The geometry of consumer preference aggregation*

Fedor Sandomirskiy (California Institute of Technology)†
Philip Ushchev (ECARES, Université libre de Bruxelles)

January 1, 2023
the latest version: **link**

Abstract

This paper revisits a classical question in economics: how do individual preferences and incomes of consumers shape aggregate behavior? We develop a method that reduces the hard problem of aggregation to simply computing a weighted average. The method applies to populations with homothetic preferences. The key idea is to handle aggregation in the space of logarithmic expenditure functions.

We demonstrate the power of this method by (*i*) characterizing classes of preferences invariant with respect to aggregation, i.e., such that any population of heterogeneous consumers with preferences from the class behaves as if it were a single aggregate consumer from the same class; (*ii*) characterizing classes of aggregate preferences generated by popular preference domains such as linear or Leontief; (*iii*) describing indecomposable preferences, i.e., those that do not correspond to aggregate behavior of any non-trivial population; (*iv*) representing any preference as an aggregation of indecomposable ones.

We discuss connections and applications of our findings to robust welfare analysis, information design, stochastic discrete choice, pseudo-market mechanisms, and preference identification.

*The paper has benefited from our discussions with (in alphabetic order) Michael Amior, Alexander W. Bloedel, Anna Bogomolnaia, Peter Caradonna, Tommaso Denti, Bram De Rock, Federico Echenique, David Genesove, Ben Golub, Aram Grigoryan, Yoram Halevy, Eran Hoffmann, Matthew O. Jackson, Peter Klibanoff, Ilan Kremer, Alexey Kushnir, Arthur Lewbel, Kiminori Matsuyama, John Moore, Hervé Moulin, Ichiro Obara, Luciano Pomatto, Joseph Root, Ariel Rubinstein, Kota Saito, Itay Saporta-Eksten, Ilya Segal, Roberto Serrano, Denis Shishkin, Ali Shourideh, Joel Sobel, Ran Spiegler, Ross Starr, Omer Tamuz, Ina A. Taneva, Laura Taylor, Alexis Akira Toda, Aleh Tsyvinski, Richard Van Weelden, Shoshana Vasserman, Xavier Vives, Adam Wierman, Kai Hao Yang, and William Zame.

†Fedor thanks the Linde Institute at Caltech and National Science Foundation (grant CNS 1518941) for their support.

1 Introduction

Although economic decisions are made by individuals, it is often the aggregate behavior which matters. We examine the relationship between individual and population behavior in the context of consumer choice. Whenever one makes assumptions about incomes and preferences of individual consumers, it is important to understand the restrictions these assumptions impose on aggregate demand. The reverse question is also relevant, especially for welfare analysis: if we know the aggregate demand, what can be said about the individual characteristics of consumers which could give rise to this aggregate demand? Since Sonnenschein (1973), the profession has been quite pessimistic regarding the extent to which the above questions can be answered. The following quote from Kreps (2020) provides an accurate summary of the "anything goes" consensus:

> *"So what can we say about aggregate demand based on the hypothesis that individuals are preference/utility maximizers? Unless we are able to make strong assumptions about the distribution of preferences or income throughout the economy (e.g., everyone has the same preferences) there is little we can say."*

Our paper argues that there is a middle ground: a rich enough setting where aggregation questions are tractable. We develop a method for aggregate demand analysis applicable to populations of consumers with homothetic preferences. For example, what can be said about aggregate demand if all goods are perfect substitutes at the individual level, so that consumers have linear (but not identical) preferences? Or, can a CES aggregate demand with complements be obtained by aggregating Leontief preferences? Our method delivers precise and exhaustive answers to these questions among many others and has immediate applications to robust welfare analysis and the algorithmic complexity of markets.

The key idea of our method is simple: instead of looking at direct utilities of individual consumers, we look at logarithms of their expenditure functions (logarithmic expenditure functions in what follows).[1] We find that the allegedly hard problem of demand aggregation boils down to simply computing a weighted average of logarithmic expenditure functions, with weights equal to relative incomes. This weighted average corresponds to a homothetic preference of the aggregate consumer whose demand coincides with the demand of the original population for all prices.[2]

[1] For a given preference, the expenditure function is a function of prices equal to the minimal budget that an agent with this preference needs to achieve a given utility level.

[2] In contrast to the textbook definition of a representative consumer by Gorman (1961), the preference of the aggregate consumer is allowed to depend on income distribution. As a result, the aggregate consumer is well-defined for all populations with homothetic preferences (Eisenberg, 1961) while Gorman's representative fails to exist for such populations unless all preferences are identical.

The problem of demand aggregation reduces to the problem of preference aggregation: understanding how the preference of the aggregate consumer depends on preferences and incomes of the underlying population. Our method provides insights into economics and the geometry of this dependence:

- *Aggregation-invariant classes of preferences.* Consider a class of preferences, e.g., linear, Leontief, or any other subset of homothetic preferences. We refer to such classes as domains. Let us call a domain invariant with respect to aggregation if the aggregate preference always belongs to this domain. For example, the domain of all homothetic preferences, a domain containing just one preference, and the domain of Cobb-Douglas preferences are invariant. Invariant domains are important as in such domains the aggregate behavior of a population has the same structure as that of an individual.

  We characterize all invariant domains by the convexity property of the set of associated logarithmic expenditure functions. This characterization allows us to construct simple parametric invariant domains and describe the minimal invariant domains containing popular ones.

- *Characterization of feasible aggregate behaviors for a given class of preferences.* Suppose we know a domain to which individual consumers' preferences belong. What aggregate behaviors can be micro-founded? This question boils down to understanding what preferences can be obtained by aggregation of individual preferences from this domain. We call the set of all such aggregate preferences the domain completion. The notion of completion is closely related to invariance: the completion of a domain is the minimal invariant collection of preferences containing the domain. The characterization of the invariance implies that the completion of a domain can be found by computing the convex hull of the set of logarithmic expenditure functions.

  We describe the completion explicitly for the domains of linear and Leontief preferences.[3] A viable conjecture would be that the completion of linear preferences gives all preferences exhibiting substitutability among goods. We show that this guess is correct in the case of two goods only. For Leontief preferences, the completion turns out to be a proper subset of preferences with complementarity even for two goods.

- *Decomposition of preferences.* Consider the inverse to the problem of aggregation: given a preference from a particular domain, represent it as an aggregation of preferences from the same domain. There is always a trivial representation since we can take a population

[3]Surprisingly, this problem happens to be connected to several branches of economics and mathematics such as additive random utility models (ARUM), completely monotone functions, the Stieltjes transform, and even complex analysis.

where each agent has the same given preference. We call those preferences that can only be represented by themselves, indecomposable.

Geometrically, indecomposable preferences correspond to extreme points in the space of logarithmic expenditure functions. As any point of a convex set can be represented as a convex combination of extreme points, indecomposable preferences play the role of elementary building blocks: any preference can be represented as an aggregation of indecomposable ones.

For example, linear and Leontief preferences are indecomposable in the domain of all homothetic preferences. We show that the set of indecomposable preferences is much bigger and contains all Leontief preferences on linear composite goods. In particular, aggregation of linear and Leontief preferences together does not give the whole domain of homothetic preferences. We also explore indecomposable preferences in the domains with substitutability or complementarity.

We illustrate how our approach to aggregation can be applied in several economic environments:

- *Robust welfare analysis.* An analyst observing aggregate behavior aims to estimate a quantity depending on the structure of the population on the micro-level, e.g., the change in welfare induced by a change in prices. As the same aggregate behavior can be compatible with different populations, it can also be compatible with a range of values for welfare. We demonstrate that this range is non-trivial even for standard welfare measures such as the equivalent variation. We show that this range can be computed via information-design tools. As a corollary, we obtain that the utility of a representative consumer — the standard proxy for the population's welfare — gives the most pessimistic estimate on the actual welfare gains. This conclusion suggests a possible explanation for unexpectedly low gains from trade as measured in recent quantitative literature (Arkolakis et al., 2012, 2019).

- *Domain complexity, Fisher markets, and bidding languages for pseudo-market mechanisms.* For invariant domains, the aggregate behavior is as simple as that of a single agent. Since the completion of a domain is the minimal invariant domain containing it, the completion reflects the complexity of possible aggregate behavior.

  We formalize this intuition in application to Fisher markets: simple exchange economies where consumers with fixed incomes face a fixed supply of goods. Such markets are essential for the pseudo-market (or competitive) approach to fair allocation of resources (Moulin, 2019; Pycia, 2022) and serve as a benchmark model for equilibrium computation in algorithmic economics (Nisan et al., 2007). Computing an equilibrium of a Fisher market turns out to be a challenging problem even in a seemingly innocent case of linear preferences thus limiting the applicability of pseudo-market mechanisms. We explore the origin of the complexity and demonstrate that computing equilibria can be hard even in small parametric domains if their completion is large. We show how to construct domains with small completion and describe an algorithm making use of this smallness. The choice of a domain is interpreted as bidding language design.

- *Preference identification.* Given a domain of individual preferences, we ask whether observing the price dependence of market demand is enough to identify the distribution of preferences and income over the population. We relate the possibility of identification with the geometric simplex property of the domain meaning that there is a unique way to represent each preference as an aggregation of indecomposable ones. Examples of domains where identification is possible include Leontief and linear preferences over two goods.

The methodological importance of the link between aggregation and weighted averages is that it brings new tools — the most important of which are convexification and extreme-point techniques of Choquet theory — to consumer demand literature. These tools are increasingly popular in other branches of economic theory such as information economics and mechanism design (Kleiner et al., 2021); our paper demonstrates their power for the analysis of aggregate demand. We also uncover a connection between aggregation and modern literature in convex geometry on the geometric mean of convex sets (Milman and Rotem, 2017; Böröczky et al., 2012). This connection not only enables geometric tools in our economic problem but also leads to new insights about the geometric mean of convex sets suggested by the economic interpretation.

1.1 Related literature

As suggested by the quote from (Kreps, 2020), the existing results on demand aggregation have fallen into one of the two extremes. One extreme stems from the classical general equilibrium literature dealing with economies where agents have general convex preferences and earn money by trading their endowments. This literature concludes that the aggregate demand inherits no properties of individual behavior (Sonnenschein, 1973; Mantel, 1974; Debreu, 1974; Chiappori and Ekeland, 1999). The opposite extreme is given by the representative-agent literature aiming to replace the population with a single rational agent whose preferences are independent of the income distribution (Gorman, 1953, 1961); see also earlier results by Antonelli and Nataf discussed by (Shafer and Sonnenschein, 1982). The independence requirement is so restrictive that it is fair to say that Gorman's representative almost never exists. Exceptions are very special cases, e.g., when the whole population has identical homothetic preferences. The profession has been divided on how seriously the non-existence should be taken. Applied researchers often postulate the existence of a representative consumer (e.g., Chamley, 1986; Rogoff, 1990) but this approach is criticized as lacking micro-foundations (e.g., Kirman, 1992; Carroll, 2000).

As demonstrated by Jackson and Yariv (2019), tweaking Gorman's notion of the representative consumer while maintaining an analog of income independence does not substantially alter the non-existence conclusion. We escape this conclusion by allowing the aggregate consumer to depend on the income distribution. The existence of such an aggregate consumer was pointed out by Eisenberg and Gale (1959) for populations with linear preferences and, by Eisenberg (1961) and Chipman and Moore (1979), in the whole domain of homothetic preferences (see a survey by Shafer and Sonnenschein (1982)).[4] In the modern literature, this insight has gone largely unnoticed with the exception of algorithmic economics and fair allocation mechanisms; see Section 4.3. The converse statement to Eisenberg's result was obtained by Jerison (1984) who showed that homotheticity is necessary for the existence provided that incomes are fixed.

Demand aggregation has been studied in the context of household behavior, where heterogeneous agents redistribute their incomes so that the resulting individual consumption maximizes the household's welfare. Under mild assumptions, such households behave like a single representative agent (Samuelson, 1956; Varian, 1984; Jerison, 1994) confirming the common wisdom that the unrestricted endogeneity of incomes is more important for the negative Sonnenschein-Mantel-Debreu results than the generality of preferences (Mantel, 1976; Hildenbrand, 2014). An active empirical literature aims to link household consumption with individual characteristics of its members; see (Browning and Chiappori, 1998; Lewbel and Pendakur, 2009; Browning et al., 2013) and references therein.

The relation between individual preferences and the representative preference of a welfare-maximizing household has been studied by Chambers and Hayashi (2018) for egalitarian welfare functionals. Their analysis suggests that the role played by the geometric mean of convex sets in our setting with independent consumers (Section 3.1) is played by the Minkowski sum in egalitarian households.

The robust welfare analysis developed in our paper is close in spirit to that of Kang and Vasserman (2022) and Kocourek et al. (2022). The three approaches address different aspects of robustness and are complementary. We are interested in the setting where the aggregate demand is not a sufficient statistic for welfare. Kang and Vasserman (2022) assume that aggregate demand is a sufficient statistic but the number of distinct observations of the demand is not enough to pin it down. In the paper by Kocourek et al. (2022), the aggregate demand is perfectly observed and is a sufficient statistic, but consumers may not be fully aware of prices. Curiously, all the papers rely on auxiliary

[4] Eisenberg and Gale (1959) were motivated by the question of probabilistic forecast aggregation and introduced an auxiliary exchange economy of bets, a "prediction market" in modern terms. A closely related idea for belief aggregation is known in the financial-market literature as the Negishi approach (Jouini and Napp, 2007).

Bayesian persuasion problems of different forms and origins.

Our results on the identification of preference distributions contribute to broad econometric literature on non-parametric identification of stochastic choice models; see surveys (Matzkin, 2007, 2013). This literature has mostly focused on identifying an unknown deterministic part of the decision maker's utility whereas our problem can be interpreted as identifying the noise distribution when the deterministic component is known; see Section 4.1 for a formal relation between additive random utility models and market demand for populations with linear preferences. Preference identification has also been studied in the literature on household behavior and identification has been obtained either for small populations, e.g., two-agent households (Chiappori, 1988) or under the assumption of preferences "orthogonality" (Chiappori and Ekeland, 2009). Our results do not restrict the size of populations and allow agents to have closely aligned preferences.

2 Preliminaries

This section is about notation and basic concepts from consumer demand theory.

Notation. We use $\mathbb{R}$ for the set of all real numbers, $\mathbb{R}_+$ and $\mathbb{R}_-$ for non-negative/non-positive ones, and $\mathbb{R}_{++}$ and $\mathbb{R}_{--}$ for strictly positive/negative ones. Ratios of the form $t/0$ with $t \geqslant 0$ are assumed to be equal to $+\infty$.

Bold font is used for vectors, e.g., $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$. For a pair of vectors of the same dimension, we write $\mathbf{x} \geqslant \mathbf{y}$ if the inequality holds component-wise, i.e., $x_i \geqslant y_i$ for all i. The scalar product of $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ is denoted by $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i \cdot y_i$.

For subsets of $\mathbb{R}^n$, multiplication by a scalar and summation are defined element-wise: $\alpha \cdot X = \{\alpha \cdot \mathbf{x} : \mathbf{x} \in X\}$ and $X + Y = \{\mathbf{x} + \mathbf{y} : \mathbf{x} \in X, \mathbf{y} \in Y\}$ (the Minkowski sum of sets). The standard $(n-1)$-dimensional simplex is denoted by $\Delta_{n-1} = \{\mathbf{x} \in \mathbb{R}_+^n : x_1 + \ldots + x_n = 1\}$.

The gradient of a function $f = f(\mathbf{x})$ is the vector of its partial derivatives $\nabla f = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right)$.

Preferences and demand. Consider a consumer who is endowed with a budget $b \in \mathbb{R}_{++}$ and has a preference $\succsim$ over vectors $\mathbf{x} \in \mathbb{R}_+^n$ interpreted as bundles of $n \geqslant 1$ divisible goods. We assume that preferences satisfy the following standard requirements:

- **homotheticity:** $\mathbf{x}' \succsim \mathbf{x}$ implies $\alpha \cdot \mathbf{x}' \succsim \alpha \cdot \mathbf{x}$ for any $\alpha \geqslant 0$
- **convexity:** for any $\mathbf{x}, \mathbf{x}' \in \mathbb{R}_+^n$ between which the consumer is indifferent, $\lambda \mathbf{x} + (1-\lambda)\mathbf{x}' \succsim \mathbf{x}$ for any $\lambda \in [0, 1]$
- **monotonicity:** if $\mathbf{x}, \mathbf{x}' \in \mathbb{R}_+^n$ and $\mathbf{x}' \geqslant \mathbf{x}$, then $\mathbf{x}' \succsim \mathbf{x}$

- **continuity:** for any $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^n_+$ and a convergent sequence $\mathbf{z}^{(l)}$, $l = 1, 2, 3, \ldots$, such that $\mathbf{x}' \gtrsim \mathbf{z}^{(l)} \gtrsim \mathbf{x}$, we have $\mathbf{x}' \gtrsim \lim_{l\to\infty} \mathbf{z}^{(l)} \gtrsim \mathbf{x}$

- **non-degeneracy:** there exist $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^n_{++}$ such that the consumer is not indifferent between them.

For brevity, we will refer to all preferences satisfying these assumptions as homothetic preferences.

Given a vector of prices $\mathbf{p} \in \mathbb{R}^n_{++}$, the budget set of the consumer is the set of affordable bundles $\{\mathbf{x} \in \mathbb{R}^n_+ : \langle \mathbf{p}, \mathbf{x}\rangle \leqslant b\}$. The demand of the consumer consists of her most preferred bundles from the budget set:

$$D(\mathbf{p}, b) = \operatorname*{argmax}_{\mathbf{x}\in\mathbb{R}^n_+ \,:\, \langle \mathbf{p},\mathbf{x}\rangle \leqslant b} \gtrsim .$$

The demand is a non-empty closed convex subset of the budget set. The demand correspondence satisfies homogeneity with respect to budgets: $D(\mathbf{p}, b) = b \cdot D(\mathbf{p}, 1)$. It is a singleton (one-element set) for almost all $\mathbf{p}$, which allows us to think of the demand as a single-valued function of $\mathbf{p}$ defined almost everywhere; see Appendix A.

Representations of preferences. We will use several ways to represent homothetic preferences.

Any homothetic preference $\gtrsim$ can be represented by a utility function $u = u_{\gtrsim}(\mathbf{x})$ so that $u(\mathbf{x}) \geqslant u(\mathbf{x}')$ if and only if $\mathbf{x} \gtrsim \mathbf{x}'$. This utility function can be selected to be continuous, non-decreasing, concave, homogeneous ($u(\alpha \cdot \mathbf{x}) = \alpha \cdot u(\mathbf{x})$ for all bundles $\mathbf{x}$ and $\alpha \geqslant 0$), non-negative, and not identically zero. Utility functions satisfying all these requirements are called homogeneous in what follows. Any homogeneous utility function defines a homothetic preference and each homothetic preference pins down a unique homogeneous utility function up to a multiplicative factor.

A homothetic preference is determined by its upper contour set $\{\mathbf{x} \in \mathbb{R}^n_+ : u(\mathbf{x}) \geqslant 1\}$. A set $X \subset \mathbb{R}^n$ is called upward-closed if $\mathbf{x} \in X$ implies that all vectors $\mathbf{x}' \in \mathbb{R}^n$ such that $\mathbf{x}' \geqslant \mathbf{x}$ also belong to X. The upper contour set is a closed convex subset of $\mathbb{R}^n_+$ that does not contain 0 and is upward-closed. Any set with these properties corresponds to a homothetic preference. Hence, we can use such sets as another representation for homothetic preferences keeping in mind that, for a given preference, the set is defined up to a homothetic transformation inheriting the freedom in the choice of the multiplicative factor in the utility function.

A dual representation of preferences through expenditure functions will be the most convenient for our analysis. For a consumer with a preference $\gtrsim$, the expenditure function $E = E_{\gtrsim}(\mathbf{p})$ is defined by

$$E(\mathbf{p}) = \min_{\mathbf{x}\in\mathbb{R}^n_+ \,:\, u(\mathbf{x})\geqslant 1} \langle \mathbf{p}, \mathbf{x}\rangle, \tag{1}$$

i.e., the expenditure function is the minimal budget that the consumer needs to achieve the unitary utility level.[5] An expenditure function $E\colon \mathbb{R}_+^n \to \mathbb{R}$ is continuous, non-decreasing, concave, homogeneous, non-negative, and not identically equal to zero. Conversely, any function with these properties is an expenditure function for some homothetic preference. We get yet another way of representing homothetic preferences. Similarly to utility functions, the expenditure function for a given preference is defined up to a multiplicative factor.

Expenditure functions are related to indirect utilities. The indirect utility is the maximal utility that a consumer can achieve as a function of prices and her budget:

$$v(\mathbf{p}, b) = \max_{\mathbf{x} \in \mathbb{R}_+^n \,:\, \langle \mathbf{p}, \mathbf{x} \rangle \leqslant b} u(\mathbf{x}). \tag{2}$$

It can be expressed through the expenditure function by

$$v(\mathbf{p}, b) = \frac{b}{E(\mathbf{p})}. \tag{3}$$

For homothetic preferences, Shephard's lemma implies the following identity:[6]

$$D(\mathbf{p}, b) = b \cdot \nabla \ln\left(E(\mathbf{p})\right), \tag{4}$$

i.e., the demand is proportional to the gradient of the logarithm of the expenditure function (the *logarithmic expenditure function* in what follows). The identity holds for all prices $\mathbf{p} \in \mathbb{R}_{++}^n$ where E is differentiable. This set of prices has full measure; see Appendix A.

Consider the expenditure share function $\mathbf{s}(\mathbf{p})$ whose i-th component $s_i(\mathbf{p})$ is the fraction of the budget that the consumer spends on good $i = 1, \ldots, n$ given the prices, i.e.,

$$s_i(\mathbf{p}) = p_i \cdot \frac{D_i(\mathbf{p}, b)}{b} = p_i \cdot D_i(\mathbf{p}, 1). \tag{5}$$

We treat $\mathbf{s}$ as a single-valued vector function taking values in the standard simplex Δ_{n-1} and defined on the set of $\mathbf{p} \in \mathbb{R}_{++}^n$ of full measure where the demand is a singleton. By (4), expenditure shares can be computed as the elasticities of the expenditure function with respect to prices

$$s_i(\mathbf{p}) = p_i \cdot \frac{\partial \ln\left(E(\mathbf{p})\right)}{\partial p_i} = \frac{\partial \ln\left(E(\mathbf{p})\right)}{\partial \ln(p_i)}. \tag{6}$$

For two goods, preferences can be represented via expenditure share functions using the following characterization.[7] For any homothetic preference $\succsim$ over $\mathbb{R}_+^2$, the expenditure share of the first good

[5] Usually, the expenditure function is considered to be a function of two variables: prices and the utility level. For homothetic preferences, the dependence on the utility level is redundant and we normalize the level to be equal to one.

[6] In this form, the result can be found in (Samuelson, 1972); see Appendix A for a derivation.

[7] To the best of our knowledge, this characterization has not appeared in the literature.

takes the form

$$s_1(p_1, p_2) = \frac{1}{1 + Q\left(\frac{p_1}{p_2}\right) / \frac{p_1}{p_2}} \tag{7}$$

for some non-decreasing non-negative function $Q: \mathbb{R}_{++} \to \mathbb{R}_+ \cup \{+\infty\}$. Moreover, for any such function Q, there is a unique homothetic preference; see Lemma 5 in Appendix C.7.

By plugging a function Q with an infinite number of jumps in (7), we see that, rather counter-intuitively, $s_1(p_1, p_2)$ may change monotonicity infinitely many times as p_2 increases, i.e., the consumer starts spending more on the first good as the price of the second one goes up, then less, then more again, and so on.[8]

Substitutes and complements. The two important subdomains of homothetic preferences are free from the non-monotone behavior of expenditure shares described above.

A preference $\succsim$ is said to exhibit substitutability among the goods if the expenditure share $s_i(\mathbf{p})$ is a non-decreasing function of p_j for each pair of goods $i \neq j$. For differentiable expenditure shares,

$$\frac{\partial s_i(\mathbf{p})}{\partial p_j} > 0 \quad \text{for all } i \neq j. \tag{8}$$

The intuition is that whenever the price of a good increases, the consumer starts spending more on other goods since this good can be substituted. The canonical example is given by linear preferences that correspond to utility functions

$$u(\mathbf{x}) = \langle \mathbf{v}, \mathbf{x} \rangle$$

for some vector of values $\mathbf{v} \in \mathbb{R}^n_+ \setminus \{0\}$. An elementary computation gives the expenditure function and formula (4) provides expenditure shares

$$E(\mathbf{p}) = \min_{i=1,\ldots,n} \frac{p_i}{v_i} \qquad \text{and} \qquad s_i(\mathbf{p}) = \begin{cases} 1, & \text{if } \frac{v_i}{p_i} > \frac{v_j}{p_j} \text{ for all } j \neq i, \\ 0, & \text{otherwise} \end{cases}. \tag{9}$$

As we see, under linear preferences, the consumer spends her whole budget on the good with the highest value-to-price ratio.

A preference $\succsim$ exhibits complementarity among goods if $s_i(\mathbf{p})$ is a non-increasing function of p_j for each pair of goods $i \neq j$. Each of the complementary goods is essential for consumer's satisfaction and so, when one good becomes more expensive, more money is spent on it and less on other goods. The standard example is given by the Leontief preferences which correspond to the following utility function

$$u(\mathbf{x}) = \min_{i=1,\ldots,n} \frac{x_i}{v_i}$$

[8] For example, expenditure shares change monotonicity infinitely many times for Leontief preferences over an infinite number of linear composite goods discussed in Section 5.1.

for some vector of values $\mathbf{v} \in \mathbb{R}^n_+ \backslash \{0\}$. Note that the utility function has the same functional form as the expenditure function for linear preferences. By duality, the expenditure function for Leontief preferences is linear

$$E(\mathbf{p}) = \langle \mathbf{v}, \mathbf{p} \rangle \qquad \text{and} \qquad s_i(\mathbf{p}) = \frac{v_i \cdot p_i}{\langle \mathbf{v}, \mathbf{p} \rangle}. \tag{10}$$

The intersection of the domains of preferences exhibiting substitutability and complementarity consists of those preferences $\gtrsim$ for which expenditure shares are constant, i.e., there is a fixed vector $\mathbf{a} \in \Delta_{n-1}$ such that $\mathbf{s}(\mathbf{p}) = \mathbf{a}$ for any $\mathbf{p}$. Expenditure shares determine the expenditure function by (4)

$$E(\mathbf{p}) = \prod_{i=1}^{n} p_i^{a_i} \tag{11}$$

and the corresponding preference is given by the Cobb-Douglas utility function

$$u(\mathbf{x}) = \prod_{i=1}^{n} x_i^{a_i}. \tag{12}$$

Leontief, Cobb-Douglas, and linear preferences are contained as limit cases in a widely used parametric family of preferences with constant elasticity of substitution (CES). A preference $\gtrsim$ is a CES preference with elasticity of substitution $\sigma \in \mathbb{R}_{++} \backslash \{1\}$ if the corresponding utility function has the form

$$u(\mathbf{x}) = \left(\sum_{i=1}^{n} (a_i \cdot x_i)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \tag{13}$$

for some vector $\mathbf{a} \in \mathbb{R}^n_{++}$. The corresponding expenditure functions and expenditure shares are given by

$$E(\mathbf{p}) = \left(\sum_{i=1}^{n} \left(\frac{p_i}{a_i} \right)^{1-\sigma} \right)^{\frac{1}{1-\sigma}} \qquad \text{and} \qquad s_i(\mathbf{p}) = \frac{\left(\frac{p_i}{a_i} \right)^{1-\sigma}}{\sum_{j=1}^{n} \left(\frac{p_j}{a_j} \right)^{1-\sigma}}. \tag{14}$$

CES preferences exhibit substitutability for $\sigma > 1$ and complementarity for $\sigma \in (0, 1)$. Leontief, Cobb-Douglas, and linear preferences are the limiting cases as σ goes, respectively, to 0, 1, and $+\infty$. The limits are taken with respect to the topology that we discuss next.

