A Frequentist Approach to Revealed Preference Analysis*

Charles Gauthier† Raghav Malhotra‡ Agustín Troccoli Moretti§

February 7, 2026

Abstract

This paper develops a framework to study the statistical power of revealed-preference tests. With randomly sampled budgets and mild smoothness of demand, statistical learning implies that any model consistent with the data must approximate true choice behaviour. We interpret this result as follows: passing a revealed-preference test is informative only to the extent that the data are sufficiently rich to rule out economically meaningful departures from the maintained model. We make this precise by linking sample size and confidence to the magnitude of detectable departures, and by characterising how power rises with additional observations. Extending our approach beyond revealed-preference inequalities to smooth functional restrictions yields practical tests, even when exact revealed-preference tests are computationally infeasible. We also provide confidence intervals for smooth functionals of demand, including welfare effects. Simulations show that standard sample sizes can generate widely different power across models, contextualizing why some conditions "rarely reject" in practice.

Keywords: revealed preference, finite data, preferences, hypothesis testing
JEL classification: C12, C91, D0, D11, D12

*The authors thank Herakles Polemarchakis, Costas Cavounidis, Robert Akerlof, Roy Allen, Matt Polisson, Victor Aguiar, Motty Perry, Daniele Condorelli, Ben Lockwood, Sharun Mukand, Carlo Perroni, and Ludovic Renou for helpful discussions, comments, and encouragement, as well as audiences at CRETA (Warwick) and the Bristol Econometric Study Group Conference 2025. Any remaining errors, as well as the opinions expressed herein, are our own.

†Universitat de Barcelona; charles.gauthier@ub.edu
‡University of Leicester; r.malhotra@leicester.ac.uk
§Universitat Pompeu Fabra and BSE; agustin.troccoli@upf.edu

1 Introduction

Analysts often seek to verify whether a decision maker (DM) is consistent with a given model, using finite choice data, to ensure that policy predictions about behavior and welfare are reliable. A common approach is revealed preference (RP) analysis, which provides a set of exhaustive inequalities that are satisfied if and only if the DM's choices are consistent with a given class of preferences.[1] RP tests typically assume that data consist of a finite number of non-stochastic observations. A subtle issue arises when these observations pass the RP test. Namely, the test cannot distinguish between a DM who is truly consistent with the model and one who is not, but for whom there were insufficient data to detect a deviation.

The frequentist approach to addressing this problem in the context of hypothesis testing focuses on statistical power, the probability of rejecting a decision maker's choices under the assumption that the DM does not adhere to the model. This approach requires constructing parsimonious alternative hypotheses. However, as noted by Adams and Crawford [2015], *"The difficulty is that there are many alternatives to rational choice models and no obvious benchmark."* In other words, there is no straightforward way to assess the power of RP tests.[2] This paper shows how results from statistical learning theory can be used to circumvent this limitation and enable power analysis.

To illustrate the problem, suppose an analyst observes a consumer making grocery purchases on ten different occasions, each time facing different prices. The consumer's choices satisfy all the revealed preference conditions—there are no apparent "mistakes". Should the analyst conclude that the consumer is rational? Not necessarily. Ten observations may simply be too few to detect irrationality, even if it is present. With only ten price–consumption pairs, many non-rational choice patterns can slip through undetected, much as a student who answers randomly on a ten-question multiple-choice exam might still pass by chance.

This raises a natural question: how many observations n are needed to detect with high probability a consumer who deviates from rationality by an amount ε? Standard revealed preference analysis cannot answer this question: it tells us whether observed choices could have come from a rational consumer, not whether the available data are

[1] Following the terminology of Varian [1983], we refer to a set of inequalities as exhaustive if they are necessary and sufficient for the data set to pass the RP test.

[2] For example, Bronars [1987] proposes a power test based on the assumption that choices are uniformly random on the budget frontier.

sufficient to reliably detect deviations.

Our paper fills this gap using a simple idea. Suppose we observe a consumer's choices at n randomly selected price points. Because demand functions are smooth (satisfying a Lipschitz condition), any function that fits these observations must be close to the consumer's actual choice function everywhere—not just at the observed points—once n is sufficiently large. Consequently, if the true choice function is far from rational, so will any function consistent with the data, and the revealed preference test will reject rationality. If the consumer is rational, the test will never reject. By quantifying the number of observations needed as a function of the detection threshold ε, we show that explicit power guarantees are possible.

Formally, we consider an economic model $g \in \mathcal{G}$, where $\mathcal{G}$ is a class of models satisfying a Lipschitz condition, which maps an exogenous variable (prices) into an endogenous variable (demand). The analyst observes exogenous prices p_t and demand x_t related by $x_t = g(p_t)$. Consider some subclass $\mathcal{F} \subset \mathcal{G}$ and define the ball of size $\varepsilon > 0$ around $\mathcal{F}$ as

$$\mathcal{F}(\varepsilon) := \Big\{ g \in \mathcal{G} : \inf_{g' \in \mathcal{F}} d(g, g') \leq \varepsilon \Big\}.$$

Suppose the analyst can sample $\{p_t\}_{t=1}^n$ independently at random. This generates a dataset of the form

$$D_n = \{(p_k, g(p_k))\}_{k=1}^n. \tag{1}$$

Note that although g is non-stochastic, the data are stochastic because prices are sampled at random. This aligns with the frequentist tradition in which the state of the world is deterministic and all stochasticity arises from the sampling procedure.

For given $\delta, \varepsilon > 0$, we aim to find a computable function $n(\delta, \varepsilon)$ such that, whenever a dataset has more than $n(\delta, \varepsilon)$ observations, one can construct tests of size zero and power greater than δ to distinguish between the null hypothesis $\mathbf{H}_0$ and the alternative $\mathbf{H}_1$:

$$\mathbf{H}_0 : g \in \mathcal{F} \quad \text{vs.} \quad \mathbf{H}_1 : g \in \mathcal{G} \setminus \mathcal{F}(\varepsilon). \tag{2}$$

We show that if the class $\mathcal{G}$ is Lipschitz and the analyst has access to a revealed preference characterization of $\mathcal{F}$, such tests and function $n(\delta, \varepsilon)$ can be constructed for any choice of ε. Furthermore, we show that when the dataset size n is fixed, one can compute a lower bound on the power $\delta(n, \varepsilon)$ and a lower bound on the closeness of preferences from the model $\varepsilon(n, \delta)$.

We then generalize our main results by studying cases where the analyst does not

have access to RP characterizations, but rather some functional restriction $\mathcal{R} : \mathcal{G} \to \mathbb{R}$ of $\mathcal{F}$ such that:

$$\mathcal{R}(g) = 0 \iff g \in \mathcal{F}.$$

Under some regularity conditions on $\mathcal{R}$, we can conduct the same analysis as above by constructing

$$\mathcal{F}(\varepsilon) = \{g' \in \mathcal{G} : \mathcal{R}(g) \leq \varepsilon\},$$

and then forming the hypothesis test in the manner described in (2). In that case, the size of the test is not zero, but some computable function of $\mathcal{G}$, $\mathcal{R}$, and $\mathcal{F}$.

Our method applies to preference classes where revealed preference characterizations are unknown or computationally infeasible. For Lipschitz demands, we find that the computational issues of several RP tests stem mostly from their knife-edge nature which have size zero. We demonstrate that this approach can be used to construct confidence intervals for smooth functionals of demand, such as changes in cost-of-living indices. Finally, we show how to integrate additional assumptions on the underlying preference classes and adaptive sampling to improve power and confidence intervals.

We emphasize that the results in this paper are primarily existence results that establish what can be learned from finite choice data, rather than providing readily applicable procedures. Our theorems characterize the sample sizes sufficient to achieve desired power levels and the precision attainable for welfare inference, but operationalizing these bounds requires knowledge of quantities—such as the Lipschitz constant of demand and the sampling distribution of prices—that may be difficult to determine in practice. We view our contribution as providing the theoretical foundation for power analysis in revealed preference settings; developing practical implementations to estimate or bound these quantities from data is an important direction for future work.

Section 2 describes our setup and notation. Section 3 shows how to construct our tests with fixed power, sample size, and distance from the model (Theorems 3.1–3.3), followed by power simulations for RP tests. Section 4 extends our testing results to functional characterizations (Theorem 4.1) and shows how to conduct inference on smooth estimands such as welfare changes (Theorem 4.2). The section also provides functional characterizations of common classes of preferences, such as substitutes and complements, and of choice under uncertainty (Propositions 1–5). Section 5 studies extensions of our approach. Section 6 concludes.

1.1 Related Literature

The RP approach tests the consistency of a finite dataset with utility maximization. The approach originated with Samuelson [1938] and was refined by Afriat [1967] and Varian [1982]. In particular, Afriat [1967] established that the Generalized Axiom of Revealed Preference (GARP) is a necessary and sufficient condition for any finite dataset to be consistent with utility maximization. The appeal of GARP stems from the ease with which the rationality of a dataset can be verified through simple algebraic inequalities.

The idea of constructing an alternative against which to assess the power of revealed-preference (RP) tests dates back to Becker [1962], who considered consumers choosing randomly on their budget lines. Building on this, Bronars [1987] examined how often RP tests would detect violations when choices were generated uniformly at random. More recently, Andreoni et al. [2013] proposed a more realistic benchmark by drawing from the observed distribution of choices, thereby calibrating the alternative hypothesis to match empirical patterns.

Because a parsimonious alternative to fully rational choice is hard to specify, several papers assess predictive power without positing an explicit behavioural model. Andreoni et al. [2013] define the Afriat Power Index for datasets with no revealed-preference violations. Reporting the smallest budget perturbation needed to induce a violation—small adjustments indicate a sensitive test, large adjustments a weak one. A related approach, based on Selten [1991] and applied by Beatty and Crawford [2011], considers a method which uses the set of all theory-consistent behaviors with respect to all possible behaviors.

Since those seminal contributions, RP theory has explored various extensions, including functional-form restrictions and intertemporal models. Significant contributions include Kubler, Selden, and Wei [2014] for expected utility with objective (known) probabilities and Echenique and Saito [2015] for subjective probabilities. A similar problem for "translation-invariant" preferences is considered in Chambers, Echenique, and Saito [2016]. Polisson, Quah, and Renou [2020] give a general method to construct RP inequalities that applies to several classes of preferences over risk and uncertainty.[3] We take the set of Lipschitz demands as the alternative, thus covering these subclasses and rationality itself.

[3]For recent extensive reviews, see Crawford and De Rock [2014], Echenique [2019], and Demuynck and Hjertstrand [2019].

The functional approach for testing decision-makers [Slutsky, 1915, Antonelli, 1971, Deaton and Muellbauer, 1980] relies on the entire demand function and assumes access to infinite data. We build on this literature by extending these tests to finite data. In this context, our work is related to Aguiar and Serrano [2018], who adapt the measure of bounded rationality from Aguiar and Serrano [2017] to finite data. While their focus is on measuring and classifying bounded rationality, our approach provides tests for general restrictions under preference monotonicity, making it applicable to a broader range of datasets, including pairwise comparisons.

Papers on extrapolating fundamentals from finite data use regularity conditions similar to ours. Chambers et al. [2021] imposes a structural restriction on the space of preferences that is closely related to Lipschitz continuity, but it allows sample complexity to depend on the underlying (unknown) preference, which we rule out by construction. Related Lipschitz-type upper bounds for utilities are also used in Chambers et al. [2023] to deliver finite-sample guarantee results. Finally, Chambers and Echenique [2025] shows that economic models that approximately satisfy axioms can be rationalized by utilities close to those that exactly rationalize the model, providing microfoundations for using our approach to test axiomatic decision models.

The RP literature focuses on exhaustive restrictions in finite datasets that characterize a model.[4] In some instances, those tests can pose a significant computational challenge. Indeed, Echenique [2014] and Cherchye, Demuynck, De Rock, and Hjertstrand [2015] show that testing for weak separability is NP-Hard. Accordingly, these tests rely on non-polynomial-time algorithms, such as mixed-integer programming [Cherchye, Demuynck, De Rock, and Hjertstrand, 2015, Hjertstrand, Swofford, and Whitney, 2020]. We show that the computational issues can be mitigated by permitting tests with nonzero size.

The deterministic RP approach has recently been extended to stochastic environments by Kitamura and Stoye [2018] and Aguiar and Kashaev [2021]. In both cases, RP analysis is embedded within a random utility framework, although the latter treats measurement error as the source of randomness. By contrast, in our framework, all randomness arises from the sampling procedure rather than from the decision-maker. Consequently, our approach is closer in spirit to standard RP tests conducted on individual-level data, while still treating individual observations as random.[5]

[4]The only revealed preference test that is not a characterization which we are familiar with is for probabilistic sophistication [Epstein, 2000].

[5]In a different direction, Allen et al. [2024] propose a statistical consumer model who chooses

2 Setup

We consider a DM whose choices are observed a finite number of times. Let $Y \subset \mathbb{R}^{\ell}_{\geq 0}$ be a compact and convex consumption space, where ℓ denote the number of goods. A DM faces prices $p \in \mathcal{P}$ and income $I \in \mathcal{I}$, where $\mathcal{P}$ is a compact subset of $\mathbb{R}^{\ell}_{>0}$ and $\mathcal{I} = [a, b] \subset \mathbb{R}_{>0}$. We denote the budget corresponding to a tuple (p, I) by

$$B(p, I) = \Big\{ y \in Y : p \cdot y \leq I \Big\},$$

and denote the set of all budget sets by $\mathcal{D}$. The DM has a choice function $x(p, I)$ with $x(p, I) \in B(p, I)$. Let $\mathcal{G}$ denote the class of all choice functions $x : \mathcal{P} \times \mathcal{I} \to Y$ that satisfy the following Lipschitz property: there exists a constant $L > 0$, known to the analyst, such that for all $(p, I), (p', I') \in \mathcal{P} \times \mathcal{I}$,

$$\|x(p, I) - x(p', I')\| \leq L\, d\big((p, I)^{\top}, (p', I')^{\top}\big),$$

where $\|\cdot\|$ is the Euclidean norm and d is the Euclidean distance.

In several applications, we restrict attention to choice functions arising from utility maximisation. A preference relation $\succeq \subseteq Y \times Y$ is *rational* if it is complete and transitive. We further assume that preferences are continuous, strictly monotone and strictly convex.[6] Given a rational preference $\succeq$, let $u_{\succeq} : Y \to \mathbb{R}$ be a (continuous) utility representation and define the associated Marshallian demand by

$$x_{\succeq}(p, I) \;:=\; \underset{y \in B(p,I)}{\arg\max}\, u_{\succeq}(y).$$

Under strict convexity this maximiser is unique, so $x_{\succeq}$ is single-valued. Accordingly, we say that a choice function $x \in \mathcal{G}$ is an *admissible rational demand* if there exists a rational preference $\succeq$ such that $x = x_{\succeq}$. In what follows, we use the term *choice function* for an arbitrary element of $\mathcal{G}$, and the term *demand function* for an admissible rational demand.

A dataset consists of budget-choice pairs $D_n = \{(B_k, x_k)\}_{k=1}^{n}$. A choice function x

a distribution of demands given prices. They show that standard RP conditions apply to average demands under a weak mean-expenditure condition, so our method should extend to their framework as well.

[6] Strict monotonicity means that for any $x, y \in Y$, if $x_i \geq y_i$ for all $i = 1, \ldots, \ell$ and $x \neq y$, then $x \succ y$. Strict convexity means that for any $x \neq y$ in Y and any $\alpha \in (0, 1)$, $x \succeq y$ implies $\alpha x + (1 - \alpha) y \succ y$.

rationalizes a dataset D_n if it exactly reproduces the observed choices $x_k = x(B_k)$ for all $k = 1, \ldots, n$.

Let $g \in \mathcal{G}$ denote the true choice function, which we may refer to as the underlying model. The analyst draws n price–income pairs independently at random from some distribution $\mu \in \Delta(\mathcal{P} \times \mathcal{I})$, generating a dataset

$$D_n(x^*) \;=\; \big\{(B_k,\, g(B_k))\big\}_{k=1}^{n}.$$

Let μ^n denote the product measure on $(\mathcal{P} \times \mathcal{I})^n$.[7] Since $g \in \mathcal{G}$ is fixed, μ^n induces a distribution μ_g^n on datasets of size n by mapping each sampled sequence $\big((p_k, I_k)\big)_{k=1}^{n}$ to its corresponding budget-choice sequence $\big(B_k, g(B_k)\big)_{k=1}^{n}$. By treating the dataset as random draws from a fixed model, we can study the size and power of tests over repeated sampling. The timing of this conceptual framework is illustrated in Figure 1.

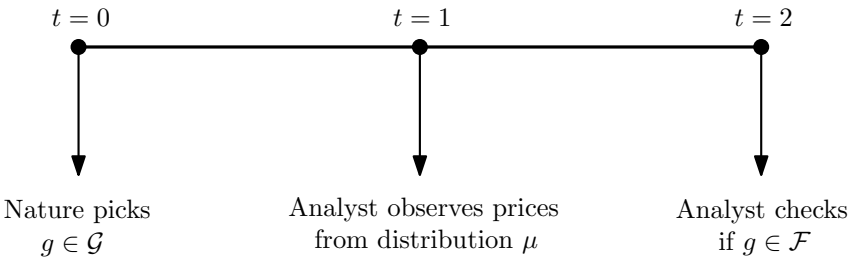


Figure 1: Depiction of the timing in our framework.

Let $\mathbf{H}_0$, $\mathbf{H}_1 \subset \mathcal{G}$ denote the null and alternative hypotheses. We have $g \in \mathbf{H}_0$ if the null hypothesis holds and $g \in \mathbf{H}_1$ if the alternative hypothesis holds. A *test* is a measurable function $\mathcal{T} : \mathcal{D} \to \{0, 1\}$ that takes as input a dataset and returns whether the null hypothesis should be rejected, where by convention $\mathcal{T}(D) = 1$ means the null is rejected and $\mathcal{T}(D) = 0$ means the null is not rejected.

Definition 1 (Size and Power)**.** Let $\mathcal{T}$ be a test and $\mathcal{T}^{-1}(1)$ denote the set of datasets rejected by $\mathcal{T}$.

(i) The *size* of $\mathcal{T}$ is the largest probability of incorrectly rejecting when the null is

[7] Formally, if $\mu \in \Delta(\mathcal{P} \times \mathcal{I})$ is the sampling distribution of price–income pairs, then μ^n is the product measure on $(\mathcal{P} \times \mathcal{I})^n$, defined for any measurable set $A \subseteq (\mathcal{P} \times \mathcal{I})^n$ as

$$\mu^n(A) \;=\; \int \mathbf{1}_A\big((p_1, I_1), \ldots, (p_n, I_n)\big)\, \mathrm{d}\mu(p_1, I_1) \cdots \mathrm{d}\mu(p_n, I_n).$$

true:

$$Size(n) \;=\; \sup_{x \in \mathbf{H}_0} \mu_x^n\big(\mathcal{T}^{-1}(1)\big)\,.$$

(ii) The *power* of $\mathcal{T}$ is the probability of correctly rejecting the null when the alternative is true:

$$Power(n; g) \;=\; \mu_g^n\big(\mathcal{T}^{-1}(1)\big) \quad \text{for a given } g \in \mathbf{H}_1.$$

The distinction between size and power lies in the domain over which probabilities are evaluated: size is defined uniformly over all models in the null, taking the supremum over $\mathbf{H}_0$, while power is evaluated at a particular alternative $g \in \mathbf{H}_1$. For brevity, we will henceforth omit the explicit dependence of μ_g^n on g whenever there is no risk of confusion, and simply write μ^n.

