Random Choice and Differentiation

Junnan He
and
Paulo Natenzon*

First posted: July 2023; Updated: December 2023

Abstract

The degree to which consumers treat different options as distinct or differentiated is a key determinant of market competition and pricing. To facilitate the measurement of differentiation, we develop a flexible yet tractable model of random choice in a multi-attribute setting. We show the analyst can separately identify vertical and horizontal differentiation from binary comparison data alone. We characterize the binary choice rules that arise from our model using four easily understood axioms. In multinomial choice, we show that the intersection of our model with the classic random utility framework yields random coefficients with an elliptical distribution. We provide applications to consumer demand with differentiated products and to measuring the complexity faced by an agent in individual decision-making problems.

Keywords: discrete choice; differentiation; random coefficients; complexity.

*Junnan He, Department of Economics, Sciences Po. Paulo Natenzon, Olin Business School, Washington University in St. Louis. Contact email: `pnatenzon@wustl.edu`. We are grateful to David Ahn, Brad Larsen, and seminar audiences at the London School of Economics, Ohio State, Johns Hopkins, University of Michigan, SAET Paris and UC Berkeley for helpful comments and suggestions. Please check `http://pnatenzon.wustl.edu` for the latest version.

1 Introduction

The degree to which consumers treat different options as distinct or differentiated is a key determinant of market competition and pricing. Measuring differentiation is therefore important for determining the substitutability of different products in demand analysis; evaluating the impact of mergers and acquisitions; assessing the welfare consequences of product innovation; and understanding how the competitive environment of an industry evolves over time.

To facilitate the measurement of differentiation, in this paper, we develop and analyze a model of random choice in a multi-attribute setting. We make two main contributions: first, we provide a flexible framework in which vertical and horizontal product differentiation are uniquely revealed from observed choice frequencies. Second, we establish behavioral foundations for applying our framework by providing two characterization theorems that determine the testable implications of the model for binary comparison choice and for multinomial choice.

We start with the analysis of random choice in binary comparison problems. In our *linear differentiation model*, which we introduce in detail in Section 3, differentiation takes the form of a generalized Euclidean distance between options. This distance is parameterized by a positive definite matrix, which flexibly accommodates the degree to which each observable attribute contributes to choice options being treated as more or less differentiated by decision makers. Our formulation allows us to obtain a unique decomposition of the differentiation between the options into its vertical and horizontal components (Proposition 1) and to fully identify these components and the remaining parameters of the model from binary choice data (Proposition 2). In Theorem 1, we fully characterize the binary choice behavior that arises from this model using four easily understood axioms.

Next, we extend the analysis from binary choice to multinomial choice. In Theorem 2, we characterize the intersection of the linear differentiation model with the classic random coefficients framework of discrete choice estimation. Formally, we combine the multinomial choice postulates from the random expected utility theory of Gul and Pesendorfer (2006) with our binary choice postulates, and we show this yields a very convenient elliptical distribution for the random coefficients representation, which is unique up to scaling.

We offer two applications of our results. Our first application is to aggregate consumer demand with differentiated products. In this setting, the random choice model captures the heterogeneity of tastes in a population of standard rational consumers, and choice probabilities represent the market shares of different products. Our theory offers a direct application for empiricists in that it provides a new interpretation of typical estimated parameters from demand models. Using the automobile demand estimates of Berry et al. (1995), we illustrate how our vertical and horizontal differentiation measures can shed light on the product differentiation strategies employed by firms.

Our second application contributes to a growing literature that models imprecision and noise in individual decision making (see Woodford, 2020, for a recent review). In this application, the choice probabilities of our random choice model reflect the variability of individual choices due to imperfect information, mistakes, or rationally inattentive behavior. We show that our vertical and horizontal differentiation decomposition resolves the existing theoretical and empirical ambiguity concerning the fundamental relationship between the ability of a decision maker to correctly discriminate between options and the distance between those options in the attribute space. In particular, we show that our horizontal differentiation measure offers a precise and quantifiable notion of the complexity of the tradeoffs faced by an agent in binary decision-making problems. We illustrate the measurement of complexity in a simple application using the experimental choice data from Tversky and Russo (1969), where subjects compare rectangles of different sizes.

1.1 Related Literature

The axiomatic literature on random choice has largely focused on the shortcomings of the classic logit model (Luce, 1959), proposing generalizations to address its limitations (Gul, Natenzon and Pesendorfer, 2014; Fudenberg, Iijima and Strzalecki, 2015; Echenique and Saito, 2019; Dutta, 2020; Horan, 2021; Faro, 2023; Chambers, Cuhadaroglu and Masatlioglu, 2023). We share with this literature the goal of accounting for the role of product differentiation in choice, which is assumed away by the classic logit model. We take a different starting point, however, in that we offer a generalization of the classic probit model (Thurstone, 1927). A second important departure from this literature is that we explicitly model the observable attributes of the choice options, and our axioms directly relate to the observable characteristics of the choice objects. This allow us to provide a novel measurement of product differentiation uniquely decomposed into its vertical differentiation and horizontal differentiation components, and to directly relate our results to applications with real-world data.

Our elliptical random coefficients model belongs to the class of random parameter models studied by Apesteguia and Ballester (2018), who show this class satisfies desirable monotonicity properties.

Our Theorem 2 builds on Gul and Pesendorfer (2006), who characterize a random expected utility model with four axioms in the context of risky choice. In Theorem 2, we bring their postulates from the choice over lotteries environment to choice options that are vectors in an Euclidean space. We impose three additional postulates to obtain our elliptical coefficient representation. These three additional postulates put restrictions solely on binary choice comparisons, and yield the representation that allows us to pin down vertical and horizontal differentiation uniquely from choice data.

He and Natenzon (2023) study binary choice in an abstract setting, that is, without any observable attributes. They show that a wide variety of random choice models are special cases of the moderate utility binary choice formula proposed in Halff (1976). They prove that the moderate utility formula is equivalent to the moderate transitivity postulate, which is one of the key assumptions in our Theorems 1 and 2. Hence, both our linear differentiation model and our random elliptical coefficients model are special cases of the very general moderate utility formula.

Our results also relate to the demand estimation literature in industrial organization. Accounting for the degree of differentiation between choice options is essential for the analysis of demand in several markets such as automobiles (Berry et al., 1995), ready-to-eat cereal (Nevo, 2001), online newspapers (Gentzkow, 2007), and health plans (Einav et al., 2013). Formally, we study what the industrial organization literature calls a pure characteristics model (Berry and Pakes, 2007). Our Theorem 2 offers a complete non-parametric characterization of the pure characteristics models with elliptically distributed coefficients. In Proposition 4, we also characterize the important special case of independent Gaussian coefficients commonly assumed in the literature.

Often, empiricists estimate these models with additional iid error terms to utility, both in an effort to account for unobserved factors and for computational convenience (McFadden and Train, 2000). Lu and Saito (2022) study the effects of these additional iid error terms, and their analysis implies predictable deviations from our random choice postulates. Those differences aside, our results allow us to re-interpret some of the demand estimates obtained in this literature. In Section 5.1 we use the demand estimates from Berry et al. (1995) to illustrate how our vertical and horizontal differentiation measures can provide insight into the product characteristics decisions of firms.

Finally, our paper also relates to the behavioral economics literature that studies stochastic individual choice due to imperfect information, imperfect perception, and noisy cognition— e.g. Natenzon (2019) and He (2023); see Woodford (2020) for a recent review.

In particular, we touch on a theoretical and experimental literature that attempts to quantity the complexity faced by decision makers, and to relate measures of complexity to stochastic choice behavior (e.g. Oprea, 2022; Enke et al., 2023; Enke and Graeber, 2023; Puri, 2023). Puri (2023), for example, shows the number of prizes in the support of a lottery is a quantitatively meaningful indicator of the effect of complexity on choice behavior. Our results contribute to this literature by providing foundations for a new measure of complexity, endogenously obtained from choice data. In Section 5.2, we show our horizontal differentiation measure captures the complexity of evaluating tradeoffs between two options, independent from utility. We also relate our complexity measure to the literature of endogenous information acquisition, aka rational inattention (Sims, 2003; Matejka and McKay, 2015; Hebert and Woodford, 2021; Pomatto et al., 2023; Dean and Neligh, 2023).

The rest of the paper is organized as follows. Section 2 introduces the setup. Section 3 presents the linear differentiation model; introduces our vertical and horizontal differentiation decomposition; and contains the identification and characterization results for binary choice. Section 4 introduces the elliptical random coefficients model and contains the analysis of multinomial choice. Section 5 contains our applications to aggregate consumer demand with differentiated products and to measuring the complexity faced by an agent in individual choice problems.

2 Model

We consider a setting in which choice options are differentiated across multiple attributes. Each option $x = (x_1, \dots, x_n)$ is identified with its location in $\mathbb{R}^n$. A decision problem is a finite subset $A \subset \mathbb{R}^n$. The primitive observed by the analyst is a *choice rule* ρ mapping each option $x \in \mathbb{R}^n$ and each decision problem A to a probability $0 \leq \rho(x, A) \leq 1$. A choice rule assigns $\rho(x, A) > 0$ only if $x \in A$ and satisfies

$$\sum_{x \in A} \rho(x, A) = 1$$

for every decision problem A.

In the population interpretation of random choice, ρ describes market shares. That is, $\rho(x, A)$ is the proportion of consumers that choose x when the set of available options in the market is A. Alternatively, ρ may describe the stochastic behavior of a single agent, where $\rho(x, A)$ is the probability that the individual chooses x from A.

Binary choice problems are common in the experimental literature and play a special role in the theoretical analysis that follows. For a binary decision problem $A = \{x, y\}$, with $x \neq y$, we will write $\rho(x, y)$ instead of $\rho(x, \{x, y\})$. When a choice rule ρ is restricted to binary choice problems, we will call it a *binary choice rule*. Let $\mathcal{D} = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n : x \neq y\}$ be the set of all ordered pairs of distinct options in $\mathbb{R}^n$. Then, a binary choice rule is a function $\rho : \mathcal{D} \to [0, 1]$ such that $\rho(x, y) + \rho(y, x) = 1$ for every $x \neq y$. In the population interpretation of random choice, $\rho(x, y)$ is the proportion of a heterogeneous population that preferes x to y. In the individual stochastic choice behavior interpretation, $\rho(x, y)$ is the probability that the individual chooses x in a binary comparison against y. We start with the analysis of the special but important binary choice case, and we leave the general multinomial choice analysis to Section 4.

3 Binary choice

We now introduce the *linear differentiation model*, a parametric representation for a binary choice rule ρ. A linear function $U(x)$ captures the utility of each option, and a generalized Euclidean distance $\|x - y\|$ measures the differentiation between the options. An option x is chosen over y with probability

$$\rho(x, y) = F\left(\frac{U(x) - U(y)}{\|x - y\|}\right) \tag{1}$$

where F is a continuous, strictly increasing transformation. Note that since $\rho(x, y) + \rho(y, x) = 1$, we must have $F(t) + F(-t) = 1$ for all t and, in particular, $F(0) = 1/2$.

The linear differentiation model (1) writes the strength with which x is preferred to y as a ratio: strength of preference is directly proportional to utility difference, and inversely proportional to differentiation. The transformation F maps strength of preference to choice probabilities. It follows from (1) that two options $x \neq y$ are chosen equally often in a binary comparison if and only if they have the same utility. For example, in Figure 1 options x and $\hat{x}$ lie on the same indifference curve and therefore $\rho(x, \hat{x}) = 1/2$.

Note the important role of differentiation in the denominator: for any fixed utility difference $U(x) - U(y)$, increasing the differentiation $\|x - y\|$ between x and y drives choice probabilities closer to 1/2, capturing the fact that more differentiated options are less substitutable and harder to compare. For example, in Figure 1 option y has lower utility than x and $\hat{x}$, and therefore $\rho(x, y) > 1/2$ and $\rho(\hat{x}, y) > 1/2$. However, y is more differentiated from x than from $\hat{x}$, that is, $\|x - y\| > \|\hat{x} - y\|$. A larger differentiation between x and y means that choice is less decisive, that is, $\rho(\hat{x}, y) > \rho(x, y) > 1/2$.