Topology on preferences. Convergence of preferences, closed and open sets, and the Borel structure are understood with respect to the following metric. We define the distance between preferences $\gtrsim$ and $\gtrsim'$ with expenditure functions E and E' by

$$d(\gtrsim, \gtrsim') = \sup_{\mathbf{p} \in \Delta_{n-1} \cap \mathbb{R}^n_{++}} \left| \frac{(\ln E(\mathbf{p}) - \ln E(\mathbf{e})) - (\ln E'(\mathbf{p}) - \ln E'(\mathbf{e}))}{(1 + \max_i |\ln p_i|)^2} \right|, \tag{15}$$

where $\mathbf{e} = (1, \ldots, 1)$. The main advantage of this way to introduce the distance is that it makes the set of all homothetic preferences a compact metric space.[9] In particular, the distance between any pair of preferences is finite and bounded by 2. See Appendix B for the intuition behind the definition.

3 Preference aggregation

Consider $m \geqslant 1$ consumers $k = 1, \ldots, m$. Consumer k has a positive budget $b_k \in \mathbb{R}_{++}$ and a homothetic preference $\succsim_k$ over bundles of $n \geqslant 1$ divisible goods as in Section 2. For any vector of prices $\mathbf{p}$, this population generates the market demand equal to the sum of individual demands $D_1(\mathbf{p}, b_1) + \ldots + D_m(\mathbf{p}, b_m)$. To study the market demand, we aim to replace the population of m consumers with a single aggregate consumer generating the same demand. The following definition plays a central role in this methodology.

Definition 1. *A preference $\succsim_{\text{aggregate}}$ is the aggregate preference for a population of consumers with preferences $\succsim_1, \ldots, \succsim_m$ and budgets $b_1, \ldots, b_m$ if*

$$D_{\text{aggregate}}\left(\mathbf{p},\ \sum_{k=1}^{m} b_k\right) = D_1(\mathbf{p}, b_1) + \ldots + D_m(\mathbf{p}, b_m) \tag{16}$$

for any price vector $\mathbf{p} \in \mathbb{R}^n_{++}$. A consumer with preference $\succsim_{\text{aggregate}}$ is referred to as the aggregate consumer.

In other words, the market demand generated by the population of consumers coincides with the demand of the aggregate consumer endowed with the total budget. We stress that the aggregate consumer is selected for a given collection of budgets $b_1, \ldots, b_m$ of individual consumers, and so, for a different distribution of incomes over the population, we may end up with a different aggregate consumer. This is an important distinction between Definition 1 and the approach of Gorman (1961) who insist on the independence of the aggregate preference on the income distribution which can be achieved in knife-edge cases only.

Example 1. Consider $m = n$ single-minded consumers: $u_i(\mathbf{x}) = x_i$, i.e., consumer i only cares about good i. Hence, no matter what the prices are, consumer i spends her total budget b_i on good i. This observation helps to guess the aggregate consumer without any computations. Indeed, the aggregate consumer spends the amount b_i out of her total budget $b_1 + \ldots + b_n$ on good i independently of prices. In other words, the expenditure share of each good i for the aggregate

[9] Economic literature has considered compact topologies on the set of preferences, e.g., the closed convergence topology of upper contour sets (Hildenbrand, 2015; Bridges and Mehta, 2013). To the best of our knowledge, an explicit metric structure giving compactness has not appeared in the literature.

consumer is price-independent and equal to $s_{\text{aggregate},i}(\mathbf{p}) = b_i/(b_1 + \ldots + b_n)$. Hence, the aggregate consumer must have the Cobb-Douglas preferences (12) with $a_i = b_i/(b_1 + \ldots + b_n)$. One can verify this guess directly by checking that the demand identity (16) holds. Alternatively, the result can be deduced immediately from Theorem 1 below and explicit formulas for expenditure functions of Cobb-Douglas and linear preferences.

The existence of an aggregate preference was established by Eisenberg (1961) for any population of consumers with homothetic preferences. Denote by B the total budget of the population and by β_k the relative fraction of consumer k's budget

$$B = \sum_{k=1}^{m} b_k, \qquad \beta_k = \frac{b_k}{B}. \tag{17}$$

Eisenberg (1961) showed that the aggregate preference corresponds to the utility function obtained as the solution to the following optimization problem

$$u_{\text{aggregate}}(\mathbf{x}) = \max\left\{\prod_{k=1}^{m}\left(\frac{u_k(\mathbf{x}_k)}{\beta_k}\right)^{\beta_k} \quad : \quad \mathbf{x}_k \in \mathbb{R}_+^n, \quad k = 1, \ldots, m, \quad \sum_{k=1}^{m} \mathbf{x}_k = \mathbf{x}\right\}. \tag{18}$$

In other words, the utility for the aggregate preference at a bundle $\mathbf{x}$ is equal to the maximal weighted Nash social welfare where the maximum is taken over all possible allocations of $\mathbf{x}$ over the consumers and consumer's weight is equal to her relative budget.[10] The optimization problem (18) is called the Eisenberg-Gale problem since it is similar to a problem studied by Eisenberg and Gale (1959) in the context of probabilistic forecast aggregation.

To determine the utility of an aggregate consumer, one needs to solve the Eisenberg-Gale problem (18) for each $\mathbf{x} \in \mathbb{R}_+^n$. Except for special cases such as Cobb-Douglas preferences, it does not admit an explicit solution and is not easy to work with both analytically and computationally; see Section 4.3.

We observe that the question of describing the aggregate consumer substantially simplifies if we use the dual representation of preferences via expenditure functions.

Theorem 1. *Consider a population of consumers with homothetic preferences $\succsim_1, \ldots, \succsim_m$ and budgets $b_1, \ldots, b_m$. The preference of the aggregate consumer is described by the expenditure function $E_{\text{aggregate}}$ satisfying*

$$\ln E_{\text{aggregate}}(\mathbf{p}) = \sum_{k=1}^{m} \beta_k \cdot \ln E_k(\mathbf{p}), \tag{19}$$

where the weights β_k are given by (17).

[10] The welfare function equal to the product of consumer's utilities is dubbed the Nash social welfare or the Nash product as this welfare function naturally arises in the context of axiomatic bargaining studied by Nash (1950).

Hence, preference aggregation is equivalent to taking convex combinations of individual logarithmic expenditure functions. The simplicity of this operation will allow us to describe domains invariant with respect to aggregation (Section 4) and to study the decomposition of a given preference as an aggregation of elementary ones (Section 5).

The identity (19) becomes almost immediate if we take into account the relation (4) between the demand and the gradient of the expenditure function: $D(\mathbf{p}, b) = b \cdot \nabla \ln\left(E(\mathbf{p})\right)$. The definition of the aggregate consumer implies the equality

$$B \cdot \nabla \ln E_{\text{aggregate}}(\mathbf{p}) = \sum_{k=1}^{m} b_k \cdot \nabla \ln E_k(\mathbf{p}), \tag{20}$$

which must hold at all points of differentiability of the expenditure functions. As any concave function is differentiable almost everywhere with respect to the Lebesgue measure (Section A), the equality (20) holds on the set of full measure and can be integrated resulting in the identity (19). Integration constants get absorbed by the expenditure functions since they are defined up to multiplicative factors.

In Appendix C.1, we prove Theorem 1 using an approach similar to the one used by Eisenberg (1961) and not relying on formula (4). This alternative proof clarifies that Theorem 1 is dual to Eisenberg's result.[11]

3.1 Connection to the geometric mean of convex sets

Theorem 1 links preferences aggregation and recent attempts to define the geometric mean of convex sets; see a survey by Milman and Rotem (2017). Recall that the support function of a convex set $X \subset \mathbb{R}^n$ is defined by

$$h_X(\mathbf{p}) = \sup_{x \in X} \langle \mathbf{p}, \mathbf{x} \rangle.$$

Böröczky et al. (2012) define the weighted geometric mean of convex sets by taking the usual weighted geometric mean in the space of support functions.[12] Formally, the weighted geometric mean of convex sets X and Y with weights $(\lambda, 1-\lambda)$, $\lambda \in [0,1]$, is the convex set Z denoted by $X^\lambda \otimes Y^{1-\lambda}$ such that

$$|h_{X^\lambda \otimes Y^{1-\lambda}}| = |h_X|^\lambda \cdot |h_Y|^{1-\lambda}. \tag{21}$$

[11] The connection between Eisenberg's theorem and Theorem 1 can be seen as a version of duality for the inf-convolution (Rockafellar and Wets, 2009, Chapter 1.H).

[12] Defining algebraic operations on convex sets through the usual algebraic operations on their support functions is a standard approach. For example, the Minkowski addition of convex sets corresponds to the pointwise summation of their support functions.

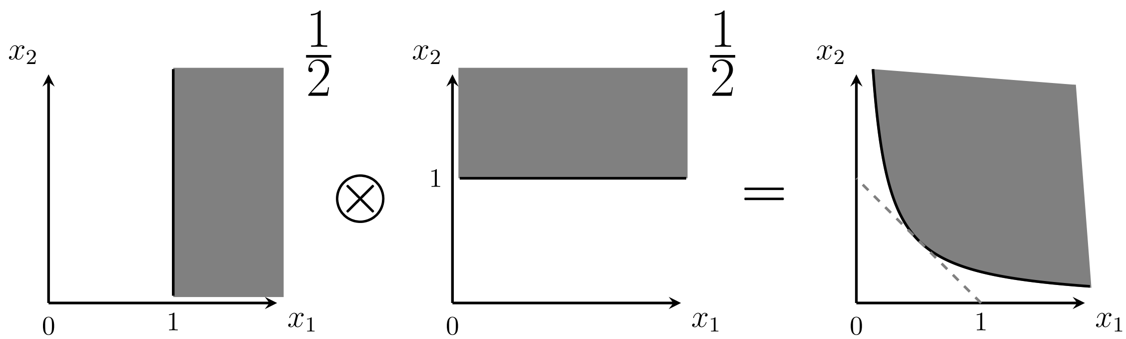

Figure 1: **Geometry:** the set bounded by the hyperbola is the geometric mean of the two orthogonal halfspaces. **Economics:** an aggregation of two extreme linear preferences where each consumer cares only about her own good gives a Cobb-Douglas preference.

The weighted geometric mean extends to any number of convex sets in a straightforward manner.[13]

To see the connection between preference aggregation and the geometric mean, note that the expenditure function E is equal to the support function of the upper contour set up to a sign:

$$E(\mathbf{p}) = -h_X(-\mathbf{p}) \qquad \text{where} \quad X = \{u(\mathbf{x}) \geqslant 1\}.$$

We obtain the following equivalent version of Theorem 1.

Corollary 1. *An upper contour set of the aggregate consumer's preferences* $\{u_{\text{aggregate}}(\mathbf{x}) \geqslant 1\}$ *is the weighted geometric mean of individual upper contour sets with budget-proportional weights:*

$$\left\{u_{\text{aggregate}}(\mathbf{x}) \geqslant 1\right\} = \left\{u_1(\mathbf{x}) \geqslant 1\right\}^{\beta_1} \otimes \left\{u_2(\mathbf{x}) \geqslant 1\right\}^{\beta_2} \otimes \ldots\ldots \otimes \left\{u_m(\mathbf{x}) \geqslant 1\right\}^{\beta_k}. \qquad (22)$$

In Example 1, we saw that Cobb-Douglas preferences over n goods originate as an aggregation of n extreme linear preferences. Figure 1 illustrates the corresponding identity of convex sets for $n = 2$ and equal budgets.

Corollary 1 highlights a peculiar property of the class of convex sets that can be obtained as upper contour sets of homothetic preferences. From formula (21), it is not evident that the geometric mean is well-defined, i.e., that we can always find a convex set whose support function is equal to $h_{X^\lambda \otimes Y^{1-\lambda}}$. A byproduct of Corollary 1 is that the weighted geometric mean is well-defined within the class of all convex subsets of $\mathbb{R}_+^n$ that do not contain zero and are upward-closed. Indeed,

[13] Böröczky et al. (2012) refer to $X^\lambda \otimes Y^{1-\lambda}$ as the logarithmic sum of convex sets to distinguish it from other definitions of the geometric mean. Since we do not consider other definitions, we call $X^\lambda \otimes Y^{1-\lambda}$ the weighted geometric mean.

any such set is an upper contour set of some homothetic preference and the weighted geometric mean is an upper contour set of the aggregate consumer's preference. Contrast this observation with the case of bounded convex sets which mathematical literature has mostly focused on. The weighted geometric mean is well-defined for bounded convex sets containing zero; however, sets that do not contain zero are problematic as the support function can be negative and the definition of the weighted geometric mean requires ad hoc modifications.

3.2 Application: robust welfare analysis as information design

Understanding how the population's welfare changes as a function of prices is crucial for economic policy evaluation. Consider an analyst who observes market demand as a function of prices and, based on this information, aims to estimate a certain aggregate measure of individual well-being, e.g., the change in the population's welfare before and after introducing a policy affecting prices.

The standard approach is postulating a representative agent and then using the representative's utility as a proxy for the population's welfare. Implicitly, this approach assumes that market demand is a sufficient statistic for welfare.[14] However, the same market demand — hence, the same aggregate preference — can be generated by different populations of consumers. As a result, the same aggregate behavior may be compatible with a range of welfare levels.

The aggregate behavior may not be a sufficient statistic for welfare even for standard welfare measures and preference domains. Suppose prices change from $\mathbf{p^0}$ to $\mathbf{p^1}$. The standard way of quantifying the corresponding change in individual well-being is by the change in income that makes the consumer indifferent between the two prices. Depending on whether the indifference is considered with respect to the old prices or the new ones, we get welfare measures known as the equivalent variation EV and the compensating variation CV, respectively (Mas-Colell et al., 1995, Chapter 3.I). For a consumer with income b and homothetic preference $\succsim$ represented by an expenditure function E, the two measures take the form

$$\mathrm{EV}_{\mathbf{p^0}\to\mathbf{p^1}}(\succsim, b) = b \cdot \left(\frac{E(\mathbf{p^0})}{E(\mathbf{p^1})} - 1\right) \qquad \text{and} \qquad \mathrm{CV}_{\mathbf{p^0}\to\mathbf{p^1}}(\succsim, b) = b \cdot \left(1 - \frac{E(\mathbf{p^1})}{E(\mathbf{p^0})}\right). \tag{23}$$

For a population of consumers, the change in welfare is defined as the sum of individual changes, e.g., for the equivalent variation EV, we get

$$W\big[(\succsim_k, b_k)_{k=1,2,\ldots}\big] = \sum_k \mathrm{EV}_{\mathbf{p^0}\to\mathbf{p^1}}(\succsim_k, b_k). \tag{24}$$

[14] For example, the recent quantitative literature on gains from trade uses representative consumer's utility as an aggregate welfare measure; see (Costinot and Rodríguez-Clare, 2014) for a survey. Hence, there is a one-to-one mapping between market demand and welfare. As pointed out by Arkolakis et al. (2012) and, more recently, by Arkolakis et al. (2019), this approach leads to surprisingly low gains from trade.

The following toy example illustrates that the aggregate behavior may not only be compatible with a range of welfare levels but may even fail to determine the direction of the welfare change.

Example 2. In a two-good economy, a population of Cobb-Douglas consumers generates market demand equal to $1/p_1$ for the first good and $2/p_2$ for the second. In other words, this population behaves like a single Cobb-Douglas consumer with utility $u_{\text{aggregate}}(\mathbf{x}) = x_1^{1/3} \cdot x_2^{2/3}$ and budget $B = 3$. The government considers introducing new tariffs changing prices from $\mathbf{p^0} = (1, 8)$ to $\mathbf{p^1} = (5, 5)$. Will this policy be beneficial for the population?

Assume that the well-being of an individual is measured by the equivalent variation and so the population's welfare is given by (24). Recall that the expenditure function for a Cobb-Douglas consumer with $u(\mathbf{x}) = x_1^{\alpha} \cdot x_2^{1-\alpha}$ equals $E(\mathbf{p}) = p_1^{\alpha} \cdot p_2^{1-\alpha}$.

Consider a population of consumers with identical preferences $\succsim_k = \succsim_{\text{aggregate}}$ for any k and total budget B. For such a population, W is equal to the equivalent variation for the aggregate consumer:

$$W = \mathrm{EV}_{\mathbf{p^0} \to \mathbf{p^1}}(\succsim_{\text{aggregate}}, B) = 3 \cdot \left(\frac{1^{1/3} \cdot 8^{2/3}}{5^{1/3} \cdot 5^{2/3}} - 1\right) = -\frac{3}{5} < 0.$$

On the other hand, the same aggregate demand can be generated by a population of two single-minded consumers: one with utility $u_1(\mathbf{x}) = x_2$ and budget $b_1 = 2$ and the other with $u_2(\mathbf{x}) = x_1$ and $b_2 = 1$; see Example 1. For this heterogeneous population, we get

$$W = 2 \cdot \left(\frac{1^0 \cdot 8^1}{5^0 \cdot 5^1} - 1\right) + 1 \cdot \left(\frac{1^1 \cdot 8^0}{5^1 \cdot 5^0} - 1\right) = \frac{2}{5} > 0.$$

We conclude that the same aggregate behavior can be compatible with a range of values for W and, moreover, this range can contain zero, i.e., the aggregate agent consumer may not even predict whether the change in prices is beneficial for the underlying population or not.

We propose a robust approach to welfare analysis. Since the aggregate behavior may not pin down the exact value of the population's welfare (or of any other functional of interest depending on the population's structure on the micro-level), we aim to compute the range of possible values being fully agnostic about the specific decomposition of the market demand into individual demands. For this purpose, we combine Theorem 1 with insights from information design.

Let us describe the problem formally. Since market demand determines the aggregate consumer, we assume that the aggregate preference $\succsim_{\text{aggregate}}$ and the total income B in the economy are given. In addition, we assume that the analyst knows that individual consumers have preferences from some subset $\mathcal{D}$ of homothetic preferences; we will refer to $\mathcal{D}$ as a domain.[15] The goal is to find the range $[\underline{W}, \overline{W}]$ of values of a functional

$$W = W\big[(\succsim_k, b_k)_{k=1,2,\ldots}\big] \tag{25}$$

[15]Examples of domains include all homothetic preferences, linear preferences, preferences exhibiting complementarity, any finite collection of preferences, or any other subset of homothetic preferences.

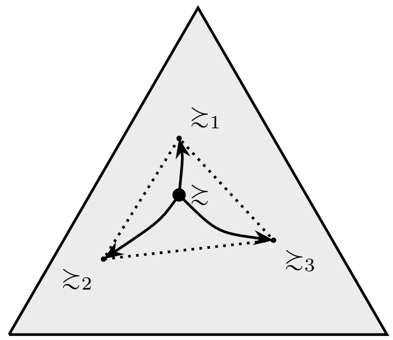

Figure 2: In the space of logarithmic expenditure functions, all populations with aggregate behavior captured by $\succsim = \succsim_{\text{aggregate}}$ correspond to all possible ways to represent $\succsim$ as a weighted average of other points.

over all finite populations of consumers $k = 1, 2, \ldots$ with preferences $\succsim_1, \succsim_2, \ldots$ and incomes $b_1, b_2, \ldots$ such that the individual preferences aggregate up to $\succsim_{\text{aggregate}}$, incomes sum up to $B = \sum_k b_k$, and individual preferences belong to $\mathcal{D}$.

With a domain $\mathcal{D}$, we associate the set $\mathcal{L}_{\mathcal{D}}$ of all logarithmic expenditure functions of preferences from $\mathcal{D}$:

$$\mathcal{L}_{\mathcal{D}} = \Big\{ f \ : \ \mathbb{R}^n_{++} \to \mathbb{R} \quad : \quad f = \ln E_{\succsim}, \quad \succsim \in \ \mathcal{D} \Big\}.$$

The set $\mathcal{L}_{\mathcal{D}}$ inherits the freedom in the choice of expenditure functions: if $f \in \mathcal{L}_{\mathcal{D}}$, then $f + \text{const}$ is also in $\mathcal{L}_{\mathcal{D}}$ and corresponds to the same preference; we will call such elements of $\mathcal{L}_{\mathcal{D}}$ equivalent.

We can think of preferences as classes of equivalent points in $\mathcal{L}_{\mathcal{D}}$. The key observation that makes the problem tractable is that, in this space, populations with given aggregate behavior $\succsim_{\text{aggregate}}$ correspond to all the different ways of representing $\ln E_{\text{aggregate}}$ as a weighted average of logarithmic expenditure functions $\ln E_{\text{aggregate}} = \sum_k \beta_k \ln E_k$; see Figure 2. Hence, to find the range of values of W, we need to minimize or maximize it over all such representations. A similar optimization problem arises in the context of Bayesian persuasion (Aumann et al., 1995; Kamenica and Gentzkow, 2011) and indeed there is a formal connection between the two problems inspiring our analysis and described in Appendix D.

To obtain an explicit formula for the range of values, we assume a particular functional form of W:

$$W = \sum_k b_k \cdot w(\succsim_k). \tag{26}$$

This functional form captures both the equivalent and the compensating variations (24). In particular, the individual measure of well-being w can also depend on other parameters — e.g., prices before and after a market intervention — but such dependence does not affect our analysis and hence omitted. Functionals of the form (26) can also capture distributional objectives with no welfare interpretations, e.g., the total income of agents having preferences from a certain subset $\mathcal{D}' \subset \mathcal{D}$, which corresponds to $w = \mathbb{1}_{\mathcal{D}'}$.

For a function f defined on a subset X of a linear space, its concavification $\mathrm{cav}_X[f]$ is the smallest concave function larger than f on X. It can be computed as follows:

$$\mathrm{cav}_X[f](x) = \sup\left\{\sum_k \beta_k \cdot f(x_k) \quad : \quad x = \sum_k \beta_k \cdot x_k, \quad x_k \in X, \quad \beta_k \geqslant 0, \quad \sum_k \beta_k = 1\right\}, \qquad (27)$$

where the supremum is taken over all finite convex combinations of points in X that average to x. Note that the set X may not be convex but the concavification is well-defined on its convex hull. One can similarly define convexification $\mathrm{vex}_X[f]$ as the biggest convex function smaller than f on X, i.e., $\mathrm{vex}_X[f] = -\mathrm{cav}_X[-f]$.

Proposition 1. *For W of the form (26), the range of values $[\underline{W}, \overline{W}]$ compatible with an aggregate preference $\succsim_{\text{aggregate}}$, total income B, and individual domain of preferences $\mathcal{D}$ is given by*

$$[\underline{W}, \overline{W}] = \Big[B \cdot \mathrm{vex}_{\mathcal{L}_{\mathcal{D}}}[w]\big(\succsim_{\text{aggregate}} \big), \quad B \cdot \mathrm{cav}_{\mathcal{L}_{\mathcal{D}}}[w]\big(\succsim_{\text{aggregate}} \big)\Big], \qquad (28)$$

where $\mathcal{L}_{\mathcal{D}}$ denotes the set of logarithmic expenditure functions corresponding to the domain $\mathcal{D}$.

We stress that the function $w(\succsim)$ in (34) is treated as a functional on the space of logarithmic expenditure functions and is concavified and convexified over this space. The proposition is proved in Appendix D, and we now illustrate how the proposition applies to the setting from Example 2.

Example 3. Let $\mathcal{D}$ be the domain of Cobb-Douglas preferences over two goods. The corresponding logarithmic expenditure are given by $\ln E(\mathbf{p}) = \alpha \ln p_1 + (1-\alpha)\ln p_2$. As in Example 2, suppose that the aggregate behavior corresponds to $\alpha = 1/3$ and total budget $B = 3$, and prices change from $\mathbf{p^0} = (1, 8)$ to $\mathbf{p^1} = (5, 5)$. The goal is to estimate the range of the welfare change measured by the equivalent variation (24). The corresponding functional $w(\succsim)$ takes the form

$$w(\succsim) = \frac{E(\mathbf{p^0})}{E(\mathbf{p^1})} - 1 = \frac{1^\alpha \cdot 8^{1-\alpha}}{5^\alpha \cdot 5^{1-\alpha}} - 1 = \frac{1}{5}8^{-\alpha} - 1$$

Since the set $\mathcal{L}_D$ of logarithmic expenditure functions can be identified with the interval $\alpha \in [0, 1]$ via a linear transformation, convexification and concavification over $\mathcal{L}_D$ boil down to the same operations with respect to α. Denote $f(\alpha) = \frac{1}{5}8^{-\alpha} - 1$. The function f is convex and so $\mathrm{vex}_{[0,1]}[f] = f$ and $\mathrm{cav}_{[0,1]}[f](\alpha) = \alpha f(0) + (1-\alpha) f(1)$. We obtain the following answer for the range of values of W:

$$[\underline{W}, \overline{W}] = \big[3 \cdot \mathrm{vex}_{[0,1]} f(1/3), \quad 3 \cdot \mathrm{cav}_{[0,1]} f(1/3)\big] = \left[\frac{-3}{5}, \quad \frac{2}{5}\right].$$

In particular, the two populations described in Example 2 give the lowest and the highest possible values for the welfare change.