Definition 2 (RP Characterization)**.** Let $\mathcal{F} \subset \mathcal{G}$ be a class of preferences. An *RP characterization* $\mathcal{T}^*$ of $\mathcal{F}$ is a test with the following property:

(i) $\mathcal{T}^*$ never rejects a dataset rationalisable by some $x \in \mathcal{F}$ (size zero).

(ii) For any other test $\mathcal{T}$ satisfying (i), if $\mathcal{T}$ rejects a dataset D_n, then $\mathcal{T}^*$ also rejects D_n (maximal power).

In words, an RP characterization never rejects a dataset consistent with $\mathcal{F}$, and has maximal power among all such size-zero tests.

3 Finite-Sample Theory

This section develops finite–sample power bounds that connect the revealed preference framework to the statistical theory of PAC learning. We proceed in three steps. First, we determine the required sample size to achieve a specified power and accuracy level, which may inform experimental design. Second, for a given sample size and accuracy, we derive the attainable power level, helping to gauge the effectiveness of RP tests. Finally, we characterize the minimal separation from $\mathcal{G}$ that is detectable given a sample size and desired power.

3.1 Preference Learnability

This subsection introduces the notion of PAC learnability and states the main result we use as a building block for our procedure. First, define the generalization error between two choice functions x, $x' : \mathcal{P} \times \mathcal{I} \to Y$ as:

$$\mathsf{erf}_\mu(x, x') = \int_{\mathcal{P}\times\mathcal{I}} \|x(p, I) - x'(p, I)\| \, \mathrm{d}\mu,$$

where $\|\cdot\|$ denotes the Euclidean norm. The next definition introduces the concept of PAC learnability.

Definition 3 (PAC Learnability)**.** A class of demand functions $\mathcal{G}$ is *probably approximately correct* (PAC) learnable by a hypothesis class $\mathcal{H}$ if the following holds. For any ε, $\delta > 0$, any choice function $x \in \mathcal{G}$, and any distribution $\mu \in \Delta(\mathcal{P} \times \mathcal{I})$, there exists an algorithm L such that, given a sample of size

$$m_L = m_L(\varepsilon, \delta) = \mathrm{Poly}\left(\tfrac{1}{\varepsilon}, \tfrac{1}{\delta}\right),$$

the algorithm returns a hypothesis $h \in \mathcal{H}$ satisfying $\mathsf{erf}_\mu(x, h) < \varepsilon$ with probability at least $1 - \delta$.[8]

In words, $\mathcal{G}$ is PAC learnable by $\mathcal{H}$ if there exists an algorithm that can approximate every demand function in $\mathcal{G}$ by some hypothesis $h \in \mathcal{H}$ within error ε, with probability at least $1 - \delta$. Here, a hypothesis h corresponds to a candidate demand function approximating the true choice function. While in the general learning-theory literature the hypothesis class $\mathcal{H}$ may differ from the target class $\mathcal{G}$, we restrict our attention to the case $\mathcal{H} = \mathcal{G}$. We can now state a key result for our derivations.

Lemma 1. *Let $\mathcal{G}$ be the class of choice functions over the compact set $\mathcal{P} \times \mathcal{I}$ with Lipschitz constant L. Then $\mathcal{G}$ is learnable by hypothesis class $\mathcal{H} = \mathcal{G}$ with sample complexity:*

$$m_L(\varepsilon, \delta) = O\left(\frac{1}{\varepsilon^2}\left(\ln^2\left(\frac{1}{\varepsilon}\right) \cdot \left(\frac{L}{\varepsilon}\right)^{\ell+1} + \ln\left(\frac{1}{\delta}\right)\right)\right).$$[9]

[8] We write $\mathrm{Poly}(1/\varepsilon, 1/\delta)$ to indicate that the required sample size is bounded by a polynomial in $1/\varepsilon$ and $1/\delta$.

[9] If, in addition, the choice rules are homogeneous of degree zero in (p, I) and I is bounded away from zero, we may normalise by income and obtain the same bound with exponent ℓ.

This lemma guarantees that, for any distribution μ, the true demand function can be approximated within any pre-specified accuracy ε, with confidence at least $1-\delta$, from a sample size that grows only polynomially in $1/\varepsilon$ and $1/\delta$. Our proof follows from a simple extension of Beigman and Vohra [2006], who prove that Lipschitz demand functions are learnable under the above sample guarantees. However, their proof relies on a simple packing-number argument that does not use the fact that the hypothesis class is assumed to consist of demand functions. They also provide a constructive algorithm that achieves this bound, thereby delivering both sample efficiency and polynomial-time computation.[10]

For each (ε, δ), let $n(\varepsilon, \delta)$ denote the *minimal* sample size such that there exists *some* learning algorithm (possibly depending on (ε, δ)) that (ε, δ)–learns $\mathcal{G}$ in the sense of Definition 3. In particular, for any specific algorithm L satisfying Definition 3 with sample size $m_L(\varepsilon, \delta)$, we have $n(\varepsilon, \delta) \leq m_L(\varepsilon, \delta)$. Moreover, $n(\varepsilon, \delta)$ is weakly decreasing in ε for each fixed δ.

3.2 Finite–Sample Power Bounds

Our first result presents the "forward" formulation that characterises the sample size required to ensure that an RP test has power at least $1-\delta$ against all alternatives that are ε–separated from the null class.

Theorem 3.1. *Let $\mathcal{F} \subset \mathcal{G}$ be a class of models-, and suppose the analyst has access to an RP characterization $\mathcal{T}_{\mathcal{F}}$ for $\mathcal{F}$. Then, for every ε, $\delta > 0$, if the sample size n satisfies $n \geq n(\varepsilon, \delta)$,[11] the test based on $\mathcal{T}_{\mathcal{F}}$ applied to a dataset of size n has power at least $1-\delta$ for the hypotheses*

$$\mathbf{H}_0 : x^* \in \mathcal{F} \qquad \textit{vs.} \qquad \mathbf{H}_1 : x^* \in \mathcal{G} \setminus \mathcal{F}(\varepsilon),$$

where

$$\mathcal{F}(\varepsilon) \;:=\; \left\{ x \in \mathcal{G} : \inf_{x' \in \mathcal{F}} \mathsf{erf}_\mu(x, x') \;\leq\; \varepsilon \right\}.$$

In words, this result guarantees that the test rejects any demand strictly more than ε away from $\mathcal{F}$ with probability at least $1-\delta$ if the analyst collects any sample

[10] Kubler, Malhotra, and Polemarchakis [2020] show that when the sampling distribution μ is uniform over $\mathcal{P}$, the assumption of a known Lipschitz constant can be relaxed, but they do not provide an explicit procedure to compute the relevant constants.

[11] In particular, by Lemma 1 it suffices to take $n = m_L(\varepsilon, \delta)$.

size of $n \geq n(\varepsilon, \delta)$ observations.[12] Here $n(\varepsilon, \delta)$ is the minimal sample complexity, and it suffices to take $n = m_L(\varepsilon, \delta)$. In particular, it is worth noting that Theorem 3.1 allows to study power of rational preferences against an alternative Lipschitz demand. Indeed, one can simply take $\mathcal{F}$ to be the set of all choice functions that arise from utility maximisation, and the RP characterization for this class reduces to the standard Afriat inequalities. The geometric structure of the argument in Theorem 3.1 is conveyed in Figure 2.

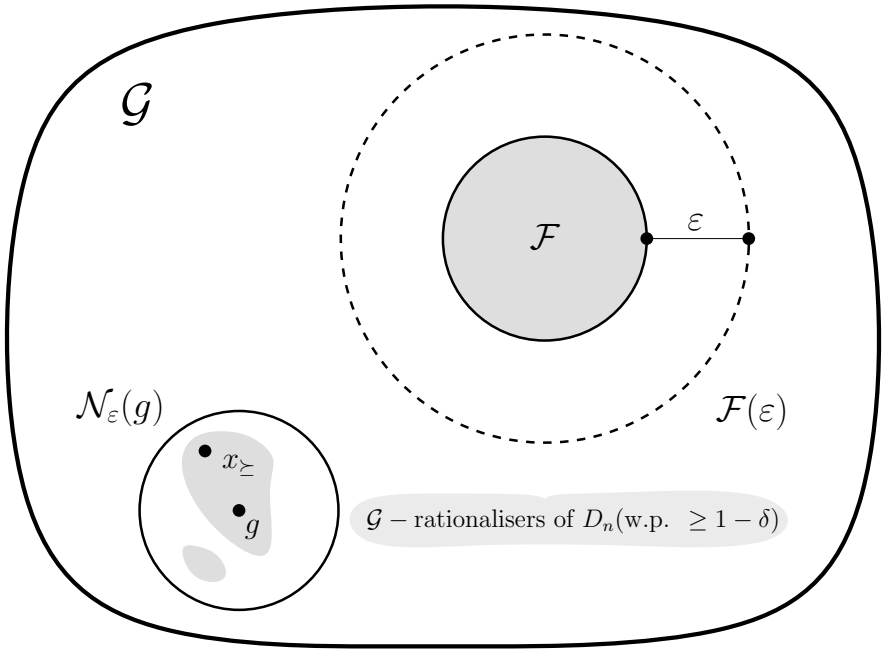


Figure 2: Geometric sketch of the proof. When $n \geq n(\varepsilon, \delta)$, with probability at least $1-\delta$, the $\mathcal{G}$-rationalisers of the dataset D_n are contained in the ε-neighbourhood $\mathcal{N}_\varepsilon(g)$ of the true model (shaded region). Under the alternative $g \notin \mathcal{F}(\varepsilon)$, this neighbourhood is disjoint from $\mathcal{F}$. Hence no element of $\mathcal{F}$ can rationalise the data, and an RP characterisation rejects with probability at least $1-\delta$.

There, the set of $\mathcal{G}$-rationalising demands is represented by the neighbourhood $\mathcal{N}_\varepsilon(g)$. When $n \geq n(\varepsilon, \delta)$, with probability at least $1-\delta$, every $\mathcal{G}$-rationaliser of the observed dataset lies in $\mathcal{N}_\varepsilon(g)$. Under the alternative hypothesis, the true model satisfies $g \notin \mathcal{F}(\varepsilon)$, equivalently $\inf_{f \in \mathcal{F}} \mathsf{erf}_\mu(g, f) > \varepsilon$, so $\mathcal{N}_\varepsilon(g) \cap \mathcal{F} = \emptyset$. Consequently, with probability at least $1-\delta$, no element of $\mathcal{F}$ can rationalise the observed dataset. Since an RP characterisation rejects precisely when no element of $\mathcal{F}$ rationalises the

[12] The distance $\mathsf{erf}_\mu(\cdot, \cdot)$ is computed using the same distribution μ that governs the sampling of price–income pairs. Thus, both the measure of approximation error and the test are based on the same underlying data distribution.

data, it follows that the test rejects under the alternative with probability at least $1-\delta$.

While Theorem 3.1 provided the forward bound, the next result states the corresponding inverse bound: given a fixed sample size n and tolerance ε, it determines the confidence level $1-\delta(\varepsilon, n)$ that can be guaranteed. The difference is only in which parameters are treated as primitive, so the proofs mirror each other. To avoid trivialities, for fixed $\varepsilon > 0$ and $n \in \mathbb{N}$ we assume that there exists some $\delta \in (0,1)$ such that $n(\varepsilon, \delta) \leq n$.

Theorem 3.2. *Suppose the minimal sample complexity of (ε, δ)-learning $\mathcal{G}$ is denoted by $n(\varepsilon, \delta)$. For each $\varepsilon > 0$ and $n \in \mathbb{N}$, define*

$$\delta(\varepsilon, n) \;:=\; \inf\{\delta \in (0,1) : n(\varepsilon, \delta) \leq n\}.$$

Let $\mathcal{T}$ be an RP characterisation of $\mathcal{F}$. Then, for any alternative $g \in \mathcal{G}$ with

$$\inf_{f \in \mathcal{F}} \mathsf{erf}_\mu(g, f) \;\geq\; \varepsilon,$$

we have

$$\mu_g^n\big(\,\mathcal{T}^{-1}(1)\,\big) \;\geq\; 1 - \delta(\varepsilon, n).$$

In words, at tolerance ε and sample size n, the test $\mathcal{T}$ has power at least $1-\delta(\varepsilon, n)$ against all alternatives at distance at least ε from $\mathcal{F}$.

Algebraically, Theorem 3.2 amounts to inverting the sample–complexity map: instead of choosing n large enough to guarantee a target confidence level $1-\delta$, we fix (ε, n) and identify the *best* confidence level compatible with n, namely $1-\delta(\varepsilon, n)$, where $\delta(\varepsilon, n)$ is the smallest $\delta \in (0,1)$ such that $n(\varepsilon, \delta) \leq n$. Geometrically, the idea is the same as in Figure 2: with probability at least $1-\delta(\varepsilon, n)$, all $\mathcal{G}$-rationalisers of the dataset lie in the ε–ball around the true model, and if this ball is disjoint from $\mathcal{F}$ the test rejects.

Having stated the forward bound (fix (ε, δ), solve for n) and its inverse (fix (ε, n), solve for $\delta(\varepsilon, n)$), it is natural to ask the complementary question: given a sample size n and confidence level $1-\delta$, what separation ε from $\mathcal{F}$ can the test reliably detect? The next result answers this by introducing the minimal detectable separation scale $\varepsilon(n, \delta)$. To avoid trivialities, we fix $n \in \mathbb{N}$ and $\delta \in (0,1)$ such that the set $\{\varepsilon > 0 :\; n(\varepsilon, \delta) \leq n\}$ is nonempty.

Theorem 3.3. *Let $n(\varepsilon, \delta)$ denote the minimal sample complexity of (ε, δ)-learning $\mathcal{G}$. For $n \in \mathbb{N}$ and $\delta \in (0, 1)$, define*

$$\varepsilon(n, \delta) \;:=\; \inf\{\varepsilon > 0 :\; n(\varepsilon, \delta) \leq n\}.$$

Let $\mathcal{T}$ be an RP characterisation of $\mathcal{F}$. Then, for every $\gamma > 0$ and every alternative $g \in \mathcal{G}$ with

$$\inf_{f \in \mathcal{F}} \mathsf{erf}_\mu(g, f) \;\geq\; \varepsilon(n, \delta) + \gamma,$$

we have

$$\mu_g^n\big(\mathcal{T}^{-1}(1)\big) \;\geq\; 1 - \delta.$$

In words, for any slack $\gamma > 0$, at confidence level $1 - \delta$ and sample size n, the test $\mathcal{T}$ has power at least $1 - \delta$ against all alternatives at distance at least $\varepsilon(n, \delta) + \gamma$ from $\mathcal{F}$.

The argument mirrors the proofs of Theorems 3.1 and 3.2 and relies on the same geometry as in Figure 2. The key difference is that, for fixed (n, δ), the boundary tolerance $\varepsilon(n, \delta)$ is defined as an *infimum* and need not be attained. Accordingly, the result is stated with an arbitrary slack $\gamma > 0$.

Conceptually, Theorem 3.3 turns the sample–complexity map on its head: for a given sample size n and confidence level $1 - \delta$, it identifies a detectability frontier in terms of separation from $\mathcal{F}$. Any alternative whose distance from $\mathcal{F}$ exceeds $\varepsilon(n, \delta)$ by a strictly positive margin γ is guaranteed to be detected with probability at least $1 - \delta$ by any RP characterisation of $\mathcal{F}$. Thus, $\varepsilon(n, \delta)$ summarises the finite-sample resolution of the frequentist revealed-preference test: larger samples push the frontier down, while higher confidence shifts it up.

Remark. Notice that the power computation depends only on the sample complexity of PAC learning for $\mathcal{G}$. Appendix 5 shows this can be exploited to improve power by presenting methods that accelerate learning rates.

Remark. It is easy to show that if observed demands are subject to additive mean-zero measurement error with finite variance, then the class of learnable demand functions $\mathcal{G}$ remains learnable [Bartlett et al., 1994]. Consequently, our previous results remain valid under additive mean-zero measurement error.

3.3 Power Analysis

In what follows, we study the finite-sample power of revealed preference tests for static utility maximization (SUM), homotheticity (H), weak separability (WS), and expected utility (EU) against smooth violations of integrability. We consider a setup with four goods and, in the EU case, two equally probable states of the world. For the tests, we follow Varian [1983] and Green and Srivastava [1986]. For weak separability, we only use necessary conditions since exact tests based on mixed-integer programming are NP-complete [Cherchye et al., 2015].

We consider simulated demands given by $y(p, I) = x(p, I) + \varepsilon f(p, I)$, where $x(p, I)$ are Cobb-Douglas demands, $\varepsilon > 0$ controls the deviation from rationality, and $f(p, I)$ is a perturbation that preserves homogeneity of degree zero and budget neutrality, but violates Slutsky symmetry.[13] We normalize f such that the Frobenius norm of the skew-symmetric part of the Slutsky matrix equals one, ensuring comparability of ε across alternative hypotheses. See Appendix C for full details.

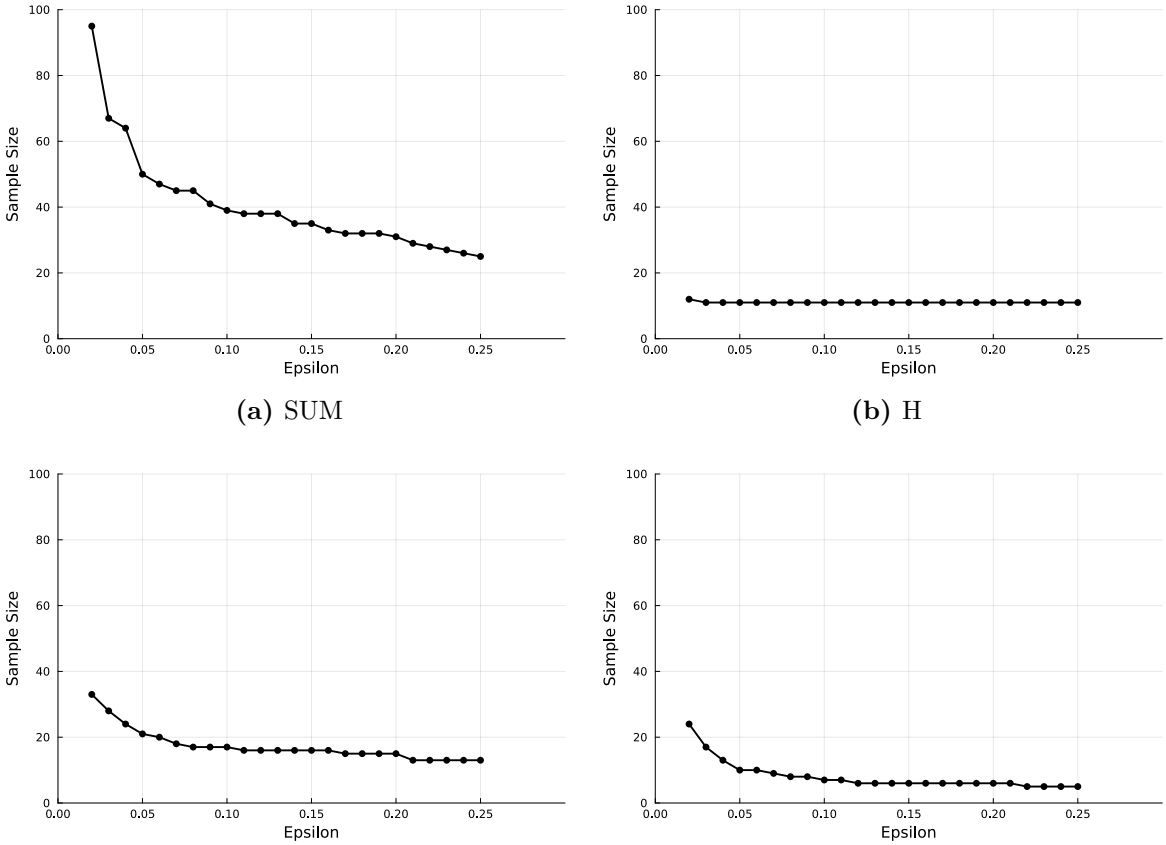
We simulate 500 perturbed demands from Cobb-Douglas preferences drawn uniformly from the unit interval and scaled such as to sum up to one. Prices and income are drawn uniformly from $(0.01, 10.00)$ and $(1.00, 6.00)$, respectively. This level of price variation exceeds what is typically observed in empirical datasets. As a result, the RP tests are given favorable conditions for detecting deviations from rationality. For each value of $\epsilon \in [0.02, 0.03, \ldots, 0.25]$, we compute the smallest sample size required to achieve a rejection probability of at least 90%. The resulting power curves are reported in Figure 3.[14]

The figure shows that for small deviations from rationality, SUM and WS require large sample sizes to achieve the target rejection rate. In contrast, H and EU only require modest sample sizes to achieve the target rejection rate. For large deviations from rationality, every RP test achieves the target rejection rate with small to moderate sample sizes.