The differentiation $\|x-y\|$ in the linear differentiation model is given by a generalized Euclidean distance. This means the norm $\|\cdot\|$ is generated by an inner product, and there is a unique $n \times n$ symmetric positive-definite matrix Σ such that, for every x, y,

$$\|x - y\| = \sqrt{(x - y)'\Sigma(x - y)}.$$

Moreover, the utility function U is linear, so there is a unique utility vector $u \in \mathbb{R}^n$ such that $U(x) = u'x$ for every x. Writing the parameters as a triple (u, Σ, F), the linear differentiation model becomes

$$\rho(x, y) = F\left(\frac{u'(x - y)}{\sqrt{(x - y)'\Sigma(x - y)}}\right). \tag{2}$$

3.1 Vertical and horizontal differentiation

Our framework allows us to obtain a unique decomposition of the difference $x-y$ between any two options x, y into two components, which we call the *vertical* and *horizontal* differentiation, following the language used in the spatial differentiation literature in industrial organization (see e.g. Lancaster, 1979).

We say that x and y are (purely) *vertically differentiated* if they solve the choice probability maximization problem $\max_{x,y} \rho(x, y)$. In other words, vertical differentiation maximizes the agreement among consumers over the ranking of two options. Conversely, we say that x and y are (purely) *horizontally differentiated* when the agreement among consumers is minimized, that is, when $\rho(x, y) = 1/2$. The interpretation is that vertical differentiation reflects quality differences that generate consensus, while horizontal differentiation reflects variety instead of quality, with the ranking of options depending on the tastes or location of each particular consumer.

Our definitions of vertical and horizontal differentiation are behavioral, written solely in terms of observable choices given by ρ. Our first result relates vertical and horizontal differentiation to the parameters (u, Σ, F) of the linear differentiation model.

Proposition 1. *Let ρ be a linear differentiation model with parameters (u, Σ, F). Then:*

(i) *x and y are vertically differentiated if and only if $x - y = \alpha\Sigma^{-1}u$ for some $\alpha \neq 0$;*

(ii) *x and y are horizontally differentiated if and only if $u'x = u'y$;*

(iii) *x, y vertically and w, z horizontally differentiated implies $(x - y)'\Sigma(w - z) = 0$;*

(iv) *Any $x - y$ is uniquely decomposed into vertical and horizontal differentiation by:*

$$x - y = \left[\frac{\Sigma^{-1}uu'}{u'\Sigma^{-1}u}\right](x - y) + \left[I - \frac{\Sigma^{-1}uu'}{u'\Sigma^{-1}u}\right](x - y).$$

We prove Proposition 1 in the Appendix. This result shows that in the model the differentiation between any two options x and y admits a unique decomposition into vertical and horizontal differentiation components and, moreover, that these two components are always orthogonal according to the inner product that generates the differentiation norm.

To illustrate the vertical and horizontal differentiation decomposition, consider again options x and y in Figure 1. Option x has a higher value than option y and lies on a higher indifference curve. An orthogonal projection of y onto the indifference curve for x using the inner product given by Σ yields option $\hat{x}$. To see this, note that differentiation from y is represented by the gray ellipsoids centered around y. Option $\hat{x}$ lies on the ellipsoid that is exactly tangent to the indifference curve containing x. Hence it has the smallest differentiation among the options indifferent to x.

This, in turn, means that $\hat{x}$ and y in Figure 1 are vertically differentiated, that is, $\rho(\hat{x}, y)$ attains maximum choice probability. To see why, note that along the indifference curve containing x, every option has the same difference in utility compared to y, that is, the same numerator in equation (2). However, the options along the indifference curve containing x have varying levels of differentiation from y, corresponding to different denominators in (2). Since $\hat{x}$ yields the minimum denominator, it achieves the maximum choice probability against y. By Proposition 1 we have

$$\|x - y\| = \sqrt{\|x - \hat{x}\|^2 + \|\hat{x} - y\|^2}$$

where $\|\cdot\|$ denotes the inner product norm generated by Σ. Next, we show how vertical and horizontal differentiation can be uniquely identified from binary choice data.

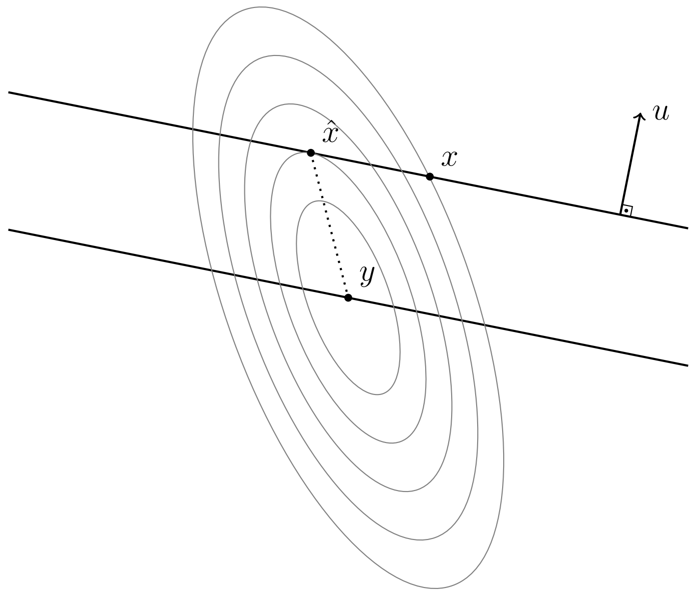

Figure 1: Linear differentiation model. Straight black lines represent indifference curves. Each ellipsoid centered on y represents options that are equally differentiated from y. Among the options indifferent to x, there is a unique option $\hat{x}$ that is closest to y. The difference $x - y$ admits a unique orthogonal decomposition: $\hat{x} - y$ maximizes agreement among consumers (pure vertical differentiation), while $x - \hat{x}$ minimizes agreement among consumers (pure horizontal differentiation).

3.2 Identification

The parameters (u, Σ, F) of the linear differentiation model are unique up to scaling by three positive constants:

Proposition 2 (Uniqueness)**.** *Let (u, Σ, F) represent linear differentiation model ρ with $u \neq 0$ and let $T = \sqrt{u'\Sigma^{-1}u}$. Then $(\hat{u}, \hat{\Sigma}, \hat{F})$ also represent ρ iff there exist $A, B, C > 0$ such that:*

(i) $\hat{u} = A\,u$;

(ii) $\hat{\Sigma} = B^2\,(\Sigma - uu'/T^2) + C^2\,(uu'/T^2)$;

(iii) $\hat{F}(t) = F\left(tB/\sqrt{A^2 + (B^2 - C^2)(t/T)^2}\right)$ *for all* $t \in [-AT/C, AT/C]$.

Proposition 2 determines the permissible transformations of the model parameters. Item (i) says utility is unique up to scaling by a constant $A > 0$. Item (ii) says the differentiation metric is unique up to scaling by two constants $B, C > 0$. The constant

$B > 0$ scales the horizontal differentiation component, while the constant $C > 0$ scales the vertical differentiation component. Item (iii) determines the only permissible transformations of F, providing an explicit formula for how to obtain $\hat{F}$ from F and the scaling factors $A, B, C > 0$. In particular, when $B = C$ we obtain a simple rescaling of the domain $\hat{F}(t) = F(tB/A)$.

One main takeaway from Proposition 2 is that all the parameters of the model are uniquely pinned down with three straightforward normalizations. First, we must normalize the scale of the utility vector by imposing, for example, the normalization $u_1^2 + \cdots + u_n^2 = 1$. Second, we must normalize the scale of the differentiation metric along the horizontal and vertical dimensions, for example, by imposing $u'\Sigma^{-1}u = 1$ and $\mathrm{Trace}(\Sigma) = 2$. Proposition 2 then yields that no further transformations are permissible, and (u, Σ, F) are uniquely pinned down.

A second important takeaway from Proposition 2 is that horizontal differentiation and vertical differentiation can be uniquely measured from binary comparison choice data, each one with their own scale. Separate scales for horizontal and vertical differentiation means that statements such as "x and y are more horizontally differentiated than w and z", or "x and y are more vertically differentiated than w and z" are meaningful. But to compare the total amount of differentiation across pairs of options, we need additional information that pins down the relative scales of horizontal and vertical differentiation.

To pin down these relative scales, we may use a *special regressor*. This is an observable attribute, such as price, that unambiguously affects the utility of the options but does not affect differentiation. If a special regressor is available, Proposition 2 implies the relative scale of horizontal and vertical differentiation is uniquely pinned down. We illustrate this with a classic example of automobile demand estimates in Section 5.1.

3.3 Characterization

We provide four straightforward postulates on a binary choice rule ρ which we use to fully characterize the linear differentiation model in Theorem 1. First, it is immediate from the formula in (1) that every linear differentiation model is *continuous* in the domain $\mathcal{D}$. Second, the assumptions that utility is a linear function and that differentiation is a norm imply the model is *linear*, that is, we have

$$\rho(x, y) = \rho(\alpha x + (1 - \alpha)z, \alpha y + (1 - \alpha)z)$$

for any options x, y, z and $0 < \alpha < 1$. Third, the assumption that the differentiation norm comes from an inner product implies the model is *balanced*: whenever $\rho(x, y) = 1/2$, $\rho(x, z) > \rho(y, z) > 1/2$, and $1/2 < \alpha < 1$, we have

$$\rho(\alpha x + (1 - \alpha)y, z) > \rho(\alpha y + (1 - \alpha)x, z).$$

Finally, the model is also *moderately transitive*: if $\rho(x, y) \geq 1/2$ and $\rho(y, z) \geq 1/2$, then we must have $\rho(x, z) > \min\{\rho(x, y), \rho(y, z)\}$ or else $\rho(x, z) = \rho(x, y) = \rho(y, z)$. These four postulates are not only necessary but also sufficient:

Theorem 1. *A binary choice rule ρ is a linear differentiation model if and only if ρ is continuous, linear, balanced and moderately transitive.*

We prove Theorem 1 in the Appendix. This characterization result shows that four straightforward properties exhaust the testable implications of the linear differentiation model. Continuity and linearity are familiar postulates from the random choice literature (e.g. Gul and Pesendorfer, 2006). Moderate transitivity is one of several possible transitivity postulates for binary random choice. While stronger and weaker transitivity postulates have been extensively used, moderate transitivity provides a useful compromise between flexibility and predictive power (He and Natenzon, 2023). Finally, the postulate that ρ is balanced is, to the best of our knowledge, novel in the random choice literature and therefore deserves some additional discussion and motivation.

To interpret the postulate, suppose x and y are two political candidates that divide the population of voters equally $\rho(x, y) = 1/2$ by having the same overall quality but very distant political platforms and proposals. Now suppose z is a candidate of lower quality than x and y but much closer to x in terms of political platform and proposals. Being of higher quality, both x and y would beat z in an runoff election. The contest between y and z involves significant tradeoffs in terms of platforms, making the election more divisive, and y beats z with a small margin. However, in a contest between x and z the platform tradeoffs are much smaller and x's higher quality makes the choice easier for many voters. Hence x beats z by a larger margin. In other words, $\rho(x, z) > \rho(y, z) > 1/2$.

The balanced postulate requires the inequality $\rho(x, z) > \rho(y, z)$ to be preserved when both x and y move their platforms towards their midpoint $x/2 + y/2$. In fact, for each mixture weight $0 < \beta < 1$ we have

$$\beta x + (1-\beta)\left(\frac{x}{2}+\frac{y}{2}\right) = \left(\frac{1}{2}+\frac{\beta}{2}\right)x + \left(\frac{1}{2}-\frac{\beta}{2}\right)y = \alpha x + (1-\alpha)y$$
$$\beta y + (1-\beta)\left(\frac{x}{2}+\frac{y}{2}\right) = \left(\frac{1}{2}+\frac{\beta}{2}\right)y + \left(\frac{1}{2}-\frac{\beta}{2}\right)x = \alpha y + (1-\alpha)x$$

where $\alpha = 1/2 + \beta/2 > 1/2$ for every β.

Voting is more decisive in favor of x against z than in favor of y against z. The postulate requires voting to remain more decisive when we substitute the mixture $\alpha x + (1-\alpha)y$ which gives more weight to x for x, and the mixture $\alpha y + (1-\alpha)x$ which gives more weight to y for y. This intuitive property holds because the norm in the linear differentiation model comes from an inner product. Figure 2 shows an example where relaxing the assumption of inner product norm in the linear differentiation model generates a binary choice rule ρ that fails to be balanced.