Let us now discuss the general implications of Proposition 1. In Example 3, the functional w corresponding to the equivalent variation turned out to be a convex function of the logarithmic expenditure function. This is true in general. Indeed, the ratio

$$\frac{E(\mathbf{p^0})}{E(\mathbf{p^1})} = \exp\left(\ln E(\mathbf{p^0}) - \ln E(\mathbf{p^1})\right)$$

is a convex function over the space of logarithmic expenditure functions as the composition of a convex function $\exp(t)$ and a linear functional evaluating the difference between the values of $\ln E$ at two different points $\mathbf{p}$. Therefore, $\mathrm{vex}_{\mathcal{L}_\mathcal{D}}[w] = w$ for the equivalent variation. Similarly, one concludes that the compensated variation is concave, and we obtain the following corollary of Proposition 1.

Corollary 2. *If the change in welfare is measured by the equivalent variation, then $\underline{W}$ equals $\mathrm{EV}(\succsim_{\text{aggregate}}, B)$, i.e., the representative-agent approach gives the most pessimistic prediction for the welfare change of the population. On the other hand, if the compensating variation is used, $\overline{W}$ is equal to $\mathrm{CV}(\succsim_{\text{aggregate}}, B)$ and so the representative consumer provides the most optimistic estimate.*

The trade literature relies on the equivalent variation and, hence, the representative agent approach used by this literature gives the lower bound on the actual welfare change possibly explaining the low gains from trade discussed by Arkolakis et al. (2012).

A common assumption in industrial organization and trade literature is that a population behaves like a single CES consumer. The following example illustrates how to compute the range of values for the equivalent variation for CES aggregate behavior provided that individual preferences exhibit substitutability.

Example 4. Let the domain of individual preferences $\mathcal{D}$ be all the preferences over two goods exhibiting substitutability. Assume that the aggregate consumer has a CES preference corresponding to

$$u_{\text{aggregate}}(\mathbf{x}) = \left((a_1 \cdot x_1)^{\frac{\sigma-1}{\sigma}} + (a_2 \cdot x_2)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} \qquad \text{with} \qquad \sigma > 1. \tag{29}$$

Our goal is to find the range of values $[\underline{W}, \overline{W}]$ for welfare change measured by the equivalent variation (24). By Corollary 2, the pessimistic bound $\underline{W}$ is given by $\mathrm{EV}_{\mathbf{p^0}\to\mathbf{p^1}}(\succsim_{\text{aggregate}}, B)$ and so we focus on computing $\overline{W}$.

To concavify a convex function f over a convex set X, it is enough to maximize in (27) over extreme points of X. We will need the following facts established in subsequent sections: the set

$\mathcal{L}_\mathcal{D}$ for two substitutes is convex (Section 4) and linear preferences correspond to its extreme points (Section 5.2). Formula (46) gives a unique way to represent a CES preference as an aggregation of linear ones: the marginal rate of substitution $\mathrm{MRS} = v_1/v_2$ of a linear preference $u(\mathbf{x}) = v_1 \cdot x_1 + v_2 \cdot x_2$ has to be distributed over the population according to the distribution μ with the following cumulative distribution function:

$$\mu\big([0,\, t)\big) = \frac{(a_2 \cdot t)^{\sigma-1}}{(a_1)^{\sigma-1} + (a_2 \cdot t)^{\sigma-1}}. \tag{30}$$

Since EV is a convex functional, we conclude that its concavification corresponds to representing $\ln E_{\text{aggregate}}$ as a convex combination of extreme points of $\mathcal{L}_\mathcal{D}$ which correspond to linear preferences. Using formula (9) for the expenditure function of a linear preference $\succsim$ with $\mathrm{MRS} = t$, we get

$$\mathrm{EV}_{\mathbf{p^0}\to\mathbf{p^1}}(\succsim, b) = b\left(\frac{p_2^0}{p_2^1} \cdot \frac{\min\{p_1^0/p_2^0,\ t\}}{\min\{p_1^1/p_2^1,\ t\}} - 1\right). \tag{31}$$

Hence, the maximal value of the welfare change compatible with the CES aggregate preference $\succsim_{\text{aggregate}}$ is given by averaging (31) with respect to μ:

$$\overline{W} = B \cdot \int_0^\infty \left(\frac{p_2^0}{p_2^1} \cdot \frac{\min\{p_1^0/p_2^0,\ t\}}{\min\{p_1^1/p_2^1,\ t\}} - 1\right) \mathrm{d}\mu(t).$$

Except for particular values of τ such as $\tau = 1$, this integral cannot be computed in elementary functions because of the contribution of the interval $t \in \big[\min_i p_1^i/p_2^i,\ \max_i p_1^i/p_2^i\big]$ leading to the hypergeometric function. This integral can be computed approximately provided that the change in prices is small. Instead of performing this computation, we explain below the general technique to approximate $\overline{W}$. For this purpose, we first discuss functionals for which the aggregate behavior is a sufficient statistic.

For a general objective W of the form (26), the use of the representative agent approach is justified if the interval $[\underline{W}, \overline{W}]$ is, in fact, a singleton, i.e., the convexification of w coincides with the concavification. The two coincide only for affine functionals.

Corollary 3. *Market demand is a sufficient statistic for the value of W if the measure of individual well-being $w(\succsim)$ is an affine functional of the logarithmic expenditure function $\ln(E_\succsim)$. If it is not affine, there is an aggregate preference $\succsim_{\text{aggregate}}$ and two populations with the same total budget whose preferences aggregate up to $\succsim_{\text{aggregate}}$ but levels of W are different.*

The equivalent and compensating variations as well as the total income of agents with preferences from $\mathcal{D}' \subset \mathcal{D}$ are not affine. An example of an affine functional is given by the area variation

$$\mathrm{AV}_{\mathbf{p^0}\to\mathbf{p^1}}(\succsim, b) = b \cdot \int_{\mathbf{p^1}}^{\mathbf{p^0}} D\big(\mathbf{p},\, b\big)\, \mathrm{d}\mathbf{p},$$

where the integration is over a curve connecting $\mathbf{p^1}$ and $\mathbf{p^0}$ in the space of prices.[16] Since the demand is proportional to the gradient of the logarithmic expenditure function (4), we get

$$\mathrm{AV}_{\mathbf{p^0}\to\mathbf{p^1}}(\succsim, b) = b \cdot \big(\ln E(\mathbf{p^1}) - \ln E(\mathbf{p^0})\big). \tag{32}$$

Hence, w is indeed an affine functional of the logarithmic expenditure function.

Let us discuss the behavior of the welfare change provided that the change in prices $\delta\mathbf{p} = \mathbf{p^1} - \mathbf{p^0}$ is small. We focus on the case where the welfare change is measured by the equivalent variation (24). Let us first derive an approximate formula for the equivalent variation $\mathrm{EV}(\succsim, b)$ of a single agent. Denote $\ln E(\mathbf{p}^1) - \ln E(\mathbf{p}^0)$ by $\delta \ln E$. Using the Taylor formula, we get

$$\mathrm{EV}(\succsim, b) = b \cdot (\exp(-\delta \ln E) - 1) \simeq b \cdot \left(-\delta \ln E + \frac{1}{2}(\delta \ln E)^2\right) = \mathrm{AV}(\succsim, b) + \frac{b}{2}(\delta \ln E)^2, \tag{33}$$

where $\simeq$ means equality up to the third order of magnitude in $\delta\mathbf{p}$.[17] Approximating the small change in the argument of $\ln E$ via the gradient and expressing the gradient through the demand (4), we rewrite

$$b \cdot \big(\delta \ln E\big)^2 \simeq b \cdot \big\langle \delta\mathbf{p},\, \nabla \ln E(\mathbf{p^0})\big\rangle^2 = b \cdot \big\langle \delta\mathbf{p},\, D\left(\mathbf{p^0},\, 1\right)\big\rangle^2 = \left\langle \delta\mathbf{p},\, D\left(\mathbf{p^0},\, \sqrt{b}\right)\right\rangle^2.$$

Taking into account that the area variation is an affine function, we get the following corollary of Proposition 1.

Corollary 4. *If individuals have preferences from a domain $\mathcal{D}$, the welfare change is measured by the equivalent variation, and the change in prices $\delta\mathbf{p}$ is small, then*

$$\overline{W} \simeq \mathrm{AV}_{\mathbf{p^0}\to\mathbf{p^1}}(\succsim_{\mathrm{aggregate}}, B) + \frac{1}{2} \cdot \mathrm{cav}_{\mathcal{L}_D}\left[\left\langle \delta\mathbf{p},\, D\left(\mathbf{p^0},\, \sqrt{B}\right)\right\rangle^2\right](\succsim_{\mathrm{aggregate}}).$$

Formula (33) allows one to express AV through EV. Hence, Corollary 4 implies

$$\overline{W} - \underline{W} \simeq \frac{1}{2} \cdot \left(\mathrm{cav}_{\mathcal{L}_D}\left[\left\langle \delta\mathbf{p},\, D\left(\mathbf{p^0},\, \sqrt{B}\right)\right\rangle^2\right](\succsim_{\mathrm{aggregate}}) - \left\langle \delta\mathbf{p},\, D_{\mathrm{aggregate}}\left(\mathbf{p^0},\, \sqrt{B}\right)\right\rangle^2\right). \tag{34}$$

One can check that formula (34) also remains valid if the welfare change is measured by the compensating variation.

Formula (34) admits a simple interpretation. Note that the concavification $\mathrm{cav}_X[f](x)$ can be seen as the maximal expectation of $\mathbb{E}_y[f(y)]$ with $y \sim \mu$ over all distributions μ on X such that $\mathbb{E}_y[y] = x$ with a finite number of atoms.

[16] The area variation is also known as the consumer surplus. It does not have a straightforward welfare interpretation except for the case of the quasilinear domain where AV = EV = CV. Beyond the quasilinear domain, AV is often used as an approximation to the equivalent or compensating variation since AV is easier to compute in practice thanks to the direct observability of the Marshallian demand D (Willig, 1976). Note that CV $\leqslant$ AV $\leqslant$ EV for homothetic preferences which follows from the chain of inequalities $1 - \frac{1}{t} \leqslant \ln t \leqslant t + 1$ with $t = E(\mathbf{p^0})/E(\mathbf{p^1})$.

[17] Formally, $\mathrm{EV} - \mathrm{AV} - \frac{b}{2}(\delta \ln E)^2 = O(||\delta\mathbf{p}||^3)$ as $||\delta\mathbf{p}|| \to 0$.

Corollary 5. *Assume that individuals have preferences from $\mathcal{D}$, the welfare change is measured by either* EV *or* CV, *and the change in prices* $\delta\mathbf{p}$ *is small. The range of possible values for the change in welfare can be expressed through the variance of demand as follows:*

$$\overline{W} - \underline{W} \simeq \frac{1}{2} \cdot \sup_{(\succsim_k, b_k)_{k=1,\ldots}} \mathrm{Var}_{\succsim}\left[\left\langle \delta\mathbf{p},\, D_{\succsim}\left(\mathbf{p^0},\, \sqrt{B}\right)\right\rangle\right], \tag{35}$$

where the maximum is taken over all populations with preferences from $\mathcal{D}$ and given aggregate behavior $\succsim_{\text{aggregate}}$, and $\succsim$ is a random preference equal to $\succsim_k$ with probability proportional to b_k.

An agent with budget $\sqrt{B}$ cannot buy more than $\sqrt{B}/p_i^0$ units of a good i at prices $\mathbf{p^0}$. Hence, the range for small $\delta\mathbf{p}$ is bounded from above by

$$\overline{W} - \underline{W} \lesssim \frac{B}{2} \cdot \max_i \left(\frac{\delta p_i}{p_i^0}\right)^2.$$

We conclude that both for the equivalent and compensating variations, the range of values $\underline{W} - \overline{W}$ has the second order of magnitude in $\mathbf{p^1} - \mathbf{p^0}$ when the change in prices is small.

Example 5. Consider the setting from Example 4: a CES aggregate consumer (29) and individuals having preferences from the domain $\mathcal{D}$ of substitutes. Our goal is to compute the range $\overline{W} - \underline{W}$ explicitly provided that $\delta\mathbf{p} = \mathbf{p^1} - \mathbf{p^0}$ is small.

Note that $\left\langle \delta\mathbf{p},\, D_{\succsim}\left(\mathbf{p^0},\, \sqrt{B}\right)\right\rangle^2$ is a convex functional on logarithmic expenditure functions. Hence, as in Example 4, maximal value in (34) and (35) is attained when $\succsim_{\text{aggregate}}$ is generated by a population of linear consumers. The distribution of MRS $= v_1/v_2$ over the population is given by (30). A linear agent with MRS $= t$ spends all her money on the first good if $\frac{p_1^0}{p_2^0} < t$ and on the second if the inequality is reversed. Hence,

$$\left\langle \delta\mathbf{p},\, D_{\succsim}\left(\mathbf{p^0},\, \sqrt{B}\right)\right\rangle = \sqrt{B}\left(\frac{\delta p_1}{p_1^0} \cdot \mathbb{1}_{\{p_1^0/p_2^0 < t\}} + \frac{\delta p_1}{p_1^0} \mathbb{1}_{\{p_1^0/p_2^0 \geqslant t\}}\right).$$

Thus the variance is equal to the variance of a random variable that equals $\sqrt{B}\delta p_1/p_1^0$ or $\sqrt{B}\delta p_2/p_2^0$ with probabilities

$$\gamma_1 = \mu\big((p_1^0/p_2^0,\, \infty)\big) = \frac{(a_1 \cdot p_2^0)^{\sigma-1}}{(a_1 \cdot p_2^0)^{\sigma-1} + (a_2 \cdot p_1^0)^{\sigma-1}} \qquad \text{or} \qquad \gamma_2 = 1 - \gamma_1.$$

We obtain the following formula

$$\overline{W} - \underline{W} \simeq \frac{B}{2}\left(\left(\frac{\delta p_1}{p_1^0}\right)^2 \gamma_1 + \left(\frac{\delta p_2}{p_2^0}\right)^2 \gamma_2 - \left(\frac{\delta p_1}{p_1^0}\gamma_1 + \frac{\delta p_2}{p_2^0}\gamma_2\right)^2\right)$$

for the range of welfare change. It is applicable both to the equivalent and compensating variations.

4 Invariant domains

In Section 3, we saw that a population of consumers with homothetic preferences can be replaced with a single aggregate consumer. Here we study domains of homothetic preferences invariant with respect to aggregation: if each consumer's preference belongs to the domain, so does the aggregate preference. Tools developed in the previous section reduce invariance to the convexity of the set of logarithmic expenditure functions and yield a flexible procedure for constructing invariant domains.

Recall that we refer to an arbitrary subset of homothetic preferences as a domain.

Definition 2. *A domain $\mathcal{D}$ of homothetic preferences over $\mathbb{R}^n_+$ is invariant with respect to aggregation if for any $m \geqslant 2$ and any population of m consumers with preferences $\succsim_k \in \mathcal{D}$ and budgets $b_k \in \mathbb{R}_{++}$, $k = 1, \ldots, m$, the aggregate preference $\succsim_{\text{aggregate}}$ also belongs to $\mathcal{D}$.*

The set of all homothetic preferences and a domain containing just one preference $\mathcal{D} = \{\succsim\}$ are elementary examples of invariant domains.

Note that it is enough to check the condition of invariance for populations of $m = 2$ consumers. Indeed, aggregation for a population of $m > 2$ consumers reduces to aggregation for pairs by adding consumers one by one sequentially. Hence, if the outcome of aggregation belongs to the domain for any pair, the outcome will belong to this domain for any population.

Recall that $\mathcal{L}_{\mathcal{D}}$ denotes the set of all logarithmic expenditure functions of preferences from $\mathcal{D}$, i.e., $\mathcal{L}_{\mathcal{D}} = \Big\{f \;:\; \mathbb{R}^n_{++} \to \mathbb{R} \quad : \; f = \ln E_{\succsim}, \; \succsim \in \; \mathcal{D}\Big\}$. Each preference $\succsim \in \mathcal{D}$ corresponds to a family of functions from $\mathcal{L}_D$ that differ by a constant. The following result is a direct corollary of Theorem 1.

Corollary 6. *A domain $\mathcal{D}$ is invariant with respect to aggregation if and only if the set of logarithmic expenditure functions $\mathcal{L}_{\mathcal{D}}$ is a convex set of functions on $\mathbb{R}^n_{++}$.*

In other words, $\mathcal{D}$ is invariant whenever, for any pair of preferences $\succsim', \succsim'' \in \mathcal{D}$ with expenditure functions E' and E'' and $\lambda \in (0, 1)$, the preference $\succsim$ corresponding to the expenditure function E defined by

$$\ln E = \lambda \cdot \ln E' + (1 - \lambda) \cdot \ln E'' \tag{36}$$

also belongs to $\mathcal{D}$.

For example, the domain of Cobb-Douglas preferences (12) satisfies the requirement (36) and, hence, is invariant. The domains of preferences exhibiting substitutability or complementarity are also invariant. Indeed, expenditure shares can be obtained by differentiating logarithmic expenditure functions (6) and so the monotonicity conditions defining these domains are preserved under convex combinations.

Corollary 6 not only characterizes invariant domains in geometric terms but also gives a handy tool to construct invariant domains containing a given one. Suppose $\mathcal{D}$ is not invariant. How to complete it to an invariant domain? Of course, $\mathcal{D}$ is contained in the domain of all homothetic preferences which is invariant. To exclude such a trivial answer we require the completion to be minimal with respect to set inclusion.

Definition 3. *For a domain $\mathcal{D}$, its completion $\mathcal{D}^{\text{complete}}$ is the minimal closed domain that is invariant with respect to aggregation and contains $\mathcal{D}$.*

The closure is defined with respect to the metric structure (15) on preferences. The closedness assumption helps to get a tractable answer for infinite domains. As we will see, taking closure is equivalent to enriching $\mathcal{D}^{\text{complete}}$ by aggregate preferences of non-atomic populations with preferences from $\mathcal{D}$.

The completion $\mathcal{D}^{\text{complete}}$ exists since it can be obtained as the intersection of all closed invariant domains containing $\mathcal{D}$ and there is at least one such domain, namely, the domain of all homothetic preferences. Corollary 6 implies a geometric characterization of $\mathcal{D}^{\text{complete}}$.

Corollary 7. *For any domain $\mathcal{D}$ of homothetic preferences, its completion $\mathcal{D}^{\text{complete}}$ is equal to the set of all preferences corresponding to logarithmic expenditure functions from the closed convex hull of $\mathcal{L}_{\mathcal{D}}$:*

$$\mathcal{D}^{\text{complete}} = \left\{ \succsim \ : \ \ln(E_{\succsim}) \in \operatorname{conv}\Big[\mathcal{L}_{\mathcal{D}}\Big]\right\},$$

where $\operatorname{conv}[X]$ *denotes the minimal closed convex set containing X.*

This corollary assumes that the choice of the topology on preferences is aligned with that on logarithmic expenditure functions. This requirement is satisfied by the topology from Appendix B.

Note that $\operatorname{conv}[X]$ can be obtained as the closure of the set of all convex combinations of finite collections of elements from X. For a finite domains $\mathcal{D} = \{\succsim_1, \ldots, \succsim_q\}$, looking at combinations of at most $q = |\mathcal{D}|$ elements is enough and, hence, Corollary 7 is especially easy to apply. For such $\mathcal{D}$, the completion $\mathcal{D}^{\text{complete}}$ consists of all preferences $\succsim$ with expenditure functions of the form $\ln E(\mathbf{p}) = \sum_{k=1}^{q} t_k \cdot \ln E_k(\mathbf{p})$ with $\mathbf{t} \in \Delta_{q-1}$. Reinterpreting Example 1, we conclude that Cobb-Douglas preferences over n goods is the completion of $\mathcal{D} = \left\{ \succsim_1, \ldots, \succsim_n \right\}$ where $\succsim_i$ corresponds to the utility function $u_i(\mathbf{x}) = x_i$.

To compute the completion for infinite domains $\mathcal{D}$, we need to take the closure of the set of preferences $\succsim$ corresponding to all finite convex combinations of logarithmic expenditure functions

$$\ln E(\mathbf{p}) = \sum_{k=1}^{q} t_k \cdot \ln E_k(\mathbf{p}),$$

where $q \geqslant 1$, a vector $\mathbf{t} \in \Delta_q$, and E_k represents some preference $\succsim_k$ from $\mathcal{D}$. It is convenient to think about this convex combination as a result of integration with respect to the atomic distribution μ placing weight t_k on preference $\succsim_k$:

$$\ln E(\mathbf{p}) = \int_{\mathcal{D}} \ln E_{\succsim'}(\mathbf{p})\, \mathrm{d}\mu(\succsim'). \tag{37}$$

It turns out that taking closure is equivalent to allowing arbitrary probability measures μ in (37), not necessarily atomic. For parametric domains such as linear or Leontief preferences discussed below, the integral in (37) can be seen as the integral over the space of parameters and, hence, passing to an arbitrary μ is straightforward. In Appendix B, we explain how to define (37) for any domain $\mathcal{D}$ and measure μ.

Theorem 2. *The completion of a domain $\mathcal{D}$ consists of all preferences $\succsim$ such that their expenditure function E can be represented as*

$$\ln E(\mathbf{p}) = \int_{\overline{\mathcal{D}}} \ln E_{\succsim'}(\mathbf{p})\, \mathrm{d}\mu(\succsim') \tag{38}$$

with some Borel probability measure μ supported on the closure $\overline{\mathcal{D}}$ of $\mathcal{D}$.

With general μ, representation (38) can be interpreted as the result of preference aggregation where non-atomic populations are allowed and μ plays the role of preference distribution over the population. In what follows, we refer to (38) as continuous aggregation.

A generalization of Theorem 2 is proved in Appendix C.3 together with Theorem 3 formulated in the next section. Both results rely on Choquet theory which studies extreme points of compact convex sets in topological vector spaces (Phelps, 2001). Application of this theory requires careful choice of a topology and a measurable structure. For the proof to work, it is crucial that the sets of preferences and logarithmic expenditure functions endowed with the distance (15) are compact and admit an isometric embedding in a Banach space.

4.1 ARUM and the completion of linear preferences

Consider the domain $\mathcal{D}$ of all linear preferences. Our goal is to characterize its completion $\mathcal{D}^{\text{complete}}$. This problem turns out to be tightly related to stochastic discrete choice theory. The intuition behind this connection is as follows. By Theorem 2, finding the completion boils down to taking the average of logarithmic expenditure functions with respect to some measure μ. This average can be thought of as the expectation over random preferences of a single decision maker and expenditure shares can be interpreted as probabilities of choosing one of n possible alternatives.

In the additive random utility model (ARUM), there is a single decision maker choosing between one of n alternatives. Her utility for alternative i is equal to $w_i + \varepsilon_i$, where w_i is a deterministic component and ε_i is a stochastic shock. The vector $\mathbf{w} \in \mathbb{R}^n$ and the joint distribution of shocks $(\varepsilon_1, \ldots, \varepsilon_n) \in \mathbb{R}^n$ are given. For each realization of the shocks, the agent selects the alternative with the highest utility. Hence, the expected utility of the decision maker and the probability that she chooses alternative i are equal to[18]

$$U(\mathbf{w}) = \mathbb{E}\left[\max_{i=1,\ldots n}(w_i + \varepsilon_i)\right] \quad \text{and} \quad S_i(\mathbf{w}) = \mathbb{E}\left[\{w_i + \varepsilon_i > w_j + \varepsilon_j \ \forall j \neq i\}\right],$$

where $\mathbb{E}$ and $\mathbb{P}$ denote the expectation and the probability with respect to the shock distribution.

Proposition 2. *A preference $\succsim$ with an expenditure function E belongs to the completion of the domain of all linear preferences over n goods if and only if there is a distribution of shocks such that*

$$U(\mathbf{w}) = -\ln\left(E(e^{-w_1}, \ldots, e^{-w_n})\right) \tag{39}$$

is the expected utility in ARUM with deterministic utilities $\mathbf{w} \in \mathbb{R}^n$.

Taking the gradient on both sides of (39) gives a version of the statement for expenditure shares:[19] $\succsim$ is in the completion of linear preferences whenever $\mathbf{s}(e^{-w_1}, \ldots, e^{-w_n})$ is the vector of choice probabilities for some additive random utility model, i.e., there exists a distribution of shocks such that

$$s_i(e^{-w_1}, \ldots, e^{-w_n}) = \mathbb{P}\left[\{w_i + \varepsilon_i > w_j + \varepsilon_j \ \forall j \neq i\}\right] \tag{40}$$

for all $i = 1, \ldots, n$ and Lebesgue almost all $\mathbf{w} \in \mathbb{R}^n$.

Substituting formula (9) for expenditure functions of linear preferences into Theorem 2, we see that the completion of linear preferences consists of all preferences $\succsim$ whose expenditure functions E can be represented as

$$\ln E(\mathbf{p}) = \int_{\mathbb{R}^n_+} \ln\left(\min_{i=1,\ldots,n} \frac{p_i}{v_i}\right) \mathrm{d}\mu(\mathbf{v}) \tag{41}$$

for some measure μ such that the integral converges. To get (39), it is enough to change variables in (41). Denote $\varepsilon_i = \ln v_i$ and interpret $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ as a random vector by assuming that $\mathbf{v}$ is sampled from distribution μ. Plugging in $p_i = e^{-w_i}$, we get $-\ln\left(\min_{i=1,\ldots,n} \frac{p_i}{v_i}\right) = \max_{i=1,\ldots,n}(w_i +$

[18]The formula for the choice probabilities holds for $\mathbf{w}$ such that the probability of a tie $w_i + \varepsilon_i = w_j + \varepsilon_j$ is zero. This requirement is satisfied for Lebesgue almost all $\mathbf{w}$ no matter what the distribution of the shocks is.

[19]The fact that the choice probabilities $S_i(\mathbf{w})$ can be obtained as partial derivatives of decision maker's utility is known as the Williams–Daly–Zachary theorem and its classical version requires regularity assumptions on the distribution of shocks (McFadden, 1981). The possibility to drop all the assumptions and get the conclusion for Lebesgue almost all $\mathbf{w}$ is a recent result (Sørensen and Fosgerau, 2022). The connection between ARUM and aggregation of linear preferences makes this result a corollary of general formula (6) expressing expenditure shares as the gradient of logarithmic expenditure functions for almost all prices.

ε_i) and, hence, (41) is equivalent to

$$-\ln\left(E(e^{-w_1},\dots,e^{-w_n}\right) = \mathbb{E}\left[\max_{i=1,\dots n}(w_i+\varepsilon_i)\right].$$

As the right-hand side has the form of the expected utility in ARUM, we obtain Proposition 2.