Next, we summarize the relationship between the distance from rationality and the

[13] Although the perturbed demands may be negative, revealed preference theory still applies [Deb et al., 2023], so they are inconsequential for our simulations.

[14] We obtain qualitatively similar power curves with an alternative data generating process featuring a different pattern of deviations from Slutsky symmetry, indicating that the relative ranking of tests is robust across alternative hypotheses.



(a) SUM **(b)** H

(c) WS **(d)** EU

Figure 3: Power curves for SUM, H, WS, and EU.

sample size needed to achieve 90% power using a log-log regression:

$$\log(\overline{n}^m) = \beta_0^m + \beta_1^m \log \varepsilon^m + \omega^m,$$

where $\overline{n}^m$ is the smallest sample size needed to achieve 90% power for model $m \in \{\text{SUM}, \text{H}, \text{WS}, \text{EU}\}$, ϵ^m is the distance from rationality, and ω^m captures residual variation due to the simulation process. The intercepts (β_0) reflect baseline differences in the sample size needed to reach a given power level when the deviation from rationality is one. The slopes (β_1) reflect the elasticity of the required sample size with respect to deviations from rationality. Notice that by Lemma 1, fixing δ to 0.1, the order of β_1 should be smaller than $2 + l$. We find this empirically, and the results are presented in Table 1.[15]

[15]Standard errors are omitted because the estimates are derived from Monte Carlo simulations rather than empirical sampling. We show two further regressions in Appendix C for completeness.

Table 1: Log-log regression estimates from power curves

Parameter	SUM	H	WS	EU
β_0	2.6448	2.3719	2.0862	0.7681
β_1	-0.4573	-0.0135	-0.3363	-0.5589

Table 1 shows that the required sample size decreases progressively from SUM to H, then WS, and finally EU. The table also shows that SUM and EU have higher sensitivity of required sample sizes to deviations from rationality than H and WS. Since sample size is modeled on a log scale, EU has a steeper slope (β_1) in terms of relative changes even though its curve appears flatter in absolute sample size due to a lower baseline sample size (β_0) compared with WS.

4 Functional Tests

The previous section applied our framework to analyze the power of nonparametric revealed-preference tests. This section extends the analysis to tests based on functional restrictions, which may increase power or prove useful when nonparametric characterisations are difficult to implement.

4.1 Functional Restrictions

To develop functional tests, we first introduce *functional restrictions* that single out subclasses of preferences.

Definition 4 (Functional Restriction)**.** Let $\mathcal{F} \subseteq \mathcal{G}$ be a class of demand functions. A map $\mathcal{R} : \mathcal{G} \to \mathbb{R}_{\geq 0}$ is called a *functional restriction* for $\mathcal{F}$ if $\mathcal{R}(g) = 0 \iff g \in \mathcal{F}$.

Suppose we want to describe a hypothesis test of $\mathcal{F}$ as in the preceding section:

$$\mathbf{H}_0 : g \in \mathcal{F} \qquad \text{vs.} \qquad \mathbf{H}_1 : g \in \mathcal{G} \setminus \mathcal{F}(\varepsilon),$$

where

$$\mathcal{F}(\varepsilon) := \{g \in \mathcal{G} : \inf_{g' \in \mathcal{F}} \mathsf{erf}_\mu(g, g') \leq \varepsilon\},$$

as in Theorem 3.1. As the examples below show, the definition of functional restriction imposes no finite-sample bounds on size and power without additional restrictions.

Example 1 (Size Bounds). Let $\mathcal{F} \subset \mathcal{G}$ and define the functional restriction:

$$\mathcal{R}_{\text{ind}}(g) \;:=\; \mathbf{1}\{g \notin \mathcal{F}\}.$$

Consider any (measurable) test $\mathcal{T} : \mathcal{D}^n \to \{0, 1\}$ whose decision rule depends only on $\mathcal{R}_{\text{ind}}$ (e.g., "reject iff $\mathcal{R}_{\text{ind}}(g) = 1$"). Notice that for any $\varepsilon > 0$, we can pick some $f \in \mathcal{F}$ near the boundary such that there exists $g \notin \mathcal{F}$ with $d(f, g) \leq \varepsilon$. Hence,

$$\sup_{g \in \mathbf{H}_0} \mu_g(\mathcal{T}^{-1}(1)) = 1.$$

Example 2 (Power Bounds). Let $\mathcal{F} \subset \mathcal{G}$ and define the functional restriction

$$\mathcal{R}_{\text{lin}}(g) = \gamma\, d(g, \mathcal{F}), \quad \gamma > 0.$$

Observe that no power bounds can be computed without knowledge of γ. To see this, suppose the analyst observes a dataset D_n, picks a rationalization $x \in \mathcal{G}$, and computes $\mathcal{R}(x) = k$. By definition, $d(x, \mathcal{F}) = k/\gamma$. If γ is small, then $d(x, \mathcal{F})$ is large and x appears to lie deep in the alternative; if γ is large, then $d(x, \mathcal{F})$ is small and x appears close to the null. Since the same observed value k is consistent with either conclusion, the analyst cannot determine whether to reject without a bound on γ.

As the above examples illustrate, functional restrictions alone may be too weak unless they satisfy additional properties that control their behavior under small perturbations of the data. Intuitively, we want two properties: *(i)* if the true preference lies in $\mathcal{F}$, then any nearby preference should nearly satisfy the restriction (continuity), and *(ii)* if a preference is far from satisfying the restriction, then this separation should be detectable in terms of our error metric (regularity). We now formalise these requirements.

Definition 5 (Uniform Continuity)**.** A restriction $\mathcal{R}$ defining $\mathcal{F}$ is *uniformly continuous* if there exists a function $\gamma : (0, \infty) \to (0, \infty)$ with $\gamma(\varepsilon) \to 0$ as $\varepsilon \to 0$, such that for every $g, g' \in \mathcal{G}$ and every $\varepsilon > 0$,

$$\mathsf{erf}_\mu(g, g') \;<\; \varepsilon \quad \Longrightarrow \quad \left|\mathcal{R}(g) - \mathcal{R}(g')\right| \;<\; \gamma(\varepsilon).$$

Definition 6 (Regularity)**.** A restriction $\mathcal{R}$ defining $\mathcal{F}$ is *regular* if there exists a

computable strictly increasing function[16] $\lambda : (0,\infty) \to (0,\infty)$ such that, for every $g \in \mathcal{G}$ and every $\varepsilon > 0$,

$$\inf_{f \in \mathcal{F}} \mathsf{erf}_\mu(g, f) \;>\; \varepsilon \quad\Longrightarrow\quad \mathcal{R}(g) \;>\; \lambda(\varepsilon).$$

Putting everything together, a functional restriction is simply a condition $\mathcal{R}$ that defines a subclass $\mathcal{F} \subset \mathcal{G}$ and induces a natural hypothesis test. To make such tests well behaved in our setting, we impose uniform continuity and regularity. The former guarantees the stability of the restriction under small perturbations, while the latter guarantees detectability of violations with respect to our error metric.

Remark. The functions $\gamma(\varepsilon)$ and $\lambda(\varepsilon)$ depend only on the restriction $\mathcal{R}$ and the domain $\mathcal{P}$, not on the specific preference relation or the realised sample. This uniformity ensures that the properties of continuity and regularity can be used in constructing feasible tests without requiring knowledge of the true underlying preference.

4.2 Testing Algorithm

The preceding subsection established the role of functional restrictions in defining well–behaved subclasses $\mathcal{F} \subset \mathcal{G}$ and showed how uniform continuity and regularity provide the discipline needed for asymptotic analysis. We now turn to the construction of an explicit testing algorithm. The idea is to exploit the functional restriction $\mathcal{R}$ to obtain a computable statistic that separates the null $\mathbf{H}_0 : g \in \mathcal{F}$ from the alternative $\mathbf{H}_1 : g \in \mathcal{G} \setminus \mathcal{F}(\varepsilon)$.

The algorithm below makes concrete the approach developed in Section 3.2. Rather than treating revealed preference tests, it delivers a constructive procedure with explicit finite–sample power bounds. It also illustrates how functional restrictions can serve as a practical testing device in settings where algebraic characterisations of the null are either unknown or computationally prohibitive.

[16] Here "computable" means that the function $\lambda(\varepsilon)$ can be obtained explicitly (for example, in closed form or via an algorithm) from the specification of the restriction $\mathcal{R}$ and the underlying environment $\mathcal{P} \times \mathcal{I}$.

Algorithm 1: Functional–Restriction Test

1. Fix a uniformly continuous restriction $\mathcal{R}$ defining a subclass $\mathcal{F} \subset \mathcal{G}$, with Lipschitz constant L. Take parameters ε, $\delta > 0$, and let the analyst sample a dataset of size n.

2. Choose $n = n(\varepsilon/t, \delta)$, where t is large enough that $\lambda(\varepsilon) - \gamma(\varepsilon/t) > \gamma(\varepsilon/t)$.[a]

3. Pick any rationalising choice function x and compute $\mathcal{R}(x)$. The decision rule is

$$\text{Decision} = \begin{cases} \text{Reject}, & \mathcal{R}(x) > \gamma(\varepsilon/t), \\ \text{Accept}, & \text{otherwise.} \end{cases}$$

[a] Such a t exists as $\gamma(0) = 0$.

The algorithm operationalises functional restrictions by providing a statistic $\mathcal{R}(x)$ capable of distinguishing the null from the alternative. Step 2 ensures that the sample size is sufficiently large relative to the tolerance parameters so that the separation implied by regularity dominates the continuity margin. Step 3 then reduces the test to a simple decision rule based on whether the restriction is violated beyond the uniform–continuity threshold. In this way, the procedure turns the abstract properties of functional restrictions into a concrete finite–sample test. The next result uses this algorithm to construct a test based on functional restrictions.

Theorem 4.1. *Following Algorithm 1, the induced test based on a uniformly continuous and regular restriction $\mathcal{R}$ satisfies:*

1. *Size. Under $\mathbf{H}_0 : g \in \mathcal{F}$, the probability of rejection is at most δ.*

2. *Power. Under any $\mathbf{H}_1 : g \in \mathcal{G} \setminus \mathcal{F}(\varepsilon)$, the probability of rejection is at least $1 - \delta$.*

Theorem 4.1 follows directly from the properties of uniform continuity and regularity. Under the null, Algorithm 1 ensures that any rationalising preference $\succeq_n$ lies within ε/t of the true preference with probability at least $1 - \delta$. Uniform continuity then implies $\mathcal{R}(x_n) \leq \gamma(\varepsilon/t)$, so rejection occurs with probability at most δ. Under the alternative, the true preference is separated from $\mathcal{F}$ by at least ε, and regularity guarantees $\mathcal{R}(g) \geq \lambda(\varepsilon)$. With probability at least $1-\delta$, the rationalising preference $\succeq_n$ selected by the analyst lies within ε/t of the true preference, so $\mathcal{R}(x_{\succeq_n}) \geq \lambda(\varepsilon) - \gamma(\varepsilon/t)$,

which exceeds the critical value by construction. Hence, rejection occurs with probability at least $1-\delta$. Taken together, Algorithm 1 and Theorem 4.1 show how functional restrictions yield implementable frequentist tests with explicit finite–sample bounds.

Remark. It is possible to show that our approach depends only on the learnability of the induced demand system, not on the particular type of choice sets. Once learnability is assured, the same continuity arguments and finite–sample bounds apply, so any functional restriction $\mathcal{R}$ usable with budget sets can be applied equally well to data generated from other admissible families of choice sets.

4.3 Inference for Smooth Estimands

This section outlines how to construct confidence intervals for smooth estimands of demand. To fix ideas, let $\mathcal{R}$ denote a functional restriction that maps a demand $x_{\succeq}$ into a real number summarising the change in a cost–of–living index induced by the preference $\succeq$. The analyst's goal is to use the observed data to estimate the true welfare value $\Delta W^* = \mathcal{R}(g)$, where g is the demand generated by the true underlying preference. The following result shows that our methodology allows one to obtain confidence intervals on welfare estimates.

Theorem 4.2. *Let $\mathcal{R} : \mathcal{G} \to \mathbb{R}$ be a uniformly continuous restriction with modulus γ, and let the true demand be $g \in \mathcal{G}$ with welfare value $\Delta W^* = \mathcal{R}(g)$. Fix $\varepsilon > 0$ and $\delta \in (0,1)$. By sampling $n \geq n(\varepsilon, \delta)$ observations, the analyst can select any rationalizing demand x, define $\Delta W = \mathcal{R}(x)$ and construct the interval*

$$\Big[\Delta W - \gamma(\varepsilon),\ \Delta W + \gamma(\varepsilon)\Big],$$

which contains the true value ΔW^* with probability at least $1 - \delta$.

Theorem 4.2 shows that when $\mathcal{R}$ is uniformly continuous, x is ensured to lie within ε of g, such that the values of the functional differ by at most ε. With probability at least $1 - \delta$, the true value of the estimand lies within $\gamma(\varepsilon)$ of the computed ΔW. The attainable precision depends only on the modulus of continuity of $\mathcal{R}$ and on the PAC sample complexity $n(\cdot, \cdot)$ of the class $\mathcal{G}$.

Remark. If the analyst wishes to achieve a target welfare precision $\eta > 0$, it suffices to choose any $\varepsilon > 0$ satisfying $\gamma(\varepsilon) \leq \eta$ and sample $n \geq n(\varepsilon, \delta)$ observations. Such

an ε exists because $\gamma(\varepsilon) \to 0$ as $\varepsilon \to 0$. In particular, when γ is strictly increasing—as it is for all the restrictions constructed in Section 4.4—the analyst can simply set $\varepsilon = \gamma^{-1}(\eta)$ and sample $n \geq n(\gamma^{-1}(\eta), \delta)$.

Example 3. As an illustration, define the indirect utility function $V_{\succeq}(p, I) = \max u_{\succeq}(y)$ subject to $y \in B(p, I)$. For a price change p_1 to p_2, the *equivalent variation* (EV) is the income adjustment after the price change that restores utility to its initial level, $V_{\succeq}(p_1, I) = V_{\succeq}\big(p_2,\, I - \mathsf{EV}_{p_1,p_2}(x_{\succeq})\big)$. By Hausman [1981], the EV is the solution to a differential equation in Marshallian demand. If $x_{\succeq}$ is twice continuously differentiable, then the EV functional is continuous in $x_{\succeq}$. Since the space $\mathcal{G}$ is compact, EV is in fact uniformly continuous, so Theorem 4.2 applies directly.

4.4 Functional Characterization and Extensions

Recall from Section 2 that $\mathcal{P} \times \mathcal{I}$ is compact and $\mathcal{G}$ denotes the class of income-Lipschitz demands. In this section we restrict attention to the subclass $\mathcal{G}^R \subseteq \mathcal{G}$ of (rational) models whose Marshallian demand $x : \mathcal{P} \times \mathcal{I} \to Y$ is twice continuously differentiable in (p, I). This additional regularity is used only to justify the derivative-based restriction operators $\mathcal{R}$. The finite-sample size and power bounds from Sections 3 are otherwise unchanged.

Throughout this section, we write $\mathsf{D}_p x(p, I) \in \mathbb{R}^{\ell \times \ell}$ for the Jacobian with respect to prices, with entries $(\mathsf{D}_p x)_{ij}(p, I) = \partial x_i(p, I)/\partial p_j$, and we write $\partial_I x(p, I) \in \mathbb{R}^\ell$ and $\partial_I^2 x(p, I) \in \mathbb{R}^\ell$ for the first and second derivatives with respect to income.[17] We also use the Slutsky matrix $S(x)(p, I) := \mathsf{D}_p x(p, I) + \big(\partial_I x(p, I)\big) x(p, I)^\top$, with entries $S_{ij}(x)(p, I) = (\mathsf{D}_p x)_{ij}(p, I) + x_j(p, I) \cdot \partial_I x_i(p, I)$.

4.4.1 Homotheticity

Homotheticity can be tested via the linearity of Engel curves. Under our assumptions, homothetic preferences $\succeq$ are equivalent to linearity of Marshallian demand in income at every fixed price. Equivalently, the second derivative of demand with respect to income vanishes. This yields a simple derivative-based restriction.

Definition 7. A preference $\succeq$ on Y is *homothetic* if for all $x, y \in Y$ and all $t > 0$, $x \succeq y$ if and only if $tx \succeq ty$.

[17] Whenever the dependence on (p, I) is clear, we omit the arguments.

Proposition 1. *Let $x_{\succeq} : \mathcal{P} \times \mathcal{I} \to Y$ be a C^2 Marshallian demand on a compact $\mathcal{P} \times \mathcal{I}$ arising from the rational preference $\succeq$. Then the following are equivalent:*

1. *$\succeq$ is homothetic.*
2. *$\partial_I^2 x_{\succeq}(p, I) \equiv 0$ on $\mathcal{P} \times \mathcal{I}$.*

Consequently, the restriction operator $\mathcal{R}^{\mathrm{hom}}(x_{\succeq}) := \partial_I^2 x_{\succeq}$ vanishes under the null of homotheticity and is uniformly continuous into L^1_μ by Proposition 6.[18]

Homotheticity admits a transparent derivative restriction: $\mathcal{R}^{\mathrm{hom}}(x_{\succeq}) := \partial_I^2 x_{\succeq}$ must vanish under the null. Since $\mathcal{R}^{\mathrm{hom}}$ is uniformly continuous (Proposition 6), it fits directly into our testing recipe: estimate $\mathcal{R}^{\mathrm{hom}}$ from data on (p, I, x), form the empirical score, and apply Algorithm 1 to obtain finite-sample size control and power against non-homothetic alternatives.

4.4.2 Weak Separability

We now turn to weakly separable preferences across a given partition of goods. The revealed preference tests for this class are NP–hard, which motivates our functional approach. For simplicity, we define weak separability directly via a utility representation, as axiomatic characterisations in terms of preferences are well known (Debreu [1960], Koopmans [1972]).

