cdot x + y \le m(\varepsilon) \right\}$$

be the maximizer of choices when the values (η,ε) are realized by the random variables. Here $y^{*,t}(\eta,\varepsilon) = m(\varepsilon) - p^t \cdot x^{*,t}(\eta,\varepsilon)$.

Conditioning on the realization of the random variables, we can compare this to the $x^{*,s}(\eta,\varepsilon)$ when purchased at prices p^t. It follows that

$$U\left(x^{*,t}(\eta,\varepsilon),\eta\right) - p^t x^{*,t}(\eta,\varepsilon) \;\ge\; U\left(x^{*,s}(\eta,\varepsilon),\eta\right) - p^t x^{*,s}(\eta,\varepsilon)$$

where finiteness of the utility numbers is ensured by the existence of maximizers. Still conditioning on (η,ε), we can look at any sequence $\{t_m\}_{m=1}^M$ with $t_m \in T$ and get that

$$\sum_{m=1}^{M} p^{t_m}\cdot\left(x^{*,t_m}(\eta,\varepsilon) - x^{*,t_{m+1}}(\eta,\varepsilon)\right) \;\le\; 0$$

where $t_{M+1} = t_1$. Taking expectations over (η,ε), it follows that

$$\sum_{m=1}^{M} p^{t_m}\cdot\left(\mathbb{E}\left[x^{*,t_m}(\eta,\varepsilon)\right] - \mathbb{E}\left[x^{*,t_{m+1}}(\eta,\varepsilon)\right]\right) \;\le\; 0$$

where $t_{M+1} = t_1$ by linearity of expectations. This holds for any dataset generated by a random quasilinear utility and income maximizer.

To see that mean acyclicity is satisfied, suppose by contradiction that there is a cycle. It follows that there exists a sequence $\{\tilde{t}_m\}_{m=1}^M$ with $\tilde{t}_m \in T$ where

$$p^{\tilde{t}_m}\cdot\mathbb{E}\left[x^{*,\tilde{t}_{m+1}}(\eta,\varepsilon)\right] \;\le\; p^{\tilde{t}_m}\cdot\mathbb{E}\left[x^{*,\tilde{t}_m}(\eta,\varepsilon)\right]$$

with at least one inequality strict where $\tilde{t}_{M+1} = \tilde{t}_1$. Summing these inequalities up yields

$$\sum_{m=1}^{M} p^{t_m}\cdot\left(\mathbb{E}\left[x^{*,t_m}(\eta,\varepsilon)\right] - \mathbb{E}\left[x^{*,t_{m+1}}(\eta,\varepsilon)\right]\right) \;>\; 0$$

which contradicts that the data is generated by a random quasilinear utility. $\square$

Appendix B First order stochastic dominance

Here 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 $\ell = 1, \ldots, L$. The distribution μ first order stochastically dominates ν, or $\mu \succeq \nu$, whenever $\int f(x) d\mu(x) \geq \int f(x) d\nu(x)$, 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 B.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

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

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

See Lemma 4 in Kamae and Krengel (1978) for the proof. We say that μ strictly dominates ν, and denote it by $\mu \succ \nu$, if $\mu \succeq \nu$ and $\mu \neq \nu$. Using Lemma B.1, it is easy to show that we have $\mu \succ \nu$ if and only if there are random variables $X_\mu, X_\nu : \Omega \to X$ such that $X_\mu(\omega) \geq X_\nu(\omega)$, for all $\omega \in \Omega$, where the inequality is strict for all ω in some measurable set F such that $\tau(F) > 0$. We now prove a series of lemmas.

Lemma B.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.*[20]

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

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

is increasing. Since the simple function is defined on a measurable set, it is measurable. However, it must be that $\int \chi_D(x)d\mu(x) = \mu(D) < \nu(D) = \int \chi_D(x)d\nu(x)$, 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 $\{y \in \mathbb{R}^L : f(y) > r\}$ and $\{y \in \mathbb{R}^L : f(y) < r\}$ are upward and downward comprehensive, respectively. Moreover, they are both measurable, by measurability of f. This implies that

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

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 $x'_\ell \geq x_\ell$, for all $\ell = 1, \ldots, L$, and $x'_\ell > x_\ell$, for some ℓ, implies $f(x') > f(x)$, for any $x, x' \in X$.

Lemma B.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 $\int f(x)d\nu(x) > \int f(x)d\mu(x)$.*

Proof. Given that $\mu \succeq \nu$, Lemma B.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 B.1). For any strictly increasing $f : X \to \mathbb{R}$, we have

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

This completes the proof. □

Lemma B.4. *Suppose that $X+\mathbb{R}^L_+ \subseteq X$. For any measure $\mu \in \Delta_X$ and its 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 B.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

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

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

References

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

Agranov, M., P. J. Healy, and K. Nielsen (2020): "Stable randomization," Working Paper.

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

——— (2020): "Ranges of Preferences and Randomization," Working Paper.

Aguiar, V. H. and N. Kashaev (2020): "Stochastic revealed preferences with measurement error," *Review of Economic Studies*, forthcoming.

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

Allen, R., P. Dziewulski, and J. Rehbeck (2021): "Making sense of monkey business," Working paper.

Allen, R. and J. Rehbeck (2019): "Revealed Stochastic Choice with Attributes," *Available at SSRN 2818041*.

——— (2020a): "Counterfactual and welfare analysis with an approximate Model," ArXiv preprint arXiv:2009.03379.

——— (2020b): "Satisficing, aggregation, and quasilinear utility," *Available at SSRN 3180302*.

Banerjee, A. V., S. Chassang, S. Montero, and E. Snowberg (2020): "A theory of experimenters: Robustness, randomization, and balance," *American Economic Review*, 110, 1206–30.

Barten, A. P. (1969): "Maximum likelihood estimation of a complete system of demand equations," *European economic review*, 1, 7–73.

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

Brown, D. J. and C. Calsamiglia (2007): "The nonparametric approach to applied welfare analysis," *Economic Theory*, 31, 183–188.

Cerreia-Vioglio, S., D. Dillenberger, P. Ortoleva, and G. Riella (2019): "Deliberately Stochastic," *American Economic Review*, 109, 2425–2445.

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*, Cambridge University Press.

Deaton, A. and J. Muellbauer (1980): "An almost ideal demand system," *American Economic Review*, 70, 312–326.

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

Debreu, G. (1974): "Excess demand functions," *Journal of Mathematical Economics*, 1, 15–21.

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

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.

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

Feldman, P. and J. Rehbeck (2020): "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, 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, 2371–2409.

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

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

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, 245–264.

Pesaran, M. H. and A. S. Deaton (1978): "Testing non-nested nonlinear regression models," *Econometrica*, 677–694.

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

Rockafellar, R. T. (1970): *Convex analysis*, Princeton university press.

Smeulders, B., L. Cherchye, and B. De Rock (2019): "Nonparametric analysis of random utility models: Computational tools for statistical testing," *FEB Research Report Department of Economics DPS19. 08*.

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

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

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

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

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