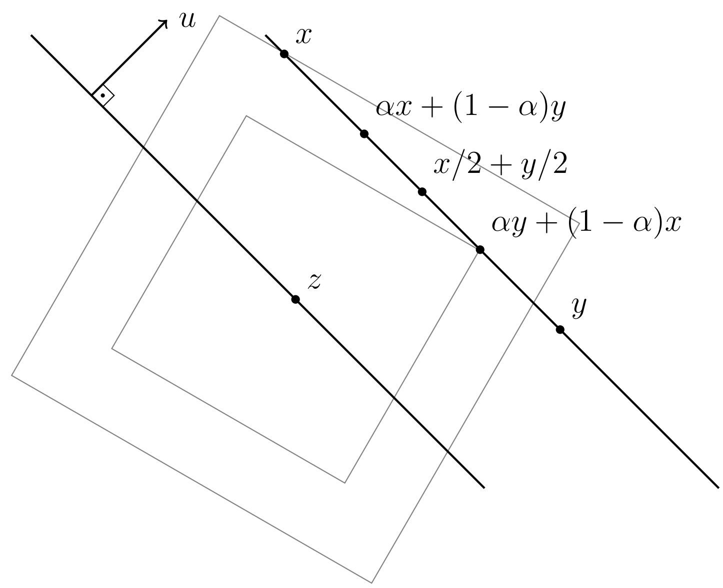

Figure 2: Relaxing the inner product norm assumption in the linear differentiation model allows binary choice rules which fail to be balanced. Each rectangle represents options that are equally differentiated from option z. The rectangular shape is compatible with a norm that cannot be generated from an inner product. In this norm, x and z are less differentiated than y and z, but the mixture $\alpha x + (1-\alpha)y$ and z are more differentiated than the opposite mixture $\alpha y + (1-\alpha)x$ and z.

Together with Propositions 1 and 2, Theorem 1 provides behavioral foundations for measuring vertical and horizontal differentiation using binary choice data. Binary choice data is commonly collected in lab experiments. However, many applications in economics involve choice data from menus with more than two options. In the next Section, we extend the analysis to multinomial choice.

4 Multinomial choice

A choice rule ρ is a *random coefficients model* when there exists an n-dimensional random vector β such that, for each menu A and each option $x \in A$,

$$\rho(x, A) = \mathbb{P}\left\{\beta' x \geq \beta' y \quad \forall y \in A\right\}. \tag{3}$$

Note that for (3) to hold, the random vector β must satisfy the following regularity condition:

$$\mathbb{P}\left\{\beta' x = \beta' y\right\} = 0, \text{ for all } x \neq y. \tag{4}$$

In other words, utility ties always have zero probability in a random coefficients model, for otherwise there would be some x, y with $\rho(x, y) + \rho(y, x) > 1$ and the represented ρ would fail to be a choice rule. This requirement rules out, for example, that the distribution of β has any atoms.

The random coefficients model is a workhorse in industrial organization (e.g. Hausman and Wise, 1978; Berry et al., 1995; Nevo, 2000), where the random vector β captures heterogeneity of tastes in a population of rational, utility maximizing consumers. In the setting of choice under risk, this formulation yields the random expected utility model (Gul and Pesendorfer, 2006). Our next Theorem characterizes the intersection of our linear differentiation model (1) with the random coefficients framework (3), and pins down the distribution of the random coefficients vector β.

A k-dimensional random vector ε has a *spherical distribution* when $\Gamma\beta$ and β have the same distribution for every orthogonal $k \times k$ matrix Γ. If a k-dimensional ε has a spherical distribution, u is a vector in $\mathbb{R}^n$ for some $n \geq k$, and Λ is an $n \times k$ full-rank matrix, then $u + \Lambda\varepsilon$ has an *elliptical distribution.* The distribution is *unbounded* when it has an unbounded support. The class of elliptical distributions with unbounded support contains the multivariate Gaussian, multivariate Cauchy, multivariate t, and multivariate exponential as special cases.

Definition. *A random coefficients model ρ in $\mathbb{R}^n$ is an* elliptical coefficients model *when the random coefficient vector $\beta = u + \Lambda\varepsilon$ is elliptical and the spherical vector ε is*

$(n-1)$*-dimensional and unbounded.*

The next result shows that if two random coefficient models represent the same choice rule ρ, then they must have the same distribution for β up to a simple positive scaling.

Proposition 3 (Uniqueness)**.** *The distribution of the random coefficients vector β is unique up to scaling in the elliptical coefficients model.*

Since the elliptical coefficients model is a random coefficients model (3), it inherits four postulates for a multinomial choice rule from the Gul and Pesendorfer (2006) random expected utility representation: ρ is continuous, linear, extreme and monotone. We now define and discuss each one.

The continuity postulate in Gul and Pesendorfer (2006) extends the continuity postulate from Theorem 1 from binary to multinomial choice rules. Note that ρ maps each finite menu of options in $\mathbb{R}^n$ to a probability measure on the Borel sigma-algebra in $\mathbb{R}^n$. We endow the set of all finite subsets of $\mathbb{R}^n$ with the topology induced by the Hausdorff metric, and we endow the set of probability measures on the Borel sigma-algebra with the topology of weak convergence. A multinomial choice rule ρ is *continuous* when this induced mapping is continuous.

A multinomial choice rule ρ is *linear* if for each $x \in A$, $y \in \mathbb{R}^n$ and $0 < \alpha < 1$,

$$\rho(x, A) = \rho(\alpha x + (1-\alpha)y, \{\alpha z + (1-\alpha)y : z \in A\}).$$

This definition is a straightforward extension of our linearity postulate from binary choice to multinomial choice.

A choice rule ρ is *monotone* if $\rho(x, A) \geq \rho(x, B)$ whenever $A \subseteq B$. This means the market share of a product x can only decrease when new competing products are introduced. This postulate ties together the choices from menus with different numbers of options, and therefore imposes no restrictions when only binary choices are available.

Finally, a choice rule ρ is *extreme* if $\rho(x, A) > 0$ implies x is an extreme point of A. It is easy to see that both options are extreme in any binary menu, hence this postulate only restricts behavior from menus with three options or more.

We provide a characterization of the model under the assumption that choice probability one is achieved for some pair of options. A choice rule ρ is *full* if $\rho(x, y) = 1$ or some x, y. Together with linearity, this postulate implies there exists a direction in the attribute space —such as a price decrease, for example— which all consumers agree is desirable.

Theorem 2. *A multinomial choice rule ρ is an elliptical coefficients model if and only if ρ is continuous, linear, monotone, extreme, balanced, moderately transitive, and full.*

We prove this result in the appendix. The four postulates from Gul and Pesendorfer (2006) yield their random coefficients representation for multinomial choice; only two of them (ρ is continuous and linear) put restrictions on binary choice. Adding two additional postulates on binary choice (ρ is balanced and moderately transitive) we can invoke Theorem 1 to yield the linear differentiation representation for binary choices. The main step in the proof uses this linear differentiation representation to show the distribution of the coefficients in the random coefficients representation must be elliptical.

The proof also reveals the mapping between the two parametric representations. Recall the linear differentiation representation has three parameters (u, Σ, F). By Proposition 2, it is a permissible transformation to rescale the vertical differentiation distance by a constant $C > 0$; a corresponding adjustment in the transformation F maintains the binary choice behavior the same. Taking the limit $C \to 0$ we obtain a new triple $(u, \hat{\Sigma}, \hat{F})$ where the new $\hat{\Sigma}$ now has rank $n-1$ and still represents an inner product norm along the horizontal differentiation dimensions, while giving the vertical differentiation dimension length zero. The new $\hat{F}$ obtained using the formula in Proposition 2 with $C \to 0$ is a strictly increasing cumulative distribution function with unbounded support on the real line. These parameters describe the distribution of the elliptical β coefficients: we have $\beta = u + \Lambda\varepsilon$ where $\hat{\Sigma} = \Lambda\Lambda'$ and $\hat{F}$ is the marginal cumulative distribution function that fully characterizes the spherical ε.

Theorem 2 shows the extension of the linear differentiation model to multinomial choice using the random coefficients framework yields random coefficients with an elliptical distribution. The result provides behavioral foundations for the measurement of vertical and horizontal differentiation in multinomial choice. In the next Section we provide a natural application of this result to model aggregate consumer demand.

5 Applications

Our first application is to aggregate consumer demand with differentiated products, using the population interpretation of ρ. Our second application uses the individual random choice interpretation of ρ to offer a measure of the complexity faced by an agent in binary decision-making problems.

Some applications have a natural restriction of the domain of choice options to strict subsets of $\mathbb{R}^n$. For example, prices are positive reals in consumer demand, and lotteries are restricted to a probability simplex in risky choice. We start with a lemma showing that our results apply to many kinds of restricted domains without any loss of generality.

Lemma 1. *If $K \subset \mathbb{R}^n$ has a non-empty interior and the choice rule ρ on K is linear, then ρ has a unique linear extension to all of $\mathbb{R}^n$.*

Proof. Let the choice rules ρ' and ρ'' be two linear extensions of ρ to $\mathbb{R}^n$, let A be a finite subset of $\mathbb{R}^n$ and let $x \in A$. We must show that $\rho'(x, A) = \rho''(x, A)$. Since K has a non-empty interior, there exists an option z in the interior of K. That is, there exists an open ball centered on z and contained in K. Hence, for $0 < \alpha < 1$ sufficiently small we have $\alpha A + (1 - \alpha)z$ contained in K. By linearity, we have

$$\begin{aligned} \rho'(x, A) &= \rho'(\alpha x + (1-\alpha)z, \alpha A + (1-\alpha)z) \\ &= \rho(\alpha x + (1-\alpha)z, \alpha A + (1-\alpha)z) \\ &= \rho''(\alpha x + (1-\alpha)z, \alpha A + (1-\alpha)z) \\ &= \rho''(x, A). \end{aligned}$$

□

5.1 Consumer demand

To analyze consumer demand data, we assume the analyst observes the price $p_x \geq 0$ for each option x, in addition to observing the vector of n product characteristics, and we write each option $x = (x_1, \dots, x_n, p_x)$ as an $n + 1$ dimensional vector in $\mathbb{R}^n \times \mathbb{R}_+$.

Consider a demand system described by a multinomial choice rule ρ satisfying the postulates of Theorem 2. In this application ρ is full as a consequence of assuming that every consumer prefers to pay less for the same product, that is,

$$p_x < p_y \text{ and } x_i = y_i \text{ for } i = 1, \dots, n-1 \text{ implies } \rho(x, y) = 1. \tag{5}$$

By Theorem 2, ρ admits an elliptical coefficients representation, in which the indirect utility of each option $x = (x_1, \dots, x_n, p_x)$ can be written as

$$V(x) = \sum_{i=1}^{n} \beta_i x_i + \beta_{n+1} p_x \tag{6}$$

where $\beta = u + \Lambda'\varepsilon$ is the elliptical vector of random coefficients, and ε is spherical with unbounded support and of dimension n.

The coefficient $-\beta_{n+1}$ captures the distribution of the marginal utility of income in the population. Assumption (5) implies $\beta_{n+1} < 0$ almost surely. Moreover, since ε has unbounded support we must have $\beta_{n+1} < 0$ is in fact constant, and the covariance matrix for random coefficients takes the form

$$\Lambda\Lambda' = \begin{bmatrix} M & 0 \\ 0 & 0 \end{bmatrix}$$

where M is a symmetric $n \times n$ positive definite matrix.

Even though the coefficient β_{n+1} is constant, the model still accommodates a heterogeneous marginal utility of income in the population. To see this, let the true distribution of marginal utility of income be given by any random variable γ with $\gamma > 0$ almost surely. Multiplying every β_i by $-\gamma/\beta_{n+1}$ we obtain an equivalent random coefficients model where $-\gamma$ is the random coefficient on price. Hence, under the postulates of Theorem 2, the heterogeneity of marginal utility of income is absorbed into the n elliptical random coefficients $\beta_1, \ldots, \beta_n$ without loss of generality.