Definition 8 (Weak separability)**.** Fix a partition G_1, G_2 of the commodity index set $\{1, \dots, \ell\}$. A rational preference $\succeq$ on Y is *weakly separable across* (G_1, G_2) if there exist subutility indices $v_1 : \mathbb{R}_+^{|G_1|} \to \mathbb{R}$, $v_2 : \mathbb{R}_+^{|G_2|} \to \mathbb{R}$ and an aggregator $w : \mathbb{R}^2 \to \mathbb{R}$ such that $u_{\succeq}(x) = w\big(v_1(x_{G_1}), v_2(x_{G_2})\big)$ represents $\succeq$.

For a C^2 demand $x : \mathcal{P} \times \mathcal{I} \to Y$, the Slutsky matrix is $\mathcal{S}(x) := \mathsf{D}_p x + x\,(\partial_I x)^\top$, with entries $\mathcal{S}_{ij}(x) = \partial x_i/\partial p_j + x_j\,\partial x_i/\partial I$. Before stating the main result, we introduce some notation. For (p, I), write the between–group Slutsky block

$$\mathcal{S}^{12}(x_{\succeq})(p, I) := \big(\mathcal{S}_{ij}(x_{\succeq})(p, I)\big)_{i \in G_1,\, j \in G_2} \in \mathbb{R}^{|G_1| \times |G_2|},$$

[18] This object is defined formally in Proposition 6 in the appendix. For completeness, $L^1_\mu(\mathcal{P} \times \mathcal{I})$ denotes the space of μ-integrable functions on $\mathcal{P} \times \mathcal{I}$, equipped with the norm $\|f\|_{L^1_\mu} := \int_{\mathcal{P}\times\mathcal{I}} \|f(p, I)\|\, d\mu(p, I)$.

and the income–effect vectors

$$a(p,I) := \big(\partial_I x_{\succeq,i}(p,I)\big)_{i\in G_1}, \qquad b(p,I) := \big(\partial_I x_{\succeq,j}(p,I)\big)_{j\in G_2}.$$

We also impose a two–sided income–effect nondegeneracy condition: for every (p,I) there exist $i \in G_1$ and $j \in G_2$ with $\partial_I x_{\succeq,i}(p,I) \neq 0$ and $\partial_I x_{\succeq,j}(p,I) \neq 0$.

Proposition 2. *Fix a partition $G_1, G_2 \subset \{1,\dots,\ell\}$ and let $x_\succeq \in C^2(\mathcal{P}\times\mathcal{I};Y)$ be the Marshallian demand generated by a rational preference $\succeq$ on Y. The preference $\succeq$ is weakly separable across (G_1,G_2) if an only if*

$$\mathcal{R}^{\text{sep}} := \mathcal{S}_{ij}(x_\succeq)\,\partial_I x_{\succeq,i'}\,\partial_I x_{\succeq,j'} - \mathcal{S}_{i'j'}(x_\succeq)\,\partial_I x_{\succeq,i}\,\partial_I x_{\succeq,j} = 0 \tag{3}$$

for all $i,i' \in G_1$ and all $j,j' \in G_2$.

The proposition shows that, under the nondegeneracy condition on income effects, weak separability across (G_1,G_2) is *equivalent* to the vanishing of the functional restriction $\mathcal{R}^{\text{sep}}$. In other words, the structural property of weak separability is captured exactly by the condition that the between–group Slutsky block has rank one, with income effects determining the factorisation.

To see this more concretely, consider the case of three goods with $G_1 = \{1\}$ and $G_2 = \{2,3\}$. In this case, the between–group Slutsky block reduces to the row

$$\mathcal{S}^{12}(x_\succeq)(p,I) = \big[\,\mathcal{S}_{12}(x_\succeq)(p,I) \;\; \mathcal{S}_{13}(x_\succeq)(p,I)\,\big].$$

The restriction operator becomes the scalar condition

$$\mathcal{R}^{\text{sep}}(x_\succeq)(p,I) = \mathcal{S}_{12}(x_\succeq)(p,I)\,\partial_I x_{\succeq,3}(p,I) - \mathcal{S}_{13}(x_\succeq)(p,I)\,\partial_I x_{\succeq,2}(p,I).$$

Hence, weak separability across $(\{1\},\{2,3\})$ holds if and only if $\mathcal{R}^{\text{sep}}(x_\succeq)(p,I) = 0$ for all (p,I). Where both income effects $\partial_I x_{\succeq,2}$ and $\partial_I x_{\succeq,3}$ are nonzero, this condition is equivalent to

$$\frac{\mathcal{S}_{12}(x_\succeq)(p,I)}{\mathcal{S}_{13}(x_\succeq)(p,I)} = \frac{\partial_I x_{\succeq,2}(p,I)}{\partial_I x_{\succeq,3}(p,I)}.$$

If one of the two income effects vanishes, the restriction reduces correspondingly to $\mathcal{S}_{12} = 0$ or $\mathcal{S}_{13} = 0$. This simplified case makes transparent the role of $\mathcal{R}^{\text{sep}}$: it records the precise alignment between substitution effects and income effects across groups that characterises weak separability.

4.4.3 Complementarity and Substitutability

The notions of complementarity and substitutability have been formalized in several ways in the literature.[19] Two demand–based definitions are most prominent. First, goods i and j are said to be *gross complements* (*substitutes*) whenever

$$\frac{\partial x_i(p,I)}{\partial p_j} \; < \; 0 \quad \Big(\; > \; 0 \; \Big).$$

Because this derivative includes income effects, the notion is not symmetric. Second, following Hicks and Allen [1934], goods i and j are *net complements* (*substitutes*) if the corresponding Hicksian demand satisfies

$$\frac{\partial h_i(p,u)}{\partial p_j} \; < \; 0 \quad \Big(\; > \; 0 \; \Big).$$

By the Slutsky decomposition, this is equivalent to the sign of the Slutsky term

$$\mathcal{S}_{ij}(x)(p,I) = \frac{\partial x_i(p,I)}{\partial p_j} + x_j(p,I)\,\frac{\partial x_i(p,I)}{\partial I}.$$

Unlike the gross notion, this condition is symmetric across goods.[20]

Proposition 3. *Let $x_{\succeq} : \mathcal{P} \times \mathcal{I} \to Y$ be the Marshallian demand function generated by a rational preference $\succeq$ on Y. Fix distinct goods $i, j \in \{1, \ldots, \ell\}$. A functional restriction for* gross complementarity *is*

$$\mathcal{R}_{ij}^{\text{grossC}}(x_{\succeq})(p,I) := \max\left\{0, \; \frac{\partial x_{\succeq,i}(p,I)}{\partial p_j}\right\},$$

so that i and j are gross complements if and only if $\mathcal{R}_{ij}^{\text{grossC}}(x_{\succeq}) \equiv 0$. Analogous forms obtain for gross substitutability and for net complementarity/substitutability, replacing $\partial x_{\succeq,i}/\partial p_j$ with $-\partial x_{\succeq,i}/\partial p_j$ or with the Slutsky term $\mathcal{S}_{ij}(x_{\succeq})$, respectively.

Since these functional restrictions are constructed from derivatives of uniformly continuous functions, they inherit uniform continuity away from zero. Hence, as in

[19] Early definitions based on the signs of second derivatives of utility functions (see Auspitz and Lieben [1889], Fisher [1892], Edgeworth [1897], Pareto [1909]) are not invariant to monotone transformations. For surveys, see Samuelson [1974] and Newman [1987].

[20] For a recent approach that restores intuitive appeal to these definitions, see Weinstein [2022].

the case of weak separability, they provide well–behaved finite–sample tests of strict complementarity and strict substitutability within our framework.

4.4.4 Choice under Risk

In this section, we adapt the functional form approach of Kubler et al. [2014, Section III], who provide a characterization of expected utility preferences from contingent-claim demand.[21] Our objective is to extend this analysis to the broader *betweenness class* of preferences introduced by Dekel [1986] and Chew [1989]. Betweenness preferences generalise expected utility by requiring indifference curves to be straight lines, though not necessarily parallel, and admit an implicit representation of utility.

Formally, let there be S states of nature, indexed by $s \in \{1, \ldots, S\}$. The decision maker has preferences over state–contingent consumption vectors $\mathbf{x} \in \mathbb{R}^S_{>0}$ and holds objective beliefs $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_S) \in \Delta(S)$. Preferences are represented by a strictly increasing, thrice continuously differentiable, and strictly quasi–concave utility function $V(\mathbf{x}; \boldsymbol{\pi})$. By Dekel [1986, p. 312], a preference belongs to the betweenness class if and only if there exists a Bernoulli index $u : \mathbb{R}_{>0} \times [0,1] \to \mathbb{R}$, increasing in the first argument and continuous in the second, such that

$$V(\mathbf{x}; \boldsymbol{\pi}) = \sum_{s \in S} \pi_s \, u(x_s, \, V(\mathbf{x}; \boldsymbol{\pi})). \tag{4}$$

We assume throughout that u is thrice continuously differentiable in both arguments, concave in consumption, and satisfies $u_{11} < 0$.[22]

As in Kubler et al. [2014], we assume complete markets with strictly positive prices $\mathbf{p} \in \mathbb{R}^S_{>0}$ and income $I > 0$. The Marshallian demand induced by a betweenness preference $\succeq$ is then defined by

$$x_{\succeq}(\mathbf{p}, I, \boldsymbol{\pi}) \;\in\; \underset{\mathbf{x} \in \mathbb{R}^S_{>0}}{\arg\max} \Big\{ V(\mathbf{x}; \boldsymbol{\pi}) : \mathbf{p} \cdot \mathbf{x} \leq I \Big\}. \tag{5}$$

Proposition 4. *Let $S > 2$ and let $x_{\succeq}(\mathbf{p}, I, \boldsymbol{\pi}) \in \mathbb{R}^S_{>0}$ be the Marshallian demand*

[21] See Theorem 3 in Kubler et al. [2014, p. 3475].

[22] Here and below, subscripts denote partial derivatives of the utility index u: for example $u_1(x, v) = \partial u(x, v)/\partial x$ and $u_{11}(x, v) = \partial^2 u(x, v)/\partial x^2$.

induced by a betweenness preference represented by

$$V(\mathbf{x};\boldsymbol{\pi}) = \sum_{t=1}^{S} \pi_t \, u\big(x_t, V(\mathbf{x};\boldsymbol{\pi})\big),$$

with u strictly increasing in consumption, C^3 in both arguments, concave in consumption, and $u_{11} < 0$. Define $k_r(\mathbf{p},\boldsymbol{\pi}) := (\pi_r/\pi_1)(p_1/p_r)$ for $r \geq 2$. Then for each $s \in \{3,\dots,S\}$ there exists a function

$$f : \mathbb{R}^4_{>0} \to \mathbb{R}_{>0}, \qquad (x_1, x_2, k_2, k_s) \longmapsto f(x_1, x_2, k_2, k_s),$$

strictly increasing in its last argument k_s and satisfying $f(x, x, k_2, 1) = x$ for all $x, k_2 > 0$, such that

$$x_{\succeq,s}(\mathbf{p}, I, \boldsymbol{\pi}) \;=\; f\big(x_{\succeq,1}(\mathbf{p}, I, \boldsymbol{\pi}),\, x_{\succeq,2}(\mathbf{p}, I, \boldsymbol{\pi}),\, k_2(\mathbf{p},\boldsymbol{\pi}),\, k_s(\mathbf{p},\boldsymbol{\pi})\big). \tag{6}$$

For each fixed $s \geq 3$ define the map

$$H_s(\mathbf{p}, I; \boldsymbol{\pi}) := \Big(x_{\succeq,1}(\mathbf{p}, I, \boldsymbol{\pi}),\, x_{\succeq,2}(\mathbf{p}, I, \boldsymbol{\pi}),\, k_s(\mathbf{p}, \boldsymbol{\pi},\, x_{\succeq,s}(\mathbf{p}, I, \boldsymbol{\pi}))\Big) \in \mathbb{R}^4,$$

where $k_s = (\pi_s/\pi_1)(p_1/p_s)$ and $\boldsymbol{\pi}$ is treated as a parameter. Let $J_s(\mathbf{p}, I)$ be the $4 \times (S+1)$ Jacobian of H_s with respect to $(\mathbf{p}, I)$.

Proposition 5. *Under the assumptions of Proposition 4, for each $s \geq 3$ the Jacobian J_s has rank at most 3 at every $(\mathbf{p}, I)$. Equivalently, all 4×4 minors of J_s vanish. Define*

$$\mathcal{R}_s^{\mathrm{bet}}(x_\succeq)(\mathbf{p}, I, \boldsymbol{\pi}) \;:=\; \sum_{M \in \mathscr{M}_{4\times 4}(J_s(\mathbf{p},I))} \big(\det M\big)^2,$$

where $\mathscr{M}_{4\times4}(J_s)$ is the collection of all 4×4 submatrices of J_s. Then

$$\mathcal{R}_s^{\mathrm{bet}}(x_\succeq) \equiv 0 \quad \textit{is a necessary condition for betweenness rationalisability.}$$

Proposition 5 imposes a clear testable implication. Whenever we perturb prices and income so that, to first order, x_1, x_2, and the ratio $k_s = (\pi_s/\pi_1)(p_1/p_s)$ remain unchanged, then the demand for state s must also remain unchanged. Equivalently, the only channels through which x_s is allowed to move are movements in x_1, x_2, or k_s. There is no independent direction in $(\mathbf{p}, I)$ that can change x_s while holding these

other quantities fixed. This is the differential analogue of the functional dependence in Proposition 4: under betweenness, once (x_1, x_2, k_s) are fixed, x_s is pinned down.

The key difference with expected utility is that, under betweenness, $u_1(\cdot, V)$ depends jointly on consumption and the *endogenous* fixed point V. In expected utility, $u_1(\cdot)$ is univariate, which lets the FOCs collapse to a representation $x_s = f(x_1, k_s)$ that characterises the model. In contrast, for betweenness the FOCs imply that x_s is a function of (x_1, x_2, k_2, k_s), but this *does not* guarantee that a given f arises from some $u(\cdot, \cdot)$ consistent with the fixed–point structure. Thus, necessity holds—any betweenness rationalisation must satisfy the rank condition— but sufficiency need not hold in general, as not every solution to the rank condition corresponds to some betweenness utility.

Corollary 1. *Let $\mathcal{F}_{\text{bet}} \subset \mathcal{G}$ denote the class of preferences satisfying $\mathcal{R}_s^{\text{bet}}(x_\succeq) \equiv 0$ for every $s \geq 3$. Then the class of betweenness preferences is contained in $\mathcal{F}_{\text{bet}}$.*

The corollary makes precise the one–sided nature of the restriction: any demand violating $\mathcal{R}_s^{\text{bet}} \equiv 0$ cannot come from betweenness, but some non–betweenness demands may also satisfy it.

Although $\mathcal{R}^{\text{bet}}$ is not a full characterisation, it yields a valid non–rejection test. We test

$$\mathbf{H}_0 : \; x_\succeq \in \mathcal{F}_{\text{bet}} \qquad \text{vs} \qquad \mathbf{H}_1 : \; x_\succeq \in \mathcal{G} \setminus \mathcal{F}_{\text{bet}}(\varepsilon),$$

where $\mathcal{F}_{\text{bet}}(\varepsilon)$ is defined as in Section 3. For each $s \geq 3$ and each evaluation point $(\mathbf{p}, I)$, compute the Jacobian $J_s(\mathbf{p}, I)$ of

$$H_s(\mathbf{p}, I; \boldsymbol{\pi}) = \big(x_{\succeq,1},\, x_{\succeq,2},\, k_s,\, x_{\succeq,s}\big), \qquad k_s = (\pi_s/\pi_1)(p_1/p_s),$$

with respect to $(\mathbf{p}, I)$. Form

$$\mathcal{R}_s^{\text{bet}}(x_\succeq)(\mathbf{p}, I, \boldsymbol{\pi}) = \sum_{M \in \mathscr{M}_{4\times 4}(J_s(\mathbf{p},I))} (\det M)^2,$$

and aggregate across s and $(\mathbf{p}, I)$ via the L^1_μ–integral of Section 5, obtaining the sample analogue $\widehat{\mathcal{R}}^{\text{bet}}$. Under $\mathbf{H}_0$, the restriction equals zero. By Proposition 6 and compactness, the map $x \mapsto \mathcal{R}^{\text{bet}}$ is uniformly continuous into L^1_μ, so the finite–sample bounds established earlier apply.

Operationally, one can approximate the minors of J_s by finite differences, using small perturbations $(\Delta\mathbf{p}, \Delta I)$ chosen so that x_1, x_2, and k_s are (approximately) held

constant to first order. The restriction requires the induced change in x_s to be (approximately) zero. Equivalently, the estimated 4×4 minors evaluated at these perturbations should be close to zero under $\mathbf{H}_0$.

5 Efficiency Gains: Queries and Subclasses

The sample complexity of both testing and estimation in our framework is inherited from the PAC–learnability of $\mathcal{G}$. This means that the attainable precision and confidence of welfare estimates are directly tied to the sample complexity of the underlying learning algorithm. In this section, we formalise this dependence and show how efficiency gains can be realised in two ways: by imposing additional structure on the class $\mathcal{G}$, and by allowing the analyst to select prices adaptively. The following proposition makes precise the link between improved sample complexity and tighter confidence intervals.

Propostion 5.1. *Fix a sample size n and confidence level $\delta \in (0,1)$. Let $n_1(\varepsilon, \delta)$ and $n_2(\varepsilon, \delta)$ be two valid sample complexity functions for (ε, δ)–learning $\mathcal{G}$ such that $n_1(\varepsilon, \delta) < n_2(\varepsilon, \delta)$ for all ε, δ. Then the minimal confidence interval width implied by n_1,*

$$t_1^*(\delta, n) \;:=\; \inf\{\, t > 0 : n_1(\gamma(t), \delta) \leq n \,\},$$

is weakly smaller than that implied by n_2,

$$t_2^*(\delta, n) \;:=\; \inf\{\, t > 0 : n_2(\gamma(t), \delta) \leq n \,\}.$$

Proposition 5.1 formalises the connection between sample complexity and inferential precision in our framework. The sample complexity $n(\varepsilon, \delta)$ captures how demanding it is to learn the class $\mathcal{G}$: smaller values mean that fewer observations are required to reach a given accuracy and confidence. Thus, the results states that improvements in sample complexity translate directly into tighter confidence bounds at the same (n, δ).

Example 4. Suppose two candidate sample complexities are given by

$$n_1(\varepsilon, \delta) \;=\; \frac{1}{\varepsilon^2} \cdot \frac{1}{\delta}, \qquad n_2(\varepsilon, \delta) \;=\; \frac{1}{\varepsilon^{k+2}} \cdot \frac{1}{\delta},$$

with $\gamma(t) = t$. For each $i = 1, 2$, the minimal interval width is $t_i^*(\delta, n) = \inf\{\, t > 0 :$

$n_i(\gamma(t), \delta) \leq n \}$. Solving these inequalities yields

$$t_1^*(\delta, n) \;=\; (\delta n)^{-1/2}, \qquad t_2^*(\delta, n) \;=\; (\delta n)^{-1/(k+2)}.$$

Thus the first learning algorithm delivers strictly tighter confidence intervals, since $t_1^*(\delta, n) < t_2^*(\delta, n)$ for all $k > 0$.