The class of vector functions that can arise as choice probabilities $\mathbf{S}(\mathbf{w})$ for some ARUM is well-studied in the discrete choice theory. We need the following necessary condition applicable to smooth vector functions. For any ARUM with n alternatives and any subset of distinct alternatives $i, j_1, j_2, \dots, j_q$ with $q \leqslant n-1$, the following inequality holds

$$\frac{\partial^q S_i(\mathbf{w})}{\partial w_{j_1}\partial w_{j_2}\dots\partial w_{j_q}}\cdot(-1)^q \leqslant 0$$

at any $\mathbf{w}$ where $\mathbf{S}$ is q times differentiable (Anderson et al., 1992). Taking into account the connection between expenditure shares and choice probabilities (40) and the identity $\frac{\partial}{\partial w_i} = -\frac{\partial}{\partial \ln p_i}$ for $p_i = e^{-w_i}$, we obtain the following corollary of Proposition 2.

Corollary 8. *If a preference $\succsim$ belongs to the completion of the domain of all linear preferences, then its expenditure shares satisfy the following inequalities*

$$\frac{\partial s_i(\mathbf{p})}{\partial \ln p_{j_1}\partial \ln p_{j_2}\dots\partial \ln p_{j_q}} \geqslant 0 \tag{42}$$

for any distinct goods $i, j_1, j_2, \dots, j_q$ with $q \leqslant n-1$ at any price vector $\mathbf{p} \in \mathbb{R}^n_{++}$ where $\mathbf{s}$ is differentiable q times.

For $q=1$, the condition (42) becomes the substitutability condition (8). In other words, any preference $\succsim$ from the completion of linear preferences exhibits substitutability among goods. This conclusion is not surprising as linear preferences exhibit substitutability and aggregation respects this property.

One could expect that any preference exhibiting substitutability is in the completion of linear preferences. However, for $n \geqslant 3$ goods, the condition (42) gives extra restrictions on top of substitutability by restricting the second-order derivatives.

In the following example, we construct a preference over $n = 3$ substitutes such that the second-order elasticities of $\mathbf{s}$ are negative and, hence, this preference is not in the completion of linear ones. This example is a special case of Example 4 from (Matsuyama and Ushchev, 2017).

Example 6. Consider the following expenditure function over $n = 3$ goods:

$$E(\mathbf{p}) = (p_1+p_2+p_3)^{1-\alpha}\left(p_1^{1/3}\cdot p_2^{1/3}\cdot p_3^{1/3}\right)^{\alpha}, \tag{43}$$

with $1 < \alpha < 3/2$. The corresponding expenditure share of good $i = 1, 2, 3$ is given by

$$s_i(\mathbf{p}) = \frac{\partial \ln E(\mathbf{p})}{\partial \ln p_i} = \frac{\alpha}{3} + (1-\alpha)\frac{p_i}{p_1 + p_2 + p_3}. \tag{44}$$

The restriction $\alpha < 3/2$ guarantees that $s_i(\mathbf{p}) > 0$ for all price vectors, while the restriction $\alpha > 1$ guarantees substitutability. Yet, (43) does not belong to the invariant domain of linear preferences:

$$\frac{\partial s_1(\mathbf{p})}{\partial \ln p_2 \, \partial \ln p_3} = 2(1-\alpha)\frac{p_1 \cdot p_2 \cdot p_3}{(p_1 + p_2 + p_3)^3} < 0,$$

which violates (42). It remains to verify that (43) is an expenditure function of some homothetic preference. For this purpose, we need to check that E is homogeneous, monotone, and concave. Homogeneity is straightforward. Monotonicity follows since the elasticities (44) are all positive. The concavity of E follows from quasi-concavity since E is homogeneous. To prove quasi-concavity we show that $\ln E$ is concave. To check the latter, we compute the quadratic form of the Hessian of $\ln E(\mathbf{p})$ on a vector $\mathbf{y} \in \mathbb{R}^3 \backslash \{0\}$:

$$\frac{\alpha}{3}\left(\frac{y_1^2}{p_1^2} + \frac{y_2^2}{p_2^2} + \frac{y_3^2}{p_3^2}\right) + (\alpha - 1)\left(\frac{y_1 + y_2 + y_3}{p_1 + p_2 + p_3}\right)^2 < -\left(1 - \frac{2}{3}\alpha\right)\max_{i=1,2,3}\frac{y_i^2}{p_i^2} < 0.$$

Hence, the Hessian is negative definite, which implies the concavity. Thus (43) is indeed an expenditure function.

Corollary 9. *For $n \geqslant 3$ goods, the completion of the domain of all linear preferences is a proper subset of the domain of preferences exhibiting substitutability.*

The corollary tells nothing about the case of $n = 2$ goods, which turns out to be an exception.

Proposition 3. *For $n = 2$ goods, the completion of the domain of all linear preferences coincides with the set of all preferences exhibiting substitutability.*

This result follows from an explicit construction. Given $\succsim$ such that the expenditure share of the first good $s_1(p_1, p_2) = \frac{\partial \ln E(p_1, p_2)}{\partial \ln p_1}$ is non-decreasing in p_2, we need to find a distribution μ of value vectors $\mathbf{v}$ so that the continuous aggregation (41) of linear preferences leads to $\succsim$.

To guess an explicit formula for such μ, take the partial derivative $\frac{\partial}{\partial \ln p_1}$ on both sides of (41):

$$s_1(p_1, p_2) = \mu\left(\left\{\frac{v_1}{v_2} \geqslant \frac{p_1}{p_2}\right\}\right). \tag{45}$$

The derivative exists and the identity holds for Lebesgue almost all (p_1, p_2). The ratio $\mathrm{MRS} = v_1/v_2$ is the marginal rate of substitution for the corresponding linear preference. We conclude that $1 - s_1(\,\cdot\,, 1)$ must be the cumulative distribution function of MRS and the monotonicity of s_1 makes this possible. Choosing any such distribution μ and adding atoms of the weight $1-\lim_{p_1\to+0} s_1(p_1,1)$ at $\mathbf{v}=(0,1)$ and of the weight $\lim_{p_1\to\infty} s_1(p_1,1)$ at $\mathbf{v}=(1,0)$ completes the construction.

Note that we pinned down the distribution of the MRS $= v_1/v_2$ but not the magnitude of $\mathbf{v}$. As $\mathbf{v}$ and $\lambda\cdot\mathbf{v}$ with $\lambda>0$ correspond to the same linear preference, the distribution of preferences over the population is determined uniquely and we are free to choose any normalization of $\mathbf{v}$ so that the integral in (41) converges, e.g., we can assume that μ is supported on $v_1+v_2=1$.

Corollary 10. *Any preference over two goods exhibiting substitutability can be represented as a continuous aggregation of linear preferences* (41)*. The distribution of linear preferences over the population corresponding to $\succsim$ is pinned down uniquely and admits an explicit formula: the cumulative distribution function of the marginal rate of substitution* MRS $= v_1/v_2$ *equals* $1-s_1(\,\cdot\,,1)$*.*

Example 7 (translog preferences and Benford's law)*.* To illustrate Corollary 10, let us show how a family of consumers with linear preferences over two goods can generate translog preferences, a popular class of homothetic preferences obtained as a perturbation of Cobb-Douglas preferences in the space of expenditure functions (Diewert, 1974, p.139). A preference $\succsim$ is translog if its logarithmic expenditure function has the following form

$$\ln E(p_1,p_2)=\begin{cases}\ln p_1, & \ln\left(\frac{p_1}{p_2}\right)<-\frac{1-\alpha}{\beta}\\ \alpha\ln p_1+(1-\alpha)\ln p_2-\frac{\beta}{2}\left(\ln\frac{p_1}{p_2}\right)^2, & -\frac{1-\alpha}{\beta}\leqslant\ln\left(\frac{p_1}{p_2}\right)\leqslant\frac{\alpha}{\beta}\\ \ln p_2, & \ln\left(\frac{p_1}{p_2}\right)>\frac{\alpha}{\beta}\end{cases},$$

where $\alpha\in(0,1)$ and $\beta>0$.

By elementary computations and formula (45), we obtain that a distribution of value vectors $\mathbf{v}=(v_1,v_2)$ aggregates up to the translog preference if and only if the logarithm of the marginal rate of substitution MRS $= v_1/v_2$ is distributed uniformly:

$$\ln \mathrm{MRS}\sim\mathbb{U}\left(\left[-\frac{1-\alpha}{\beta},\ \frac{\alpha}{\beta}\right]\right),$$

where $\mathbb{U}([c,d])$ denotes the uniform distribution supported on $[c,d]$. Curiously enough, this is equivalent to MRS following the so-called Benford law of digit bias (Benford, 1938).

Consider a particular case of a translog preference $\succsim$ with $\alpha=\beta=1/2$. We obtain that $\succsim$ can be represented as aggregation over the continuous population of consumers distributed uniformly on $[-1,1]$ so that consumer $\alpha\in[-1,1]$ has utility $u(x_1,x_2)=e^{\frac{\alpha}{2}}\cdot x_1+e^{-\frac{\alpha}{2}}\cdot x_2$. The corresponding identity (41) takes the following form:

$$\ln E(p_1,p_2)=\int_{-1}^{1}\ln\left(\min\left\{p_1\cdot e^{-\frac{\alpha}{2}},\ p_2\cdot e^{\frac{\alpha}{2}}\right\}\right)\,\mathrm{d}\alpha.$$

The next example shows how a CES preference over two substitutes arises as an aggregation of linear preferences.[20]

Example 8 (CES as an aggregation of linear preferences)*.* Consider a CES preference $\succsim$ over two substitutes:

$$u(\mathbf{x}) = \left((a_1 \cdot x_1)^{\frac{\sigma-1}{\sigma}} + (a_2 \cdot x_2)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} \qquad \text{with} \qquad \sigma > 1.$$

Using (45) and formula (14) for expenditure shares, we conclude that the cumulative distribution function of the marginal rate of substitution $\mathrm{MRS} = v_1/v_2$ in a population generating $\succsim$ takes the following Pareto-like form:

$$\mu\left(\{\mathrm{MRS} < t\}\right) = 1 - s_{\succsim,1}(t,1) = \frac{t^{\sigma-1}}{t^{\sigma-1} + \left(\frac{a_2}{a_2}\right)^{\sigma-1}}. \tag{46}$$

4.2 Complete monotonicity and the completion of Leontief preferences

We saw that, for $n = 2$ goods, the completion of linear preferences is the whole domain of preferences exhibiting substitutability. By contrast, the completion of all Leontief preferences turns out to be substantially narrower than the domain of all preferences with complementarity, even for $n = 2$.

By Theorem 2 and formula (10) for expenditure functions of Leontief preferences, the completion of Leontief preferences over $n \geqslant 2$ goods is the set of all preferences $\succsim$ with expenditure functions E of the following form:

$$\ln E(\mathbf{p}) = \int_{\mathbb{R}^n_+} \ln\langle \mathbf{v}, \mathbf{p}\rangle \, \mathrm{d}\mu(\mathbf{v}) \tag{47}$$

for some probability measure μ on $\mathbb{R}^n_+$ such that the integral converges.

Note that $\ln\langle \mathbf{v}, \mathbf{p}\rangle$ is an infinitely smooth function of $\mathbf{p} \in \mathbb{R}^n_{++}$. Exchanging integration and differentiation in (47), we conclude that the left-hand side must also be infinitely smooth. Thus the completion of Leontief preferences cannot contain preferences with expenditure functions and expenditure shares that are not infinitely smooth.

Corollary 11. *For any number $n \geqslant 2$ of goods, the completion of the domain of Leontief preferences is a proper subset of the domain of preferences exhibiting complementarity.*

The following example provides a concrete preference over two complements that is outside of the completion of Leontief preferences.

Example 9 (A preference over two complements outside of the completion of Leontief)*.* We aim to find a preference $\succsim$ over $n = 2$ complements such that its expenditure function E is not infinitely smooth. It is enough to find a preference such that the expenditure share of the first good

[20] With the help of Proposition 2, this example implies a 2-good version of the result by Anderson et al. (1987) connecting CES and nested logit.

$s_1(p_1,p_2) = \frac{\partial \ln E(p_1,p_2)}{\partial \ln p_1}$ has a discontinuous derivative. The existence of such $\succsim$ follows from the characterization of expenditure shares (7). We will describe $\succsim$ explicitly.

The idea is to combine two distinct preferences exhibiting complementarity so that the consumer's preference alternates between the two depending on prices. Consider $\succsim$ corresponding to the following utility function:

$$u(x_1,x_2) = \min\left\{\sqrt{x_1 \cdot x_2}, \quad x_1\right\}. \tag{48}$$

A consumer with this preference behaves as if she switches between Cobb-Douglas and Leontief preferences at $p_1 = p_2$. A simple computation gives the expenditure share:

$$s_1(p_1,p_2) = \begin{cases} \frac{1}{2}, & \frac{p_1}{p_2} < 1 \\ \frac{p_1}{p_1+p_2}, & 1 \leqslant \frac{p_1}{p_2} \end{cases}.$$

As we see, s_1 is decreasing in p_2 and, hence, $\succsim$ exhibits complementarity. However, this preference cannot be obtained as a continuous aggregation of Leontief preferences (47) since the expenditure share has a discontinuous derivative.[21]

The condition that a preference $\succsim$ belongs to the completion of Leontief preferences is substantially more restrictive than the requirement of smoothness of the expenditure function. An infinitely smooth function $f = f(\lambda)$, $\lambda \in \mathbb{R}_{++}$, is called completely monotone if [22]

$$(-1)^k \cdot \frac{\mathrm{d}^k}{\mathrm{d}\lambda^k} f \geqslant 0 \quad \text{for all } k = 0,1,2,\ldots.$$

Complete monotonicity provides a necessary condition for $\succsim$ to be in the completion of Leontief preferences.

Proposition 4. *If a preference $\succsim$ belongs to the completion of Leontief preferences, then the demand $D_i(\mathbf{p}, b)$ for a good i is a completely monotone function of p_i for each $i = 1,\ldots,n$ and $b > 0$.*

This proposition follows from the integral representation (47) of the expenditure function E and the possibility to exchange differentiation with respect to p_i and integration. The derivatives of the

[21]For complements, expenditure shares can have discontinuous derivatives but are necessarily continuous themselves. This is a simple corollary of (7). By contrast, expenditure shares for substitutes can be discontinuous, e.g., for linear preferences expenditure shares are step functions.

[22]In economics, completely monotone functions arise as the dependence of decision maker's payoff on the discount factor λ. Indeed, by Bernstein's theorem, a function f is completely monotone if and only if $f(\lambda) = \int_{\mathbb{R}_+} e^{-\lambda t} \mathrm{d}\nu(t)$ for some positive measure ν (Schilling et al., 2012, Theorem 1.4), i.e., f is the expected utility of a risk-neutral decision maker with geometric discounting for a stream of payoff ν.

Utility functions with mixed risk aversion ($u(x) = \int_{\mathbb{R}_+} (1 - e^{-\lambda x}) \mathrm{d}\nu(x)$) provide another economic context for completely monotone functions (Caballé and Pomansky, 1996).

integrand in (47) can be computed explicitly

$$\frac{\partial^{k+1}}{\partial^k p_i} \ln\langle \mathbf{v}, \mathbf{p}\rangle = (-1)^k \frac{v_i^k \cdot k!}{(\langle \mathbf{v}, \mathbf{p}\rangle)^{k+1}}.$$

Hence,

$$(-1)^k \frac{\partial^{k+1}}{\partial^k p_i} \ln E(\mathbf{p}) = k! \int_{\mathbb{R}_+^n} \frac{v_i^k}{(\langle \mathbf{v}, \mathbf{p}\rangle)^{k+1}} \,\mathrm{d}\mu(\mathbf{v}) \geqslant 0.$$

Since $D_i(\mathbf{p}, b) = b \cdot \frac{\partial}{\partial p_i} \ln E(\mathbf{p})$ by formula (4), we conclude that $D_i(\mathbf{p}, b)$ is a completely monotone functions of p_i.

For $n = 2$ goods, we are able to provide a simple criterion for a preference $\succsim$ to be in the completion of Leontief preferences. This criterion suggests that the necessary condition of complete monotonicity established in Proposition 4.2 is almost sufficient.

A function $f = f(\lambda)$, $\lambda \in \mathbb{R}_{++}$, is called a Stieltjes function if it can be represented as

$$f(\lambda) = \int_{\mathbb{R}_+} \frac{1}{\lambda + z} \mathrm{d}\nu(z) \tag{49}$$

for some positive measure ν on $\mathbb{R}_+$.[23] The Stieltjes functions are exactly those completely monotone functions that themselves can be obtained as the Laplace transform of a completely monotone density.[24]

Proposition 5. *For $n = 2$ goods, a preference $\succsim$ belongs to the completion of Leontief preferences if and only if the demand for the first good $D_1(\mathbf{p}, b)$ with $\mathbf{p} = (p_1, 1)$ and $b = 1$ is a Stieltjes function of the price p_1.*

By (47), for any $\succsim$ in the completion, we have

$$\frac{\partial}{\partial p_1} \ln E(p_1, 1) = \int_{\mathbb{R}_+^2} \frac{v_1}{v_1 p_1 + v_2} \,\mathrm{d}\mu(v_1, v_2) = \int_{\mathbb{R}_+} \frac{1}{p_1 + z} \mathrm{d}\nu(z),$$

where ν is the distribution of v_2/v_1. Since $D_1((p_1, 1), 1) = \frac{\partial}{\partial p_1} \ln E(p_1, 1)$, we get

$$D_1((\lambda, 1), 1) = \int_{\mathbb{R}_+} \frac{1}{\lambda + z} \mathrm{d}\nu(z) \tag{50}$$

[23] These functions are omnipresent in various branches of mathematics such as probability theory, spectral theory, continued fractions, and potential theory (Schilling et al., 2012).

[24] The integral operator on the right-hand side of (49) is known as the Stieltjes transform. It equals the Laplace transform applied to ν twice: $f(\lambda) = \int_{\mathbb{R}_+} e^{-\lambda t} \left(\int_{\mathbb{R}_+} e^{-tz} \mathrm{d}\nu(z) \right) \mathrm{d}t$. By Bernstein's theorem, the set of completely monotone functions is the set of all functions equal to the Laplace transform of positive measures (Schilling et al., 2012, Theorem 1.4). Hence, Stieltjes functions are those completely monotone functions that are Laplace transforms of completely monotone ones.

and conclude that $D_1((\lambda,1),1)$ is a Stieltjes function for any $\succsim$ from the completion. In the opposite direction, if f is a Stieltjes function, then $f(\lambda)=D_1((\lambda,1),1)$ for a preference $\succsim$ from the completion of Leontief preferences corresponding to μ in (47) such that the ratio v_2/v_1 is ν-distributed.

The right-hand side of (50) is called the Stieltjes transform of ν. The Stieltjes transform is invertible (Schilling et al., 2012, Chapter 2). Hence, if $\succsim$ belongs to the completion of Leontief preferences, the demand determines the distribution ν satisfying (50) uniquely. As a result, the distribution of Leontief preferences over the population that leads to $\succsim$ is pinned down uniquely. Namely, the continuous aggregation of Leontief preferences (47) with distribution μ of (v_1, v_2) such that the ratio v_2/v_1 is distributed according to ν gives $\succsim$. Leontief preferences with the same ratio coincide and so the distribution of preferences corresponding to $\succsim$ is indeed unique.

The Stieltjes transform can be inverted explicitly using tools from complex analysis. Before describing the tools, we give an example obtained with their help.

Example 10 (CES with complements as an aggregation of Leontief preferences). We show that any CES preference $\succsim$ over $n=2$ complements (13) belongs to the completion of Leontief preferences.

First, consider a particular case with the elasticity of substitution $\sigma=\frac{1}{2}$ and weights $\mathbf{a}=(1,1)$; the corresponding utility function is the harmonic mean. The utility and the demand for the first good are as follows:

$$u(x_1,x_2)=\left(\frac{1}{x_1}+\frac{1}{x_2}\right)^{-1} \qquad \text{and} \qquad D_1((p_1,p_2),b)=\frac{b}{p_1+\sqrt{p_1\cdot p_2}}.$$

By Proposition 5, finding a probability distribution ν on $\mathbb{R}_+$ such that the identity (50) holds is enough to show that $\succsim$ is in the completion of Leontief preferences. We end up with the following equation:

$$\frac{1}{\lambda+\sqrt{\lambda}}=\int_{\mathbb{R}_+}\frac{1}{\lambda+z}\mathrm{d}\nu(z).$$

One can check that ν with a density φ given by

$$\varphi(z)=\frac{1}{\pi}\frac{1}{\sqrt{z}(1+z)} \tag{51}$$

is a solution, hence, $\succsim$ is indeed in the completion. By taking any distribution μ of $\mathbf{v}=(v_1,v_2)$ such that v_2/v_1 is ν-distributed (e.g., v_1 equals 1 identically and v_2 has distribution ν), we represent $\succsim$ via a continuous aggregation of Leontief preferences (47).

The above analysis extends to any CES preference over two complements. The utility function and the expenditure share have the form

$$u(x_1,x_2)=\left((a_1\cdot x_1)^{\frac{\sigma-1}{\sigma}}+(a_2\cdot x_2)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} \qquad \text{and} \qquad D_1((p_1,p_2),b)=\frac{b}{p_1+(p_1)^{\sigma}\cdot\left(\frac{a_1}{a_2}p_2\right)^{1-\sigma}},$$

where $\sigma \in (0,1)$. The corresponding distribution ν of v_2/v_1 has to solve the equation

$$\frac{1}{\lambda + \left(\frac{a_1}{a_2}\right)^{1-\sigma} \cdot \lambda^{\sigma}} = \int_{\mathbb{R}_+} \frac{1}{\lambda + z} \mathrm{d}\nu(z).$$

One can check that ν with density

$$\varphi(z) = \frac{\sin(\pi\sigma)}{\pi} \left(\frac{1}{\left(\frac{a_2}{a_1}\right)^{1-\sigma} \cdot z^{2-\sigma} + z \cdot \cos(\pi\sigma) + \left(\frac{a_1}{a_2}\right)^{1-\sigma} \cdot z^{\sigma}} \right) \tag{52}$$

is a solution. Formula (51) is a particular case of (52) for $\sigma = 1/2$ and $a_1 = a_2$.

Formulas (51) and (52) were derived using the following observation from complex analysis. For any distribution ν on $\mathbb{R}_+$, its Stieltjes transform is defined not only for $\lambda \in \mathbb{R}_{++}$ but also for all complex values of $\lambda \in \mathbb{C}\backslash\mathbb{R}_-$, where $\mathbb{C}$ denotes the complex plane. Moreover, the function is analytic on $\mathbb{C}\backslash\mathbb{R}_-$. The values of this analytic continuation above and below the "cut" over the negative reals can be used to reconstruct ν. The answer is given by the Stieltjes-Perron formula: if f is the Stieltjes transform of a measure ν with density φ, then

$$\varphi(z) = \frac{1}{2\pi i} \cdot \lim_{\varepsilon \to 0} \left(f(-z + i\varepsilon) - f(-z - i\varepsilon) \right), \tag{53}$$

where i is the imaginary unit and ε tends to zero from above.

Combining the Stieltjes-Perron formula and Proposition 5, we get the following corollary.

Corollary 12. *If a preference $\succsim$ over $n = 2$ goods belongs to the completion of Leontief preferences, then the demand for the first good $D_1((p_1, 1), 1)$ as a function of its price p_1 admits an analytic continuation to complex prices $p_1 \in \mathbb{C}\backslash\mathbb{R}_+$. The function φ given by (53) for*

$$f(\lambda) = D_1((\lambda, 1), 1)$$

is the density of the unique distribution of v_2/v_1 such that the continuous aggregation of Leontief preferences (47) gives $\succsim$.

Note that the analytic continuation is unique if exists. Hence, Corollary 12 can be used to check whether a given preference belongs to the completion of Leontief preferences. First we check whether the expenditure share admits an analytic continuation. If it does, we compute a candidate for the distribution ν via the Stieltjes-Perron formula. Finally, we check that what we got is a probability distribution and the expenditure share can be obtained as its Stieltjes transform. A preference passes the test if and only if it is in the completion. Example 10 illustrated this approach.

4.3 Application: Fisher markets, fair division, complexity, and bidding languages

Consider a population of consumers with budgets $b_1, \ldots, b_m$ and homothetic preferences $\succsim_1, \ldots, \succsim_m$ over n goods. Let us augment this setting by adding a bundle $\mathbf{x} \in \mathbb{R}^n_{++}$ interpreted as a fixed total supply of the goods. This economy is known in algorithmic economics literature as the Fisher market[25] and is by far the most studied economy from computational perspective (Nisan et al., 2007, Chapters 5 and 6). Since the classical works of Varian (1974) and Hylland and Zeckhauser (1979), Fisher markets and their modifications are used for fair allocation of private goods without monetary transfers giving rise to a famous mechanism known as the competitive equilibrium with equal incomes (CEEI) or the pseudo-market mechanism (Moulin, 2019; Pycia, 2022).

A collection of bundles $\mathbf{x}_1, \ldots, \mathbf{x}_m$ and a price vector $\mathbf{p}$ form a competitive equilibrium (CE) of the Fisher market with preferences $\succsim_1, \ldots, \succsim_m$, budgets $b_1, \ldots, b_m$, and total supply $\mathbf{x}$ if

$$\mathbf{x}_k \in D_k(\mathbf{p}, b_k) \quad \text{for each consumer } k \qquad \text{and} \qquad \sum_{k=1}^{m} \mathbf{x}_k = \mathbf{x}, \tag{54}$$

i.e., each consumer buys the most preferred bundle within her budget and the market clears. We pinpoint that money has no intrinsic value and the Fisher market is equivalent to an exchange economy where each agent k is endowed with the fraction $\beta_k = b_k/B$ of $\mathbf{x}$ where B is the total budget.

One can think of a CE as an allocation mechanism: agents report their preferences, and the mechanism computes an equilibrium and allocates each agent her bundle $\mathbf{x}_k$. In this interpretation, budgets b_k represent agents' entitlement to the goods in the bundle $\mathbf{x}$. The case of equal entitlements $b_1 = \ldots = b_m$ (CEEI) is especially important. In this case, each agent selects her best bundle from the same budget set and, hence, the resulting allocation is envy-free in the sense that $\mathbf{x}_k \succsim_k \mathbf{x}_l$ for any pair of agents k and l. Since any CE is Pareto optimal by the first welfare theorem, CEEI gives a simple recipe to combine strong fairness and efficiency guarantees. CEEI and its variants have been applied to rent division (Goldman and Procaccia, 2015), chores allocation (Bogomolnaia et al., 2017), course allocation (Budish et al., 2017; Kornbluth and Kushnir, 2021; Soumalias et al., 2022), cloud computing (Devanur et al., 2018), school choice (Ashlagi and Shi, 2016; He et al., 2018), and other problems (Echenique et al., 2021).