Some multinomial choice models commonly used in empirical industrial organization (Hausman and Wise, 1978; Berry, Levinsohn and Pakes, 1995; Nevo, 2000) are special cases of this framework, fixing a particular elliptical distribution for the coefficients:

Example 1. *The random coefficients model of Hausman and Wise (1978) assumes a linear utility function $V(x) = \beta_1 x_1 + \cdots + \beta_n x_n$ where the vector of coefficients $\beta = (\beta_1, \ldots, \beta_n)$ is random with a joint Gaussian distribution $\beta \sim \mathcal{N}\left(\bar{\beta}, \Sigma\right)$. The probability that x is preferred to y in this model is given by*

$$\rho(x, y) = \mathbb{P}\{V(x) > V(y)\} = \Phi\left(\frac{\bar{\beta}'x - \bar{\beta}'y}{\sqrt{(x-y)'\Sigma(x-y)}}\right).$$

An important caveat, however, is that the full econometric specification used in the empirical literature often includes additional additive independent error terms to utility. Lu and Saito (2022) show these additional noise terms introduce systematic deviations from our postulates. Here, we ignore these additional error terms and focus on the main structural component of these demand estimation models, sometimes called the "pure characteristics" model (Berry and Pakes, 2007).

By Proposition 3, the parameters of Example 1 are unique up to scaling and therefore can be uniquely pinned down with one single normalization of scale. This can also be seen directly using the uniqueness result in Proposition 2. Since $F = \Phi$ is fixed in Example 1 to be the standard Gaussian cdf, item (iii) in Proposition 2 implies the only remaining permissible transformations must have three equal scaling parameters $A = B = C > 0$. Hence, a single normalization (for example setting the utility vector u to have unit norm) pins down all the remaining parameters using binary choice comparison data alone.

The estimation literature often assumes the random coefficients $\beta_1, \ldots, \beta_n$ are independent Gaussian variables (e.g. Berry, Levinsohn and Pakes (1995)), which corresponds to assuming a diagonal matrix Σ in Example 1. Our next result pins down the empirical content of this parametric restriction.

We say that a choice rule ρ is *factorable* if

$$\rho(x, \{x, y, z\}) = \rho(x, y) \cdot \rho(x, z)$$

whenever y differs from x only in the i-th attribute and price, and z differs from x only in the j-th attribute and price, with $i \neq j$.

Proposition 4. *An elliptical coefficients model ρ has independent Gaussian coefficients if and only if ρ is factorable.*

Figure 3 illustrates our decomposition of vertical and horizontal differentiation in real-world data using the automobile demand estimates in Berry, Levinsohn and Pakes (1995). The two plotted attributes are acceleration and size. Let β_1 and β_2 be the random coefficients that represent the marginal utility for acceleration and size, respectively. To plot the Figure we the estimated averages of these random coefficients $\bar{\beta}_1 = 2.833$ and $\bar{\beta}_2 = 3.460$; and their estimated standard deviations $\sigma_1 = 4.628$ and $\sigma_2 = 2.056$.

The black parallel lines in Figure 3 are estimated indifference curves, given by sets of points (x_1, x_2) for which $\bar{\beta}_1 x_1 + \bar{\beta}_2 x_2$ is constant. Two products placed on the same indifference curve (and with the same price) divide the market exactly in half in a duopoly. Likewise, changes in a product's characteristics along the same indifference curve reflect variety and maximize disagreement in the population: for each consumer that likes the change, there is another consumer that dislikes it.

The dashed line in Figure 3 is the estimated vertical differentiation dimension, which by Proposition 1 is unique and given by

$$\Sigma^{-1}\bar{\beta} = \begin{bmatrix} 1/\sigma_1 & 0 \\ 0 & 1/\sigma_2 \end{bmatrix} \cdot \begin{bmatrix} \bar{\beta}_1 \\ \bar{\beta}_2 \end{bmatrix} = \begin{bmatrix} \bar{\beta}_1/\sigma_1 \\ \bar{\beta}_2/\sigma_2 \end{bmatrix} \tag{7}$$

Each component of the vertical differentiation direction vector has an average marginal utility $\bar{\beta}_i$ divided by the variance of marginal utility σ_i in the population. A larger variance σ_i means the marginal utility of attribute i is more heterogeneous in the population. In particular, a larger σ_i increases the left tail of the distribution, so that a higher fraction of the population dislikes attribute i. Since vertical differentiation is by definition the direction of maximum agreement among consumers, a larger variance σ_i reduces the component of attribute i in (7). Conversely, a smaller variance σ_i means there is larger agreement on the desirability of attribute i and corresponds to a larger i-th component in vector (7).

The gray ellipsoids in Figure 3 represent iso-differentiation curves centered on the Honda Accord. The Ford Taurus is the most differentiated automobile from the Accord, lying on the largest ellipsoid, while the Chevrolet Cavalier is the least differentiated from the Accord, lying on the smallest ellipsoid.

Decomposing the differentiation between the products into horizontal and vertical components can provide further insight into the strategic product positioning by firms. For example, take the Ford Taurus, an automobile notorious for its size. In fact, a main point of emphasis in the 1989 "Ford Competition Today" sales training video (see Figure 4) is the comparison to the Accord along the size dimension: the video illustrates with colorful cube props that the Taurus offered 11 cubic feet of additional interior space and 3 cubic feet of additional trunk space. Is this additional space vertical or horizontal differentiation? A closer look at Figure 3 reveals the Taurus is the closest model to the Accord in vertical differentiation, and the large differentiation between the Taurus and the Accord is mostly horizontal.

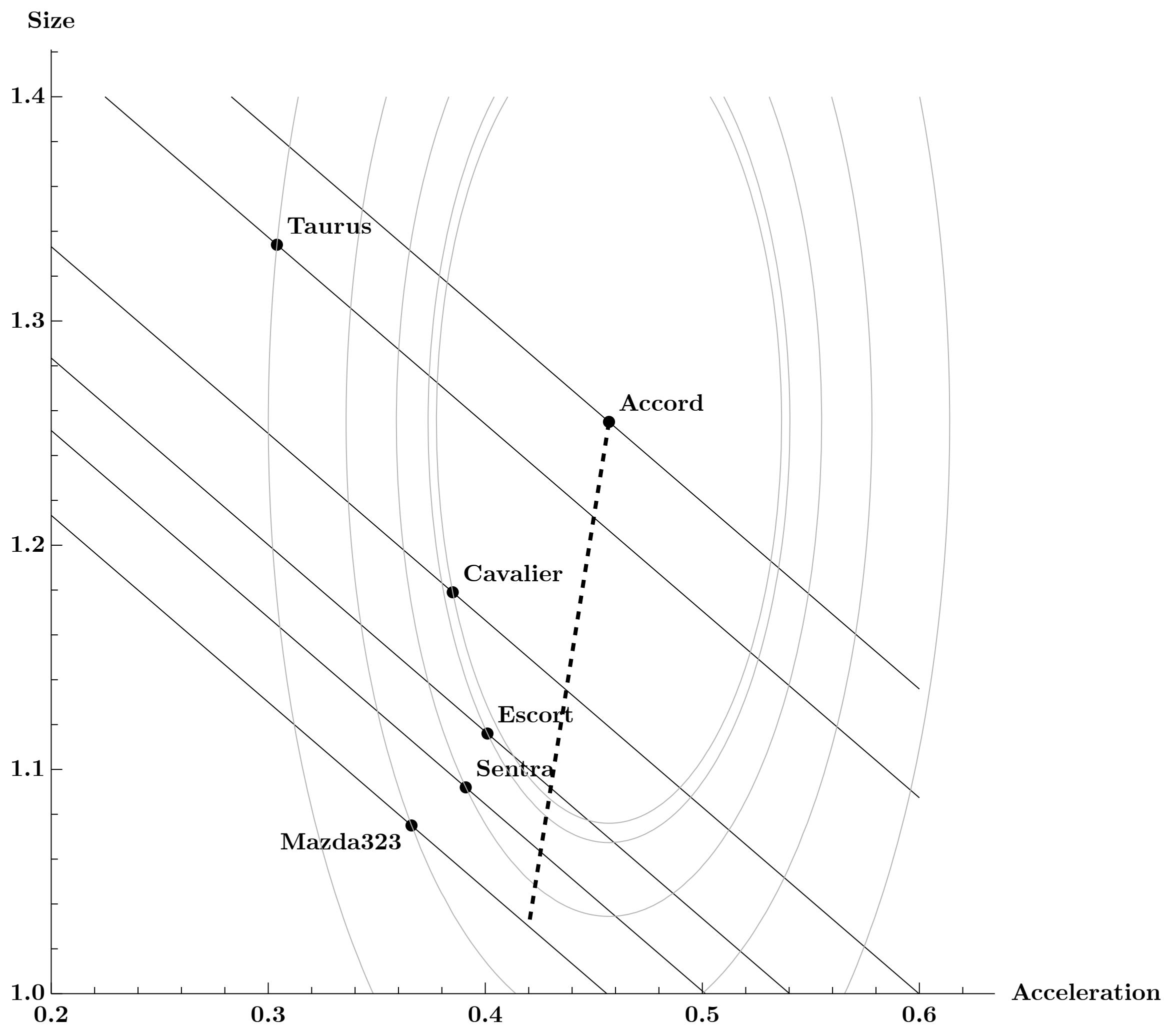

Figure 3: An illustration of estimated vertical and horizontal differentiation for six automobiles from 1990 using demand estimates in Berry et al. (1995). The horizontal axis represents a measure of acceleration and the vertical axis is size. Gray ellipses represent the iso-differentiation curves centered at the Honda Accord. Black lines are indifference curves. The dashed line is the unique vertical differentiation direction.

Figure 4: Screenshot from the sales training video "Ford Competition Today" comparing the Ford Taurus to the Honda Accord. *"Taurus has a full 11 cubic feet of passenger room volume more than Accord. Now that's this much room, room that'll really be appreciated by American buyers..."* (Ford, 1989).

5.2 Complexity in individual choice

A rapidly growing literature in economics introduces imprecision and noise into standard models of individual decision making (see e.g. Woodford (2020) for a review). A key concept from this literature is the psychometric function. This concept originates in experimental psychology and describes the relationship between the probability of correctly discriminating between two options x and y and the observable characteristics of the options. In this section, we apply our results to provide a novel, quantitatively precise answer to a fundamental question in this literature: *does the distance in characteristics between x and y make the task of discriminating between x and y easier or harder?*

The existing theoretical and empirical literature offers an ambiguous answer: the effect of increasing the distance between two objects can go either way. On the one hand, objects that are closer to each other will be closer in utility, and when two options are close to indifference it is harder to choose between them. On the other hand, objects that are more distant in the attribute space can have very different mixes of attributes and involve more complex tradeoffs, making the comparison more difficult. Therefore, the overall effect of distance remains unclear clear. However, our formulation (2) allows us to uniquely decompose the differentiation between x and y into vertical and horizontal components. We now show this decomposition fully resolves the ambiguity.

First, we fix a linear differentiation model representation (u, Σ, F). Using the formulas that we obtained in Proposition 1, we define $v(x, y)$ to be the signed vertical distance between x and y:

$$v(x, y) = \frac{u'(x-y)}{\sqrt{u'\Sigma^{-1}u}}$$

note that $v(x, y)$ can be positive or negative, and the vertical differentiation between x and y is given by its absolute value $|v(x, y)|$. Also note that $v(x, y)$ is directly proportional to the difference in utility between x and y. Using Proposition 2, we can rescale vertical differentiation, if necessary, so that without loss of generality $u'\Sigma^{-1}u = 1$ and therefore $v(x, y) = u'(x - y)$.

Next, define $h(x, y)$ to be the horizontal differentiation between x and y:

$$h(x, y) = \sqrt{(x-y)'\Sigma(x-y) - v(x, y)^2}.$$

With these definitions the linear differentiation model becomes simply

$$\rho(x, y) = F\left(\frac{v(x, y)}{\sqrt{v(x, y)^2 + h(x, y)^2}}\right)$$

by taking partial derivatives of this expression with respect to v and h, it is straightforward to check that *increasing vertical differentiation always makes discrimination easier, while increasing horizontal differentiation always makes discrimination harder.*

Figure 5 plots the probability of choosing x over y as a function of $v(x, y)$, for several fixed values of $h(x, y)$. For a fixed value of $h(x, y)$ this probability follows the typical sigmoid curve compatible with the classic Fechnerian utility model. Increasing the value of $h(x, y)$ results in a flatter sigmoid curve, meaning that discrimination between x and y is more imprecise for each utility difference $v(x, y)$. Hence, the horizontal differentiation distance $h(x, y)$ can be interpreted as a measure of the *complexity* involved in evaluating the tradeoffs between x and y. This measure of complexity is orthogonal to utility, and our results show that it can be uniquely measured from binary choice data alone.