A similar comparison applies to the attainable confidence levels when the interval width t is fixed. Define $1 - \delta_i^*(t, n) \;=\; \sup\{\, 1 - \delta : \; n_i(\gamma(t), \delta) \leq n \,\}$. Then

$$1 - \delta_1^*(t, n) \;=\; 1 - \frac{1}{nt^2}, \qquad 1 - \delta_2^*(t, n) \;=\; 1 - \frac{1}{nt^{k+2}}.$$

Here too the algorithm with lower sample complexity, n_1, yields higher confidence at fixed precision. In Appendix A, we formalize the efficiency gains that occur through adaptive sampling and further assumptions on $\mathcal{G}$.

6 Conclusion

This paper develops a frequentist framework for conducting power analysis of revealed preference tests using finite choice data and tackles a central challenge in the literature: constructing parsimonious alternative hypotheses against which the discriminatory power of rationality tests can be assessed. Our procedure is grounded in the Probably Approximately Correct (PAC) learning framework and allows us to design tests that integrate functional and finite-data approaches. This method is broadly applicable, ranging from weak separability to choice under risk, highlighting its versatility for analyzing decision-making.

We make four main contributions. First, we show that PAC learnability results for Lipschitz demand functions provide the theoretical foundation for constructing well-defined alternative hypotheses. By exploiting the fact that any rationalizing demand function lies within a learnable neighborhood of the true demand with high probability, we can compute explicit bounds on the sample size required to achieve any desired level of statistical power against alternatives that are ε-separated from the null hypothesis.

Second, we extend our framework beyond traditional revealed preference characterizations to functional restrictions. This extension is particularly valuable for preference classes—such as weakly separable preferences—where exact RP tests are computationally intractable (NP-hard). We show that by tolerating a small but controlled size, one

can construct polynomial-time tests with explicit finite-sample power guarantees.

Third, we demonstrate how to construct confidence intervals for smooth functionals of demand, such as equivalent variation and other welfare measures. The width of these intervals depends directly on the sample complexity of learning the underlying preference class, thereby linking the richness of the maintained hypothesis to the precision of welfare inference.

Fourth, our simulation results reveal an important insight: the apparent empirical success of GARP may partly reflect its limited ability to detect small deviations from rationality. In contrast, the frequent rejections of expected utility in experimental data may reflect its sensitivity to even negligible departures from the model. This suggests that comparing rejection rates across different preference classes requires careful attention to their differential statistical power.

Our approach shifts the focus from testing the dataset itself to testing the decision maker. Unlike RP tests that always reject a dataset if it cannot be rationalized by any preference from the class of interest, our tests are one-sided: they never reject a rationalizable dataset. Still, they may fail to reject a non-rationalizable one. However, the probability of failing to reject a non-rationalizable dataset decreases as more data are sampled. Our simulation exercises show that our tests perform well in finite samples and correctly reject demands that do not belong to specific classes of preferences.

References

Abi Adams and Ian Crawford. Models of revealed preference. *Emerging Trends in the Social and Behavioral Sciences*, pages 1–15, 2015.

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

Victor H Aguiar and Nail Kashaev. Stochastic revealed preferences with measurement error. *The Review of Economic Studies*, 88(4):2042–2093, 2021.

Victor H Aguiar and Roberto Serrano. Slutsky matrix norms: The size, classification, and comparative statics of bounded rationality. *Journal of Economic Theory*, 172: 163–201, 2017.

Victor H Aguiar and Roberto Serrano. Classifying bounded rationality in limited data sets: a slutsky matrix approach. *SERIEs*, 9(4):389–421, 2018.

Roy Allen, Paweł Dziewulski, and John Rehbeck. Revealed statistical consumer theory. *Economic Theory*, 77(3):823–847, 2024.

James Andreoni, Benjamin J Gillen, and William T Harbaugh. The power of revealed preference tests: Ex-post evaluation of experimental design. *Unpublished manuscript*, 2013.

G. Antonelli. On the mathematical theory of political economy. In J. Chipman, L. Hurwicz, M. Richter, and H. Sonnenschein, editors, *Preference, Utility and Demand*. Harcourt Brace Jovanavich Ltd, 1971.

Rudolf Auspitz and Richard Lieben. *Untersuchungen über die Theorie des Preises*. Duncker und Humblot, 1889.

Maria-Florina Balcan, Amit Daniely, Ruta Mehta, Ruth Urner, and Vijay V Vazirani. Learning economic parameters from revealed preferences. In *Web and Internet Economics: 10th International Conference, WINE 2014, Beijing, China, December 14-17, 2014. Proceedings 10*, pages 338–353. Springer, 2014.

Peter L Bartlett, Philip M Long, and Robert C Williamson. Fat-shattering and the learnability of real-valued functions. In *Proceedings of the seventh annual conference on Computational learning theory*, pages 299–310, 1994.

T. K. M. Beatty and I. A. Crawford. How demanding is the revealed preference approach to demand? *American Economic Review*, 101, 2011.

Gary S Becker. Irrational behavior and economic theory. *Journal of political economy*, 70(1):1–13, 1962.

Eyal Beigman and Rakesh Vohra. Learning from revealed preference. In *Proceedings of the 7th ACM Conference on Electronic Commerce*, pages 36–42, 2006.

Stephen G Bronars. The power of nonparametric tests of preference maximization. *Econometrica: Journal of the Econometric Society*, pages 693–698, 1987.

C. Chambers, F. Echenique, and K. Saito. Testing theories of financial decision making (formerly titled testable implications of ranslation invariance and homotheticity: Variational, maxmin, cara and crra preferences). *Proceedings of the National Academy of Sciences*, 113, 2016.

Christopher P Chambers and Federico Echenique. Decision theory and the" almost implies near" phenomenon. *arXiv preprint arXiv:2502.07126*, 2025.

Christopher P Chambers, Federico Echenique, and Nicolas S Lambert. Recovering preferences from finite data. *Econometrica*, 89(4):1633–1664, 2021.

Christopher P Chambers, Federico Echenique, and Nicolas S Lambert. Recovering utility. *arXiv preprint arXiv:2301.11492*, 2023.

Laurens Cherchye, Thomas Demuynck, Bram De Rock, and Per Hjertstrand. Revealed preference tests for weak separability: An integer programming approach. *Journal of Econometrics*, 186(1):129–141, 2015. doi: 10.1016/j.jeconom.2014.07. URL https://ideas.repec.org/a/eee/econom/v186y2015i1p129-141.html.

Soo Hong Chew. Axiomatic utility theories with the betweenness property. *Annals of operations Research*, 19(1):273–298, 1989.

Ian Crawford and Bram De Rock. Empirical Revealed Preference. *Annual Review of Economics*, 6(1):503–524, August 2014. URL https://ideas.repec.org/a/anr/reveco/v6y2014p503-524.html.

Angus Deaton and John Muellbauer. *Economics and Consumer Behavior*. Number 9780521296762 in Cambridge Books. Cambridge University Press, January 1980. URL https://ideas.repec.org/b/cup/cbooks/9780521296762.html.

Rahul Deb, Yuichi Kitamura, John KH Quah, and Jörg Stoye. Revealed price preference: theory and empirical analysis. *The Review of Economic Studies*, 90(2):707–743, 2023.

G. Debreu. Topological methods in cardinal utility theory. *Mathematical Methods in the Social Sciences*, Stanford University Press:16–26, 1960.

Eddie Dekel. An axiomatic characterization of preferences under uncertainty: Weakening the independence axiom. *Journal of Economic theory*, 40(2):304–318, 1986.

Thomas Demuynck and Per Hjertstrand. Samuelson's approach to revealed preference theory: Some recent advances. *Paul Samuelson*, pages 193–227, 2019.

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

Federico Echenique. Testing for separability is hard. *arXiv preprint arXiv:1401.4499*, 2014.

Federico Echenique. New developments in revealed preference theory: decisions under risk, uncertainty, and intertemporal choice. Papers 1908.07561, arXiv.org, August 2019. URL https://ideas.repec.org/p/arx/papers/1908.07561.html.

Francis Y Edgeworth. La teoria pura del monopolio. *Giornale degli economisti*, pages 13–31, 1897.

Larry G. Epstein. Are Probabilities Used in Markets ? *Journal of Economic Theory*, 91(1):86–90, March 2000. URL https://ideas.repec.org/a/eee/jetheo/v91y2000i1p86-90.html.

I Fisher. Mathematical investigations in the theory of value and prices. transactons of the connecticut academy, vol. ix. 1892.

Steven M Goldman and Hirofumi Uzawa. A note on separability in demand analysis. *Econometrica: Journal of the Econometric Society*, pages 387–398, 1964.

Richard C Green and Sanjay Srivastava. Expected utility maximization and demand behavior. *Journal of Economic Theory*, 38(2):313–323, 1986.

Jerry A Hausman. Exact consumer's surplus and deadweight loss. *The American Economic Review*, 71(4):662–676, 1981.

John R Hicks and Roy GD Allen. A reconsideration of the theory of value. part i. *Economica*, 1(1):52–76, 1934.

Per Hjertstrand, James L. Swofford, and Gerald A. Whitney. General Revealed Preference Tests of Weak Separability and Utility Maximization with Incomplete Adjustment. Working Paper Series 1327, Research Institute of Industrial Economics, March 2020. URL https://ideas.repec.org/p/hhs/iuiwop/1327.html.

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

Tjalling C. Koopmans. Representation of preference orderings with independent components of consumption. In C. B. McGuire and Roy Radner, editors, *Decision and Organization: A Volume in Honor of Jacob Marschak*, pages 57–78. North-Holland Publishing Company, Amsterdam and London, 1972.

Felix Kubler, Larry Selden, and Xiao Wei. Asset demand based tests of expected utility maximization. *American Economic Review*, 104(11):3459–3480, 2014.

Felix Kubler, Raghav Malhotra, and Herakles Polemarchakis. Identification of preferences, demand and equilibrium with finite data. Technical report, University of Warwick, Department of Economics, 2020.

Peter Newman. Substitutes and complements. *The New Palgrave: A Dictionary of Economics*, pages 545–548, 1987.

Vilfredo Pareto. *Manuel d'économie politique*, volume 38. Giard & Brière, 1909.

Matthew Polisson, John K.-H. Quah, and Ludovic Renou. Revealed Preferences over Risk and Uncertainty. *American Economic Review*, 110(6):1782–1820, June 2020. doi: 10.3886/E112146V1. URL https://ideas.repec.org/a/aea/aecrev/v110y2020i6p1782-1820.html.

P. A. Samuelson. A note on the pure theory of consumers' behavior. *Economica*, 5: 61–71, 1938.

Paul A Samuelson. Complementarity: An essay on the 40th anniversary of the hicks-allen revolution in demand theory. *Journal of Economic Literature*, 12(4):1255–1289, 1974.

Reinhard Selten. Properties of a measure of predictive success. *Mathematical social sciences*, 21(2):153–167, 1991.

E. Slutsky. Sulla teoria del bilancio del consumatore. *Giornale degli Economisti*, 51: 1–26, 1915.

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

Hal R Varian. Non-parametric tests of consumer behaviour. *The review of economic studies*, 50(1):99–110, 1983.

Jonathan Weinstein. Direct complementarity. 2022.

Appendices (Online-Only)

Appendix A Efficiency Gains: Queries and Subclasses

A.1 Subclasses

A natural way to obtain sharper precision bounds (and higher power for RP characterisations) is to impose additional structure on the preference family $\mathcal{G}$. Restricting attention to subclasses can dramatically reduce sample complexity, and hence tighten the confidence intervals described in Proposition 5.1. Balcan et al. [2014] provide sharp results for several economically relevant subclasses, including linear preferences, separable piecewise linear concave (SPLC) preferences with l segments, and CES preferences with a known parameter ρ.

To illustrate how subclass restrictions operate in practice, we recall three canonical families of preferences studied in Balcan et al. [2014]. Each imposes a different structure on the utility function, which in turn reduces the statistical complexity of learning relative to the unrestricted class $\mathcal{G}$. The cases of linear, separable piecewise linear concave (SPLC), and constant elasticity of substitution (CES) preferences are of particular interest, both because of their economic relevance and because their structural features lead to sharply different sample complexities.

Definition 9 (Linear Preference)**.** A utility function U is called *linear* if utility is linear in each good. Formally, for some $a \in \mathbb{R}^d_+$,

$$U(x) = U_a(x) = \sum_{j=1}^{d} a_j x_j,$$

with the normalisation $\sum_j a_j = 1$ taken without loss of generality.

Definition 10 (Separable Piecewise Linear Concave (SPLC))**.** A utility function U is called *SPLC* if

$$U(x) = \sum_{j=1}^{d} U_j(x_j),$$

where each $U_j : \mathbb{R}_+ \to \mathbb{R}_+$ is non-decreasing, concave, and piecewise linear. The number of segments of U_j is denoted $|U_j|$, and the kth segment is written (j, k). If

segment (j,k) has domain $[a,b] \subseteq \mathbb{R}_+$ and slope c, we set $a_{jk} = c$ and $\ell_{jk} = b - a$, with $\ell_{j|U_j|} = \infty$. Concavity implies $a_{j(k-1)} > a_{jk}$ for all $k \geq 2$. An SPLC function with $|U_j| \leq \kappa$ for all j can thus be represented by two matrices $A, L \in \mathbb{R}_+^{d\times\kappa}$, and denoted $U_{A,L}$.

Definition 11 (Constant Elasticity of Substitution (CES))**.** A utility function U is called *CES* if, for some $\rho \in (-\infty, 1]$ and $a \in \mathbb{R}_+^d$,

$$U(x) = U_{a,\rho}(x) = \Big(\sum_{j=1}^{d} a_j x_j^{\rho}\Big)^{1/\rho},$$

with the normalisation $\sum_j a_j = 1$ taken without loss of generality.

The efficiency gains from subclass restrictions can be quantified directly through the PAC–learning sample complexity of these classes. Balcan et al. [2014] derive bounds for linear, SPLC, and CES preferences, showing how the additional structure substantially reduces the number of observations needed to achieve accuracy ε and confidence $1-\delta$. Table 2 summarises their results.

Table 2: Sample Complexity by Preference Class

Preference Class	**Linear**	**SPLC (l segments)**	**CES**
Sample Complexity	$O\left(\frac{k\log(1/\varepsilon)+\log(1/\delta)}{\varepsilon}\right)$	$O\left(\frac{kl\log(1/\varepsilon)+\log(1/\delta)}{\varepsilon}\right)$	$O\left(\frac{k\log(1/\varepsilon)+\log(1/\delta)}{\varepsilon}\right)$

These bounds highlight how subclass assumptions translate into improved statistical performance within our framework. In particular, the dependence on $1/\varepsilon$ improves from quadratic in the unrestricted case to linear under these structured classes. This reduction feeds directly into the confidence bounds of Proposition 5.1: by replacing the generic sample complexity $n(\varepsilon,\delta)$ with the subclass–specific expressions from Table 2, the analyst obtains strictly tighter confidence intervals and higher power for the same dataset. We next show how adaptive sampling can generate further efficiency gains.

A.2 The Adaptive Model

A further efficiency gain arises when the analyst is allowed to choose budget sets adaptively. In this model, the analyst selects prices sequentially and observes demand

from the corresponding budget set, with each subsequent query conditioned on the data revealed in earlier rounds.

The effect of adaptivity on sample complexity is substantial. Balcan et al. [2014] show that when queries are chosen adaptively, the number of observations needed to learn certain subclasses of preferences drops dramatically relative to the non-adaptive case. Table 3 reports their bounds for linear, SPLC, and CES preferences.

Table 3: Sample Complexity by Preference Class

Preference Class	**Linear**	**SPLC (l segments)**	**CES**
Sample Complexity	$O(qd)$	$O(qkd)$	$O(1)$

Here, q denotes the bit-size of the coefficients. If utility functions are L–Lipschitz, one can interpret q as setting the granularity of an approximation: coefficients take values in $[0, L]$ with increments of $L/(2^q-1)$. The table shows that adaptivity eliminates the $1/\varepsilon$ dependence altogether: learning becomes essentially finite in the CES case, linear in dimension for linear preferences, and only modestly larger for SPLC. This highlights how adaptive sampling can deliver dramatic efficiency gains in revealed–preference analysis.

Appendix B Proofs

A continuity proposition for derivative–based restrictions

Recall from Section 2 that $\mathcal{P}\times\mathcal{I}$ is compact and $\mathcal{G}$ denotes the class of income-Lipschitz demands. In this section we restrict attention to the subclass $\mathcal{G}^R \subseteq \mathcal{G}$ of models whose Marshallian demand $x : \mathcal{P}\times\mathcal{I} \to \mathbb{R}^\ell$ is twice continuously differentiable in (p, I). This additional regularity is used only to justify the derivative-based restriction operators $\mathcal{R}$. The finite-sample size and power bounds from Sections 3–5 are otherwise unchanged.

Proposition 6. *Let $\mu \in \Delta(\mathcal{P}\times\mathcal{I})$ and $\mathcal{G} \subset C^2(\mathcal{P}\times\mathcal{I};\mathbb{R}^\ell)$. Define $\mathcal{R}_p(x) := \mathsf{D}_p x$, $\mathcal{R}_I(x) := \partial_I x$, and $\mathcal{R}_{II}(x) := \partial_I^2 x$, viewed as elements of $L^1_\mu(\mathcal{P}\times\mathcal{I})$ endowed with the norm*

$$\|f\|_{L^1_\mu} := \int_{\mathcal{P}\times\mathcal{I}} \|f(p,I)\|\,\mathrm{d}\mu(p,I).$$

Then $\mathcal{R}_p, \mathcal{R}_I : (\mathcal{G}, \|\cdot\|_{C^{1,\infty}(\mathcal{P}\times\mathcal{I})}) \to L^1_\mu(\mathcal{P}\times\mathcal{I})$ and $\mathcal{R}_{II} : (\mathcal{G}, \|\cdot\|_{C^{2,\infty}(\mathcal{P}\times\mathcal{I})}) \to L^1_\mu(\mathcal{P}\times\mathcal{I})$

are bounded linear maps with operator norm at most 1*. In particular, for all* $x, y \in \mathcal{G}$*,*[23]

$$\|\mathcal{R}_p(x) - \mathcal{R}_p(y)\|_{L^1_\mu} \le \|x - y\|_{C^{1,\infty}}, \quad \|\mathcal{R}_I(x) - \mathcal{R}_I(y)\|_{L^1_\mu} \le \|x - y\|_{C^{1,\infty}},$$
$$\|\mathcal{R}_{II}(x) - \mathcal{R}_{II}(y)\|_{L^1_\mu} \le \|x - y\|_{C^{2,\infty}}.$$

Proof. Recall that $\Omega := \mathcal{P} \times \mathcal{I}$ is compact. We equip $\mathbb{R}^\ell$ with the Euclidean norm $\|\cdot\|$, matrices in $\mathbb{R}^{\ell\times\ell}$ with the induced operator norm $\|A\| := \sup_{\|v\|=1} \|Av\|$, and higher-order derivatives with the corresponding multilinear operator norms.

For $x : \Omega \to \mathbb{R}^\ell$ define

$$\|x\|_{C^{1,\infty}(\Omega)} := \max\Big\{ \sup_{(p,I)\in\Omega} \|x(p,I)\|, \sup_{(p,I)\in\Omega} \|\mathsf{D}_p x(p,I)\|, \sup_{(p,I)\in\Omega} \|\partial_I x(p,I)\| \Big\},$$

and

$$\|x\|_{C^{2,\infty}(\Omega)} := \max\Big\{ \|x\|_{C^{1,\infty}(\Omega)}, \sup_{(p,I)\in\Omega} \|\partial_I^2 x(p,I)\|, \sup_{(p,I)\in\Omega} \|\mathsf{D}_p \partial_I x(p,I)\|, \sup_{(p,I)\in\Omega} \|\mathsf{D}_p^2 x(p,I)\| \Big\},$$

where $\big(\mathsf{D}_p \partial_I x(p,I)\big)_{ij} = \partial^2 x_i(p,I)/(\partial p_j\, \partial I)$.