Despite its attractive properties, the popularity of CEEI remains limited as computing its outcome is a challenging problem. It is known that an equilibrium allocation $\mathbf{x}_1, \ldots, \mathbf{x}_m$ can be obtained

[25] Named after Irving Fisher who introduced a hydraulic method for equilibrium price computation; see (Brainard and Scarf, 2005).

via maximizing the Nash social welfare

$$\prod_{k=1}^{m}\left(\frac{u_k(\mathbf{x}_k)}{\beta_k}\right)^{\beta_k} \tag{55}$$

over all bundles $\mathbf{x}_1,\ldots,\mathbf{x}_m$ such that $\sum_{k=1}^m \mathbf{x}_k = \mathbf{x}$. This result tightly related to the existence of an aggregate consumer was established by Eisenberg and Gale (1959) for linear preferences but holds for all homothetic preferences; see (Shafer and Sonnenschein, 1982). Although the Eisenberg-Gale problem is convex, computing its solutions is not an easy task unless n or m are small. Even in the benchmark case of linear preferences, algorithms with good theoretical performance have required more than a decade of research and dozens of papers using cutting-edge techniques; see, e.g., (Devanur et al., 2002; Orlin, 2010; Végh, 2012). Developing algorithms with good performance in practice is critical for large-scale applications of Fisher markets — e.g., to fair recommender systems (Gao and Kroer, 2022) and Internet ad markets (Conitzer et al., 2022) — but despite the recent progress this problem is yet to be solved.

We examine the question of finding a CE from the preference aggregation perspective. This perspective sheds light on why computing a CE can be challenging in seemingly innocent domains such as linear preferences and helps to identify domains where computing a CE is easy.

The find a CE, it is enough to compute the vector of equilibrium prices $\mathbf{p}$. Once we know $\mathbf{p}$, each agent is allocated her demanded bundle $\mathbf{x}_k$ at these prices.[26] Thus the essence of computing a CE is finding a vector of prices $\mathbf{p}$ such that the market demand matches the supply. In other words, we need to find $\mathbf{p}$ such that *the aggregate consumer's demand contains* $\mathbf{x}$. This simple observation combined with our insights about the structure of aggregate preferences has many implications.

Finding a CE for a population of consumers boils down to finding a CE for one aggregate consumer and we know that aggregation is easier to handle in the space of logarithmic expenditure functions. Recall that the demand is proportional to the gradient of the logarithmic expenditure function (4) and, hence, $\mathbf{p}$ is an equilibrium price vector if and only if[27]

$$\mathbf{x} = B\cdot\nabla\ln E_{\text{aggregate}}(\mathbf{p}),$$

where B is the total budget. Interpreting this identity as the first order condition and taking into account the concavity of $\ln E_{\text{aggregate}}$, we conclude that $\mathbf{p}$ is a vector of equilibrium prices whenever

$$\mathbf{p} \quad \text{is the global maximum of} \quad \langle\mathbf{x},\mathbf{p}\rangle - B\cdot\ln E_{\text{aggregate}}(\mathbf{p}). \tag{56}$$

[26] If utilities are not strictly concave, $D_k(\mathbf{p}, b)$ may not be a singleton, e.g., in the domain of linear preferences. Even in this case, once equilibrium prices are known, choosing bundles $\mathbf{x}_k$ from each agent's demand so that $\sum_k \mathbf{x}_k = \mathbf{x}$ is a simple problem, which boils down to a maximum flow computation (Devanur et al., 2002; Brânzei and Sandomirskiy, 2019).

[27] For expenditure functions that are not smooth the gradient is to be replaced with the superdifferential.

Combining this result with Theorem 1, we get the following proposition.

Proposition 6. *A vector* $\mathbf{p}$ *is a vector of equilibrium prices for a population of consumers with homothetic preferences* $\succsim_1, \ldots, \succsim_m$*, budgets* $b_1, \ldots, b_m$*, and total supply* $\mathbf{x}$ *if and only if* $\mathbf{p}$ *is the global maximum of*

$$\langle \mathbf{x}, \mathbf{p} \rangle - \sum_{k=1}^{m} b_k \cdot \ln E_k(\mathbf{p}). \tag{57}$$

This optimization problem is convex. Its particular case for linear preferences has been known and obtained as the Lagrange dual to the Eisenberg-Gale optimization problem (Cole et al., 2017; Devanur et al., 2016; Shmyrev, 2009). Our approach explains the preference-aggregation origin of this dual and provides a generalization to all homothetic preferences almost without any computations.

Optimization problem (56) indicates that to find a CE for a population of consumers with preferences from a certain domain $\mathcal{D}$ we must be able to find the market equilibrium for any preference that can be obtained as an aggregation of preferences from $\mathcal{D}$. In other words, the complexity of finding a CE is determined not by the domain $\mathcal{D}$ of individual preferences itself but rather by its completion $\mathcal{D}^{\text{complete}}$. We illustrate this point for the domain of linear preferences over two goods.

By Proposition 3, aggregation of linear preferences over $n = 2$ goods (domain $\mathcal{D}$) gives all preferences with substitutability ($\mathcal{D}^{\text{complete}}$). We will show that any algorithm computing an approximate CE for preferences from $\mathcal{D}$ can be used to compute an approximate CE for $\mathcal{D}^{\text{complete}}$. Hence, finding CE for $\mathcal{D}$ cannot be easy if it is hard for $\mathcal{D}^{\text{complete}}$. Let us call $\mathbf{p}$ an ε-equilibrium price vector if there are $\mathbf{x}_k \in D_k(\mathbf{p}, b_k)$, $k = 1, \ldots, m$ such that

$$\langle \mathbf{p}, \mathbf{e} \rangle \leqslant \varepsilon \cdot B, \qquad \text{where} \quad e_i = \left| x_i - \sum_{k=1}^{m} x_{k,i} \right|,$$

i.e., the excess demand is relatively small compared to the total budget.

Proposition 7. *Let* $\mathcal{D}$ *be the domain of linear preferences over two goods and assume we have an algorithm computing an* ε*-equilibrium price vector for any population of agents with preferences from* $\mathcal{D}$*. Then a* 3ε*-equilibrium price vector for a population of* m *agents with preferences from* $\mathcal{D}^{\text{complete}}$ *can be computed by applying the algorithm as a black box to an auxiliary population with preferences from* $\mathcal{D}$ *and the number of agents of the order of* m/ε*.*

The idea is to approximate preferences from $\mathcal{D}^{\text{complete}}$ by the aggregate preference of linear consumers so that the expenditure shares differ by at most ε. Such an approximation can be constructed via Corollary 10 and requires a number of auxiliary linear consumers of the order of

$1/\varepsilon$. As we show in Appendix C.4, if expenditure shares in two populations differ by at most ε, then ε-equilibrium price vector for one population is an $(1+n)\varepsilon$-equilibrium price vector for the other. Since $n = 2$, an ε-equilibrium price vector for the approximating population of linear consumers gives a 3ε-equilibrium price vector for the population of consumers with preferences from $\mathcal{D}^{\text{complete}}$.

The example of linear preferences demonstrates that even a simple parametric domain — if the choice of parameters is not aligned with aggregation — can have a large non-parametric completion. As a result, the simplicity of a parametric domain does not carry over to the aggregate behavior thus complicating the computation of a CE. To preserve the simplicity of a parametric domain, the choice of parameters is to be aligned with aggregation. Motivated by this concern, we consider computing a CE in parametric domains invariant with respect to aggregation.

Fix a finite family of "elementary" preferences $\succsim_1, \ldots, \succsim_q$ and consider the domain $\mathcal{D} = \mathcal{D}^{\text{complete}}$ of all preferences that can be obtained by aggregating the elementary preferences. We will call such invariant domains finitely-generated. Cobb-Douglas preferences are an example of a finitely generated domain; see Example 1. By Theorem 1, a finitely generated $\mathcal{D}$ consists of all preferences $\succsim$ whose expenditure function E can be represented as $\ln E = \sum_{l=1}^{q} t_l \ln E_l$ and, hence, the vector of coefficients $\mathbf{t} \in \Delta_{q-1}$ provides a parameterization of $\mathcal{D}$.

Proposition 8. *Consider a finitely-generated invariant domain $\mathcal{D} = \{\succsim_1, \ldots, \succsim_q\}^{\text{complete}}$ and fix $\varepsilon \geqslant 0$. Assume we have access to an algorithm finding an ε-equilibrium vector of prices for $m = 1$ agent and using at most T operations. Then an ε-equilibrium price vector for a population of $m \geqslant 1$ agents can be computed in time of the order of $m \cdot q + T$.*

The proof is straightforward. If preferences of individual agents are represented by $\mathbf{t}_1, \ldots, \mathbf{t}_m$ and $\beta_1, \ldots, \beta_m$ are relative incomes, then, by Theorem 1, the aggregate consumer corresponds to $\mathbf{t} = \sum_{k=1}^{m} \beta_k \cdot \mathbf{t}_k$. Computing $\mathbf{t}$ requires a number of operations of the order of $m \cdot q$. Applying the one-agent algorithm to the aggregate agent, we get an ε-equilibrium vector of prices for the original population in T operations.

The linear growth of running time with the number of agents m and the absence of large hidden constants suggests that finitely-generated domains can be a natural candidate for scalable fair division mechanisms. Note that the best running time for linear preferences achieved by Orlin (2010) and Végh (2012) grows as m^4.

In economic design, the choice of a preference domain corresponds to the choice of a bidding language, i.e., the information about the true — possibly substantially more complicated — preferences that the participants can report to a mechanism. Our results indicate the advantage of bidding languages corresponding to finitely-generated invariant domains.

Finitely-generated invariant domains offer enough flexibility to the designer. For example, apart from Cobb-Douglas preferences, one can consider domains generated by a finite collection of linear preferences. By adding preferences exhibiting substitutability or complementarity among certain subsets – for example, pairs — of goods to the collection of elementary preferences, we can allow agents to express both substitutability and complementarity patterns while keeping the domain narrow. The use of such domains and bidding languages in practice requires additional experimental evaluation as in (Budish and Kessler, 2022).

5 Indecomposable preferences

In this section, we study those preferences that cannot be represented as an aggregation of distinct preferences within a given domain. We call such preferences indecomposable. They play the role of elementary building blocks as any preference can be represented as an aggregation of indecomposable ones.

We already saw an example of such a representation in Section 4, where we represented any preference over two substitutes as a continuous aggregation of linear preferences. In contrast to the discussion of Sections 3 and 4 where we started by specifying "elementary" preferences and asked what can be obtained by aggregating them, now we start from a given domain and aim to identify these elementary preferences.

Definition 4. *For a given domain $\mathcal{D}$, a preference $\succsim$ from $\mathcal{D}$ is indecomposable if it cannot be represented as an aggregation of two distinct preferences $\succsim'$ and $\succsim''$ from $\mathcal{D}$. The set of all indecomposable preferences from $\mathcal{D}$ is denoted by $\mathcal{D}^{\mathrm{indec}}$.*

Recall that a point x from a subset X of a linear space is called an extreme point of X if it cannot be represented as $\alpha x' + (1-\alpha)x''$ with $\alpha \in (0,1)$ and distinct[28] $x', x'' \in X$. The set of all extreme points of X is denoted by X^{extrem}. Theorem 1 implies the following corollary.

Corollary 13. *A preference $\succsim$ is indecomposable in $\mathcal{D}$ if and only if the corresponding logarithmic expenditure function $\ln E$ is an extreme point of the set of logarithmic expenditure functions*

$$\mathcal{L}_{\mathcal{D}} = \left\{ f = \ln\left(E_{\succsim'}\right) : \ \succsim' \in \mathcal{D} \right\}.$$

The Choquet theorem states that, if X is a compact convex subset of a locally convex topological vector space, then any point $x \in X$ can be obtained as the average of its extreme points $x' \in X^{\mathrm{extrem}}$ with respect to some Borel probability measure $\mu = \mu_x$ supported on X^{extrem}:

$$x = \int_{X^{\mathrm{extrem}}} x' \, \mathrm{d}\mu(x'); \tag{58}$$

see (Phelps, 2001). Using the Choquet theorem, we obtain the following result demonstrating that indecomposable preferences can indeed be seen as elementary building blocks.

[28] Usually, one assumes that X is convex but we do not make this assumption.

Theorem 3. *If $\mathcal{D}$ is a closed domain invariant with respect to aggregation, then any preference $\succsim$ from $\mathcal{D}$ can be obtained as a continuous aggregation of indecomposable preferences from $\mathcal{D}$, i.e., there exists a Borel measure μ supported on $\mathcal{D}^{\mathrm{indec}}$ such that the expenditure function $E = E_{\succsim}$ can be represented as follows*

$$\ln E(\mathbf{p}) = \int_{\mathcal{D}^{\mathrm{indec}}} \ln E_{\succsim'}(\mathbf{p})\, \mathrm{d}\mu(\succsim') \tag{59}$$

for any vector of prices $\mathbf{p} \in \mathbb{R}^n_{++}$.

As in Theorem 2, the integral (59) is formally defined in Appendix B. Both theorems are proved in Appendix C.3. The essence of the proof is checking that the topological assumptions of the Choquet theorem are satisfied.

Representation (59) is especially useful if the set of indecomposable preferences is small relative to the whole domain $\mathcal{D}$. We will see that this is the case for substitutes but not the case for complements and the full domain.

5.1 Indecomposability in the full domain

Let $\mathcal{D}$ be the domain of all homothetic preferences. It is easy to guess some indecomposable preferences from $\mathcal{D}$: for example, linear and Leontief preferences are indecomposable. It turns out that there are many more. Let us call $\succsim$ a Leontief preference over linear composite goods if it corresponds to a utility function of the form

$$u(\mathbf{x}) = \min_{\mathbf{a} \in A} \{\chi_{\mathbf{a}}(\mathbf{x})\}, \tag{60}$$

where A is a finite or countably infinite subset of $\mathbb{R}^n_+$ and each $\mathbf{a} \in A$ defines a linear composite good $\chi_{\mathbf{a}}(\mathbf{x})$ by

$$\chi_{\mathbf{a}}(\mathbf{x}) = \sum_{i=1}^{n} a_i x_i.$$

The interpretation is that an agent treats the collection of bundles $\mathbf{a} \in A$ as perfect complements. Leontief and linear preferences are particular cases. For Leontief preferences, the bundles are, in fact, single goods and so each $\mathbf{a} \in A$ has only one non-zero coordinate. Linear preferences correspond to a single bundle $\mathbf{a}$, i.e., $A = \{\mathbf{a}\}$. Geometrically, Leontief preferences over linear composite goods are exactly those preferences that have upper contour sets with piecewise linear boundary.

Proposition 9. *For any number of goods n, Leontief preferences over linear composite goods* (60) *are indecomposable in the domain of all homothetic preferences.*

The proposition implies that linear and Leontief preferences are indeed indecomposable. Another immediate corollary is that aggregation of linear and Leontief preferences together is far from giving the full domain. Any preference of the form (60) is indecomposable and, hence, cannot be represented as an aggregation of linear or Leontief preferences unless it is linear or Leontief itself. For example, one can take a preference $\succsim$ corresponding to

$$u(\mathbf{x}) = \min\{x_1 + 2 \cdot x_2,\ 2 \cdot x_1 + x_2\}.$$

The corollary can be strengthened. Expenditure shares for $\succsim$ are not monotone, i.e., $\succsim$ exhibits neither substitutability nor complementarity.[29] Since $\succsim$ is indecomposable, we conclude that not every preference can be represented as an aggregation of preferences exhibiting substitutability and preferences exhibiting complementarity.

We see that the full domain has a lot of indecomposable preferences. To formalize this observation, note that piecewise linear concave functions are dense in the set of all concave functions. Accordingly, indecomposable preferences are dense in the full domain $\mathcal{D}$ and extreme points of the set $\mathcal{L}_{\mathcal{D}}$ of logarithmic expenditure functions are dense in this set.[30]

The main insight behind Proposition 9 is as follows. We know that describing indecomposable preferences in a domain $\mathcal{D}$ boils down to finding extreme points of the set of logarithmic expenditure functions $\mathcal{L}_{\mathcal{D}}$. Finite-dimensional linear programming intuition suggests that natural candidates for extreme points of a convex set are those points where the maximal number of constraints defining the set are active. Leontief preferences over linear composite goods are those preferences $\succsim$ for which the concavity constraint on the expenditure function E is active almost everywhere.

The formal proof of Proposition 9 is contained in Appendix C.5; we sketch the argument here. A utility function u is of the form (60) if and only if the corresponding expenditure function is also piecewise linear: $E = \min_{c \in C} \sum_{i=1}^{n} c_i \cdot p_i$ for finite or countable $C \subset \mathbb{R}_+^n$. To demonstrate indecomposability, we need to show that if $E = (E_1)^\alpha (E_2)^{1-\alpha}$, then E_1 and E_1 are proportional to each other (and thus to E). By strict concavity of the function $h(\mathbf{t}) = t_1^\alpha \cdot t_2^{1-\alpha}$ on rays not passing through the origin, E cannot be linear in regions where E_1 and E_1 are not proportional. Hence, E_1 and E_2 must be proportional in each of the linearity regions of E. As these regions cover the whole space, E_1 and E_2 are proportional everywhere implying that $\succsim$ is indecomposable.

In Appendix C.5 we also explore how close Proposition 9 is to characterizing all indecomposable preferences. We show that if there is a neighborhood of a point where the concavity constraint on E is inactive, then a preference can be decomposed (Proposition 13). The idea is that we can find small perturbation $\psi = \psi(\mathbf{p})$ vanishing outside of this neighborhood and such that $E_1 = E \cdot (1 + \psi)$

29

30 The existence of non-trivial convex sets with dense extreme points highlights that finite-dimensional intuition can be misleading in infinite-dimensional convex geometry (Poulsen, 1959). In economic literature, such sets have appeared in the context of the n-good monopolist problem with $n \geqslant 2$, where mechanisms can be identified with convex functions on $[0, 1]^n$ such that their gradients also belong to $[0, 1]^n$ (Manelli and Vincent, 2007).

and $E_2 = E/(1+\psi)$ are valid logarithmic expenditure functions. Since $\ln E = 1/2 \cdot \ln E_1 + 1/2 \cdot \ln E_2$, the preference corresponding to E can indeed be decomposed.

Intuitively, a concave function is either piecewise linear or there is a neighborhood where it is strictly concave and, hence, Propositions 9 and 13 seem to cover all possible cases. However, there is a family of pathological examples not captured by this intuition, e.g., concave functions whose second derivative is a continuous measure supported on a Cantor set. Proposition 14 formulated and proved in the appendix shows that such pathological preferences are also indecomposable.

Proposition 9 has implications for the geometric mean of convex sets. Consider the collection $\mathcal{X}$ of all closed convex subsets X of $\mathbb{R}_+^n$ that do not contain zero and are upward-closed, i.e., all those that can be obtained as upper contour sets of homothetic preferences. We call a set $X \in \mathcal{X}$ indecomposable if it cannot be represented as the geometric mean $X_1^\lambda \otimes X_2^{1-\lambda}$ with distinct X_1 and X_2 from $\mathcal{X}$ and $\lambda \in (0,1)$. Proposition 9 implies that convex polytopes (with a possibly infinite number of faces) are indecomposable. There is mathematical literature inspired by Gale (1954) and studying a similar concept of indecomposability where instead of taking weighted geometric means, one takes convex combinations with respect to the Minkowski addition.[31] In contrast to our setting, planar sets that are indecomposable in the sense of Gale form a simple parametric family (Gale, 1954; Silverman, 1973). However, in the dimension 3 and higher, Gale's indecomposability behaves similarly to ours: indecomposable sets are dense in all convex sets and one can derive some necessary and some sufficient conditions of indecomposability that almost match each other but yet no criterion is known; see, e.g., (Sallee, 1972).

5.2 The domain of substitutes and the simplex property

Consider the domain $\mathcal{D}_S$ of all homothetic preferences over n substitutes. Linear preferences belong to $\mathcal{D}_S$ and are indecomposable since they are indecomposable even in the larger domain of all homothetic preferences by Proposition 9. For two goods, there are no other indecomposable preferences in $\mathcal{D}_S$.

Proposition 10. *For $n = 2$ goods, a preference $\succsim$ is indecomposable in the domain $\mathcal{D}_S$ of homothetic preferences with substitutability if and only if $\succsim$ is linear.*

From Corollary 10, we know that any preference over $n = 2$ goods exhibiting substitutability can be obtained by aggregating linear preferences. Hence, any non-linear preference can be decomposed and we get Proposition 10.

Corollary 10 provides an explicit Choquet decomposition (59) for $\mathcal{D}_S$. Moreover, the corollary states that the decomposition is unique in the sense that the distribution of linear preferences over

[31] Gale (1954) calls such sets irreducible.

the population is pinned down uniquely. This phenomenon has the following geometric interpretation.

Consider a collection of d points in a finite-dimensional linear space such that no subset of $k \leqslant d$ points belongs to a $(k-2)$-dimensional linear subspace. The convex hull of such a collection is called a simplex. A simplex has the property that any point has the unique decomposition as a convex combination of extreme points. The uniqueness of the decomposition characterizes simplices among all other closed convex subsets of a finite-dimensional space. In the infinite-dimensional space, this property can be used to define a simplex, namely, a compact convex set is called a simplex if each point can be uniquely represented as the average of the extreme points, i.e., the measure in the Choquet integral (58) is uniquely defined (Phelps, 2001). Accordingly, we say that a closed domain of preferences is a simplex domain if there is a unique way to represent any preference as an aggregation of indecomposable ones, i.e., the measure μ in (59) is unique.

Corollary 14. *For two goods, the domain of homothetic preferences exhibiting substitutability is a simplex domain.*

For $n \geqslant 3$ goods, there are other indecomposable preferences in $\mathcal{D}_S$ except for linear ones. Indeed, by Corollary 9, aggregation of linear preferences does not give the whole domain $\mathcal{D}_S$. Since any preference can be represented as an aggregation of indecomposable ones, we conclude that there must be other indecomposable preferences. Describing them explicitly and checking whether $\mathcal{D}_S$ is a simplex domain for $n \geqslant 3$ remains an open question.

5.3 The domain of complements

Let us discuss the domain $\mathcal{D}_C$ of homothetic preferences exhibiting complementarity. Leontief preferences are indecomposable in $\mathcal{D}_C$ since they are indecomposable in the full domain. By Corollary 11, aggregation of Leontief preferences does not give the whole $\mathcal{D}_C$ even for $n = 2$ goods and, hence, there must be other indecomposable preferences. It turns out that indecomposable preferences are dense in $\mathcal{D}_C$ and their structure resembles the one for the full domain. We call a preference $\succsim$ a Leontief preference over Cobb-Douglas composite goods if it corresponds to a utility function

$$u(\mathbf{x}) = \min_{\mathbf{a}\in A} \{\chi_{\mathbf{a}}(\mathbf{x})\}, \tag{61}$$

where A is finite or countably infinite subset of $\mathbb{R}_{++} \times \Delta_{n-1}$ and each $\mathbf{a} = (a_0, a_1, \ldots, a_n) \in A$ defines a Cobb-Douglas composite good $\chi_{\mathbf{a}}(\mathbf{x})$ by

$$\chi_{\mathbf{a}}(\mathbf{x}) \equiv a_0 \cdot \prod_{i=1}^{n} x_i^{a_i}.$$

Cobb-Douglas and Leontief preferences are particular cases of (61) corresponding, respectively, to a singleton $A = \{\mathbf{a}\}$ and to $A = \{(a_0^1, \mathbf{e}_1), \ldots, (a_0^n, \mathbf{e}_n)\}$ where $\mathbf{e}_i$ is the i'th basis vector. We call a Leontief preference $\succsim$ over Cobb-Douglas composite goods non-trivial if the set A contains at least two vectors $\mathbf{a}$ and $\mathbf{a}'$ with $(a_1, \ldots, a_n) \neq (a_1', \ldots, a_n')$. Equivalently, $\succsim$ is non-trivial if it is not a standard Cobb-Douglas preference.

Proposition 11. *For $n = 2$ goods, non-trivial Leontief preferences over Cobb-Douglas composite goods are indecomposable in the domain of homothetic preferences with complementarity.*

The requirement of non-triviality is needed as standard Cobb-Douglas preferences can be decomposed as an aggregation of $\succsim_i$ corresponding to $u_i(\mathbf{x}) = x_i$; see Example 1. Note that $\succsim_i$ — which can be seen as either extreme linear or extreme Cobb-Douglas preference — is indecomposable in $\mathcal{D}_C$ since it is indecomposable even in the full domain by Proposition 9.

Proposition 11 is proved in Appendix C.6. The idea is similar to Proposition 9 dealing with indecomposability in the full domain: indecomposable preferences correspond to expenditure functions E with the maximal number of active constraints. In contrast to Proposition 9 where the concavity of E was the only constraint that matters, now we have the new monotonicity constraint on the expenditure share. Leontief preferences over Cobb-Douglas composite goods are obtained if the space is partitioned into regions where the monotonicity constraint is active (expenditure shares are constant) or the concavity constraint is active (the expenditure function is linear). The former regions correspond to hyperbolic parts of the upper contour sets and the latter regions, to cusps.

5.4 Application: preference identification

Market demand reflects individual preferences but information loss is unavoidable. For example, aggregate behavior does not allow to distinguish populations where a pair of agents swapped their preferences and incomes or where a pair of agents with identical preferences is replaced with one agent with the joint income. We can still ask whether market demand determines the distribution of preferences over the population, i.e., whether, by looking at the aggregate behavior, it is possible to determine what fraction of the population's income corresponds to agents with preferences of a particular kind.

Consider a population of consumers with homothetic preferences from some domain $\mathcal{D}$. An analyst knows neither the income distribution nor the size of the population and observes market demand generated by this population for any vector of prices. For any subset of preferences $\mathcal{D}' \subset \mathcal{D}$, the analyst aims to identify what fraction of the total income corresponds to agents in $\mathcal{D}'$.

In general, identification is impossible. For example, if $\mathcal{D}$ is the domain of Cobb-Douglas preferences, the aggregate demand corresponding to $u_{\text{aggregate}}(x_1, x_2) = x_1^{1/3} \cdot x_2^{2/3}$ can be generated by a population where each agent has the same preference $\succsim = \succsim_{\text{aggregate}}$ or, alternatively, by the population where $1/3$ of the total income is earned by agents with preference $u_1(\mathbf{x}) = x_1$ and $2/3$ by those with $u_2(\mathbf{x}) = x_2$; see Example 1.