For a concrete illustration, we now fit the linear differentiation model to the choice data in the classic experiment of Tversky and Russo (1969). We use data from their Table 2, which presents the choice frequencies for $n = 3744$ binary choice trials. Subjects compared pairs of rectangles projected on a screen, and were rewarded if they correctly identified the rectangle with the largest area in each pair.

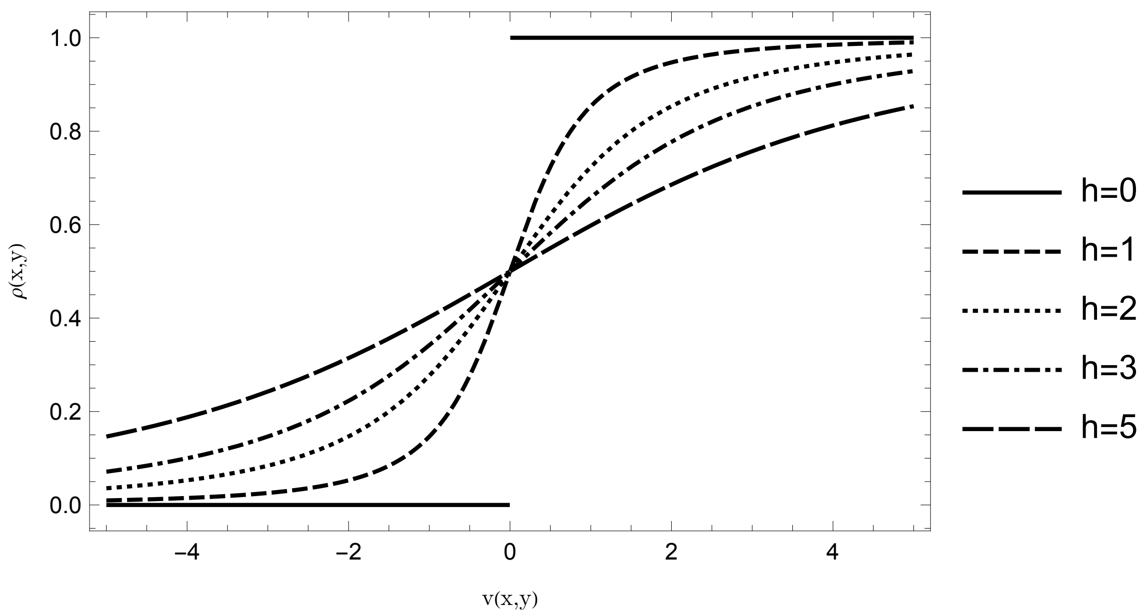

Figure 5: Probability of choosing x over y as a function of their difference in value. Each sigmoid curve is obtained by fixing a different value h for the amount of horizontal differentiation between x and y. Higher values of h are associated with more imprecise discrimination —flatter sigmoid curves— reflecting higher complexity in the comparison between x and y.

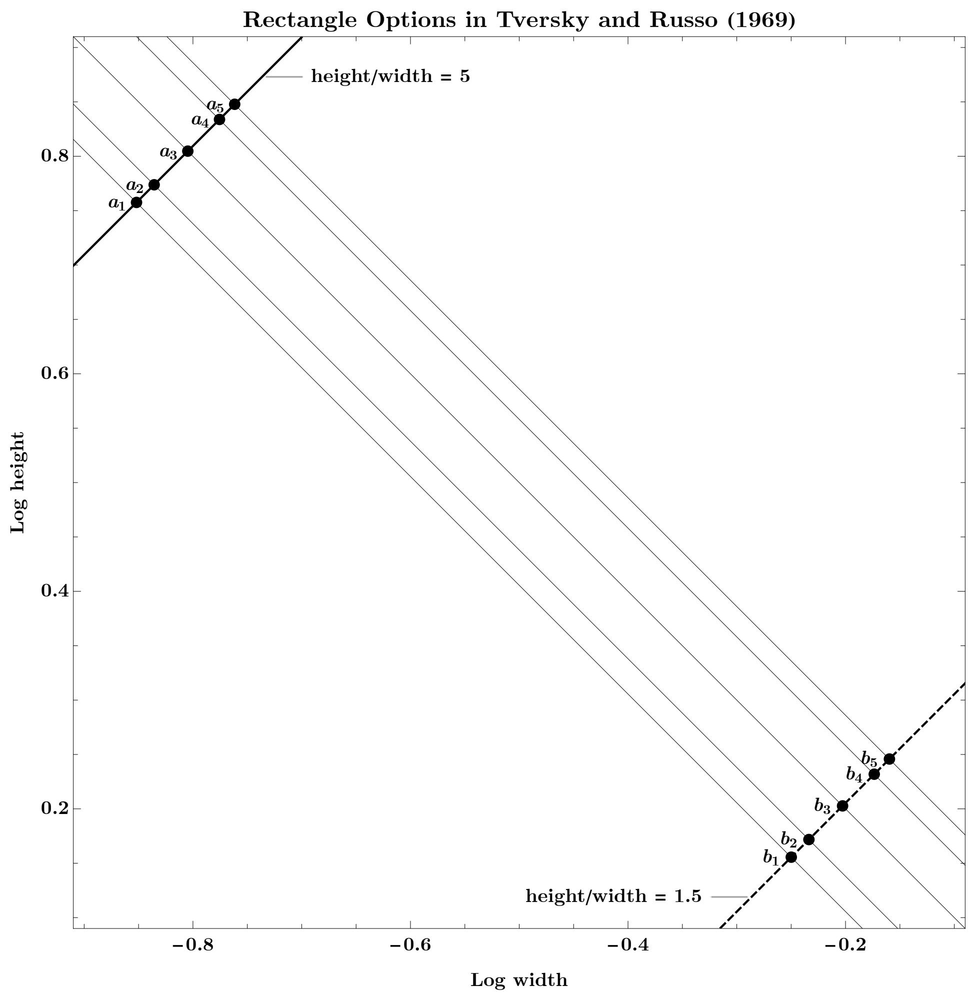

Figure 6: Log height and log width of rectangular objects shown to subjects in Tversky and Russo (1969). The height is five times the width in rectangles a_1 to a_5; whereas in rectangles b_1 to b_5 the height is just 1.5 times the width. The thin black lines represent indifference curves: the area of rectangle a_i is equal to the area of rectangle b_i for each $i = 1, \ldots, 5$. Using logs, the indifference curves are parallel straight lines, conforming to our linearity postulate.

Figure 6 shows the ten different rectangle options used for comparison, coming in five different sizes (areas), and two different height/width proportions. Rectangles a_3 and b_3 have the same area equal to 1. The data provided in Table 2 of Tversky and Russo (1969) describe the frequency with which the unit area rectangles a_3 and b_3 were chosen against each of the smaller a_1, a_2, b_1, b_2 and each of the larger a_4, a_5, b_4, b_5 rectangles.

Fitting the linear differentiation model to this type of experimental data provides a clean and straightforward illustration since the "correct" choice (the largest rectangle) is objectively observed. This type of experimental data is often used for illustration in the rational inattention literature (Dean and Neligh, 2023). We fit a simple specification where the subject maximizes a noisy signal of the true area of a rectangle x with height x_1 and width x_2:

$$U(x) = \beta_1 \log x_1 + \beta_2 \log x_2 + \beta_3 x_3$$

where we assume the random coefficients $\beta = u + \Lambda\varepsilon$ are elliptical with

$$u = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \text{ and } \Lambda = \begin{pmatrix} \sqrt{c_1^2 + c_2^2} & 0 & 0 \\ c_1 & c_2 & 0 \\ 0 & 0 & c_3 \end{pmatrix}$$

and only three free parameters $c_1, c_2, c_3 > 0$ need to be estimated. These assumptions amount to imposing that utility is given by the objectively given rectangle area and that for a given difference in area, the rate of mistakes is the minimized when rectangles have the same height/width ratio. The use of logs for height and width means indifference curves are straight parallel lines, conforming to the linearity postulate in Theorem 1. The coefficients β_1 and β_2 reflect the noisy perception of height and width, respectively.

The third attribute x_3 is an unobserved attribute that affects the differentiation between the options but not their value. In this context, we interpret x_3 to represent the location of the rectangles on the screen from left to right (e.g. $x_3 = 0$ when the rectangle is placed on the left of the screen and $x_3 = 1$ when the rectangle is placed on the right). For the purpose of our estimation, differences in the attribute x_3 are interpreted as any fixed but unobserved amount of differentiation between the two options. The coefficient β_3 is assumed to be zero-mean, uncorrelated with β_1, β_2. Since differences in x_3 are assumed to be constant across choice trials, β_3 effectively acts as an additional zero-mean error term.

To make things even more computationally simple, we assume the spherical distribution of ε is uniform on the surface of the three-dimensional unit ball. Then, it is know that the marginal distribution of each ε_i is just the uniform distribution on the interval $[-1,1]$ with a closed form cdf given by $F(t) = 1/2 + t/2$. By Lemma 11 in the proof of Theorem 2, this specification is a linear differentiation model (u, Σ, F) with $\Sigma = \Lambda\Lambda'$. It is also a random coefficients model. However, it is a departure from our elliptical random coefficients definition, since ε is full-dimensional and bounded.

The likelihood function is maximized at $(c_1, c_2, c_3) = (1.01744, 0.428679, 0.0530072)$, generating the choice probabilities given by the yellow and blue sigmoids in Figure 7. The blue sigmoid corresponds to comparisons with low complexity (low horizontal differentiation), such as a_3 versus a_1, a_2, a_4, a_5 and b_3 versus b_1, b_2, b_4, b_5.

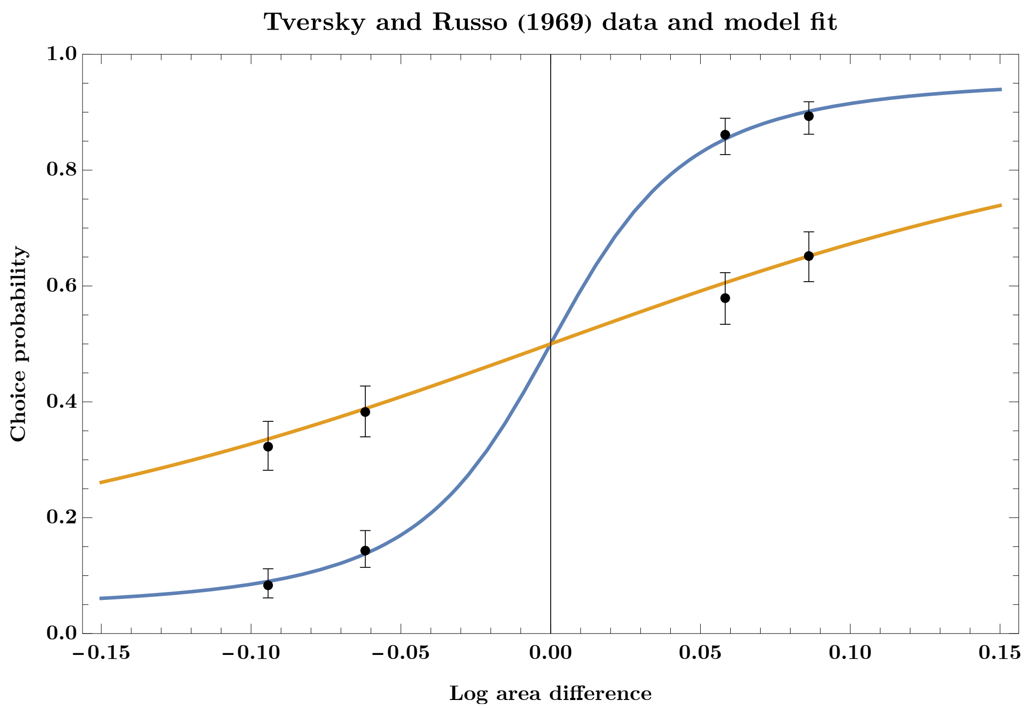

Figure 7: Black dots indicate the choice frequencies observed in the experiment (Tversky and Russo, 1969, Table 2), grouped by comparisons of rectangles with the same height/width ratio and comparsions with rectangles with different height/width ratio. The vertical bars denote 95% confidence intervals for the true choice probabilities. The two sigmoid curves depict our model's fitted choice probabilities for high complexity (yellow) and low complexity (blue) comparisons.