For any measurable vector or matrix valued function f on Ω, define

$$\|f\|_{L^1_\mu(\Omega)} := \int_\Omega \|f(p,I)\| \, \mathrm{d}\mu(p,I).$$

Since $\mu \in \Delta(\Omega)$ is a probability measure, $\mu(\Omega) = 1$, and therefore

$$\|f\|_{L^1_\mu(\Omega)} \le \|f\|_{L^\infty(\Omega)} := \sup_{(p,I)\in\Omega} \|f(p,I)\|.$$

(i) Boundedness of $\mathcal{R}_p$ *and* $\mathcal{R}_I$*.* Define the range spaces

$$\mathcal{R}_p : \mathcal{G} \to L^1_\mu(\Omega; \mathbb{R}^{\ell\times\ell}), \qquad \mathcal{R}_I : \mathcal{G} \to L^1_\mu(\Omega; \mathbb{R}^\ell),$$

[23] For $x : \mathcal{P} \times \mathcal{I} \to \mathbb{R}^\ell$, we use the Euclidean norm $\|\cdot\|$ on $\mathbb{R}^\ell$. For matrices $A \in \mathbb{R}^{\ell\times\ell}$, we use the induced operator norm $\|A\| := \sup_{\|v\|=1} \|Av\|$. The C^1 sup–norm is

$$\|x\|_{C^{1,\infty}} := \max\Big\{ \sup_{(p,I)} \|x(p,I)\|, \sup_{(p,I)} \|\mathsf{D}_p x(p,I)\|, \sup_{(p,I)} \|\partial_I x(p,I)\| \Big\},$$

and $\|x\|_{C^{2,\infty}}$ is defined analogously by additionally taking the sup–norm of $\partial_I^2 x$ (and, if needed, of the remaining second and mixed partial derivatives).

by $\mathcal{R}_p(x) := \mathsf{D}_p x$ and $\mathcal{R}_I(x) := \partial_I x$. Linearity is immediate from linearity of differentiation. For $x, y \in \mathcal{G}$,

$$\|\mathcal{R}_p(x) - \mathcal{R}_p(y)\|_{L^1_\mu(\Omega)} = \int_\Omega \|\mathsf{D}_p(x-y)(p,I)\| \, \mathrm{d}\mu(p,I) \le \|\mathsf{D}_p(x-y)\|_{L^\infty(\Omega)} \le \|x-y\|_{C^{1,\infty}(\Omega)},$$
$$\|\mathcal{R}_I(x) - \mathcal{R}_I(y)\|_{L^1_\mu(\Omega)} = \int_\Omega \|\partial_I(x-y)(p,I)\| \, \mathrm{d}\mu(p,I) \le \|\partial_I(x-y)\|_{L^\infty(\Omega)} \le \|x-y\|_{C^{1,\infty}(\Omega)}.$$

Hence $\mathcal{R}_p$ and $\mathcal{R}_I$ are bounded linear maps with operator norm at most 1.

(ii) Boundedness of $\mathcal{R}_{II}$. Define

$$\mathcal{R}_{II} : \mathcal{G} \to L^1_\mu(\Omega; \mathbb{R}^\ell), \qquad \mathcal{R}_{II}(x) := \partial_I^2 x.$$

Again, linearity is immediate. For $x, y \in \mathcal{G}$,

$$\|\mathcal{R}_{II}(x) - \mathcal{R}_{II}(y)\|_{L^1_\mu(\Omega)} = \int_\Omega \|\partial_I^2(x-y)(p,I)\| \, \mathrm{d}\mu(p,I) \le \|\partial_I^2(x-y)\|_{L^\infty(\Omega)} \le \|x-y\|_{C^{2,\infty}(\Omega)}.$$

Therefore $\mathcal{R}_{II}$ is a bounded linear map with operator norm at most 1. □

B.1 Proof of Lemma 1

Fix $L > 0$ and let $\mathcal{G}$ be the class of choice functions $x : \mathcal{P} \times \mathcal{I} \to \mathbb{R}^\ell$ that are L-Lipschitz with respect to the Euclidean distance on $\mathcal{P} \times \mathcal{I} \subset \mathbb{R}^{\ell+1}$ and the Euclidean norm on $\mathbb{R}^\ell$, that is, for all $(p, I), (p', I') \in \mathcal{P} \times \mathcal{I}$,

$$\|x(p,I) - x(p',I')\| \le L \, \|(p,I) - (p',I')\|_2.$$

We prove the stated sample-complexity bound by bounding the fat-shattering dimension of $\mathcal{G}$ via a packing-number argument and then invoking the fat-shattering sample-complexity theorem of Beigman and Vohra [2006].

Step 1. Let $\gamma > 0$ and suppose that $S = \{s_1, \dots, s_n\} \subset \mathcal{P} \times \mathcal{I}$ is γ-fat shattered by $\mathcal{G}$ in the (vector-valued) sense of Beigman and Vohra [2006, Definition 2]. Then there exist witnesses $y_1, \dots, y_n \in \mathbb{R}^\ell$ and two parallel affine hyperplanes $H_0, H_1 \subset \mathbb{R}^\ell$ with $\operatorname{dist}(H_0, H_1) > \gamma$ such that for every labelling $b \in \{0,1\}^n$ there exists $x_b \in \mathcal{G}$ with $x_b(s_i)$ lying on the b_i-designated side of the translated hyperplane $y_i + H_{b_i}$.

Fix $i \neq j$ and choose a labelling b with $b_i = 0$ and $b_j = 1$. By construction, $x_b(s_i)$ lies in the half-space determined by $y_i + H_0$ and $x_b(s_j)$ lies in the opposite half-space

determined by $y_j + H_1$. Because the two hyperplanes are parallel and at distance greater than γ, this implies

$$\|x_b(s_i) - x_b(s_j)\| > \gamma.$$

Since x_b is L-Lipschitz, we also have

$$\|x_b(s_i) - x_b(s_j)\| \leq L \|s_i - s_j\|_2.$$

Combining the two inequalities yields $\|s_i - s_j\|_2 > \gamma/L$. Since $i \neq j$ were arbitrary, the set S is (γ/L)-separated in Euclidean distance. Therefore,

$$\mathrm{fat}_{\mathcal{G}}(\gamma) \leq \mathsf{Pack}\big(\mathcal{P} \times \mathcal{I}, \|\cdot\|_2, \gamma/L\big),$$

where $\mathsf{Pack}(X, \|\cdot\|_2, r)$ denotes the r-packing number of X under Euclidean distance. This is the same separation argument used in the proof of Beigman and Vohra [2006, Theorem 6], and it uses only the Lipschitz property.

Step 2. Since $\mathcal{P} \times \mathcal{I}$ is a compact subset of $\mathbb{R}^{\ell+1}$, there exists a constant $C < \infty$ (depending only on $\mathcal{P} \times \mathcal{I}$) such that for all sufficiently small $r > 0$,

$$\mathsf{Pack}\big(\mathcal{P} \times \mathcal{I}, \|\cdot\|_2, r\big) \leq C\, r^{-(\ell+1)}.$$

Applying this with $r = \gamma/L$ and combining with Step 1 gives

$$\mathrm{fat}_{\mathcal{G}}(\gamma) \leq C \left(\frac{L}{\gamma}\right)^{\ell+1}.$$

Step 3. By Beigman and Vohra [2006, Theorem 4], the class $\mathcal{G}$ is PAC learnable with sample complexity

$$m_L(\varepsilon, \delta) = O\left(\frac{1}{\varepsilon^2}\left(\ln^2\left(\frac{1}{\varepsilon}\right)\mathrm{fat}_{\mathcal{G}}(\varepsilon) + \ln\left(\frac{1}{\delta}\right)\right)\right).$$

Substituting the bound from Step 2 at $\gamma = \varepsilon$ yields

$$m_L(\varepsilon, \delta) = O\left(\frac{1}{\varepsilon^2}\left(\ln^2\left(\frac{1}{\varepsilon}\right)\left(\frac{L}{\varepsilon}\right)^{\ell+1} + \ln\left(\frac{1}{\delta}\right)\right)\right),$$

which is the desired bound.

B.2 Proof of Theorem 3.1

By Lemma 1 (Beigman and Vohra [2006]), the class $\mathcal{G}$ is PAC learnable: for each $\varepsilon, \delta > 0$ there exists an algorithm L with sample size $m_L(\varepsilon, \delta)$ satisfying Definition 3. By the convention introduced after Definition 3, $n(\varepsilon, \delta)$ denotes the *minimal* sample complexity, hence

$$n(\varepsilon, \delta) \;\leq\; m_L(\varepsilon, \delta).$$

Fix any sample size n with $n \geq n(\varepsilon, \delta)$ (in particular, choosing $n = m_L(\varepsilon, \delta)$ suffices). Then, for any true model $g \in \mathcal{G}$,

$$\mu_g^n\Big(\mathsf{erf}_\mu(x_\succeq, g) < \varepsilon \text{ for every } x_\succeq \in \mathcal{G} \text{ that rationalises } D_n \Big) \;\geq\; 1 - \delta.$$

Suppose now that the true model satisfies the alternative such that $g \in \mathcal{G} \setminus \mathcal{F}(\varepsilon)$. Then, it follows that

$$\inf_{x \in \mathcal{F}} \big\{ \mathsf{erf}_\mu(x, g) \big\} > \varepsilon.$$

Consider any dataset D_n generated by g and suppose some $x \in \mathcal{F}$ rationalizes D_n. Since $\mathcal{F} \subseteq \mathcal{G}$, this x is among the rationalisers quantified in the PAC event above. Hence, on that event, $\mathsf{erf}_\mu(x, g) < \varepsilon$. This contradicts $\inf_{x' \in \mathcal{F}} \{\mathsf{erf}_\mu(x', g)\} > \varepsilon$: the fact that g lies at distance strictly greater than ε from $\mathcal{F}$. Hence, with probability at least $1 - \delta$, no element of $\mathcal{F}$ rationalizes D_n.

Define the size-zero test $\tilde{\mathcal{T}}_\mathcal{F}$ that rejects exactly those datasets that admit no rationalization in $\mathcal{F}$. Then, $\tilde{\mathcal{T}}_\mathcal{F}(D_n) = 1$ with probability at least $1-\delta$ under the alternative. By Definition 2, any RP characterization $\mathcal{T}_\mathcal{F}$ has maximal power among size-zero tests, and therefore $\mathcal{T}_\mathcal{F}(D_n) = 1$ whenever $\tilde{\mathcal{T}}_\mathcal{F}(D_n) = 1$. Thus

$$\mu_g^n\big(\mathcal{T}_\mathcal{F}^{-1}(1)\big) \;\geq\; 1 - \delta.$$

This establishes that $\mathcal{T}_\mathcal{F}$ has power at least $1 - \delta$ against all alternatives $g \in \mathcal{G} \setminus \mathcal{F}(\varepsilon)$. Moreover, by Definition 2 *(i)*, $\mathcal{T}_\mathcal{F}$ never rejects when $g \in \mathcal{F}$, so its size is zero.

B.3 Proof of Theorem 3.2

Fix $\varepsilon > 0$ and $n \in \mathbb{N}$. Let $\mathcal{T}$ be an RP characterisation of $\mathcal{F}$. By the PAC result for $\mathcal{G}$ (Lemma 1), for any $\delta \in (0,1)$ and any true model $g \in \mathcal{G}$, if $n \geq n(\varepsilon, \delta)$ then

$$\mu_g^n\Big(\mathsf{erf}_\mu(g,x) < \varepsilon \ \text{ for every } x \in \mathcal{G} \text{ that rationalises } D_n \Big) \ \geq \ 1-\delta. \tag{7}$$

For the fixed $\varepsilon > 0$ and n, write $S := \{\delta \in (0,1) : \ n(\varepsilon,\delta) \leq n\}$. By the non-triviality assumption stated in the text preceding the theorem, $S \neq \emptyset$. Define $\delta(\varepsilon, n) \ := \ \inf S$. By the definition of the infimum, for every $m \in \mathbb{N}$ there exists $\delta_m \in S$ such that $\delta(\varepsilon,n) \leq \delta_m < \delta(\varepsilon,n) + \frac{1}{m}$. Thus we may fix a sequence $(\delta_m)_{m\geq 1} \subset S$ with $\delta_m \downarrow \delta(\varepsilon,n)$ and

$$n(\varepsilon, \delta_m) \ \leq \ n \qquad \text{for all } m. \tag{8}$$

Fix any alternative $g \in \mathcal{G}$ satisfying the weak separation from $\mathcal{F}$,

$$\inf_{f \in \mathcal{F}} \mathsf{erf}_\mu(g,f) \ \geq \ \varepsilon.$$

For each m, since (8) implies $n \geq n(\varepsilon,\delta_m)$, applying (7) at confidence level $1-\delta_m$ yields $\mu_g^n(E_m) \ \geq \ 1-\delta_m$, where

$$E_m \ := \ \Big\{ D_n : \ \mathsf{erf}_\mu(g,x) < \varepsilon \ \text{ for every } x \in \mathcal{G} \text{ that rationalises } D_n \Big\}.$$

The remainder of the argument is identical to the corresponding step in the proof of Theorem 3.1: on E_m there cannot exist any $f \in \mathcal{F}$ that rationalises D_n, hence

$$E_m \ \subseteq \ \{D_n : \ \text{no } f \in \mathcal{F} \text{ rationalises } D_n\}.$$

Defining the canonical size–zero test $\tilde{\mathcal{T}}_{\mathcal{F}}(D_n) := \mathbf{1}\{\text{no } f \in \mathcal{F} \text{ rationalises } D_n\}$ and using the maximal-power property of RP characterisations (Definition 1 *(iii)*), we obtain for every m,

$$\mu_g^n\big(\mathcal{T}^{-1}(1) \big) \ \geq \ 1-\delta_m.$$

Letting $m \to \infty$ and using $\delta_m \downarrow \delta(\varepsilon,n)$ yields

$$\mu_g^n\big(\mathcal{T}^{-1}(1) \big) \ \geq \ 1-\delta(\varepsilon,n).$$

Since g was an arbitrary alternative with $\inf_{f\in\mathcal{F}} \mathsf{erf}_\mu(g,f) \geq \varepsilon$, this establishes the claimed power bound.

B.4 Proof of Theorem 3.3

Fix $n \in \mathbb{N}$ and $\delta \in (0,1)$ as in the non-triviality assumption stated in the text preceding the theorem, and fix an arbitrary $\gamma > 0$. Let $\mathcal{T}$ be an RP characterisation of $\mathcal{F}$. Define

$$\varepsilon(n,\delta) \;:=\; \inf\{\varepsilon > 0 :\; n(\varepsilon,\delta) \leq n\}, \qquad \bar{\varepsilon} \;:=\; \varepsilon(n,\delta) + \gamma.$$

Write $S \;:=\; \{\varepsilon > 0 :\; n(\varepsilon,\delta) \leq n\}$. By assumption, $S \neq \emptyset$. Since $\varepsilon(n,\delta) = \inf S$, by the definition of infimum there exists $\varepsilon' \in S$ such that $\varepsilon' < \bar{\varepsilon}$. In particular, $n(\varepsilon',\delta) \leq n$. Moreover, because $n(\varepsilon,\delta)$ is a *minimal* sample complexity, for each fixed δ the map $\varepsilon \mapsto n(\varepsilon,\delta)$ is weakly decreasing. Hence, since $\bar{\varepsilon} > \varepsilon'$, $n(\bar{\varepsilon},\delta) \;\leq\; n(\varepsilon',\delta) \;\leq\; n$.

By the PAC result for $\mathcal{G}$ (Lemma 1), for any true model $g \in \mathcal{G}$, if $n \geq n(\bar{\varepsilon},\delta)$ then

$$\mu_g^n\Big(\,\mathsf{erf}_\mu(g,x) < \bar{\varepsilon} \;\text{ for every } x \in \mathcal{G} \text{ rationalising } D_n\,\Big) \;\geq\; 1-\delta. \tag{9}$$

Define the event

$$E \;:=\; \Big\{D_n :\; \mathsf{erf}_\mu(g,x) < \bar{\varepsilon} \;\text{ for every } x \in \mathcal{G} \text{ rationalising } D_n\Big\}.$$

Then (9) reads $\mu_g^n(E) \geq 1-\delta$. Now fix any alternative $g \in \mathcal{G}$ satisfying

$$\inf_{f\in\mathcal{F}} \mathsf{erf}_\mu(g,f) \;\geq\; \bar{\varepsilon}. \tag{10}$$

We claim that on E there cannot exist any $f \in \mathcal{F}$ that rationalises D_n. Indeed, suppose by contradiction that for some $D_n \in E$ there exists $f \in \mathcal{F}$ rationalising D_n. Since $\mathcal{F} \subseteq \mathcal{G}$, this f is among the $\mathcal{G}$-rationalisers quantified in the definition of E, so $D_n \in E$ implies $\mathsf{erf}_\mu(f,g) < \bar{\varepsilon}$. This contradicts (10), since (10) implies $\mathsf{erf}_\mu(f,g) \geq \bar{\varepsilon}$ for every $f \in \mathcal{F}$. Hence

$$E \;\subseteq\; \{D_n :\; \text{no } f \in \mathcal{F} \text{ rationalises } D_n\}.$$

Therefore,

$$\mu_g^n\Big(\{D_n :\; \text{no } f \in \mathcal{F} \text{ rationalises } D_n\}\Big) \;\geq\; \mu_g^n(E) \;\geq\; 1-\delta.$$

Define the canonical size–zero test

$$\tilde{\mathcal{T}}_{\mathcal{F}}(D_n) \;:=\; \mathbf{1}\{\text{no } f \in \mathcal{F} \text{ rationalises } D_n\}.$$

By construction, $\tilde{\mathcal{T}}_{\mathcal{F}}$ never rejects under the null $x_{\succeq} \in \mathcal{F}$, so it has size zero. Moreover, the inequality above gives

$$\mu_g^n\big(\tilde{\mathcal{T}}_{\mathcal{F}}^{-1}(1)\big) \;\geq\; 1-\delta.$$

Finally, since $\mathcal{T}$ is an RP characterisation of $\mathcal{F}$, it has maximal power among size–zero tests, hence $\tilde{\mathcal{T}}_{\mathcal{F}}^{-1}(1) \subseteq \mathcal{T}^{-1}(1)$ and therefore

$$\mu_g^n\big(\mathcal{T}^{-1}(1)\big) \;\geq\; 1-\delta.$$

Since g was an arbitrary alternative satisfying (10), this completes the proof.

B.5 Proof of Theorem 4.1

We prove size and power separately. Throughout, let g denote the true demand generated by the true preference g, and let $x_{\succeq_n} \in \mathcal{G}$ be any rationalising demand selected by the analyst from the dataset D_n. Recall that Algorithm 1 fixes $\varepsilon, \delta > 0$, chooses $n = n(\varepsilon/t, \delta)$ with t large enough so that $\lambda(\varepsilon) - \gamma(\varepsilon/t) > \gamma(\varepsilon/t)$, and rejects whenever $\mathcal{R}(x_{\succeq_n}) > \gamma(\varepsilon/t)$.