The domain $\mathcal{D}$ of linear preferences over $n = 2$ goods is an exception. A linear preference over two goods is determined by its marginal rate of substitution $\mathrm{MRS} = v_1/v_2$. By Corollary 10, the fraction of income corresponding to consumers with MRS above a certain threshold α is equal to the fraction of income spent by the population on the first good for prices $p_1 = \alpha \cdot p_2$, i.e.,

$$\mu\left(\mathrm{MRS} \geqslant \frac{p_1}{p_2}\right) = s_{\text{aggregate},1}(\mathbf{p}) = \frac{p_1 \cdot D_{\text{aggregate},1}(\mathbf{p}, B)}{p_1 \cdot D_{\text{aggregate},1}(\mathbf{p}, B) + p_2 \cdot D_{\text{aggregate},2}(\mathbf{p}, B)}.$$

Hence, even a few observations of market demand $D_{\text{aggregate}}$ at non-collinear price vectors can give a good understanding of the preference distribution over the population.

More generally, the distribution of preferences from a domain $\mathcal{D}$ can be identified if any preference $\succsim$ obtained by aggregation of preferences from $\mathcal{D}$ cannot be decomposed over $\mathcal{D}$ in a different way. A geometric interpretation of this property relies on the notion of simplex domains from Section 5.2 and is contained in the following corollary.

Corollary 15. *If the completion $\mathcal{D}^{\text{complete}}$ of $\mathcal{D}$ is a simplex domain and $\mathcal{D}$ consists of indecomposable preferences, then the distribution of preferences over the population can be identified from price dependence of market demand.*

Apart from linear preferences over two goods, there are many other domains satisfying the requirements of Corollary 15. One can take $\mathcal{D}$ given by any finite collection of preferences $\{\succsim_1, \ldots, \succsim_q\}$ none of which can be obtained as an aggregation of the others. For example, if q equals the number of goods n and each $\succsim_k$ corresponds to $u(\mathbf{x}) = x_k$, then the income fraction of consumers with preference $\succsim_k$ is equal to the expenditure share $s_{\text{aggregate},k}(\mathbf{p})$ at any price $\mathbf{p}$; see also Example 1. We stress that just one observation of aggregate behavior at any particular vector of prices $\mathbf{p}$ turns out to be enough to determine the distribution of preferences. The origin of this phenomenon is not the orthogonality of preferences but the fact that the dimension of the domain of preferences does not exceed the dimension of the consumption space. To illustrate this point, note that if $\succsim_1, \ldots, \succsim_q$ are Cobb-Douglas preference with vectors of parameters $\mathbf{a}_1, \ldots, \mathbf{a}_q$ that are linearly independent (possible only if $q \leqslant n$), then one observation of market demand also gives a linear system enough for identification.

Another domain $\mathcal{D}$ satisfying conditions of Corollary 15 is the domain of Leontief preferences over two goods. By Corollary 12, any preference from its completion $\mathcal{D}^{\text{complete}}$ can be uniquely decomposed over Leontief preferences. Hence, $\mathcal{D}^{\text{complete}}$ is a simplex domain, and Leontief preferences are indecomposable in it. We conclude that, in theory, the distribution of Leontief preferences can be identified. A peculiarity is that the Stieltjes-Perron inversion formula underlying Corollary 12 requires continuation of demand to complex prices. Therefore, it guarantees identification but gives no practical recipe for reconstructing the distribution of preferences for an analyst who observes demand for real prices only. Instead, the analyst can use real-inversion techniques for the Stieltjes transform, e.g., (Widder, 1938; Love and Byrne, 1980).

References

C. D. Aliprantis and K. C. Border. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer-Verlag Berlin and Heidelberg GmbH & Company KG, 2013.

S. P. Anderson, A. De Palma, J.-F. Thisse, et al. The ces is a discrete choice model? *Economics Letters*, 24(2):139–140, 1987.

S. P. Anderson, A. De Palma, and J.-F. Thisse. *Discrete choice theory of product differentiation*. MIT press, 1992.

I. Arieli, Y. Babichenko, R. Smorodinsky, and T. Yamashita. Optimal persuasion via bi-pooling. *Available at SSRN 3511516*, 2019.

C. Arkolakis, A. Costinot, and A. Rodríguez-Clare. New trade models, same old gains? *American Economic Review*, 102(1):94–130, 2012.

C. Arkolakis, A. Costinot, D. Donaldson, and A. Rodríguez-Clare. The elusive pro-competitive effects of trade. *The Review of Economic Studies*, 86(1):46–80, 2019.

I. Ashlagi and P. Shi. Optimal allocation without money: An engineering approach. *Management Science*, 62(4):1078–1097, 2016.

R. J. Aumann, M. Maschler, and R. E. Stearns. *Repeated games with incomplete information*. MIT press, 1995.

F. Benford. The law of anomalous numbers. *Proceedings of the American philosophical society*, 78 (4):551–572, 1938.

A. Bogomolnaia, H. Moulin, F. Sandomirskiy, and E. Yanovskaya. Competitive division of a mixed manna. *Econometrica*, 85(6):1847–1871, 2017.

K. J. Böröczky, E. Lutwak, D. Yang, and G. Zhang. The log-Brunn–Minkowski inequality. *Advances in Mathematics*, 231(3-4):1974–1997, 2012.

W. C. Brainard and H. E. Scarf. How to compute equilibrium prices in 1891. *American Journal of Economics and Sociology*, 64(1):57–83, 2005.

S. Brânzei and F. Sandomirskiy. Algorithms for competitive division of chores. *arXiv preprint arXiv:1907.01766*, 2019.

D. S. Bridges and G. B. Mehta. *Representations of preferences orderings*, volume 422. Springer Science & Business Media, 2013.

M. Browning and P.-A. Chiappori. Efficient intra-household allocations: A general characterization and empirical tests. *Econometrica*, pages 1241–1278, 1998.

M. Browning, P.-A. Chiappori, and A. Lewbel. Estimating consumption economies of scale, adult equivalence scales, and household bargaining power. *Review of Economic Studies*, 80(4):1267–1303, 2013.

E. Budish and J. B. Kessler. Can market participants report their preferences accurately (enough)? *Management Science*, 68(2):1107–1130, 2022.

E. Budish, G. P. Cachon, J. B. Kessler, and A. Othman. Course match: A large-scale implementation of approximate competitive equilibrium from equal incomes for combinatorial allocation. *Operations Research*, 65(2):314–336, 2017.

J. Caballé and A. Pomansky. Mixed risk aversion. *Journal of economic theory*, 71(2):485–513, 1996.

C. D. Carroll. Requiem for the representative consumer? Aggregate implications of microeconomic consumption behavior. *American Economic Review*, 90(2):110–115, 2000.

C. P. Chambers and T. Hayashi. The structure of household preference. 2018.

C. Chamley. Optimal taxation of capital income in general equilibrium with infinite lives. *Econometrica: Journal of the Econometric Society*, pages 607–622, 1986.

P.-A. Chiappori. Rational household labor supply. *Econometrica: Journal of the Econometric Society*, pages 63–90, 1988.

P.-A. Chiappori and I. Ekeland. Aggregation and market demand: an exterior differential calculus viewpoint. *Econometrica*, 67(6):1435–1457, 1999.

P.-A. Chiappori and I. Ekeland. The microeconomics of efficient group behavior: Identification 1. *Econometrica*, 77(3):763–799, 2009.

J. S. Chipman and J. C. Moore. On social welfare functions and the aggregation of preferences. *Journal of economic theory*, 21(1):111–139, 1979.

R. Cole, N. Devanur, V. Gkatzelis, K. Jain, T. Mai, V. V. Vazirani, and S. Yazdanbod. Convex program duality, Fisher markets, and Nash social welfare. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pages 459–460, 2017.

V. Conitzer, C. Kroer, D. Panigrahi, O. Schrijvers, N. E. Stier-Moses, E. Sodomka, and C. A. Wilkens. Pacing equilibrium in first price auction markets. *Management Science*, 2022.

A. Costinot and A. Rodríguez-Clare. Trade theory with numbers: Quantifying the consequences of globalization. In *Handbook of international economics*, volume 4, pages 197–261. Elsevier, 2014.

G. Debreu. Excess demand functions. *Journal of mathematical economics*, 1(1):15–21, 1974.

N. R. Devanur, C. H. Papadimitriou, A. Saberi, and V. V. Vazirani. Market equilibrium via a primal-dual-type algorithm. In *The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002. Proceedings.*, pages 389–395. IEEE, 2002.

N. R. Devanur, K. Jain, T. Mai, V. V. Vazirani, and S. Yazdanbod. New convex programs for Fisher's market model and its generalizations. *arXiv preprint arXiv:1603.01257*, 2016.

N. R. Devanur, J. Garg, R. Mehta, V. V. Vaziranb, and S. Yazdanbod. A new class of combinatorial markets with covering constraints: Algorithms and applications. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2311–2325. SIAM, 2018.

W. E. Diewert. Applications of duality theory. In *Frontiers of Quantitative Economics. Volume II, ed. M.D. Intrilligator and D.A. Kendrick.*, pages 106–199. North Holland Publishing Company, 1974.

W. E. Diewert. Duality approaches to microeconomic theory. *Handbook of mathematical economics*, 2:535–599, 1982.

P. Dworczak and G. Martini. The simple economics of optimal persuasion. *Journal of Political Economy*, 127(5):1993–2048, 2019.

F. Echenique, A. Miralles, and J. Zhang. Constrained pseudo-market equilibrium. *American Economic Review*, 111(11):3699–3732, 2021.

E. Eisenberg. Aggregation of utility functions. *Management Science*, 7(4):337–350, 1961.

E. Eisenberg and D. Gale. Consensus of subjective probabilities: The pari-mutuel method. *The Annals of Mathematical Statistics*, 30(1):165–168, 1959.

D. Gale. Irreducible convex sets. *Proc. Intern. Congr. Math., Amsterdam*, 2:217–218, 1954.

Y. Gao and C. Kroer. Infinite-dimensional Fisher markets and tractable fair division. *Operations Research*, 2022.

J. Goldman and A. D. Procaccia. Spliddit: Unleashing fair division algorithms. *ACM SIGecom Exchanges*, 13(2):41–46, 2015.

W. M. Gorman. Community preference fields. *Econometrica: journal of the Econometric Society*, pages 63–80, 1953.

W. M. Gorman. On a class of preference fields. *Metroeconomica*, 13(2):53–56, 1961.

Y. He, A. Miralles, M. Pycia, and J. Yan. A pseudo-market approach to allocation with priorities. *American Economic Journal: Microeconomics*, 10(3):272–314, 2018.

W. Hildenbrand. *Market demand*. Princeton University Press, 2014.

W. Hildenbrand. Core and equilibria of a large economy.(psme-5). In *Core and Equilibria of a Large Economy.(PSME-5)*. Princeton university press, 2015.

A. Hylland and R. Zeckhauser. The efficient allocation of individuals to positions. *Journal of Political economy*, 87(2):293–314, 1979.

M. O. Jackson and L. Yariv. The non-existence of representative agents. *Available at SSRN 2684776*, 2019.

M. Jerison. Aggregation and pairwise aggregation of demand when the distribution of income is fixed. *Journal of economic Theory*, 33(1):1–31, 1984.

M. Jerison. Optimal income distribution rules and representative consumers. *The Review of Economic Studies*, 61(4):739–771, 1994.

E. Jouini and C. Napp. Consensus consumer and intertemporal asset pricing with heterogeneous beliefs. *The Review of Economic Studies*, 74(4):1149–1174, 2007.

E. Kamenica and M. Gentzkow. Bayesian persuasion. *American Economic Review*, 101(6):2590–2615, 2011.

Z. Y. Kang and S. Vasserman. Robust bounds for welfare analysis. Technical report, National Bureau of Economic Research, 2022.

A. P. Kirman. Whom or what does the representative individual represent? *Journal of economic perspectives*, 6(2):117–136, 1992.

A. Kleiner, B. Moldovanu, and P. Strack. Extreme points and majorization: Economic applications. *Econometrica*, 89(4):1557–1593, 2021.

P. Kocourek, J. Steiner, and C. Stewart. Demand in the dark. 2022.

D. Kornbluth and A. Kushnir. Undergraduate course allocation through pseudo-markets. *Available at SSRN 3901146*, 2021.

D. M. Kreps. *A course in microeconomic theory*. Princeton university press, 2020.

A. Lewbel and K. Pendakur. Tricks with hicks: The EASI demand system. *American Economic Review*, 99(3):827–63, 2009.

E. Love and A. Byrne. Real inversion theorems for generalized Stieltjes transforms. *Journal of the London Mathematical Society*, 2(2):285–306, 1980.

A. M. Manelli and D. R. Vincent. Multidimensional mechanism design: Revenue maximization and the multiple-good monopoly. *Journal of Economic theory*, 137(1):153–185, 2007.

R. R. Mantel. On the characterization of aggregate excess demand. *Journal of economic theory*, 7 (3):348–353, 1974.

R. R. Mantel. Homothetic preferences and community excess demand functions. *Journal of Economic Theory*, 12(2):197–201, 1976.

A. Mas-Colell, M. D. Whinston, J. R. Green, et al. *Microeconomic theory*, volume 1. Oxford university press New York, 1995.

K. Matsuyama and P. Ushchev. Beyond ces: three alternative classes of flexible homothetic demand systems. *Global Poverty Research Lab Working Paper*, (17-109), 2017.

R. L. Matzkin. Nonparametric identification. *Handbook of econometrics*, 6:5307–5368, 2007.

R. L. Matzkin. Nonparametric identification in structural economic models. *Annu. Rev. Econ.*, 5 (1):457–486, 2013.

D. McFadden. Econometric models of probabilistic choice. *Structural analysis of discrete data with econometric applications*, 198272, 1981.

V. Milman and L. Rotem. Non-standard constructions in convex geometry: geometric means of convex bodies. In *Convexity and concentration*, pages 361–390. Springer, 2017.

H. Moulin. Fair division in the internet age. *Annual Review of Economics*, 11:407–441, 2019.

J. F. Nash. The bargaining problem. *Econometrica: Journal of the econometric society*, pages 155–162, 1950.

N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani. *Algorithmic Game Theory*. Cambridge University Press, 2007.

J. B. Orlin. Improved algorithms for computing Fisher's market clearing prices: Computing Fisher's market clearing prices. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 291–300, 2010.

R. R. Phelps. *Lectures on Choquet's theorem*. Springer, 2001.

E. T. Poulsen. Convex sets with dense extreme points. *The American Mathematical Monthly*, 66 (7):577–578, 1959.

M. Pycia. Pseudomarkets. In *Online and Matching-Based Market Design Eds. Federico Echenique, Nicole Immorlica, and Vijay V. Vazirani*. Cambridge University Press, 2022.

R. T. Rockafellar. *Convex analysis*, volume 18. Princeton university press, 1970.

R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

K. Rogoff. Equilibrium political budget cycles. *American Economic Review*, (80):21–36, 1990.

G. Sallee. Minkowski decomposition of convex sets. *Israel Journal of Mathematics*, 12(3):266–276, 1972.

P. A. Samuelson. Social indifference curves. *The Quarterly Journal of Economics*, 70(1):1–22, 1956.

P. A. Samuelson. Unification theorem for the two basic dualities of homothetic demand theory. *Proceedings of the national academy of sciences*, 69(9):2673–2674, 1972.

R. L. Schilling, R. Song, and Z. Vondracek. Bernstein functions. In *Bernstein functions*. de Gruyter, 2012.

W. Shafer and H. Sonnenschein. Market demand and excess demand functions. *Handbook of mathematical economics*, 2:671–693, 1982.

V. I. Shmyrev. An algorithm for finding equilibrium in the linear exchange model with fixed budgets. *Journal of Applied and Industrial Mathematics*, 3(4):505–518, 2009.

R. Silverman. Decomposition of plane convex sets. I. *Pacific Journal of Mathematics*, 47(2):521–530, 1973.

H. Sonnenschein. Do Walras' identity and continuity characterize the class of community excess demand functions? *Journal of economic theory*, 6(4):345–354, 1973.

J. R.-V. Sørensen and M. Fosgerau. How McFadden met Rockafellar and learned to do more with less. *Journal of Mathematical Economics*, 100:102629, 2022.

E. Soumalias, B. Zamanlooy, J. Weissteiner, and S. Seuken. Machine learning-powered course allocation. *arXiv preprint arXiv:2210.00954*, 2022.

H. R. Varian. Equity, envy, and efficiency. *Journal of Economic Theory*, 9(1):63–91, 1974.

H. R. Varian. Social indifference curves and aggregate demand. *The Quarterly Journal of Economics*, 99(3):403–414, 1984.

L. A. Végh. Strongly polynomial algorithm for a class of minimum-cost flow problems with separable convex objectives. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 27–40, 2012.

D. Widder. The Stieltjes transform. *Transactions of the American Mathematical Society*, 43(1): 7–60, 1938.

R. D. Willig. Consumer's surplus without apology. *The American Economic Review*, 66(4):589–597, 1976.

K. H. Yang and A. K. Zentefis. Distributions of posterior quantiles and economic applications. 2022.

A Convex analysis basics

Superdifferentials. Let X be a convex subset of $\mathbb{R}^n$ and $f\colon X \to \mathbb{R}$ a concave function. Any such function f is continuous in the relative interior of X but may have discontinuities on the boundary (Aliprantis and Border, 2013, 7.24 Theorem). The superdifferential of f is defined by

$$\partial f(\mathbf{x}) = \{\mathbf{p} \in \mathbb{R}^n \ : \ f(\mathbf{y}) \leqslant f(\mathbf{x}) + \langle \mathbf{p}, \mathbf{y} - \mathbf{x} \rangle, \quad \forall \mathbf{y} \in X\}. \tag{62}$$

The superdifferential is non-empty in the relative interior of X. Recall that the gradient $\nabla f(x)$ is the vector of partial derivatives

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right).$$

The gradient is defined for any $\mathbf{x}$ where partial derivatives exist. At any such point, the superdifferential $\partial f(\mathbf{x})$ consists of just one element: $\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\}$ (Rockafellar, 1970, Theorem 25.1). By the Alexandrov theorem, a concave function is twice differentiable except for a set of zero Lebesgue; see (Rockafellar, 1970, Theorem 25.5). In particular, the gradient $\nabla f(\mathbf{x})$ is defined almost everywhere and coincides with the superdifferential.

Application to demand and expenditure functions. Consider a consumer with budget $b > 0$ and homothetic preferences $\succsim$ on $\mathbb{R}_+^n$ represented by a homogeneous utility function u. The demand $D(\mathbf{p}, b)$ considered as a function of prices and the budget is referred to as the Marshallian demand. The demand as a function of prices and the utility level $w > 0$ is referred to as the Hicksian demand:

$$H(\mathbf{p}, w) = \underset{\mathbf{x}\colon u(\mathbf{x}) \geqslant w}{\operatorname{argmin}} \langle \mathbf{p}, \mathbf{x} \rangle.$$

Sheppard's lemma for homothetic preferences gives the following identity:

$$H(\mathbf{p}, w) = w \cdot \partial E(\mathbf{p}),$$

where $\partial E(\mathbf{p})$ is the superdifferential of the expenditure function. This identity holds for all $\mathbf{p} \in \mathbb{R}_{++}^n$ including those points where E is not differentiable (Mas-Colell et al., 1995, p.141).

By the utility-maximization/cost-minimization duality (Diewert, 1982), $D(\mathbf{p}, b) = H(\mathbf{p}, w)$ if w is such that $\langle \mathbf{p}, \mathbf{x} \rangle = b$ for $\mathbf{x} \in H(\mathbf{p}, w)$. Hence, $D(\mathbf{p}, b) = w \cdot \partial E(\mathbf{p})$ for some $w = w(b)$. Choosing $w(b)$ so that each bundle from the right-hand side has the price of b, we get

$$D(\mathbf{p}, b) = b \cdot \frac{\partial E(\mathbf{p})}{E(\mathbf{p})} = b \cdot \partial \ln E(\mathbf{p}),$$

where we used the Euler identity for 1-homogeneous functions: $\langle \partial E(\mathbf{p}), \mathbf{p} \rangle = E(\mathbf{p})$.

By the Alexandrov theorem, there exists a set $A \subset \mathbb{R}_{++}^n$ such that $\mathbb{R}_{++}^n \backslash A$ has zero Lebesgue measure and E is differentiable for any $\mathbf{p} \in A$. By homogeneity of E, the set A can be selected so that $\mathbf{p} \in A \Rightarrow \alpha \cdot \mathbf{p} \in A$ for any $\alpha > 0$. For $\mathbf{p} \in A$, the superdifferential ∂E consists of just one element, the gradient $\partial E(\mathbf{p}) = \{\nabla E(\mathbf{p})\}$. Consequently, the Marshallian demand $D(\mathbf{p}, b)$ contains just one bundle on the set $\mathbf{p} \in A$ of full measure and can be thought of as the single-valued function

$$D(\mathbf{p}, b) = b \cdot \nabla \ln E(\mathbf{p})$$

defined almost everywhere.

For a bundle $\mathbf{x} \in D(\mathbf{p}, b)$, the expenditure share of a good i is defined by $x_i \cdot p_i / b$. For $\mathbf{p} \in A$, the demand is single-valued and so the expenditure share is a single-valued function of prices defined almost everywhere and satisfying the identity

$$\mathbf{s}(\mathbf{p}, b) = \left(p_1 \cdot \frac{\partial \ln E}{\partial p_1}, \ldots, p_n \cdot \frac{\partial \ln E}{\partial p_n} \right) = \left(\frac{\partial \ln E}{\partial \ln p_1}, \ldots, \frac{\partial \ln E}{\partial \ln p_n} \right).$$

B Topology on preferences and integration

There are two high-level reasons why we need a topology on preferences. The topology is necessary to define the closure of preference domains that is used in our discussion of completion but, most importantly, the topology is needed to formalize integration over preferences and to apply Choquet theory (Phelps, 2001). Recall that the Choquet theory deals with compact convex subsets of locally convex topological vector spaces. Our goal is to identify the domain of all homothetic preferences with a compact convex subset of a Banach space (complete normed and, hence, locally convex vector space).

We represent a homothetic preference $\succsim$ by its logarithmic expenditure function $\ln E$. We call two functions f and g equivalent if $f - g = \text{const}$. Since the expenditure function is defined up to a multiplicative factor, each preference corresponds to the class of equivalent logarithmic expenditure functions.

Let $\mathcal{L}$ be the set of classes of equivalent continuous functions f on $\mathbb{R}^n_{++}$ that can be obtained as logarithmic expenditure functions of homothetic preferences. The set $\mathcal{L}$ is in one-to-one correspondence with the domain of homothetic preferences. Hence, to define a topology and integration for preferences, it is enough to define them for $\mathcal{L}$. We first introduce a metric structure. To motivate the definition of a distance, we need some estimates on the magnitude of expenditure functions.

Lemma 1. *For any expenditure function E, the following inequality holds*

$$\left|\ln E(\mathbf{p}) - \ln E(\mathbf{p}')\right| \leqslant \max_i \left|\ln p_i - \ln p_i'\right| \tag{63}$$

for any pair of price vectors $\mathbf{p}$ and $\mathbf{p}'$ from $\mathbb{R}^n_{++}$.

In other words, logarithmic expenditure functions are 1-Lipshitz functions of logarithms of prices.

Proof. We need to show that

$$\min_i \frac{p_i}{p_i'} \leqslant \frac{E(p)}{E(p')} \leqslant \max_i \frac{p_i}{p_i'}.$$

It is enough to demonstrate the upper bound and the lower bound will follow by flipping the roles of $\mathbf{p}$ and $\mathbf{p}'$.

Recall that $E(\mathbf{p})$ is the minimal budget that the agent needs to achieve the unit level of utility for prices $\mathbf{p}$. Given prices $\mathbf{p}$ and $\mathbf{p}'$, define $\mathbf{p}'' = \max_i \frac{p_i}{p_i'} \cdot \mathbf{p}'$. The price of each good under $\mathbf{p}''$ is higher than for $\mathbf{p}$ and, hence, the agent needs at least as much money to achieve the same welfare level. Thus

$$E(\mathbf{p}) \leqslant E(\mathbf{p}'') = \max_i \frac{p_i}{p_i'} \cdot E(\mathbf{p}'),$$

where we used the homogeneity of the expenditure function. Dividing both sides by $E(\mathbf{p}')$, we obtain the desired inequality and complete the proof. $\square$

Denote by $\mathbf{e}$ the vector of all ones $\mathbf{e} = (1, \ldots, 1)$. By the lemma, we see that any expenditure function satisfies the following estimate

$$\left| \frac{\ln E(\mathbf{p}) - \ln E(\mathbf{e})}{1 + \max_i |\ln p_i|} \right| \leqslant 1 \tag{64}$$

for any vector of prices. The normalization in (64) suggests how to define a distance so that the set of logarithmic expenditure functions has a bounded diameter.

We define the distance between preferences $\succsim$ and $\succsim'$ or, equivalently, between the corresponding logarithmic expenditure functions $f = \ln E$ and $f' = \ln E'$ as follows:

$$d(\succsim, \succsim') = d(f, f') = \sup_{\mathbf{p} \in \Delta_{n-1} \cap \mathbb{R}^n_{++}} \left| \frac{(\ln E(\mathbf{p}) - \ln E(\mathbf{e})) - (\ln E'(\mathbf{p}) - \ln E'(\mathbf{e}))}{(1 + \max_i |\ln p_i|)^2} \right|. \tag{65}$$

The denominator in (65) is squared so that

$$\frac{\ln E(\mathbf{p}) - \ln E(\mathbf{e})}{(1 + \max_i |\ln p_i|)^2} \to 0 \quad \text{as } \mathbf{p} \text{ approaches the boundary of } \Delta_{n-1}. \tag{66}$$

Hence, the supremum is always attained at an interior point of the simplex and so can be replaced with the maximum.

Note that the ratio in (65) does not depend on the choice of a logarithmic expenditure function from the class of equivalent ones. On the other hand, the distance between any two distinct preferences or, equivalently, between two non-equivalent logarithmic expenditure functions is non-zero as the values of logarithmic expenditure functions on $\mathbb{R}^n_{++}$ are determined by their values on the interior of the simplex $\Delta_{n-1} \cap \mathbb{R}^n_{++} = \{\mathbf{p} \in \mathbb{R}^n_{++} \ : \ \sum_i x_i = 1\}$ since $E(\alpha \cdot \mathbf{p}) = \alpha \cdot E(\mathbf{p})$.

The metric structure on preferences allows one to define convergence and closed sets. A closure of a domain $\mathcal{D}$ of preference consists of all limit points of $\mathcal{D}$, i.e., of all the preferences $\succsim$ such that there exists a sequence of preferences $\succsim^{(l)} \in \mathcal{D}$ with $d(\succsim, \succsim^{(l)}) \to 0$ as $l \to \infty$. A closed domain is a domain that coincides with its closure.