In terms of measurement, Proposition 2 implies that we cannot compare the units of vertical and horizontal differentiation. For that we need to observe a special regressor, such as price in the application in Section 5.1. We can, however, make quantitatively meaningful comparisons along the horizontal differentiation dimension. For example, the horizontal differentiation between a_1 and a_3 is estimated to be 0.053 while a_1 and b_3 have an estimated horizontal differentiation of 0.269. So the task of comparing a_1 versus a_3 is roughly *five times more complex* than the the task of comparing a_1 versus b_3, despite the difference in true value (area) being the same in both comparisons.

The fit of our model to the experimental data also helps illustrate our resolution of the fundamental psychometric question posed at the start of this subsection. Compare, for instance, the choice problem a_1, a_3 and the choice problem a_2, a_3. The first pair is more differentiated than the second (Figure 6); and discrimination is easier in the more differentiated pair (Figure 7). The opposite holds, however, when we compare choice problems b_1, a_3 and a_1, a_3. Figure 6 shows the first pair is more differentiated than the second pair; but Figure 7 this time shows discrimination is easier in the *less* differentiated pair. The ambiguous role of differentiation is resolved when we decompose it into its horizontal and vertical components: in the first case, the paired comparisons a_1, a_3 and a_2, a_3 differ solely in vertical differentiation, which, as we have shown, always helps discrimination. In contrast, horizontal differentiation always makes discrimination harder, and explains why choices from b_1, a_3 are noisier than the choices from a_1, a_3.

This simple empirical illustration allowed us to obtain a remarkably good fit to the experimental data despite imposing several restrictions on the parameters (u, Σ, F). First, we fixed the utility parameter u because the area of a rectangle is objectively observed; in more general economic applications, where utility is subjective and unobserved, u can be estimated to reveal the underlying preferences.

Second, we restricted Σ such that the vertical differentiation direction —that is, the direction of maximum choice probability and fewest mistakes— is a multiple of u. In more general applications, this direction can be freely estimated.

Finally, we fixed F to be the computationally convenient cdf of the uniform distribution in $[-1, 1]$. F can be more flexibly estimated to capture variations in the overall amount of noise. For instance, suppose we ran the same experiment with two different treatments: one treatment has higher incentives and leads subjects to exert more effort. If this treatment induces more accurate choice behavior, the difference between treatments can be captured by two different levels of variance in the spherical error vector ε. Crucially, both treatments can be accommodated by the same Σ. So our measure of complexity, encoded by Σ, is a fixed feature of the architecture of the choice environment. The underlying spherical error distribution can capture variations in effort, ability, time spent contemplating the options, and other factors that affect the overall amount of information acquired by subjects before making a choice.

This last point means our complexity measure could be useful to study endogenous information acquisition. For example, in the rational inattention literature (Sims, 2003), cost functions for obtaining information based on mutual information have been commonly assumed (Matejka and McKay, 2015). A known limitation of such cost functions is that they ignore physical features of the choice environment such as the similarity between different payoff-relevant states (Hebert and Woodford, 2021; Pomatto et al., 2023; Dean and Neligh, 2023). Our complexity measure could therefore be useful for incorporating such features in the analysis of optimal information acquisition.

6 Conclusions

We provided behavioral foundations for measuring product differentiation from both binary and multinomial choice data, assuming that choice options are described by a finite-dimensional vector of attributes. Our model enables not only the measurement of differentiation but also its unique decomposition into horizontal and vertical components, respectively reflecting the variety and quality dimensions of product offerings. While the axiomatic underpinnings of random choice models constitute our central contribution, we further showcase the practical applicability of our results by demonstrating their potential in two key areas: (i) providing a new lens to examine the product differentiation strategies employed by firms, and (ii) quantifying the complexity of tradeoffs faced by individuals in decision-making situations. We think these are promising areas for further research and the development of novel applications that draw on the strengths of our theoretical framework.

A Appendix: Proofs

Proof of Proposition 1

To prove (i), consider the maximization problem

$$\max_{x,y} \rho(x,y) = \max_{x,y} F\left(\frac{u'(x-y)}{\sqrt{(x-y)'\Sigma(x-y)}}\right).$$

Since F is increasing, the maximum is achieved by maximizing the ratio

$$\frac{u'(x-y)}{\sqrt{(x-y)'\Sigma(x-y)}} = \frac{u'\Sigma^{-1/2}\Sigma^{1/2}(x-y)}{\sqrt{(x-y)'\Sigma^{1/2}\Sigma^{1/2}(x-y)}} = (\Sigma^{-1/2}u)'\left(\frac{\Sigma^{1/2}(x-y)}{\|\Sigma^{1/2}(x-y)\|_e}\right)$$

where $\|\cdot\|_e$ denotes the standard Euclidean norm. The right-hand side is just $(\Sigma^{-1/2}u)'v$ subject to the constraint $\|v\|_e = 1$. By Cauchy-Schwarz this is (uniquely) maximized when v is in the same direction as $\Sigma^{-1/2}u$, which happens if and only if the difference $x-y$ is different from zero and a multiple of $\Sigma^{-1}u$.

To prove (ii), note that by the representation $\rho(x,y) = 1/2$ for $x \neq y$ if and only if $u'(x-y) = 0$.

To prove (iii), suppose x, y are (purely) vertically differentiated and w, z are (purely) horizontally differentiated. By (i) and (ii) we have $x - y = \alpha\Sigma^{-1}u$ for some $\alpha \neq 0$ and $u'(w-z) = 0$. This implies $(x-y)'\Sigma(w-z) = \alpha(\Sigma^{-1}u)'\Sigma(w-z) = \alpha u'\Sigma^{-1}\Sigma(w-z) = 0$.

Finally, (iv) follows from (i)–(iii) by projecting $(x-y)$ onto the subspace spanned by $\Sigma^{-1}u$ to obtain the vertical differentiation component and its orthogonal residual. □

Proof of Proposition 2

Sufficiency is straightforward. To prove necessity, suppose (u, Σ, F) and $(\hat{u}, \hat{\Sigma}, \hat{F})$ are two linear differentiation model representations for the same choice rule ρ with $u \neq 0$. It is easy to see that $\hat{u}_i \geq 0$ if and only if $u_i \geq 0$ because otherwise choice behavior would differ for two options that are different only along dimension i. Since $u \neq 0$ there is some $u_i \neq 0$ and therefore $\hat{u}_i \neq 0$ with the same sign. Let $A = \hat{u}_i/u_i$. Now consider any $j \neq i$ with $u_j, \hat{u}_j \neq 0$. If we have $A \neq \hat{u}_j/u_j$ then u and $\hat{u}$ would fail to represent the same stochastic indifference $\rho(x,y) = 1/2$ along the i,j subspace. Hence (i) holds.

To prove (ii), first note that for any vectors v and w we have

$$\frac{u'v}{\|v\|_\Sigma} \geq \frac{u'w}{\|w\|_\Sigma} \text{ if and only if } \frac{\hat{u}'v}{\|v\|_{\hat{\Sigma}}} \geq \frac{\hat{u}'w}{\|w\|_{\hat{\Sigma}}}$$

since both (u, Σ, F) and $(\hat{u}, \hat{\Sigma}, \hat{F})$ represent the same ρ. By (i) we also have

$$\frac{u'v}{\|v\|_\Sigma} \geq \frac{u'w}{\|w\|_\Sigma} \text{ if and only if } \frac{u'v}{\|v\|_{\hat{\Sigma}}} \geq \frac{u'w}{\|w\|_{\hat{\Sigma}}}$$

Restricting v, w to be vectors such that $u'v = u'w = u'\Sigma^{-1}u$ we obtain $\|w\|_\Sigma \geq \|v\|_\Sigma$ if and only if $\|w\|_{\hat{\Sigma}} \geq \|v\|_{\hat{\Sigma}}$. Now for any v with $u'v = u'\Sigma^{-1}u$ we have

$$\begin{aligned}\|v\|_{\Sigma}^2 &= \|v - \Sigma^{-1}u + \Sigma^{-1}u\|_{\Sigma}^2 \\ &= (v - \Sigma^{-1}u + \Sigma^{-1}u)'\Sigma(v - \Sigma^{-1}u + \Sigma^{-1}u) \\ &= \|v - \Sigma^{-1}u\|_{\Sigma}^2 + \|\Sigma^{-1}u\|_{\Sigma}^2\end{aligned}$$

where the cross terms are zero by the assumption $u'v = u'\Sigma^{-1}u$.

By (i) above and Proposition 1–(i) we have $\Sigma^{-1}u = B^2\hat{\Sigma}^{-1}u$ for some $B > 0$. Hence for every v with $u'v = u'\Sigma^{-1}u$ we also have $\|v\|_{\hat{\Sigma}}^2 = \|v - \Sigma^{-1}u\|_{\hat{\Sigma}}^2 + \|\Sigma^{-1}u\|_{\hat{\Sigma}}^2$. This and the last display equality imply that for every v, w with $u'v = u'w = 0$,

$$\|v\|_{\Sigma} \geq \|w\|_{\Sigma} \text{ if and only if } \|v\|_{\hat{\Sigma}} \geq \|w\|_{\hat{\Sigma}}.$$

Now for each $v \in \mathbb{R}^n$ we have ,

$$u'\left(v - \frac{\Sigma^{-1}uu'}{u'\Sigma^{-1}u}v\right) = u'\left(v - \frac{\hat{\Sigma}^{-1}uu'}{u'\hat{\Sigma}^{-1}u}v\right) = 0$$

and

$$\left\|v - \frac{\Sigma^{-1}uu'}{u'\Sigma^{-1}u}v\right\|_{\Sigma}^2 = v'\Sigma v - \frac{v'uu'v}{u'\Sigma^{-1}u} = \|v\|^2_{\Sigma - \frac{uu'}{u'\Sigma^{-1}u}}$$

while

$$\left\|v - \frac{\hat{\Sigma}^{-1}uu'}{u'\hat{\Sigma}^{-1}u}v\right\|_{\hat{\Sigma}}^2 = v'\hat{\Sigma}v - \frac{v'uu'v}{u'\hat{\Sigma}^{-1}u} = v'\hat{\Sigma}v - B^2\frac{v'uu'v}{u'\Sigma^{-1}u} = \|v\|^2_{\hat{\Sigma} - B^2\frac{uu'}{u'\Sigma^{-1}u}}$$

Therefore for any $v, w \in \mathbb{R}^2$ if

$$\|v\|_{\Sigma - \frac{uu'}{u'\Sigma^{-1}u}} = C^2\|v\|_{\hat{\Sigma} - B^2\frac{uu'}{u'\Sigma^{-1}u}}$$

for some $C > 0$, then we must also have

$$\|w\|_{\Sigma - \frac{uu'}{u'\Sigma^{-1}u}} = C^2\|w\|_{\hat{\Sigma} - B\frac{uu'}{u'\Sigma^{-1}u}}$$

which yields (ii) as desired.