(i) Size. Suppose the null holds, i.e. $g \in \mathcal{F}$. By the PAC result for $\mathcal{G}$ (Lemma 1), for the chosen $n = n(\varepsilon/t, \delta)$,

$$\mu_g^n\Big(\,\mathsf{erf}_\mu\big(x_{\succeq_n}, g\big) < \tfrac{\varepsilon}{t}\,\Big) \;\geq\; 1-\delta. \tag{11}$$

Because $g \in \mathcal{F}$, we have $\mathcal{R}(g) = 0$. Uniform continuity of $\mathcal{R}$ at 0 yields

$$\mathsf{erf}_\mu\big(x_{\succeq_n}, g\big) < \tfrac{\varepsilon}{t} \;\Longrightarrow\; \big|\mathcal{R}(x_{\succeq_n}) - \mathcal{R}(g)\big| < \gamma(\varepsilon/t), \tag{12}$$

hence $\mathcal{R}(x_{\succeq_n}) < \gamma(\varepsilon/t)$. Thus, on the event in (11), the test does *not* reject. Therefore the probability of rejection under the null is at most δ for all $g \in \mathcal{F}$. Hence, the supremum of this rejection probability must be smaller than this δ.

(ii) Power. Suppose the alternative holds, i.e. $g \in \mathcal{G} \setminus \mathcal{F}(\varepsilon)$, equivalently

$$\inf_{x_{\succeq'} \in \mathcal{F}} \mathsf{erf}_\mu\big(g, x_{\succeq'}\big) \;\geq\; \varepsilon. \tag{13}$$

By regularity of $\mathcal{R}$, condition (13) implies that the true demand must satisfy a numerical lower bound:

$$\mathcal{R}(g) \;\geq\; \lambda(\varepsilon). \tag{14}$$

That is, whenever g lies at least ε away from $\mathcal{F}$ in the erf_μ metric, the restriction functional assigns it a value no smaller than $\lambda(\varepsilon)$. Again by Lemma 1 with $n = n(\varepsilon/t, \delta)$,

$$\mu_g^n\Big(\mathsf{erf}_\mu\big(x_{\succeq_n}, g\big) < \frac{\varepsilon}{t}\Big) \;\geq\; 1 - \delta. \tag{15}$$

Uniform continuity gives, on the event in (15),

$$\big|\mathcal{R}(x_{\succeq_n}) - \mathcal{R}(g)\big| < \gamma(\varepsilon/t) \implies \mathcal{R}(x_{\succeq_n}) \;\geq\; \mathcal{R}(g) - \gamma(\varepsilon/t). \tag{16}$$

Combining (14) and (16) yields

$$\mathcal{R}(x_{\succeq_n}) \;\geq\; \lambda(\varepsilon) - \gamma(\varepsilon/t).$$

By the choice of t, $\lambda(\varepsilon) - \gamma(\varepsilon/t) > \gamma(\varepsilon/t)$, so the decision rule in Algorithm 1 rejects on the event in (15). Hence the probability of rejection under the alternative is at least $1-\delta$. The two parts establish the stated size and power bounds, completing the proof.

B.6 Proof of Theorem 4.2

Fix $\varepsilon, \delta > 0$ and let $n = n(\varepsilon, \delta)$. Let $g \in \mathcal{G}$ denote the true demand and let D_n be the dataset of size n sampled from μ under g. Let $x_\succeq \in \mathcal{G}$ be any demand that rationalises D_n. By PAC learnability of $\mathcal{G}$ (Lemma 1),

$$\mu_g^n\Big(\mathsf{erf}_\mu\big(x_\succeq, g\big) < \varepsilon \;\text{ for every } x_\succeq \in \mathcal{G} \text{ that rationalises } D_n\Big) \;\geq\; 1 - \delta.$$

Since $\mathcal{R}$ is uniformly continuous, for every pair $(x_\succeq, g)$ with $\mathsf{erf}_\mu(x_\succeq, g) < \varepsilon$ we have

$$\big|\mathcal{R}(x_\succeq) - \mathcal{R}(g)\big| \;<\; \gamma(\varepsilon).$$

Define $\Delta W := \mathcal{R}(x_\succeq)$ and $\Delta W^* := \mathcal{R}(g)$. Then

$$\mu_g^n\big(|\Delta W - \Delta W^*| < \gamma(\varepsilon)\big) \;\geq\; 1 - \delta,$$

which is the claimed finite-sample bound.

B.7 Proof of Proposition 2

If $\succeq$ is weakly separable across (G_1, G_2), then there exists a scalar $\kappa : \mathcal{P} \times \mathcal{I} \to \mathbb{R}$ such that

$$\mathcal{S}^{12}(x_\succeq)(p, I) = \kappa(p, I)\, a(p, I)\, b(p, I)^\top \qquad \forall (p, I) \in \mathcal{P} \times \mathcal{I}. \tag{17}$$

Equivalently, for all $i, i' \in G_1$ and $j, j' \in G_2$,

$$\mathcal{S}_{ij}(x_\succeq)\, \partial_I x_{\succeq,i'}\, \partial_I x_{\succeq,j'} - \mathcal{S}_{i'j'}(x_\succeq)\, \partial_I x_{\succeq,i}\, \partial_I x_{\succeq,j} = 0. \tag{18}$$

Define $\mathcal{R}^{\text{sep}}(x_\succeq)$ to collect the left–hand side of (3). Then weak separability implies $\mathcal{R}^{\text{sep}}(x_\succeq) \equiv 0$.

Conversely, suppose the nondegeneracy condition holds. If $\mathcal{R}^{\text{sep}}(x_\succeq) \equiv 0$, then there exists κ such that (17) holds. By Goldman and Uzawa [1964], this is equivalent to weak separability across (G_1, G_2). Since $\mathcal{S}^{12}(x_\succeq)$ and a, b are built from $\mathsf{D}_p x_\succeq$ and $\partial_I x_\succeq$, uniform continuity of $\mathcal{R}^{\text{sep}}$ into L^1_μ follows from Proposition 6.

B.8 Proof of Proposition 1

Throughout the proof we write x for the demand $x_\succeq$ generated by $\succeq$. By strong monotonicity the budget binds at every (p, I), and by strict convexity the solution is unique and interior. Fix $p \in \mathcal{P}$ and write $g(I) := x(p, I)$. Demand is C^2 in (p, I).

(1 $\Rightarrow$ 2). If $\succeq$ is homothetic, then for any $t > 0$ the scaled budget set satisfies $\{x : p \cdot x \leq tI\} = t\{x : p \cdot x \leq I\}$, and $x(p, tI) = t\, x(p, I)$. Fix $I > 0$ and define $h(t) := x(p, tI)$. Then $h(t) = t\, h(1)$. By the chain rule, $h'(t) = \partial_I x(p, tI) \cdot I$ and $h''(t) = \partial_I^2 x(p, tI) \cdot I^2$. Differentiating $h(t) = t\, h(1)$ gives $h'(t) = h(1)$ and $h''(t) = 0$. Evaluating at $t = 1$ yields $\partial_I^2 x(p, I) = 0$ for all $I > 0$. Continuity extends this to $I = 0$ if $0 \in \mathcal{I}$. As p was arbitrary, $\partial_I^2 x \equiv 0$ on $\mathcal{P} \times \mathcal{I}$.

(2 $\Rightarrow$ 1). Assume $\partial_I^2 x \equiv 0$. For each fixed p, $I \mapsto x(p, I)$ is affine, so there exist $a(p), b(p) \in \mathbb{R}^\ell$ with

$$x(p, I) = I\, a(p) + b(p) \quad \text{for all } (p, I).$$

At income $I = 0$, the only affordable bundle under strictly positive prices is 0, and strong monotonicity implies $x(p, 0) = 0$. Hence $b(p) = 0$ and $x(p, I) = I\, a(p)$. Therefore $x(p, tI) = t\, x(p, I)$ for all $t > 0$, i.e., Marshallian demand scales linearly with income at fixed p. Consider the revealed preference induced by $x(\cdot, \cdot)$. If $x(p, I)$ is (directly) revealed preferred to some affordable y at (p, I), then at (p, tI) the feasible set is t times larger and the chosen bundle is $x(p, tI) = t\, x(p, I)$, while $t\, y$ is feasible. Thus the revealed preference respects radial scaling. By standard rationalisation results on compact domains, there exists a continuous, monotone, homothetic preference that rationalises x. Hence $\succeq$ is homothetic.

Uniform continuity of $\mathcal{R}^{\text{hom}}$ follows from Proposition 6. For completeness, for any x, y we have

$$\|\mathcal{R}^{\text{hom}}(x) - \mathcal{R}^{\text{hom}}(y)\|_{L^1_\mu} = \int_{\mathcal{P}\times\mathcal{I}} \left\|\partial_I^2 (x-y)(p, I)\right\| \mathrm{d}\mu(p, I) \le \|\partial_I^2 (x-y)\|_{L^\infty} \le \|x-y\|_{C^{2,\infty}},$$

so $\mathcal{R}^{\text{hom}} : (\mathcal{G}, \|\cdot\|_{C^{2,\infty}}) \to L^1_\mu$ is a bounded linear map (norm ≤ 1), hence uniformly continuous.

B.9 Proof of Proposition 2

Throughout fix $(p, I) \in \mathcal{P} \times \mathcal{I}$. For $i \in G_1$ and $j \in G_2$ recall

$$\mathcal{S}_{ij}(x_\succeq) \;=\; \frac{\partial x_{\succeq,i}}{\partial p_j} \;+\; x_{\succeq,j}\, \frac{\partial x_{\succeq,i}}{\partial I}, \qquad a_i \;:=\; \partial_I x_{\succeq,i}, \qquad b_j \;:=\; \partial_I x_{\succeq,j},$$

and let $\mathcal{S}^{12}(x_\succeq)$ denote the between–group Slutsky block with entries $\mathcal{S}_{ij}(x_\succeq)$ for $i \in G_1$, $j \in G_2$. We will suppress the explicit (p, I)–dependence to lighten notation.

(Necessity). Assume $\succeq$ is weakly separable across (G_1, G_2), that is, there exist subutility indices $v_1 : \mathbb{R}_+^{|G_1|} \to \mathbb{R}$, $v_2 : \mathbb{R}_+^{|G_2|} \to \mathbb{R}$ and an aggregator $w : \mathbb{R}^2 \to \mathbb{R}$ such that

$$u(x) \;=\; w\big(v_1(x_{G_1}),\, v_2(x_{G_2})\big) \quad \text{represents} \;\succeq .$$

Under our C^2 and interiority assumptions, the Hicksian compensated problem is well defined and the Slutsky matrix is the Jacobian of Hicksian demand: $\mathcal{S} = \big(\partial h/\partial p\big)^\top$. By the classical Goldman–Uzawa factorisation for weak separability, the between–group block of the Hicksian substitution matrix has rank one, with entries proportional to

the cross–group income effects,[24] hence there exists a scalar function $\kappa = \kappa(p, I)$ such that for all $i \in G_1$, $j \in G_2$,

$$\mathcal{S}_{ij}(x_{\succeq}) \;=\; \kappa\, a_i\, b_j. \tag{19}$$

In matrix form, (19) reads $\mathcal{S}^{12}(x_{\succeq}) = \kappa\, a\, b^{\top}$, which is (17). Moreover, fixing arbitrary $i, i' \in G_1$ and $j, j' \in G_2$ and using (19),

$$\mathcal{S}_{ij}\, a_{i'}\, b_{j'} \;-\; \mathcal{S}_{i'j'}\, a_i\, b_j \;=\; \kappa\, a_i b_j a_{i'} b_{j'} \;-\; \kappa\, a_{i'} b_{j'} a_i b_j \;=\; 0,$$

so each component collected by $\mathcal{R}^{\mathrm{sep}}(x_{\succeq})$ vanishes. This proves the necessity claims.

We record the elementary equivalence we will use in sufficiency.

Lemma 2. *Let $A \in \mathbb{R}^{m\times n}$, $u \in \mathbb{R}^m$, $v \in \mathbb{R}^n$. Suppose u and v are not identically zero and that for all row indices $r, r' \in \{1, \dots, m\}$ and column indices $c, c' \in \{1, \dots, n\}$,*

$$A_{rc}\, u_{r'}\, v_{c'} \;-\; A_{r'c'}\, u_r\, v_c \;=\; 0. \tag{20}$$

If there exist r_0, c_0 with $u_{r_0} \neq 0$ and $v_{c_0} \neq 0$, then setting $\lambda := A_{r_0c_0}/(u_{r_0}v_{c_0})$ yields $A = \lambda\, u\, v^{\top}$.

Proof. Fix any r, c. Plugging $(r', c') = (r_0, c_0)$ into (20) gives

$$A_{rc}\, u_{r_0}\, v_{c_0} \;=\; A_{r_0c_0}\, u_r\, v_c.$$

Since $u_{r_0}v_{c_0} \neq 0$, we may divide both sides to obtain $A_{rc} = \lambda\, u_r v_c$ with $\lambda = A_{r_0c_0}/(u_{r_0}v_{c_0})$. This holds for all r, c, hence $A = \lambda\, u\, v^{\top}$. □

(Sufficiency). Assume the two–sided nondegeneracy condition: for the given (p, I) there exist $i_0 \in G_1$ and $j_0 \in G_2$ with $a_{i_0} \neq 0$ and $b_{j_0} \neq 0$. Assume further that $\mathcal{R}^{\mathrm{sep}}(x_{\succeq}) \equiv 0$, i.e., for all $i, i' \in G_1$ and $j, j' \in G_2$,

$$\mathcal{S}_{ij}(x_{\succeq})\, a_{i'}\, b_{j'} \;-\; \mathcal{S}_{i'j'}(x_{\succeq})\, a_i\, b_j \;=\; 0. \tag{21}$$

[24] A clean way to see this is to (i) define the group expenditure functions $e_1(p_{G_1}, y_1) := \min\{p_{G_1} \cdot x_{G_1} : v_1(x_{G_1}) \geq y_1\}$ and $e_2(p_{G_2}, y_2)$; (ii) write the overall expenditure function as $e(p, u) = \min_{y_1,y_2}\{e_1(p_{G_1}, y_1) + e_2(p_{G_2}, y_2) : w(y_1, y_2) \geq u\}$; (iii) apply the envelope theorem and the FOCs for the (y_1, y_2)–minimisation to express the cross–group compensated derivatives $\partial h_i/\partial p_j$ as a common scalar (depending on p, I) times a product of group–specific terms. Identifying those group–specific terms with a_i and b_j (via Roy/Slutsky identities) yields the display below.

Apply Lemma 2 with $A = \mathcal{S}^{12}(x_{\succeq})$, $u = a$, $v = b$, and $(r_0, c_0) = (i_0, j_0)$. The hypothesis (20) is exactly (21). By nondegeneracy, $a_{i_0} b_{j_0} \neq 0$, so Lemma 2 yields a scalar

$$\kappa \;:=\; \frac{\mathcal{S}_{i_0 j_0}(x_{\succeq})}{a_{i_0}\, b_{j_0}} \quad \text{such that} \quad \mathcal{S}^{12}(x_{\succeq}) \;=\; \kappa\, a\, b^{\top}.$$

Equivalently, for all $i \in G_1$, $j \in G_2$, $\mathcal{S}_{ij}(x_{\succeq}) = \kappa\, a_i b_j$, which is (17).

By Goldman and Uzawa [1964], the rank–one factorisation of the between–group block $\mathcal{S}^{12}(x_{\succeq}) = \kappa\, a\, b^{\top}$ for all (p, I) is equivalent to weak separability across (G_1, G_2). This completes the sufficiency part.

Finally, $\mathcal{R}^{\text{sep}}$ is built from finite sums of products of terms of the form $\mathsf{D}_p x_{\succeq}$ and $\partial_I x_{\succeq}$. By Proposition 6, the maps $x \mapsto \mathsf{D}_p x$ and $x \mapsto \partial_I x$ are bounded linear maps into L^1_μ, hence uniformly continuous. On a compact domain, pointwise products and finite sums of uniformly continuous maps are uniformly continuous into L^1_μ, so $\mathcal{R}^{\text{sep}}$ is uniformly continuous into L^1_μ as claimed.

B.10 Proof of Proposition 3

Let $(\cdot)^+ : \mathbb{R} \to \mathbb{R}_{\geq 0}$ denote the positive–part map, $t^+ := \max\{0, t\}$. We will use the elementary equivalence

$$t^+ = 0 \quad \Longleftrightarrow \quad t \leq 0, \tag{22}$$

valid for every $t \in \mathbb{R}$.

(Gross complementarity). Fix distinct $i \neq j$. By definition, goods i and j are *gross complements* iff

$$\frac{\partial x_{\succeq,i}(p, I)}{\partial p_j} \;\leq\; 0 \qquad \text{for all } (p, I) \in \mathcal{P} \times \mathcal{I}. \tag{23}$$

The proposed functional restriction is

$$\mathcal{R}^{\text{grossC}}_{ij}(x_{\succeq})(p, I) := \left(\tfrac{\partial x_{\succeq,i}(p,I)}{\partial p_j}\right)^{+} = \max\Big\{0,\; \tfrac{\partial x_{\succeq,i}(p,I)}{\partial p_j}\Big\}.$$

($\Rightarrow$) If (23) holds, then for each fixed (p, I) we have $\partial x_{\succeq,i}/\partial p_j \leq 0$, hence, by (22), $\big(\partial x_{\succeq,i}/\partial p_j\big)^{+} = 0$. Therefore $\mathcal{R}^{\text{grossC}}_{ij}(x_{\succeq})(p, I) = 0$ for all (p, I), i.e. $\mathcal{R}^{\text{grossC}}_{ij}(x_{\succeq}) \equiv 0$.

($\Leftarrow$) Conversely, if $\mathcal{R}^{\text{grossC}}_{ij}(x_{\succeq}) \equiv 0$, then for each (p, I) we have $\big(\partial x_{\succeq,i}/\partial p_j\big)^{+} = 0$. By (22), this implies $\partial x_{\succeq,i}/\partial p_j \leq 0$ pointwise, proving (23). Hence i and j are gross complements.

Combining the two implications yields

$$i, j \text{ are gross complements} \quad \Longleftrightarrow \quad \mathcal{R}_{ij}^{\mathrm{grossC}}(x_{\succeq}) \equiv 0.$$

(Analogous forms).