Open sets are complements of closed ones and so the metric defines a topology. Once the topology is defined, one constructs the Borel measurable structure in the standard way (Aliprantis and Border, 2013, Section 4.4). Hence, we can write integrals of the form

$$\int_{\mathcal{D}} G(\succsim) \, \mathrm{d}\mu(\succsim) = \int_{\mathcal{L}} G(f) \, \mathrm{d}\mu(f)$$

formally where G is a Borel-measurable function and μ is a Borel measure (as usual, we identify functions and measures on preferences and on logarithmic expenditure functions). In all our examples, the integrated function G is continuous and, hence, measurable.

By (64), the diameter of $\mathcal{L}$ does not exceed 2. Hence, $\mathcal{L}$ is a bounded convex set. To fit the assumptions of the Choquet theory, we need to show that $\mathcal{L}$ is compact and can be thought of as a subset of a Banach space. We achieve both goals by constructing an isometric compact embedding of $\mathcal{L}$ into a Banach space.

Consider the Banach space $\mathcal{C}(\Delta_{n-1})$ of all continuous functions on the simplex endowed with the standard sup-norm $\|h\| = \sup_{\mathbf{p}\in\Delta_{n-1}} |h(\mathbf{p})|$.

Lemma 2. *Let T be a map that maps a logarithmic expenditure function $f = \ln E$ to a function $T[f]$ on Δ_{n-1} given by*

$$T[f](\mathbf{p}) = \begin{cases} \dfrac{\ln E(\mathbf{p}) - \ln E(\mathbf{e})}{(1+\max_i |\ln p_i|)^2}, & \mathbf{p} \in \Delta_{n-1} \cap \mathbb{R}^n_{++} \\ 0, & \textit{otherwise} \end{cases}. \tag{67}$$

Then E is an isometric embedding of the set $\mathcal{L}$ of logarithmic expenditure functions in the Banach space of continuous functions $\mathcal{C}(\Delta_{n-1})$ and the image $T[\mathcal{L}]$ is a compact convex set.

Proof. The function $T[f]$ is continuous in the interior of the simplex by the continuity of expenditure functions and it is continuous on the boundary by (66). Hence, $T[f]$ belongs to $\mathcal{C}(\Delta_{n-1})$. By the definition of the distance (65) and the norm in $\mathcal{C}(\Delta_{n-1})$, we get $d(f, f') = \|T[f] - T[f']\|$. Hence, T preserves the distance and, in particular, $f \neq f'$ implies $E[f] \neq T[f']$. Thus T is an isometric embedding of $\mathcal{L}$ in $\mathcal{C}(\Delta_{n-1})$.

The diameter of the image $T[\mathcal{L}]$ of $\mathcal{L}$ does not exceed 2 by (64). Hence, $T[\mathcal{L}]$ is a bounded subset of $\mathcal{C}(\Delta_{n-1})$. By Lemma (63), functions from $T[\mathcal{L}]$ are uniformly equicontinuous. Applying the Arcellà-Ascoli theorem, we conclude that the closure of $T[\mathcal{L}]$ is compact.[32]

It remains to show that $T[\mathcal{L}]$ is closed and convex. The set $\mathcal{L}$ is convex by Theorem 1 and E maps convex combinations to convex combinations, hence $T[\mathcal{L}]$ is convex. To show that it is closed, consider a sequence of functions $h^{(l)} \in T[\mathcal{L}]$ converging to some h and show that the limit belongs to $T[\mathcal{L}]$. Convergence in $\|\cdot\|$ implies pointwise convergence and hence h is equal to zero at the boundary of the simplex. At any $\mathbf{p}$ from the interior, we obtain that the sequence of expenditure functions $E^{(l)}(\mathbf{p})/E^{(l)}(\mathbf{e})$ corresponding to $h^{(l)}$ converges to $g(\mathbf{p}) = \exp\big((1+\max_i |\ln p_i|)^2 \cdot h(\mathbf{p})\big)$. As concavity is preserved under pointwise limits, g is a non-negative concave function on $\Delta_{n-1} \cap \mathbb{R}^n_{++}$ and, hence, there is a preference with an expenditure function $E = g$. Therefore, $h = T[\ln E]$ and so $T[\mathcal{L}]$ is closed. $\square$

By Lemma 2, one can think of $\mathcal{L}$ and the set of all homothetic preferences as a closed convex subset of $\mathcal{C}(\Delta_{n-1})$ and thus can use Choquet theory.

[32]Consider a subset $\mathcal{T}$ of the set $\mathcal{C}(X)$ of continuous functions on a compact set X with sup-norm. Arcellà-Ascoli theorem states that the closure $\overline{\mathcal{T}}$ of $\mathcal{T}$ is compact in $\mathcal{C}(X)$ if $\mathcal{T}$ is bounded and functions from $\mathcal{T}$ are uniformly equicontinuous.

C Proofs

C.1 Proof of Theorem 1

Proof. By the result of Eisenberg (1961), we know that the aggregate consumer exists and her preference corresponds to the following utility function

$$u_{\text{aggregate}}(\mathbf{x}) = \max\left\{\prod_{k=1}^{m}\left(\frac{u_k(\mathbf{x}_k)}{\beta_k}\right)^{\beta_k} \quad : \quad \mathbf{x}_k \in \mathbb{R}_+^n, \quad k = 1, \ldots, m, \quad \sum_{k=1}^{m} \mathbf{x}_k = \mathbf{x}\right\}. \tag{68}$$

Our goal is to compute the corresponding expenditure function $E_{\text{aggregate}}(\mathbf{p})$ and check that it satisfies the identity

$$\ln\left(E_{\text{aggregate}}(\mathbf{p})\right) = \sum_{k=1}^{m} \beta_k \cdot \ln\left(E_k(\mathbf{p})\right). \tag{69}$$

As an intermediate step, we compute the indirect utility of the aggregate consumer. Recall that the indirect utility of a consumer with a direct utility function u is given by

$$v(\mathbf{p}, b) = \max\left\{u(\mathbf{x}) : \mathbf{x} \in \mathbb{R}_+^n, \ \langle \mathbf{x}, \mathbf{p}\rangle \leqslant b\right\} = \frac{b}{E(\mathbf{p})}. \tag{70}$$

For the aggregate consumer, we get

$$v_{\text{aggregate}}(\mathbf{p}, b) = \max\left\{\prod_{k=1}^{m}\left(\frac{u_k(\mathbf{x}_k)}{\beta_k}\right)^{\beta_k} \quad : \quad \mathbf{x}_k \in \mathbb{R}_+^n, \quad k = 1, \ldots, m, \quad \left\langle \mathbf{p}, \ \sum_{k=1}^{m} \mathbf{x}_k \right\rangle \leqslant b\right\}. \tag{71}$$

Plug in $b = 1$ and consider an optimal collection of bundles $\mathbf{x}_k$, $k = 1, \ldots, m$ in (71). Denote their prices $\langle \mathbf{p}, \mathbf{x}_k\rangle$ by α_k. Our goal is to show that $\alpha_k = \beta_k$. The argument is along the lines of Eisenberg (1961). Rescale each bundle $\mathbf{x}_k$ to make its price equal to β_k. We obtain a new collection of bundles $\mathbf{x}_k' = \frac{\beta_k}{\alpha_k} \cdot \mathbf{x}_k$, which also satisfies the aggregate budget constraint $\langle \mathbf{p}, \ \sum_{k=1}^m \mathbf{x'}_k\rangle \leqslant 1$. By the optimality of $\mathbf{x}_k$, the product of utilities $\prod_k \left(u_k(\mathbf{x}_k)\right)^{\beta_k}$ is at least as big as $\prod_k \left(u_k(\mathbf{x}_k')\right)^{\beta_k}$. By homogeneity of utilities, this inequality of the products can be rewritten as follows:

$$1 \leqslant \prod_{k=1}^{m}\left(\frac{\alpha_k}{\beta_k}\right)^{\beta_k}.$$

Taking logarithm, we get an equivalent inequality

$$0 \leqslant \sum_{k=1}^{m} \beta_k \cdot \ln\frac{\alpha_k}{\beta_k}. \tag{72}$$

The concavity of the logarithm implies an upper bound on the right-hand side

$$\sum_{k=1}^{m} \beta_k \cdot \ln\frac{\alpha_k}{\beta_k} \leqslant \ln\left(\sum_k \beta_k \cdot \frac{\alpha_k}{\beta_k}\right) = \ln\left(\sum_k \alpha_k\right) \leqslant \ln(1) = 0. \tag{73}$$

Inequalities (72) and (73) can only be compatible if they are, in fact, equalities. As the logarithm is strictly concave, the equality between the first two expressions in (73) implies that the ratio $\frac{\alpha_k}{\beta_k}$ is a constant independent of k. Since the average value of the logarithms is zero, this constant equals one. We conclude that $\alpha_k = \beta_k$.

We proved that $\langle \mathbf{p}, \mathbf{x}_k \rangle = \beta_k$, for any optimal collection of bundles $\mathbf{x}_k$, $k = 1, \ldots, m$, from the optimization problem (71). In particular, the inequality $\langle \mathbf{p}, \mathbf{x}_k \rangle \leqslant \beta_k$ always holds at the optimum. Therefore, we can replace the budget constraint of the aggregate consumer $\langle p, \sum_{k=1}^m \mathbf{x}_k \rangle \leqslant 1$ with a stronger requirement of individual budget constraints $\langle \mathbf{p}, \mathbf{x}_k \rangle \leqslant \beta_k$, $k = 1, \ldots, m$, and this modification will not alter the value:

$$v_{\text{aggregate}}(\mathbf{p}, B) = \max \left\{ \prod_{k=1}^{m} \left(\frac{u_k(\mathbf{x}_k)}{\beta_k} \right)^{\beta_k} \quad : \quad \mathbf{x}_k \in \mathbb{R}_+^n, \quad \langle \mathbf{p}, \mathbf{x}_k \rangle \leqslant \beta_k, \quad k = 1, \ldots, m \right\}.$$

The maximization of the product reduces to maximizing each term $u_k(\mathbf{x}_k)$ separately over the corresponding budget set $\langle \mathbf{p}, \mathbf{x}_k \rangle \leqslant \beta_k$, which gives the indirect utility of consumer k:

$$v_{\text{aggregate}}(\mathbf{p}, B) = \prod_{k=1}^{m} \left(\frac{\max \left\{ u_k(\mathbf{x}_k) \quad : \quad \mathbf{x}_k \in \mathbb{R}_+^n, \quad \mathbf{p} \cdot \mathbf{x}_k \leqslant \beta_k \right\}}{\beta_k} \right)^{\beta_k} = \prod_{k=1}^{m} \left(\frac{v_k(\mathbf{p}, \beta_k)}{\beta_k} \right)^{\beta_k}.$$

Expressing each indirect utility through expenditure functions via formula (70), we end up with the following equality: $E_{\text{aggregate}}(\mathbf{p}) = \prod_{k=1}^m (E_k(\mathbf{p}))^{\beta_k}$. Taking the logarithm of both sides, we obtain identity (69) completing the proof. □

C.2 Proof of Proposition 1

Proof. Denote by $\underline{W}$ and $\overline{W}$ the infimum and the supremum of $W = \sum_k b_k \cdot w(\succsim_k)$ over all finite populations of agents with an aggregate preference $\succsim_{\text{aggregate}}$, total income B, and individual domain of preferences $\mathcal{D}$. Our goal is to show that

$$\underline{W} = B \cdot \operatorname{vex}_{\mathcal{L}_\mathcal{D}}[w]\big(\succsim_{\text{aggregate}} \big), \qquad \overline{W} = B \cdot \operatorname{cav}_{\mathcal{L}_\mathcal{D}}[w]\big(\succsim_{\text{aggregate}} \big),$$

and for any W' such that $\underline{W} < W' < \overline{W}$, there is a finite population with welfare W'.

Let us prove the formula for $\overline{W}$. By Theorem 1, populations of agents with preferences from $\mathcal{D}$ compatible with the aggregate preference $\succsim_{\text{aggregate}}$ correspond to all the different ways to represent $\ln E_{\text{aggregate}}$ as a finite convex combination $\ln E_{\text{aggregate}} = \sum_k \beta_k \ln E_k$ with $\ln E_k$ from $\mathcal{L}_\mathcal{D}$. Hence,

$$\overline{W} = \sup \left\{ B \cdot \sum_k \beta_k \cdot w(\succsim_k) \quad : \quad \ln E_{\text{aggregate}} = \sum_k \beta_k \ln E_k, \quad \ln E_k \in \mathcal{L}_\mathcal{D}, \quad \beta_k \geqslant 0, \quad \sum_k \beta_k = 1 \right\}.$$

Comparing this formula to the definition of concavification (27), we conclude that

$$\overline{W} = B \cdot \operatorname{cav}_{\mathcal{L}_\mathcal{D}}[w]\big(\succsim_{\text{aggregate}} \big).$$

A mirror argument for $\underline{W}$ is omitted.

It remains to construct a finite population with welfare $\underline{W} < W' < \overline{W}$. We already now that there are populations with welfare arbitrary close to $\underline{W}$ and $\overline{W}$. Hence, we can find W^l, $l = 1, 2$, such that there are populations $(\succsim_k^l, b_k^l)_{k=1,\ldots,K^l}$ with welfare W^l and $W^1 < W' < W^2$. Express W' as a convex combination $\alpha W^1 + (1-\alpha)W^2$, and consider the weighted union of the populations corresponding to W^1 and W^2:

$$\left(\succsim_k^1,\ \alpha \cdot b_k^1\right)_{k=1,\ldots,K^1} \bigcup \left(\succsim_k^2,\ (1-\alpha) \cdot b_k^2\right)_{k=1,\ldots,K^2}.$$

The constructed population has the desired welfare equal to W'. □

C.3 Generalizations and proofs of Theorems 2 and 3

We first prove Theorem 3 and then formulate and prove a general result containing both Theorem 2 and Theorem 3 as particular cases.

Recall that $\mathcal{D}^{\text{complete}}$ is the completion of a domain $\mathcal{D}$, i.e., the minimal closed domain invariant with respect to aggregation and containing $\mathcal{D}$. The set of indecomposable preferences in $\mathcal{D}$ is denoted by $\mathcal{D}^{\text{indec}}$ and contains all preferences $\succsim \in \mathcal{D}$ that cannot be represented as an aggregation of two distinct preferences $\succsim', \succsim'' \in \mathcal{D}$.

For the reader's convenience, we repeat the statement of Theorem 3.

Theorem. *If $\mathcal{D}$ is a closed domain such that $\mathcal{D} = \mathcal{D}^{\text{complete}}$, then a preference $\succsim$ belongs to $\mathcal{D}$ if and only if there exists a Borel probability measure μ supported on $\mathcal{D}^{\text{indec}}$ such that the expenditure function $E = E_{\succsim}$ can be represented as follows*

$$\ln E(\mathbf{p}) = \int_{\mathcal{D}^{\text{indec}}} \ln E_{\succsim'}(\mathbf{p})\, \mathrm{d}\mu(\succsim') \tag{74}$$

for any vector of prices $\mathbf{p} \in \mathbb{R}^n_{++}$.

The requirement that μ is supported on $\mathcal{D}^{\text{indec}}$ means that the complement of this set of preferences has μ-measure zero. The topology on preferences and logarithmic expenditure functions is described in Appendix B. The Borel structure is defined by this topology.

We will need the Choquet theorem formulated below. Consider a (not necessarily convex) subset X of a linear space. A point $x \in X$ is an extreme point of X if it cannot be represented as $\alpha x' + (1-\alpha)x''$ with $\alpha \in (0, 1)$ and distinct $x', x'' \in X$. All extreme points of X are denoted by X^{extrem}.

Theorem (Choquet's theorem; see Phelps (2001), Section 3)**.** *If X is a metrizable compact convex subset of a locally convex space, then a point x belongs to X if and only if there is a Borel probability* *measure μ on X supported on X^{extrem} such that*

$$x = \int_{X^{\mathrm{extrem}}} x' \,\mathrm{d}\mu(x'). \tag{75}$$

In our application, X will be a subset of the Banach space of continuous functions with the sup-norm. A Banach space is a complete separable normed space. Each such space is locally convex and metrizable via the metric induced by the norm.

The identity (75) is to be understood in the weak sense, i.e., for any continuous linear functional F

$$F[x] = \int_{X^{\mathrm{extrem}}} F[x'] \,\mathrm{d}\mu(x').$$

Proof of Theorem 3. A homothetic preference $\succsim$ is represented by a family of equivalent logarithmic expenditure functions which differ by a constant. Let $\mathcal{L}$ be the set of all classes of equivalent logarithmic expenditure functions corresponding to homothetic preferences. Denote by $\mathcal{L}_{\mathcal{D}}$ the subset of $\mathcal{L}$ corresponding to the domain $\mathcal{D}$. The set $\mathcal{L}_{\mathcal{D}}$ is closed and convex. Indeed, convexity follows from invariance of $\mathcal{D}$ by Corollary 7 and closedness of $\mathcal{D}$ is inherited by $\mathcal{L}_{\mathcal{D}}$ as the topologies on preferences and logarithmic expenditure functions are aligned.

By Lemma 2, the set $\mathcal{L}$ admits an affine isometric compact embedding T in the Banach space $\mathcal{C}\left(\Delta_{n-1}\right)$ of continuous functions on the simplex Δ_{n-1} with the sup-norm. Since $\mathcal{L}_{\mathcal{D}}$ is a closed convex subset of $\mathcal{L}$, the embedding $T[\mathcal{L}_{\mathcal{D}}]$ is a compact convex subset of $\mathcal{C}\left(\Delta_{n-1}\right)$.

Applying the Choquet theorem to $X = T[\mathcal{L}_{\mathcal{D}}]$, we conclude that a logarithmic expenditure function $\ln E$ belongs to $\mathcal{L}_{\mathcal{D}}$ if and only if there is a measure μ supported on X^{extrem} such that $x = T[\ln E]$ is given by the integral of the form (75).

As there is a natural bijection between $\mathcal{D}$ and X, we can assume that μ is a measure on $\mathcal{D}$. By Theorem 1, a preference $\succsim'$ is indecomposable in $\mathcal{D}^{\mathrm{indec}}$ if and only if its logarithmic expenditure function $\ln E_{\succsim'}$ cannot be represented as a convex combination of two non-equivalent expenditure functions from $\mathcal{D}$ (Corollary 13). Hence, $\succsim'$ belongs to $\mathcal{D}^{\mathrm{indec}}$ if and only if $T[\ln E_{\succsim'}]$ is in X^{extrem}. We obtain that $\succsim$ with an expenditure function E is contained in the completion $\mathcal{D}^{\mathrm{complete}}$ if and only if

$$T[\ln E] = \int_{\mathcal{D}^{\mathrm{indec}}} T[\ln E_{\succsim'}] \,\mathrm{d}\mu(\succsim') \tag{76}$$

for some μ supported on $\mathcal{D}^{\mathrm{indec}}$. To get the desired pointwise identity (74), it remains to apply an appropriate linear functional on both sides.

Let $F_{\mathbf{p}}$ be the functional on $\mathcal{C}\left(\Delta_{n-1}\right)$ evaluating a function at some $\mathbf{p} \in \Delta_{n-1}$. This functional is continuous and the family of such functionals with $\mathbf{p} \in \Delta_{n-1} \cap \mathbb{R}^n_{++}$ separates points, i.e., if two functions are not equal, there is a functional taking different values on them. Hence, (76) is equivalent to the following identity

$$F_{\mathbf{p}}\Big[T[\ln E]\Big] = \int_{\mathcal{D}^{\text{indec}}} F_{\mathbf{p}}\Big[T[\ln E_{\succsim'}]\Big]\,\mathrm{d}\mu(\succsim')$$

for all $\mathbf{p} \in \Delta_{n-1} \cap \mathbb{R}^n_{++}$. Plugging in the explicit form (67) of the embedding T, we conclude that $\succsim \in \mathcal{D}$ if and only if its logarithmic expenditure function $\succsim$ can be represented as follows

$$\frac{\ln E(\mathbf{p}) - \ln E(\mathbf{e})}{\left(1 + \max_i |\ln p_i|\right)^2} = \int_{\mathcal{D}^{\text{indec}}} \frac{\ln E_{\succsim'}(\mathbf{p}) - \ln E_{\succsim'}(\mathbf{e})}{\left(1 + \max_i |\ln p_i|\right)^2}\,\mathrm{d}\mu(\succsim')$$

for all $\mathbf{p} \in \Delta_{n-1} \cap \mathbb{R}^n_{++}$. Multiplying both sides by the denominator, we get

$$\ln E(\mathbf{p}) = \text{const} + \int_{\mathcal{D}^{\text{indec}}} \ln E_{\succsim'}(\mathbf{p})\,\mathrm{d}\mu(\succsim') \tag{77}$$

for $\mathbf{p} \in \Delta_{n-1} \cap \mathbb{R}^n_{++}$. Since $E(\alpha \cdot \mathbf{p}) = \alpha \cdot E(\mathbf{p})$, the identity extends to $\mathbb{R}^n_{++}$. As expenditure functions that differ by a constant correspond to the same preference, the constant in (77) can be absorbed by $\ln E$. This completes the proof of Theorem 3. $\square$

With the help of Theorem 3, we can prove that, without any assumptions on the domain $\mathcal{D}$, the completion $\mathcal{D}^{\text{complete}}$ is obtained by continuous aggregation of preferences from $\overline{\mathcal{D}}^{\text{indec}}$, i.e., of indecomposable preferences from the closure of $\mathcal{D}$. This result extends both Theorem 3 and Theorem 2.

Theorem 4. *For any domain $\mathcal{D}$, a preference $\succsim$ belongs to its completion $\mathcal{D}^{\text{complete}}$ if and only if the expenditure function E of $\succsim$ admits the following representation*

$$\ln E(\mathbf{p}) = \int_{\overline{\mathcal{D}}^{\text{indec}}} \ln E_{\succsim'}(\mathbf{p})\,\mathrm{d}\mu(\succsim'), \qquad \mathbf{p} \in \mathbb{R}^n_{++}, \tag{78}$$

for some Borel probability measure μ supported on $\overline{\mathcal{D}}^{\text{indec}}$.

Theorem 3 corresponds to closed invariant domains $\mathcal{D}$ and Theorem 2 is a corollary since $\overline{\mathcal{D}}^{\text{indec}}$ is a subset of $\overline{\mathcal{D}}$.

Our proof of Theorem 4 relies on already proved Theorem 3 and on Milman's converse to the Krein-Milman theorem. Recall that $\mathrm{conv}[Z]$ denotes the closed convex hull of a set Z.

Proposition 12 (Milman; see Phelps (2001), Proposition 1.5)**.** *If X is a compact convex subset of a locally convex space and $X = \mathrm{conv}[Z]$, then extreme points X^{extrem} are contained in the closure $\overline{Z}$.*

From the definition of extreme points, it is immediate that if $Z \subset X$, then any extreme point of X contained in Z is an extreme point of Z. Hence, the conclusion of Proposition 12 can be strengthened as $X^{\text{extrem}} \subset \overline{Z}^{\text{extrem}}$.

Since the completion corresponds to taking closed convex hull (Corollary 7) and indecomposable preferences correspond to extreme points, Proposition 12 implies the following corollary.

Corollary 16. *For any preference domain $\mathcal{D}$, all indecomposable preferences of its completion are contained in indecomposable preferences of its closure, i.e.,* $(\mathcal{D}^{\text{complete}})^{\text{indec}} \subset \overline{D}^{\text{indec}}$.

With this corollary, Theorem 4 follows from Theorem 3 almost immediately.

Proof of Theorem 4. Apply Theorem 3 to the closed invariant domain $\mathcal{D}' = \mathcal{D}^{\text{complete}}$. We get that $\succsim$ is in $\mathcal{D}^{\text{complete}}$ if and only if there is a measure supported $(\mathcal{D}^{\text{complete}})^{\text{indec}}$ such that

$$\ln E(\mathbf{p}) = \int_{(\mathcal{D}^{\text{complete}})^{\text{indec}}} \ln E_{\succsim'}(\mathbf{p})\, \mathrm{d}\mu(\succsim').$$

By Corollary 16, $(\mathcal{D}^{\text{complete}})^{\text{indec}} \subset \overline{D}^{\text{indec}}$, which completes the proof. □

C.4 Proof of Proposition 7

The proof relies on the following lemma showing that if preferences in two populations have expenditure shares that are close, then ε-equilibrium price vectors are close as well.

Lemma 3. *Consider two populations of m consumers with the same budgets $b_1, \ldots, b_m$ but different preferences over n goods: $\succsim_1, \ldots, \succsim_m$ in the first population and $\succsim'_1, \ldots, \succsim'_m$ in the second one. Assume that the expenditure shares $s_{k,i}(\mathbf{p})$ and $s'_{k,i}(\mathbf{p})$ differ by at most some $\delta > 0$ for any consumer k, good i, and price $\mathbf{p}$. Then any ε-equilibrium price vector for one population is an $(\varepsilon + n\delta)$-equilibrium price vector for the other.*

Proof of Lemma 3. The demand of an agent k for a good i can be expressed through expenditure shares as follows:

$$D_{k,i}(\mathbf{p}, b_k) = s_{k,i}(\mathbf{p}) \cdot \frac{b_k}{p_i}. \tag{79}$$

Let $\mathbf{x}_1 + \ldots + \mathbf{x}_m$ and $\mathbf{x'}_1 + \ldots + \mathbf{x'}_m$ be market demands of the two populations from the statement of the lemma at some vector of prices $\mathbf{p}$. By (79) and the assumption that expenditure shares differ by at most δ,

$$\sum_{i=1}^{n} p_i \cdot \left| \sum_{k=1}^{m} x_{k,i} - \sum_{k=1}^{m} x'_{k,i} \right| \leqslant \sum_{i=1}^{n} \sum_{k=1}^{m} b_k \cdot \max_{i,k} \left| s_{k,i}(\mathbf{p}) - s'_{k,i}(\mathbf{p}) \right| \leqslant n \cdot B \cdot \delta.$$

Hence, if $\mathbf{p}$ is an ε-equilibrium price vector for $\succsim'_1, \ldots, \succsim'_m$, it is an $(\varepsilon + n\delta)$-equilibrium price vector for $\succsim_1, \ldots, \succsim_m$. □

To prove the proposition, it remains to show that any preference $\succsim$ over two substitutes can be approximated by the aggregate preference $\succsim' = \succsim_{\text{aggregate}}$ of an auxiliary population with linear preferences so that the expenditure shares differ by at most ε at any vector of prices and the number of auxiliary agents is of the order of $1/\varepsilon$.