To prove (iii), for each x with $u'x = u'\Sigma^{-1}u$, define $t(x) = (\hat{u}'x)/\|x\|_{\hat{\Sigma}} > 0$. Proposition 1 and items (i) and (ii) above imply for each such x

$$\|x\|_{\hat{\Sigma}}^2 = \|x - \Sigma^{-1}u\|_{\hat{\Sigma}}^2 + \|\Sigma^{-1}u\|_{\hat{\Sigma}}^2 = B^2\|x - \Sigma^{-1}u\|_{\Sigma}^2 + C^2\|\Sigma^{-1}u\|_{\Sigma}^2.$$

Let $T := F^{-1}(\max_{x,y} \rho(x,y)) = [u'\Sigma^{-1}u]/\|\Sigma^{-1}u\|_{\Sigma}$. Substituting and rearranging we obtain for each x with $u'x = u'\Sigma^{-1}u$,

$$\frac{\|x - \Sigma^{-1}u\|_{\Sigma}^2}{\|\Sigma^{-1}u\|_{\Sigma}^2} = \frac{(AT)^2 - C^2t(x)^2}{B^2t(x)^2}.$$

By linearity, for each $0 < t \leq AT/C$ we have $t = t(x)$ for some x with $\hat{u}'x = \hat{u}'\Sigma^{-1}u$, thus

$$\begin{aligned}
\hat{F}(t) = \hat{F}(t(x)) &= \rho(x,0) \\
&= F\left(\frac{u'\Sigma^{-1}u - u'0}{\sqrt{\|x - \Sigma^{-1}u\|_{\Sigma}^2 + \|\Sigma^{-1}u - 0\|_{\Sigma}^2}}\right) \\
&= F\left(T/\sqrt{\frac{\|x - \Sigma^{-1}u\|_{\Sigma}^2}{\|\Sigma^{-1} - 0\|_{\Sigma}^2} + 1}\right) \\
&= F\left(T/\sqrt{\frac{(AT)^2 - C^2t(x)^2}{B^2t(x)^2} + 1}\right) \\
&= F\left(\frac{Bt}{\sqrt{A^2 + (B^2 - C^2)(t/T)^2}}\right).
\end{aligned}$$

and the results follows since $\hat{F}(t) = 1 - \hat{F}(-t)$ for all t. □

Proof of Theorem 1

To show necessity, let $U : \mathbb{R}^n \to \mathbb{R}$ be linear, let $\|\cdot\|$ be a norm generated by an inner product $\|x\| = \sqrt{\langle x, x\rangle}$, and let F be a strictly increasing, continuous transformation such that the representation (1) holds.

First, ρ must be continuous outside the diagonal since (i) U is linear; (ii) $\|\cdot\|$ is a norm hence $\|x - y\| > 0$ for $x \neq y$; and (iii) F is continuous.

Second, ρ is linear since

$$\begin{aligned}\rho(\alpha x+(1-\alpha)z,\alpha y+(1-\alpha)z) &= F\left(\frac{U(\alpha x+(1-\alpha)z)-U(\alpha y+(1-\alpha)z)}{\|\alpha x+(1-\alpha)z-[\alpha y+(1-\alpha)z]\|}\right)\\ &= F\left(\frac{\alpha[U(x)-U(y)]}{\alpha\|x-y\|}\right)\\ &= \rho(x,y)\end{aligned}$$

whenever $0<\alpha<1$ and $x\neq y$, and the equality holds trivially when $x=y$.

To see ρ is balanced, suppose $\rho(x,y)=1/2$ and $\rho(x,z)>\rho(y,z)>1/2$, and let $1>\alpha>1/2$. By (1) we have $U(x)=U(y)>U(z)$ and $\|x-z\|<\|y-z\|$. Then,

$$\begin{aligned}\|\alpha x+(1-\alpha)y-z\|^2 &= \alpha^2\|x-z\|^2+2\alpha(1-\alpha)\langle x-z,y-z\rangle+(1-\alpha)^2\|y-z\|^2\\ &< \alpha^2\|y-z\|^2+2\alpha(1-\alpha)\langle x-z,y-z\rangle+(1-\alpha)^2\|x-z\|^2\\ &= \|\alpha y+(1-\alpha)x-z\|^2\end{aligned}$$

and by (1) we have $\rho(\alpha x+(1-\alpha)y,z)>\rho(\alpha y+(1-\alpha)x,z)$ as desired.

Finally, the proof that ρ is moderately transitive follows from Theorem 1 in He and Natenzon (2023), since restricted to any three options x,y,z the linear differentiation model is a special case of the abstract moderate utility model.

To show sufficiency, let the binary choice rule ρ on $\mathbb{R}^n$ be linear, continuous, balanced, and moderately transitive. The result is trivial for constant ρ, so we now consider the case in which ρ is not constant. For the remainder of the proof, we choose and fix two options $\bar{x},\bar{y}$ with $\rho(\bar{x},\bar{y})>1/2$.

Define the relation $\succcurlyeq$ by $x\succcurlyeq y$ if and only if $\rho(x,y)\geq 1/2$. Since ρ is moderately transitive, this $\succcurlyeq$ is complete and transitive. Since ρ is linear and continuous, $\succcurlyeq$ satisfies all the vNM axioms and admits an expected utility representation. Let $U:\mathbb{R}^n\to\mathbb{R}$ be a linear function representing $\succcurlyeq$.

For each lottery x, let $I(x):=\{y\in\mathbb{R}^n:\rho(x,y)=1/2\}$ denote the set of lotteries that are stochastically indifferent to x. Note that $I(x)$ is an affine subspace of dimension $n-1$. By linearity, ρ is entirely determined by the values of the mapping $x\mapsto\rho(x,\bar{y})$ for $x\in I(\bar{x})$, where $\bar{x},\bar{y}$ are the two lotteries with $\rho(\bar{x},\bar{y})>1/2$ that we fixed above. For each $1/2<p\leq 1$, define the upper contour sets $B(p):=\{x\in I(\bar{x}):\rho(x,\bar{y})\geq p\}$.

Lemma 2. *$B(p)$ is convex for all $1/2<p\leq 1$.*

Proof. Let $x,x'\in B(p)$ and let $0<\alpha<1$. Since $I(\bar{x})$ is an affine subspace, $\alpha x+(1-$

$\alpha)x' \in I(\bar{x})$. Linearity implies $\rho(\alpha x + (1-\alpha)x', \alpha\bar{y} + (1-\alpha)x') = \rho(x, \bar{y}) \geq p$. Linearity also implies $\rho(\alpha\bar{y} + (1-\alpha)x', \bar{y}) = \rho(x', \bar{y}) \geq p$. Then, moderate transitivity implies $\rho(\alpha x + (1-\alpha)x', \bar{y}) \geq p$. $\square$

Lemma 3. *$B(p)$ is compact for all $1/2 < p \leq 1$.*

Proof. $B(p)$ is closed by continuity. Let $|\cdot|$ denote the standard Euclidean metric, not necessarily equal to the metric we are going to construct for the representation. If $B(p)$ were not bounded, there would exist a sequence $x(k)$ in $B(p)$ with $|x(k) - \bar{y}| \geq k$ for all $k \in \mathbb{N}$. For each k, by linearity $\rho(\bar{y} + (x(k) - \bar{y})/|x(k) - \bar{y}|, \bar{y}) = \rho(x(k), \bar{y}) \geq p$. By Bolzano-Weierstrass the sequence $\bar{y} + (x(k) - \bar{y})/|x(k) - \bar{y}|$ would have a subsequence converging to some $z \neq \bar{y}$. By the linearity of U we would have $U(z) = U(\bar{y})$ and $\rho(z, \bar{y}) = 1/2$, contradicting continuity. Hence $B(p)$ must also be bounded. $\square$

Lemma 4. *The mapping $x \mapsto \rho(x, \bar{y})$ has a unique maximizer $\hat{x}$ on $I(\bar{x})$.*

Proof. Since $\rho(\bar{x}, \bar{y}) > 1/2$ we have $B(p) \neq \emptyset$ for some $p > 1/2$. Since ρ is continuous, the mapping $x \mapsto \rho(x, \bar{y})$ is continuous on $I(\bar{x})$. $B(p)$ is compact by Lemma 3, hence the maximum $\rho(\hat{x}, \bar{y}) = \bar{p}$ is attained at some $\hat{x} \in B(p)$. Hence $B(\bar{p})$ is not empty, and by the previous lemmas it is compact and convex. Since ρ is balanced, $B(\bar{p})$ must be a singleton. Otherwise, by Lemmas 2 and 3 there would exist a nontrivial segment $[\hat{x}, \hat{x}']$ contained in $B(\bar{p})$ with $\hat{x}'$ on the boundary of $B(\bar{p})$, so that the point $x'' = \hat{x}' + (1/2)(\hat{x}' - \hat{x})$ lies outside $B(\bar{p})$, that is $\rho(x'', \bar{y}) < \bar{p} = \rho(\hat{x}, \bar{y})$. But then the point $(2/3)\hat{x} + (1/3)x'' = (1/2)\hat{x} + (1/2)\hat{x}' \in B(\bar{p})$ and the point $(1/3)\hat{x} + (2/3)x'' = \hat{x}' \in B(\bar{p})$, with $\rho((1/3)\hat{x} + (2/3)x'', \bar{y}) = \bar{p} = \rho((2/3)\hat{x} + (1/3)x'', \bar{y})$ contradicting that ρ is balanced. $\square$

For the rest of the proof, we fix $\hat{x}$ to be the unique maximizer of $x \mapsto \rho(x, \bar{y})$ on $I(\bar{x})$.

Lemma 5. *$x \in I(\bar{x})$ and $\rho(x, \bar{y}) = p$ implies $\rho(2\hat{x} - x, \bar{y}) = p$.*

Proof. The statement trivially holds if $x = \hat{x}$, so suppose $x \neq \hat{x}$. First note $2\hat{x} - x = \hat{x} + (\hat{x} - x) \in I(\bar{x})$. If $\rho(\hat{x} + (\hat{x} - x), \bar{y}) < p$, by continuity there is a sufficiently small $\varepsilon > 0$ such that $\rho(\hat{x} + (1-\varepsilon)(\hat{x} - x), \bar{y}) < p$. Taking $\alpha = 1/(2-\varepsilon)$ we have $\hat{x} = \alpha(\hat{x} + (1-\varepsilon)(\hat{x} - x)) + (1-\alpha)x$. By Lemma 4 we have $\rho(\alpha(\hat{x} + (1-\varepsilon)(\hat{x} - x)) + (1-\alpha)x, \bar{y}) > \rho((1-\alpha)(\hat{x} + (1-\varepsilon)(\hat{x} - x)) + \alpha x, \bar{y})$ contradicting that ρ is balanced. Hence $\rho(2\hat{x} - x, \bar{y}) \geq p$. An entirely analogous argument shows that $\rho(2\hat{x} - x, \bar{y}) \leq p$. $\square$

Recall that $\hat{x}$ is the unique maximizer $\rho(\hat{x}, \bar{y}) = \bar{p}$ on $I(\bar{x})$. Let $B = B(p) - \hat{x}$ for some fixed $p \in (1/2, \bar{p})$. We first define an auxiliary norm $\|\cdot\|_B$ on the $n-1$ dimensional subspace $I(\bar{x}) - \hat{x}$ using B as the unit ball.

Lemma 6. $\|x\|_B := \inf\{\lambda \geq 0 : x \in \lambda B\}$ *is a norm on* $I(\bar{x}) - \hat{x}$.

Proof. The *Minkowski functional* $\|\cdot\|_B$ defined above is a norm when B is a symmetric, convex set such that each line through zero meets B in a non-trivial, closed, bounded segment (Thompson, 1996). By definition $\|x\|_B \geq 0$ for all x. Moreover, if $\|x\|_B = 0$ then $x \in \lambda B$ for all $\lambda > 0$ and therefore $x = 0$. Now for each $\alpha \geq 0$ we have $x \in \lambda B$ if and only if $\alpha x \in \alpha\lambda B$ and therefore $\alpha\|x\|_B = \|\alpha x\|_B$. Lemma 5 implies $x \in \lambda B$ if and only if $-x \in \lambda B$ and therefore $\|x\|_B = \|-x\|_B$. To verify the triangle inequality, note that B is closed by Lemma 3, and therefore $x/\|x\|_B \in B$ for all $x \neq 0$. B is also convex by Lemma 2, and therefore

$$\frac{x + x'}{\|x\|_B + \|x'\|_B} = \left(\frac{\|x\|_B}{\|x\|_B + \|x'\|_B}\right)\frac{x}{\|x\|_B} + \left(\frac{\|x'\|_B}{\|x\|_B + \|x'\|_B}\right)\frac{x'}{\|x'\|_B} \in B.$$

Thus,

$$\left\|\frac{x + x'}{\|x\|_B + \|x'\|_B}\right\|_B \leq 1$$

and the triangle inequality $\|x + x'\|_B \leq \|x\|_B + \|x'\|_B$ holds. ◻

Lemma 7. *If* $\bar{p} \geq p \geq q > 1/2$ *then* $B(p) = \hat{x} + \lambda\left[B(q) - \hat{x}\right]$ *for some* $0 \leq \lambda \leq 1$.