- *Gross substitutability.* Define $\mathcal{R}_{ij}^{\mathrm{grossS}}(x_{\succeq})(p, I) := \big(- \partial x_{\succeq,i}(p, I)/\partial p_j \big)^+$. Repeating the previous argument with $t = -\partial x_{\succeq,i}/\partial p_j$ shows

$$i, j \text{ are gross substitutes} \quad \Longleftrightarrow \quad \mathcal{R}_{ij}^{\mathrm{grossS}}(x_{\succeq}) \equiv 0.$$

- *Net complementarity/substitutability.* Let $h(p, u)$ denote Hicksian demand and $\mathcal{S}_{ij}(x_{\succeq})(p, I) := \partial x_{\succeq,i}/\partial p_j + x_{\succeq,j}\, \partial x_{\succeq,i}/\partial I$ the Slutsky term. The Hicks–Allen definition says that i is a net complement (substitute) of j iff $\partial h_i(p, u)/\partial p_j \leq 0$ (resp. ≥ 0) for all (p, u). By the Slutsky decomposition, $\partial h_i/\partial p_j = \mathcal{S}_{ij}(x_{\succeq})$ evaluated at the Marshallian pair (p, I) corresponding to (p, u). Hence the two sign conditions are equivalent to $\mathcal{S}_{ij}(x_{\succeq})(p, I) \leq 0$ and $\mathcal{S}_{ij}(x_{\succeq})(p, I) \geq 0$, respectively, for all (p, I). Define

$$\mathcal{R}_{ij}^{\mathrm{netC}}(x_{\succeq})(p, I) := \big(\mathcal{S}_{ij}(x_{\succeq})(p, I)\big)^+, \qquad \mathcal{R}_{ij}^{\mathrm{netS}}(x_{\succeq})(p, I) := \big(- \mathcal{S}_{ij}(x_{\succeq})(p, I)\big)^+.$$

  Applying (22) pointwise yields

$$\begin{aligned} i, j \text{ are net complements} \quad &\Longleftrightarrow \quad \mathcal{R}_{ij}^{\mathrm{netC}}(x_{\succeq}) \equiv 0, \\ i, j \text{ are net substitutes} \quad &\Longleftrightarrow \quad \mathcal{R}_{ij}^{\mathrm{netS}}(x_{\succeq}) \equiv 0. \end{aligned}$$

Finally, note that on compact $\mathcal{P} \times \mathcal{I}$, the map $t \mapsto t^+$ is 1–Lipschitz, so composing it with $(p, I) \mapsto \partial x_{\succeq,i}/\partial p_j$ or $(p, I) \mapsto \mathcal{S}_{ij}(x_{\succeq})$ preserves uniform continuity. Since these derivatives are uniformly continuous into L^1_μ by Proposition 6, the same holds for $\mathcal{R}_{ij}^{\mathrm{grossC}}$, $\mathcal{R}_{ij}^{\mathrm{grossS}}$, $\mathcal{R}_{ij}^{\mathrm{netC}}$, and $\mathcal{R}_{ij}^{\mathrm{netS}}$. This completes the proof.

B.11 Proof of Proposition 4

Throughout, let $u : \mathbb{R}_{>0} \times [0,1] \to \mathbb{R}$ be the Bernoulli index in the betweenness representation. For a function of two variables $u(x,v)$ we use the standard notation:

$$u_1(x,v) = \frac{\partial u}{\partial x}(x,v), \qquad u_2(x,v) = \frac{\partial u}{\partial v}(x,v),$$

and

$$u_{11}(x,v) = \frac{\partial^2 u}{\partial x^2}(x,v), \quad u_{12}(x,v) = \frac{\partial^2 u}{\partial x \partial v}(x,v), \quad u_{22}(x,v) = \frac{\partial^2 u}{\partial v^2}(x,v),$$

whenever these derivatives exist.

Fix $(\mathbf{p}, I, \boldsymbol{\pi})$ and write $\mathbf{x} = x_{\succeq}(\mathbf{p}, I, \boldsymbol{\pi}) \in \mathbb{R}^S_{>0}$ for the interior maximiser of $V(\mathbf{x}; \boldsymbol{\pi})$ subject to $\mathbf{p}\cdot\mathbf{x} \leq I$. Monotonicity of u implies the budget binds, and strict quasi–concavity yields uniqueness.

Define $F(\mathbf{x}, V; \boldsymbol{\pi}) := \sum_{t=1}^S \pi_t\, u(x_t, V) - V$. By (4), $F(\mathbf{x}, V(\mathbf{x}; \boldsymbol{\pi}); \boldsymbol{\pi}) = 0$. Differentiate this identity with respect to x_s. Since only the term $t = s$ depends directly on x_s, we obtain

$$\pi_s\, u_1(x_s, V) \;+\; \Big(\sum_{t=1}^{S} \pi_t\, u_2(x_t, V)\Big)\, V_{x_s} \;-\; V_{x_s} \;=\; 0.$$

Hence

$$V_{x_s} = \frac{\pi_s\, u_1(x_s, V)}{1 - \sum_{t=1}^S \pi_t\, u_2(x_t, V)}\,. \tag{24}$$

At an interior optimum the first–order conditions imply $V_{x_s} = \lambda p_s$ for some Lagrange multiplier $\lambda > 0$, so V_{x_s} is finite and strictly positive. Therefore the denominator in (24) is nonzero at the optimum.

From $V_{x_s} = \lambda p_s$ and (24), taking the ratio s to 1 cancels the common denominator and yields

$$\frac{\pi_s\, u_1(x_s, V)}{\pi_1\, u_1(x_1, V)} = \frac{p_s}{p_1} \quad \Longleftrightarrow \quad \frac{u_1(x_1, V)}{u_1(x_s, V)} = \frac{\pi_s}{\pi_1}\frac{p_1}{p_s} =: k_s. \tag{25}$$

In particular, for $s = 2$ we have

$$\frac{u_1(x_1, V)}{u_1(x_2, V)} = k_2. \tag{26}$$

By Dekel [1986, p. 314], the fixed point $V(\mathbf{x}; \boldsymbol{\pi})$ is unique for each $(\mathbf{x}, \boldsymbol{\pi})$. Hence the

unique value $V = V(\mathbf{x}; \boldsymbol{\pi})$ at the optimum satisfies (26). Next, define

$$\phi(v; x_1, x_2, k_2) := u_1(x_1, v) - k_2\, u_1(x_2, v).$$

At the realised optimum, $v = V(\mathbf{x}; \boldsymbol{\pi})$ solves $\phi(v; x_1, x_2, k_2) = 0$ by (26). Assume the following local regularity holds at the realised point:

$$\partial_v \phi\big(V(\mathbf{x}; \boldsymbol{\pi}); x_1, x_2, k_2\big) = \; u_{12}(x_1, V) - k_2\, u_{12}(x_2, V) \; \neq \; 0. \tag{27}$$

This is a mild pointwise condition that ensures the root of $\phi(\cdot; x_1, x_2, k_2)$ is locally unique as a function of (x_1, x_2, k_2). Under (27), the Implicit Function Theorem gives a C^1 map

$$V^* = V^*(x_1, x_2, k_2) \quad \text{with} \quad \phi\big(V^*(x_1, x_2, k_2); x_1, x_2, k_2\big) = 0$$

in a neighbourhood of the realised (x_1, x_2, k_2). In particular, at the optimum $V^*(x_1, x_2, k_2) = V(\mathbf{x}; \boldsymbol{\pi})$. Using (25) with $s \geq 3$ and substituting V^*,

$$\frac{u_1(x_s, V^*)}{u_1(x_2, V^*)} = \frac{k_2}{k_s}. \tag{28}$$

For each fixed v, the map $x \mapsto u_1(x, v)$ is strictly decreasing and continuous because $u_1 > 0$ and $u_{11} < 0$. Hence it admits a continuous inverse in its first argument, denoted $h(\cdot\,; v)$. Invert (28) in x_s to obtain

$$x_s = h\left(\frac{k_2}{k_s}\, u_1(x_2, V^*(x_1, x_2, k_2)) \;;\; V^*(x_1, x_2, k_2)\right) =: f(x_1, x_2, k_2, k_s).$$

Monotonicity in k_s follows because $k_s \mapsto (k_2/k_s)$ is strictly decreasing and $r \mapsto h(r; v)$ is strictly decreasing, so the composition is strictly increasing in k_s. For the normalisation, if $x_1 = x_2 = x$ and $k_s = 1$, then $u_1(x_s, V^*) = u_1(x, V^*)$ and therefore $x_s = x$, which gives $f(x, x, k_2, 1) = x$. Since the construction holds at each $(\mathbf{p}, I, \boldsymbol{\pi})$, we have

$$x_{\succeq,s}(\mathbf{p}, I, \boldsymbol{\pi}) = f\big(x_{\succeq,1}(\mathbf{p}, I, \boldsymbol{\pi}),\, x_{\succeq,2}(\mathbf{p}, I, \boldsymbol{\pi}),\, k_2(\mathbf{p}, \boldsymbol{\pi}),\, k_s(\mathbf{p}, \boldsymbol{\pi})\big)$$

with f strictly increasing in k_s and satisfying $f(x, x, k_2, 1) = x$.

B.12 Proof of Proposition 5

Fix $s \geq 3$. By Proposition 4, locally $x_{\succeq,s} = f\big(x_{\succeq,1},\, x_{\succeq,2},\, k_2,\, k_s\big)$, with f as constructed in the proof of the result. Consider the set of admissible perturbations of $(\mathbf{p}, I)$ that keep k_2 fixed to first order. Formally, let $\Psi : \mathbb{R}^m \to \mathbb{R}^{S+1}$ be a C^1 local parametrisation of this subspace, with coordinates $\theta \in \mathbb{R}^m$, so that $(\mathbf{p}, I) = (\mathbf{p}(\theta), I(\theta))$ and $\frac{\mathrm{d}}{\mathrm{d}x} k_2(\mathbf{p}(\theta), \boldsymbol{\pi})\big|_{\theta=0} = 0$ in every coordinate direction. Define

$$H_s(\theta) := \big(x_{\succeq,1}(\mathbf{p}(\theta), I(\theta), \boldsymbol{\pi}),\ x_{\succeq,2}(\mathbf{p}(\theta), I(\theta), \boldsymbol{\pi}),\ k_s(\mathbf{p}(\theta), \boldsymbol{\pi}),\ x_{\succeq,s}(\mathbf{p}(\theta), I(\theta), \boldsymbol{\pi})\big).$$

Let J_s be the $4 \times m$ Jacobian of H_s with respect to θ at $\theta = 0$. By the chain rule applied to $x_{\succeq,s} = f(x_{\succeq,1}, x_{\succeq,2}, k_2, k_s)$ and using $\mathrm{d}k_2 = 0$ along Ψ,

$$\mathrm{d}x_{\succeq,s} = f_{x_1}\, \mathrm{d}x_{\succeq,1} \ + \ f_{x_2}\, \mathrm{d}x_{\succeq,2} \ + \ f_{k_s}\, \mathrm{d}k_s.$$

Therefore the fourth row of J_s is a linear combination of the first three rows with coefficients $(f_{x_1}, f_{x_2}, f_{k_s})$. It follows that $\operatorname{rank}(J_s) \leq 3$, so every 4×4 minor of J_s vanishes. Defining

$$\mathcal{R}_s^{\mathrm{bet}}(x_{\succeq})(\mathbf{p}, I, \boldsymbol{\pi}) := \sum_{M \in \mathscr{M}_{4\times 4}(J_s)} (\det M)^2$$

gives $\mathcal{R}_s^{\mathrm{bet}} \equiv 0$ as a necessary condition for betweenness rationalisability when the Jacobian is evaluated along directions with $\mathrm{d}k_2 = 0$.

Appendix C Simulations

In this section, we first present the pertubation we used in our simulations and prove its properties. As stated in Section 3.2, our results apply with mean-zero measurement error. As such, we include simulations and power curves for this extension.

C.1 Perturbation of Cobb–Douglas Demand

We consider demands of the form $y = x + \varepsilon f$, where f is a perturbation normalized so that the Frobenius norm of the induced skew–symmetric Slutsky component satisfies

$$\| \operatorname{Skew}(S^f) \|_F = 1.$$

Thus ε indexes the distance from Slutsky symmetry, while different choices of f allow us to explore different patterns of violation. Let

$$f^A(p,I) = \begin{pmatrix} g_{12}(p,I) \\ -\frac{p_1}{p_2} g_{12}(p,I) \\ 0 \\ 0 \end{pmatrix}, \qquad f^B(p,I) = \begin{pmatrix} 0 \\ 0 \\ g_{34}(p,I) \\ -\frac{p_3}{p_4} g_{34}(p,I) \end{pmatrix},$$

with

$$g_{12}(p,I) = \left(\frac{I}{p_1}\right)^2 \frac{p_2}{p_1+p_2}, \qquad g_{34}(p,I) = \left(\frac{I}{p_3}\right)^2 \frac{p_4}{p_3+p_4}.$$

In the simulations presented in the main text, we considered the perturbation $f = f^A + f^B$. As a robustness check, we also considered the perturbation $f = \alpha f^A + (1-\alpha) f^B$ with $\alpha = 0.1$. In what follows, we prove the properties for the perturbation and detail the normalization.

Homogeneity of degree zero. Each block is homogeneous of degree zero jointly in (p, I). Indeed, for any $\lambda > 0$, we have $g_{12}(\lambda p, \lambda I) = g_{12}(p, I)$ and $g_{34}(\lambda p, \lambda I) = g_{34}(p, I)$ such that $f^A(\lambda p, \lambda I) = f^A(p, I)$, $f^B(\lambda p, \lambda I) = f^B(p, I)$, and $f(\lambda p, \lambda I) = f(p, I)$.

Budget neutrality. Each block is budget neutral:

$$p \cdot f^A = p_1 g_{12} - p_1 g_{12} = 0, \qquad p \cdot f^B = p_3 g_{34} - p_3 g_{34} = 0.$$

Hence, for $f = f^A + f^B$ and $f = \alpha f^A + (1-\alpha) f^B$, we have $p \cdot f = 0$.

Violation of Slutsky symmetry. Consider the perturbed demand

$$y(p,I) = x(p,I) + f(p,I),$$

where $x(p, I)$ are the base Cobb-Douglas demands and $f(p, I)$ is the perturbation designed to violate integrability. The corresponding Slutsky matrix is

$$S(p,I) = \mathsf{D}_p y(p,I) + y(p,I) \left(\partial_I y(p,I)\right)^\top,$$

where $\mathsf{D}_p y$ and $\partial_I y$ denote the derivatives of y with respect to prices and income, respectively. For the first perturbation, the induced asymmetries in the Slutsky matrix are

$$S^A_{12} - S^A_{21} = \frac{I^2(\alpha_1 + \alpha_2 - 1)}{p_1^2(p_1 + p_2)}, \qquad S^B_{34} - S^B_{43} = \frac{I^2(\alpha_3 + \alpha_4 - 1)}{p_3^2(p_3 + p_4)}.$$

All other off-diagonal entries of the skew part vanish. Therefore, the Slutsky matrix is generically asymmetric, and the asymmetry is split between $(1, 2)$ and $(3, 4)$. For the second perturbation, the argument is identical but with the scale α and $1 - \alpha$ on the asymmetry $(1, 2)$ and $(3, 4)$, respectively.

Normalization to control the magnitude of deviation. Each perturbation f is scaled so that the Frobenius norm of the induced skew–symmetric Slutsky component equals one:

$$\hat{f} = \frac{f}{\| \operatorname{Skew}(S^f) \|_F},$$

where S^f denote the Slutsky matrix for demand with perturbation f and

$$\operatorname{Skew}(S^f) = \frac{1}{2}\big(S^f - (S^f)^\top\big)$$

is its skew–symmetric part. After this rescaling, we replace f by $\tilde{f}$ in the perturbed demand:

$$y = x + \varepsilon \tilde{f}.$$

After this step, ε measures the total magnitude of the violation from rationality in the same way across perturbations.

C.2 Additional Regressions

This section includes a log-linear regression and a log-linear regression with reciprocal predictor $(1/\varepsilon)$ to complement Table 1. The results for the log-linear regression are presented in Table 4

Table 4: Log-linear regression estimates from power curves

Parameter	SUM	H	WS	EU
β_0	4.2168	2.4133	3.2303	2.6499
β_1	-4.2316	-0.0870	-3.0240	-4.8806

Table 4 shows that the required sample size decreases progressively from SUM to WS, then EU, and finally HARP. The table also shows that SUM and EU have higher sensitivity of required sample sizes to deviations from rationality than H and WS. Since sample size is modeled on a log scale, EU has a steeper slope (β_1) in terms of relative changes even though its curve appears flatter in absolute sample size due to a lower baseline sample size (β_0) compared with WS. Next, the results for a log-linear regression with a reciprocal preditor are presented in Table 5.

Table 5: Log-linear regression reciprocal estimates from power curves

Parameter	SUM	H	WS	EU
β_0	3.3181	2.3873	2.5749	1.5717
β_1	0.02791	0.001209	0.02107	0.03574

Table 5 shows that SUM WS and EU require substantially larger sample sizes to detect small deviations from rationality compared to H which is relatively insensitive to the size of deviations.

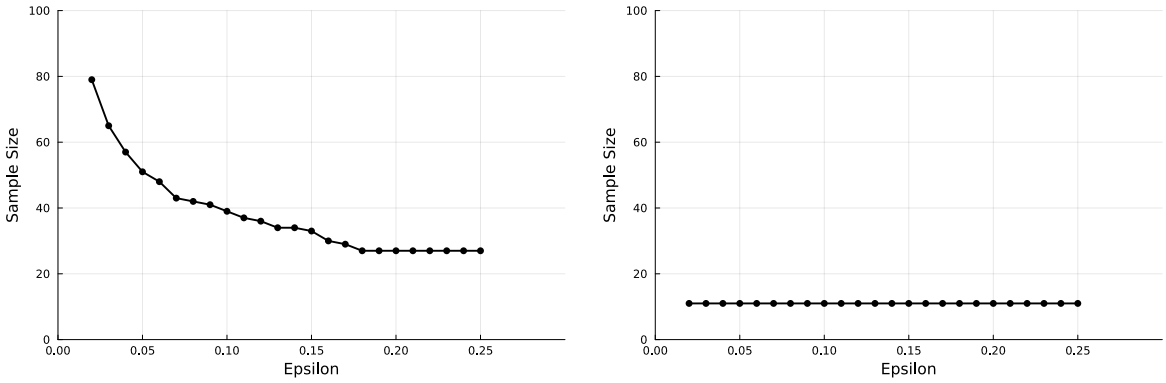
C.3 Power with Additive Measurement Error

To model measurement error in observed demands, we add mean-zero additive noise to each demand component:

$$\tilde{x}_k = x_k + m_k,$$

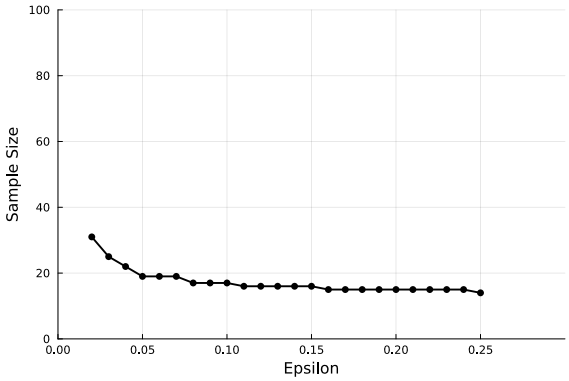
where the measurement error m_k is drawn independently according to a truncated normal distribution with standard deviation 0.25 and support set to $\pm 50\%$ of the demand x_k. This support implies that the noise is heteroskedastic and proportional to the magnitude of the demand. The noisy demands are scaled back to ensure the budget constraint holds. The data generating process and the perturbation are otherwise identical to that in the main text. In particular, the perturbation is computed for the demand x_k. The power curves are reported in Figure 4

The power curves are qualitatively similar to those in the main text, thus empirically confirming that the presence of mean-zero measurement error is inconsequential for our results.

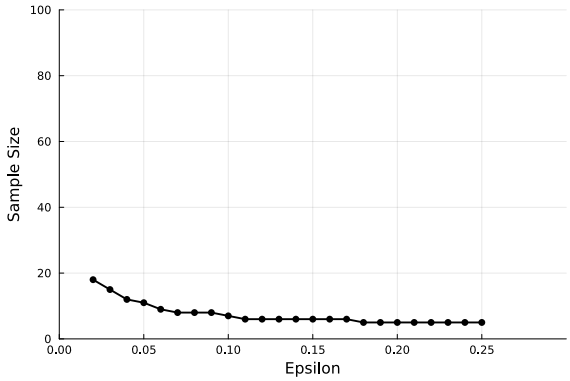


(a) SUM

(b) H



(c) WS



(d) EU

Figure 4: Power curves for SUM, H, WS, and EU.