Proof of Proposition 7. Since $s_{\succsim,1} + s_{\succsim,2} = 1$ and expenditure shares depend on the ratio of prices only, it is enough to ensure that $\left|s_{\succsim,1}(p_1, 1) - s_{\text{aggregate},1}(p_1, 1)\right| \leqslant \varepsilon$ for any $p_1 \in \mathbb{R}_{++}$. As $\succsim$ exhibits substitutability, $s_{\succsim,1}(\,\cdot\,, 1)$ is a non-increasing function with values in $[0, 1]$. For any such function f, there is a piecewise-constant function f_ε with at most $1/\varepsilon + 1$ jumps such that the two functions differ by at most ε; indeed, one can take $f_\varepsilon = \varepsilon \cdot [f/\varepsilon]$, where $[t]$ denotes the integer part of a real number t. By Corollary 10, any piecewise-constant non-decreasing function with values in $[0, 1]$ is equal to $s_{\text{aggregate},1}(\,\cdot\,, 1)$ for a population of linear consumers with the number of consumers equal to the number of jumps, marginal rates of substitution given by positions of the jumps, and budgets determined by jumps' magnitude. We conclude that $1/\varepsilon + 1$ linear consumers are enough to approximate expenditure shares of any preference exhibiting substitutability with precision ε. Combined with Lemma 3, this observation completes the proof. □

Note that constructing the approximation in a computationally efficient way requires solving the equation $s_{\succsim,1}(p_1, 1) = \varepsilon \cdot l$ multiple times for various $\succsim$ and l. Provided that there is an oracle computing expenditure shares, one can use binary search for this task.

C.5 Indecomposability in the full domain and proof of Proposition 9

We will need the following simple lemma.

Lemma 4. *Consider a function $h(\mathbf{t}) = t_1^\alpha \cdot t_2^{1-\alpha}$ where $\alpha \in (0, 1)$ and $\mathbf{t} \in \mathbb{R}^2_{++}$. If $\mathbf{t} \neq \text{const} \cdot \mathbf{t}'$, then*

$$h\big(\lambda \mathbf{t} + (1-\lambda)\mathbf{t}'\big) > \lambda h(\mathbf{t}) + (1-\lambda) h(\mathbf{t}')$$

for any $\lambda \in (0, 1)$.

Proof. The result follows from strict concavity of $g(\lambda) = h\big(\lambda \mathbf{t} + (1-\lambda)\mathbf{t}'\big)$. To demonstrate strict concavity, it is enough to show that the second derivative $g''(\lambda) < 0$. After a linear change of variable, this requirement boils down to negativity of the second derivative of $\gamma^\alpha (1+\gamma)^{1-\alpha}$ with respect to γ. We omit the elementary computation. □

With the help of this lemma, we prove Proposition 9.

Proof of Proposition 9. A utility function u is of the form (60) if and only if the corresponding expenditure function is also piecewise linear:

$$E = \min_{c \in C} \left(\sum_{j=1}^{n} c_j p_j \right), \tag{80}$$

where $C \subset \mathbb{R}^n_+$ is finite or countable. We need to show that preferences with such expenditure functions are indecomposable. Towards a contradiction, assume that E of the form (80) can be represented as

$$\ln E = \alpha \ln E_1 + (1-\alpha)\ln E_2,$$

where E_1 and E_2 are expenditure functions representing distinct homothetic preferences and $\alpha \in (0,1)$. Hence, E_1 and E_2 are not proportional to each other, i.e., the ratio $E_1/E_2 \neq \text{const}$. By continuity of expenditure functions, this means that there is a linearity region of E where $E_1/E_2 \neq$ const. Therefore, we can find $\mathbf{p}, \mathbf{p}' \in \mathbb{R}^n_{++}$ from the same linearity region of E such that

$$\frac{E_1(\mathbf{p})}{E_2(\mathbf{p})} \neq \frac{E_1(\mathbf{p}')}{E_2(\mathbf{p}')}. \tag{81}$$

By homogeneity of expenditure functions, we can assume that $\mathbf{p}$ and $\mathbf{p}'$ are normalized so that $E(\mathbf{p}) = E(\mathbf{p}') = 1$. Since $\mathbf{p}$ and $\mathbf{p}'$ belong to the same linearity region, the value of E at the mid-point $\mathbf{p}'' = (\mathbf{p}+\mathbf{p}')/2$ is also equal to 1. Therefore,

$$1 = E\left(\mathbf{p}''\right) = E_1(\mathbf{p}'')^\alpha E_2(\mathbf{p}'')^{1-\alpha} \geqslant \left(\frac{1}{2}E_1(\mathbf{p}) + \frac{1}{2}E_1(\mathbf{p}')\right)^\alpha \left(\frac{1}{2}E_2(\mathbf{p}) + \frac{1}{2}E_2(\mathbf{p}')\right)^{1-\alpha},$$

where we used concavity of E_1 and E_2. By Lemma 4, the right-hand side admits the following lower bound

$$\left(\frac{1}{2}E_1(\mathbf{p}) + \frac{1}{2}E_1(\mathbf{p}')\right)^\alpha \left(\frac{1}{2}E_2(\mathbf{p}) + \frac{1}{2}E_2(\mathbf{p}')\right)^{1-\alpha} > \frac{1}{2}E_1(\mathbf{p})^\alpha E_2(\mathbf{p})^{1-\alpha} + \frac{1}{2}E_1(\mathbf{p}')^\alpha E_2(\mathbf{p}')^{1-\alpha}.$$

The right-hand side can be rewritten as

$$\frac{1}{2}E_1(\mathbf{p})^\alpha E_2(\mathbf{p})^{1-\alpha} + \frac{1}{2}E_1(\mathbf{p}')^\alpha E_2(\mathbf{p}')^{1-\alpha} = \frac{1}{2}E(\mathbf{p}) + \frac{1}{2}E(\mathbf{p}') = 1.$$

We end up with a contradictory inequality $1 > 1$. Therefore, E cannot be represented as a convex combination (81) and we conclude that the corresponding preference is indecomposable. □

Let us explore whether there are other indecomposable preferences in the full domain. For simplicity, we focus on the case of $n = 2$ goods. As opposed to preferences with piecewise linear u and E considered in Proposition 9, we examine preferences with expenditure functions E that are strictly concave in a neighborhood of a certain point.

We say that a function of one variable $h = h(t)$ is strictly concave in the neighborhood of $t = t_0$ if there is $\varepsilon > 0$ and $\delta > 0$ such that the second derivative of $h''(t) < -\delta$ for almost all t in the ε-neighborhood $[t_0 - \varepsilon, t_0 + \varepsilon]$ of t_0. We note that the second derivative exists almost everywhere for any concave function by Alexandrov's theorem.

Proposition 13. *Consider a preference $\succsim$ over two goods with expenditure function E. If there is a point $\mathbf{p}_0 \in \mathbb{R}^2_{++}$ and a direction $\mathbf{r} \in \mathbb{R}^2 \backslash \{0\}$ such that $g(t) = E(\mathbf{p}_0 + t \cdot \mathbf{r})$ is strictly concave in the neighborhood of $t = 0$, then $\succsim$ is not indecomposable.*

Proof. Since $E(\alpha \cdot \mathbf{p}) = \alpha \cdot E(\mathbf{p})$, the values of the expenditure function on the line $p_2 = 1$ determine its values everywhere by $E(p_1, p_2) = p_2 \cdot E(p_1/p_2,\, 1)$. Accordingly, the condition from the statement is equivalent to the existence of t_0 such that $g(t) = E(t, 1)$ is strictly concave in the neighborhood of t_0.

Let us show that if $g(t) = E(t, 1)$ is strictly concave in the neighborhood of t_0, then the preference $\succsim$ is an aggregation of some distinct $\succsim_1$ and $\succsim_2$. By strict concavity $g'' < -\delta$ on $[t_0 - \varepsilon, t_0 + \varepsilon]$ for some $\varepsilon, \delta > 0$. Let $\varphi(z)$ be a smooth function on $\mathbb{R}$ not equal to zero identically and vanishing outside of the interval $[-1, 1]$ together with all its derivatives. For example, one can take

$$\varphi(z) = \exp\left(-\frac{1}{1 - z^2}\right)$$

for $z \in (-1, 1)$ and zero outside. Define

$$g_1(t) \equiv \big(1 + \gamma \cdot \varphi\big(\varepsilon(t - t_0)\big)\big)\, g(t) \qquad \text{and} \qquad g_2(t) \equiv \frac{1}{1 + \gamma \cdot \varphi\big(\varepsilon(t - t_0)\big)} g(t)$$

for some constant $\gamma > 0$. Note that $g_1 = g_2 = g$ outside the ε-neighborhood of t_0. The second derivatives of g_1 and g_2 continuously depends on γ and, for $\gamma = 0$, the derivatives are bounded from above by $-\delta$ in the ε-neighborhood of t_0. Hence, for small enough $\gamma > 0$, the second derivative is non-positive, i.e., both g_1 and g_2 are concave.

Define $E_1(\mathbf{p}) = p_2 \cdot g_1(p_1/p_2)$ and $E_2(\mathbf{p}) = p_2 \cdot g_2(p_1/p_2)$. These are non-negative homogeneous concave functions that are not proportional to each other. Hence, E_1 and E_2 are expenditure functions of some distinct preferences $\succsim_1$ and $\succsim_2$. By the construction,

$$\ln E = \frac{1}{2} \ln E_1 + \frac{1}{2} \ln E_2$$

and thus $\succsim$ is the aggregate preference for a pair of consumers with preferences $\succsim_1$ and $\succsim_2$ and equal incomes. $\square$

It may seem that Propositions 9 and 13 cover all possible preferences in the case of two goods: intuitively, an expenditure function E is either piecewise linear or there is a point in the neighborhood of which E is strictly concave. However, there are pathological examples not captured by the two propositions.

Any concave function f on $\mathbb{R}_+$ can be represented as

$$f(t) = f(0) - \int_0^t \left(\int_0^s \mathrm{d}\nu(q)\right) \mathrm{d}s$$

for some positive measure ν on $\mathbb{R}_+$. This ν is the uniquely defined distributional second derivative of f. Abusing the notation, we will write $\nu = f''$. Note that the classical pointwise second derivative (where exists) equals the density of the absolutely continuous component of ν.

Propositions 9 and 13 address the cases where the second derivative of $E(t,1)$ is either an atomic measure with a nowhere dense set of atoms or has an absolutely continuous component with a strictly negative density on a certain small interval.

Recall that ν is called singular if there is a set of zero Lebesgue measure such that its complement has ν-measure zero. For example, atomic measures with discrete sets of atoms are singular, but there are other singular measures such as non-atomic measures supported on a Cantor set or atomic measures with everywhere dense set of atoms.

Proposition 14. *If $\succsim$ is a preference over two goods with an expenditure function E such that the second distributional derivative of $g(t) = E(t,1)$ is singular, then $\succsim$ is indecomposable in the full domain.*

We see that the set of indecomposable preferences is broader than suggested by Proposition 9. Note that, in the particular case of two goods, Proposition 9 is a direct corollary of Proposition 14.

Proof. It is enough to show that if $\succsim$ is an aggregation of two distinct preferences $\succsim_1$ and $\succsim_2$, then the second distributional derivative of g has a non-zero absolutely continuous component. In other words, we need to show that the classical derivative $g'' \neq 0$ on a set of positive Lebesgue measure.

Let E_1 and E_2 be expenditure functions of $\succsim_1$ and $\succsim_2$. Since the preferences are distinct, $E_1 \neq \text{const} \cdot E_2$. By the assumption, $E = E_1^{\alpha} \cdot E_2^{1-\alpha}$ with some $\alpha \in (0,1)$. Without loss of generality, we can assume that $\alpha = 1/2$. Indeed, if $\alpha \neq \frac{1}{2}$, one can define new expenditure functions $E_1' = E_1^{\alpha-\varepsilon} \cdot E_2^{1-\alpha+\varepsilon}$ and $E_2' = E_1^{\alpha+\varepsilon} \cdot E_2^{1-\alpha-\varepsilon}$ for some $\varepsilon < \min\{\alpha,\, 1-\alpha\}$ so that $E = \sqrt{E_1' \cdot E_2'}$.

Hence, $g = \sqrt{g_1 \cdot g_2}$ where $g_1(t) = E_1(t,1)$ and $g_2(t) = E_2(t,1)$ are non-negative concave functions not proportional to each other. Computing the classical second derivative of g, we obtain

$$g'' = \frac{g_1'' \cdot g_2 + g_2'' \cdot g_1}{2\sqrt{g_1 \cdot g_2}} - \frac{\left(g_1' \cdot g_2 - g_2' \cdot g_1\right)^2}{4(g_1 \cdot g_2)^{3/2}}.$$

Both terms are non-positive. The numerator in the second term can be rewritten as follows

$$\left(g_1' \cdot g_2 - g_2' \cdot g_1\right)^2 = \left((g_2)^2 \cdot \left(\frac{g_1}{g_2}\right)'\right)^2$$

Since the ratio g_1/g_2 is non-constant, its derivative $(g_1/g_2)'$ is non-zero on a set of positive measure. Thus the distributional derivative g'' contains a non-zero absolutely continuous component. We conclude that preferences such that g'' has no absolutely continuous component are indecomposable. $\square$

The approach from Proposition 14 extends to $n > 2$ goods. It can be used to show that if the distributional second derivative of

$$g(t) = E(p_1, \dots, p_{i-1}, t, p_{i+1}, \dots, p_n)$$

is singular for any $i = 1, \dots, n$ and any fixed

$$p_{-i} = (p_1, \dots, p_{i-1}, p_{i+1}, \dots, p_n) \in \mathbb{R}^{n-1}_{++},$$

then the corresponding preference is indecomposable. This result extends Proposition 9.

C.6 Proof of Proposition 11

Proof. Consider a Leontief preference $\succsim$ over Cobb-Douglas composite goods. It corresponds to a utility function

$$u(\mathbf{x}) = \min_{\mathbf{a} \in A} \left\{ a_0 \cdot \prod_{i=1}^{n} x_i^{a_i} \right\},$$

where A is finite or countably infinite subset of $\mathbb{R}_{++} \times \Delta_{n-1}$. Assume that $\succsim$ is non-trivial, i.e., there are $\mathbf{a}, \mathbf{a}' \in A$ such that $(a_1, \dots, a_n) \neq (a'_1, \dots, a'_n)$. The intersection of convex sets corresponds to convexification of the maximum of their support functions (Aliprantis and Border, 2013, Theorem 7.56). Since expenditure functions are support functions of upper contour sets up to a sign, the expenditure function corresponding to $\succsim$ takes the following form

$$E(\mathbf{p}) = \operatorname{cav}\left[\max_{\mathbf{a} \in A} \left\{ \frac{1}{a_0} \cdot \prod_{i=1}^{n} p_i^{a_i} \right\} \right],$$

where cav denotes concavification.

Let us focus on the case of $n = 2$ goods. In this case, E has a particularly simple structure. The positive orthant $\mathbb{R}^2_{++}$ is partitioned into a finite or countably infinite number of cones of two types: (I) cones where E is linear (II) cones where E coincides with the Cobb-Douglas expenditure function $1/a_0 \cdot \prod_{i=1}^{n} p_i^{a_i}$ for some $\mathbf{a} \in A$. The cones of type (I) and (II) interlace and derivatives of E change continuously. Note that there must be at least one cone of type (I) as otherwise, E would be an expenditure function of standard Cobb-Douglas preferences, which is ruled out by the non-triviality assumption.

Let us show that such a preference $\succsim$ over two goods is indecomposable. Towards a contradiction, assume that

$$\ln E = \alpha \cdot \ln E_1 + (1 - \alpha) \ln E_2, \qquad \alpha \in (0, 1), \tag{82}$$

where and E_1 and E_2 corresponds to two distinct preferences $\succsim_1$ and $\succsim_2$ exhibiting complementarity. As in the proof of Proposition 9, one shows that in each linearity region of E, the expenditure functions E_1 and E_2 are proportional to each other. In other words, $E_1 = \text{const} \cdot E_2$ in each cone of type (I), where the constant can depend on the cone.

Recall that by (6), the partial derivative of a logarithmic expenditure function with respect to $\ln p_i$ is the expenditure shares of good i. Denote the expenditure shares for $\succsim$, $\succsim_1$, and $\succsim_2$ by s_i, $s_{1,i}$, and $s_{2,i}$, respectively. Since E_1 and E_2 are proportional in cones of type (I), we obtain that $s_i = s_{1,i} = s_{2,i}$ there.

Consider cones of type (II). In these cones, s_i is constant since expenditure shares are constant for Cobb-Douglas preferences. Taking the partial derivative on both sides of (82), we get

$$s_i(p_1, p_2) = \alpha \cdot s_{1,i}(p_1, p_2) + (1 - \alpha)s_{2,i}(p_1, p_2).$$

The expenditure shares depend only on the ratio of prices. By complementarity, $s_{1,i}$ and $s_{2,i}$ must be non-increasing functions of the ratio p_{3-i}/p_i. Note that if s_i is constant and one of $s_{1,i}$ or $s_{2,i}$ increases, the other must decrease violating the monotonicity requirement. Hence, in all the cones of type (II), s_i, $s_{1,i}$ and $s_{2,i}$ are constant. These constants must be all equal. Indeed, suppose that $s_i = c$, $s_{1,i} = c_1$ and $s_{2,i} = c_2$ in some cone of type (II) with $c_1 \neq c_2$. Since s_i is continuous and coincides with $s_{1,i}$ and $s_{2,i}$ in the neighboring cone of type (I), $s_{1,i}$ and $s_{2,i}$ are discontinuous on the boundary between the two cones and at least one of these discontinuities necessarily violates the monotonicity requirement. Thus in cones of type (II), $s_i = s_{1,i} = s_{2,i}$.

We conclude that $s_i = s_{1,i} = s_{2,i}$ everywhere, i.e., partial derivatives of $\ln E$, $\ln E_1$, and $\ln E_2$ coincide. Hence, $\ln E_1 = \ln E_2 + \text{const}$, i.e., E_1 and E_2 are proportional. Thus $\succsim_1 = \succsim_2$, which contradicts the assumption that the two preferences are distinct. We conclude that, for $n = 2$, any non-trivial Leontief preference over Cobb-Douglas composite goods is indecomposable. □

C.7 Characterization of expenditure shares for two goods

For a preference $\succsim$ over $n = 2$ goods, consider the expenditure shares $s_1(\mathbf{p})$ and $s_2(\mathbf{p})$ of these goods. We explore what functions one can get as expenditure shares. Since $s_1 + s_2 = 1$, we can focus on the expenditure share s_1 of the first good. As $s_1(\alpha \cdot \mathbf{p}) = s_1(\mathbf{p})$, the expenditure share can be seen as a function of one variable $z = p_1/p_2$. Our goal is to characterize functions $h = h(z)$ such that $h = s_1$ for some homothetic preference $\succsim$.

Lemma 5. *A function $h: \mathbb{R}_{++} \to \mathbb{R}$ is the expenditure share of the first good associated with some homothetic preference $\succsim$ over two goods (i.e., $h(p_1/p_2) = s_1(\mathbf{p})$, $\mathbf{p} \in \mathbb{R}^2_{++}$) if and only if*

$$h(z) = \frac{z}{z + Q(z)}, \tag{83}$$

where $Q: \mathbb{R}_{++} \to \mathbb{R}_+ \cup \{+\infty\}$ is a non-negative non-decreasing function.

Leontief and linear preferences correspond to the two extreme cases: a constant $Q = \text{const}$ and an infinite step function

$$Q = \begin{cases} 0, & z \leqslant \alpha \\ +\infty & z > \alpha \end{cases},$$

respectively.

Note that for any non-increasing function h with values in $[0,1]$, the function

$$Q(z) = \frac{1}{h(z)} - 1$$

satisfies the requirement of Lemma 5.

Corollary 17. *Any non-increasing function h with values in $[0,1]$ is the expenditure share of the first good for some preference $\succsim$ exhibiting substitutability among the two goods.*

Proof of Lemma 5. Consider a homothetic preference $\succsim$ for $n = 2$ goods. The expenditure share $s_1(\mathbf{p})$ of the first good satisfies $s_1(\alpha \cdot \mathbf{p}) = \alpha \cdot s(\mathbf{p})$ and, hence, $s_1(\mathbf{p}) = s_1(z, 1)$, where $z = p_1/p_2$.

Let us show that, for any $\succsim$, the function $h(z) = s_1(z, 1)$ admits the representation (83). In other words, we need to show that

$$Q(z) = \frac{z}{s_1(z,1)} - z$$

is non-negative and non-decreasing. Expressing the expenditure share through the logarithmic expenditure function by (6) and denoting $\pi(z) = E(z, 1)$, we get

$$Q(z) = \frac{\pi(z)}{\pi'(z)} - z. \tag{84}$$

Note that π is a non-negative non-decreasing concave function of z. Hence, Q is non-negative as well. To show that Q is non-decreasing, let us differentiate both sides of (84). We get

$$Q'(z) = -\frac{\pi(z)}{(\pi'(z))^2} \cdot \pi''(z). \tag{85}$$

We see that $Q' \geqslant 0$ and so Q is non-decreasing.[33] We conclude that $h = s_1(z, 1)$ admits the representation (83).

To prove the converse, consider h of the form (83) with non-negative non-decreasing Q and construct the corresponding preference. The identity (84) suggests how to define the expenditure function. We get

$$\frac{\pi'(z)}{\pi(z)} = \frac{1}{z + Q(z)},$$

[33]If π' is not differentiable, the identity (85) is to be understood in the sense of distributional derivatives: π'' is a non-negative measure, and the right-hand side is a measure having density $-\pi(z)/(\pi'(z))^2$ with respect to π''.

where $\pi(z) = E(z, 1)$. Integrating this identity, we obtain

$$\pi(z) = \exp\left(\int_1^z \frac{1}{w + Q(w)}\, \mathrm{d}w\right).$$

By the construction, π is non-negative and non-decreasing. Since the identity (85) is hardwired in the definition of π and the function Q is non-decreasing, we conclude that π'' is non-positive. Hence, π is concave. Define E by

$$E(p_1, p_2) = p_2 \cdot \pi(p_1/p_2).$$

This function is homogeneous, non-negative, non-decreasing, and concave. Thus E is an expenditure function corresponding to some homothetic preference. By the construction, $s_1(z, 1) = h(z)$ completing the proof. □

D Robust welfare analysis and Bayesian persuasion: a formal connection

Let us discuss the formal connection between robust welfare analysis as described in Section 3.2 and Bayesian persuasion, a benchmark model for a situation where an informed party decides what information to reveal to an uninformed one and has an objective depending on induced beliefs (Kamenica and Gentzkow, 2011). Mathematically, persuasion boils down to solving the following optimization problem. We are given a set of states Ω, a prior belief $\mu \in \Delta(\Omega)$ where $\Delta(\Omega)$ denotes the simplex of probability distributions over Ω, and an objective function g defined on $\Delta(\Omega)$. The goal is to maximize

$$\sum_k \beta_k \cdot g(\mu_k)$$

over all possible ways to represent the prior μ as a finite convex combination $\mu = \sum_k \beta_k \cdot \mu_k$ with $\mu_k \in \Delta(\Omega)$. The persuasion problem has an elegant geometric solution: the optimal value of the persuasion problem is $\mathrm{cav}_{\Delta(\Omega)}[g](\mu)$ (Aumann et al., 1995; Kamenica and Gentzkow, 2011).

The similarity between persuasion and finding the maximal value of W of the form (26) compatible with the observed aggregate behavior must be apparent: the set $\mathcal{L}_{\mathcal{D}}$ of logarithmic expenditure functions plays the role of $\Delta(\Omega)$, the logarithmic expenditure function of the aggregate preference corresponds to the prior μ, and $w = w(\succsim)$ considered as a function on $\mathcal{L}_{\mathcal{D}}$ is the analog of informed party's objective g.

The difference is that in Bayesian persuasion, the concavification takes place over a simplex $\Delta(\Omega)$ while the set $\mathcal{L}_{\mathcal{D}}$ of logarithmic expenditure functions is not necessarily a simplex. Recall that a convex set is a simplex if the decomposition over the extreme points is unique; we call $\mathcal{D}$ a simplex domain if the corresponding set of logarithmic expenditure functions $\mathcal{L}_{\mathcal{D}}$ is a simplex; see Section 5.2.

For simplex domains, finding the maximal value of W compatible with the observed aggregate behavior is equivalent to a persuasion problem. An elementary example, where this equivalence holds is the domain $\mathcal{D}$ of Cobb-Douglas preferences; see Example 3.

Let us also illustrate the equivalence for the domain $\mathcal{D}_S$ of all preferences exhibiting substitutability over $n = 2$ goods as in Example 4. Any preference $\succsim \in \mathcal{D}_S$ can be represented as an aggregation of linear preferences and this representation is unique (Corollary 10). Linear preferences over two goods form a one-parametric family with the marginal rate of substitution $\mathrm{MRS} = v_1/v_2 \in \mathbb{R}_+ \cup \{+\infty\}$ as a parameter. A preference $\succsim \in \mathcal{D}_S$ defines a unique distribution μ of MRS by formula (45): the cumulative distribution function is equal to $1 - s_1(\,\cdot\,, 1)$. The functional $w(\succsim)$ can equivalently be thought of as a function of μ. Thus the welfare maximization problem takes the following form. We are given $\mu_{\text{aggregate}} \in \Delta(\mathbb{R}_+ \cup \{+\infty\})$ and a functional $w = w(\mu)$. The goal is to maximize

$$B \cdot \sum_k \beta_k \cdot w(\mu_k)$$

over all possible ways to represent the prior μ as a finite convex combination $\mu = \sum_k \beta_k \cdot \mu_k$ with $\mu_k \in \Delta(\mathbb{R}_+ \cup \{+\infty\})$.

We conclude that, for two substitutes, finding the maximal welfare compatible with the observed aggregate behavior is equivalent to persuasion with the set of states $\Omega = \mathbb{R}_+ \cup \{+\infty\}$, the cumulative distribution function of the prior μ equal to $1 - s_1(\,\cdot\,, 1)$, and the objective $w = w(\mu)$. Persuasion problems with a continual state space are not easy to solve analytically unless some further assumptions are made. For example, if μ is finitely supported, i.e., there is a finite number of preference "types" in the population, then the support can be taken as the new set of states reducing the problem to the well-understood case of persuasion with a finite number of states. If μ has infinite support, tractability can be gained by imposing assumptions on the objective w. Tractable cases include convex w as in Examples 4 and 5, or w depending on μ through the mean value of a given function φ, i.e., $w = w\left(\int \varphi(z) \mathrm{d}\mu(z)\right)$ as in (Dworczak and Martini, 2019; Arieli et al., 2019; Kleiner et al., 2021), or w depending on μ through a quantile of φ as in (Yang and Zentefis, 2022).