Proof. Moderate transitivity implies that, for any $x \neq \hat{x}$ in $B(p)$, the function $t \mapsto \rho(t\hat{x} + (1-t)x, \bar{y})$ is strictly increasing for $0 \leq t \leq 1$. It suffices to show that if $\rho(x^1, \bar{y}) = \rho(x^2, \bar{y})$ for $x^1, x^2 \in I(\bar{x})$ and $0 < \alpha < 1$, then $\rho(\alpha x^1 + (1-\alpha)\hat{x}, \bar{y}) = \rho(\alpha x^2 + (1-\alpha)\hat{x}, \bar{y})$. To see that equality must hold, suppose instead that $\rho(\alpha x^1 + (1-\alpha)\hat{x}, \bar{y}) < \rho(\alpha x^2 + (1-\alpha)\hat{x}, \bar{y})$. Continuity implies $\rho(\beta x^2 + (1-\beta)\hat{x}, \bar{y}) = \rho(\alpha x^1 + (1-\alpha)\hat{x}, \bar{y})$ for some $0 < \alpha < \beta < 1$. Letting

$$\begin{aligned}
z^1 &= x^1 + \frac{\beta(1-\alpha)}{\beta - \alpha}(x^2 - x^1) \\
z^2 &= x^1 + x^2 - z^1 \\
z^3 &= 2\hat{x} - z^1 \\
z^4 &= \alpha x^1 + \beta x^2 + (2 - \alpha - \beta)\hat{x} - z^1
\end{aligned}$$

we have that the line segment $[z^1, z^2]$ contains the line segment $[x^1, x^2]$; the line segment $[z^1, z^4]$ contains the line segment $[\alpha x^1 + (1-\alpha)\hat{x}, \beta x^2 + (1-\beta)\hat{x}]$ and

$$\begin{aligned}
&z^1/2 + z^2/2 = x^1/2 + x^2/2 \\
&z^1/2 + z^3/2 = \hat{x} \\
&z^1/2 + z^4/2 = (\beta x^2 + (1-\beta)\hat{x})/2 + (\alpha x^1 + (1-\alpha)\hat{x})/2
\end{aligned}$$

By Lemma 5 we have $\rho(z^3, \bar{y}) = \rho(z^1, \bar{y})$. We must also have $\rho(z^2, \bar{y}) = \rho(z^1, \bar{y})$, for otherwise $\rho(z^2, \bar{y}) \neq \rho(z^1, \bar{y})$ and $\rho(x^1, \bar{y}) = \rho(x^2, \bar{y})$ would contradict that ρ is balanced. And again since ρ is balanced we have $\rho(z^4, \bar{y}) = \rho(z^1, \bar{y})$. Now since $0 < \alpha < \beta < 1$ we have

$$0 < \frac{\alpha\beta(2-\alpha-\beta)}{\alpha(1-\alpha)+\beta(1-\beta)} < 1 < \frac{\alpha(1-\beta)+\beta(1-\alpha)}{\alpha(1-\alpha)+\beta(1-\beta)}.$$

Letting

$$z^5 = \left(\frac{\alpha\beta(2-\alpha-\beta)}{\alpha(1-\alpha)+\beta(1-\beta)}\right) z^2 + \left(1 - \frac{\alpha\beta(2-\alpha-\beta)}{\alpha(1-\alpha)+\beta(1-\beta)}\right) z^3$$

we have z^5 belongs to the segment $[z^2, z^3]$ and by Lemma 2 it must be $\rho(z^5, \bar{y}) \geq \rho(z^4, \bar{y})$. On the other hand, it is straightforward to verify the equality

$$z^5 - \hat{x} = \left[\frac{\alpha(1-\beta)+\beta(1-\alpha)}{\alpha(1-\alpha)+\beta(1-\beta)}\right](z^4 - \hat{x})$$

so z^4 lies in the interior of the segment $[z^5, \hat{x}]$. But then the mapping $t \mapsto \rho(t\hat{x} + (1-t)z^5, \bar{y})$ is not strictly increasing for $0 \leq t \leq 1$, contradicting moderate transitivity. □

Lemma 8. *$\|\cdot\|_B$ is Euclidean, i.e., $\|x\|_B = \sqrt{\langle x, x\rangle_B}$ where $\langle \cdot, \cdot \rangle_B$ is an inner product.*

Proof. We use a characterization of inner product spaces by Gurari and Sozonov (1970), who showed that a normed linear space is an inner product space if and only if

$$\|x\| = \|y\| = 1 \text{ and } 0 \leq \alpha \leq 1 \quad \text{imply} \quad \left\|\frac{1}{2}x + \frac{1}{2}y\right\| \leq \|\alpha x + (1-\alpha)y\|. \tag{8}$$

If $\|x\|_B = \|y\|_B = 1$ then x, y are on the boundary of B, hence $\rho(x+\hat{x}, \bar{y}) = \rho(y + \hat{x}, \bar{y}) = p > 1/2$ and $\rho(x+\hat{x}, y+\hat{x}) = 1/2$. A violation of condition (8) would entail $\|(1/2)x + (1/2)y\|_B > \|\alpha x + (1-\alpha)y\|_B$ where, without loss of generality $1/2 < \alpha < 1$.

Let

$$
\begin{aligned}
x^\alpha &= \alpha(x+\hat{x}) + (1-\alpha)(y+\hat{x}) \\
x^{1/2} &= (1/2)(x+\hat{x}) + (1/2)(y+\hat{x}) \\
x' &= (\alpha - 1/2)(x+\hat{x}) + (3/2-\alpha)(y+\hat{x}) \\
x'' &= (y+\hat{x}) + (1/2)(y-x)
\end{aligned}
$$

By Lemma 7 the sets $B\left(\rho(x^\alpha,\bar{y})\right) - \hat{x}$ and $B\left(\rho(x^{1/2},\bar{y})\right) - \hat{x}$ are dilations of $B(p)-\hat{x}$, hence $\rho(x^\alpha,\bar{y}) > \rho(x^{1/2},\bar{y})$. By construction, the segment $[x^\alpha, x'']$ contains the segment $[x', y+\hat{x}]$ and both have the midpoint $(\alpha/2-1/4)(x+\hat{x}) + (5/4-\alpha/2)(y+\hat{x})$. Likewise, the segment $[x+\hat{x}, x']$ contains the segment $[x^\alpha, x^{1/2}]$ and both have the midpoint $(1/4+\alpha/2)(x+\hat{x})+(3/4-\alpha/2)(y+\hat{x})$. By Lemma 2, $\rho(x',\bar{y}) \geq \rho(y+\hat{x},\bar{y})$. This last inequality must in fact be strict, for otherwise since ρ is balanced we would have $\rho(x'',\bar{y}) = \rho(x^\alpha,\bar{y}) > \rho(y+\hat{x},\bar{y})$ contradicting Lemma 2. Thus $\rho(x',\bar{y}) > \rho(y+\hat{x},\bar{y}) = \rho(x+\hat{x},\bar{y})$. But then $\rho(x',\bar{y}) > \rho(x+\hat{x},\bar{y})$ and ρ balanced would imply $\rho(x^{1/2},\bar{y}) > \rho(x^\alpha,\bar{y})$, a contradiction. Hence $\|\cdot\|_B$ satisfies (8). $\square$

We extend the inner product $\langle\cdot,\cdot\rangle_B$ on the $n-1$ dimensional subspace $I(\bar{x})-\hat{x}$ obtained in the last Lemma to an inner product $\langle\cdot,\cdot\rangle$ in $\mathbb{R}^n$. Let $v_1,\ldots,v_{n-1}$ be an orthonormal base for the subspace $I(\bar{x})-\hat{x}$ endowed with $\langle\cdot,\cdot\rangle_B$. Let $v_n := \hat{x}-\bar{y}$ and for every $1 \leq i,j \leq n-1$ let $\langle v_i, v_j\rangle = 0$ if $i \neq j$ and $\langle v_i, v_j\rangle = 1$ if $i=j$. We let the norm be induced by this inner product $\|x\| := \sqrt{\langle x,x\rangle}$ for all $x \in \mathbb{R}^n$. Next, we show that this inner product, together with the linear utility U obtained above provide an ordinal representation for ρ.

Lemma 9. *U and $\|\cdot\|$ provide an ordinal representation of ρ, that is, for any $w \neq x$ and $y \neq z$ we have*

$$
\rho(w,x) \geq \rho(y,z) \iff \frac{U(w)-U(x)}{\|w-x\|} \geq \frac{U(y)-U(z)}{\|y-z\|}.
$$

Proof. First, suppose $\rho(w,x) \geq \rho(y,z) > 1/2$. Then $w \succ x$, $y \succ z$ and since U represents

$\succcurlyeq$ we have $U(w) > U(x)$ and $U(y) > U(z)$. Let

$$w' = \bar{y} + \frac{U(\hat{x}) - U(\bar{y})}{U(w) - U(x)}(w - x)$$
$$y' = \bar{y} + \frac{U(\hat{x}) - U(\bar{y})}{U(y) - U(z)}(y - z).$$

Since U is linear, $U(w') = U(y') = U(\bar{x})$ and hence $w', y' \in I(\bar{x})$. By the linearity of ρ, $\rho(w', \bar{y}) = \rho(w, x) \geq \rho(y, z) = \rho(y', \bar{y})$. By Lemma 7 the sets $B(\rho(w', \bar{y}))$ and $B(\rho(y', \bar{y}))$ are dilations of one another and hence $\|w' - \hat{x}\|_B \leq \|y' - \hat{x}\|_B$. By construction, $\hat{x} - \bar{y}$ is orthogonal to $I(\bar{x}) - \hat{x}$, and therefore

$$\|w' - \bar{y}\|^2 = \|w' - \hat{x}\|^2 + \|\hat{x} - \bar{y}\|^2 \leq \|y' - \hat{x}\|^2 + \|\hat{x} - \bar{y}\|^2 = \|y' - \bar{y}\|^2.$$

Thus,

$$\left\| \frac{U(\hat{x}) - U(\bar{y})}{U(w) - U(x)}(w - x) \right\| = \|w' - \bar{y}\| \leq \|y' - \bar{y}\| = \left\| \frac{U(\hat{x}) - U(\bar{y})}{U(y) - U(z)}(y - z) \right\|$$

which implies

$$\frac{U(w) - U(x)}{\|w - x\|} \geq \frac{U(y) - U(z)}{\|y - z\|}.$$

Next, suppose $\rho(w, x) \geq 1/2 \geq \rho(y, z)$ with $w \neq x$ and $y \neq z$. Then $U(w) \geq U(x)$ and $U(z) \geq U(y)$ which implies

$$\frac{U(w) - U(x)}{\|w - x\|} \geq 0 \geq \frac{U(y) - U(z)}{\|y - z\|}.$$

Finally, suppose $1/2 > \rho(w, x) \geq \rho(y, z)$. Then $\rho(z, y) \geq \rho(x, w) > 1/2$ and the desired inequality follows from the first step. Reversing the argument above to show the converse is straightforward and left to the reader. $\square$

Lemma 10. *The image of ρ is an interval $[1 - \bar{p}, \bar{p}]$.*

Proof. Linearity implies ρ is entirely determined by the values of the mapping $x \mapsto \rho(x, \bar{y})$ for $x \in I(\bar{x})$. Hence, ρ achieves its maximum at $\bar{p} = \rho(\hat{x}, \bar{y})$. Linearity also implies ρ is entirely determined by the values of the mapping $x \mapsto \rho(x, \bar{y})$ for x in a unit sphere around $\bar{y}$. The continuity of ρ implies $x \mapsto \rho(x, \bar{y})$ is continuous on the unit sphere around $\bar{y}$. The result then easily follows from the intermediate value theorem. $\square$

To construct F, we first define an auxiliary function $f : [1-\bar{p}, \bar{p}] \to \mathbb{R}$. Let $f(1/2) = 0$. For each $t \neq 1/2$, let $f(t) = [U(x) - U(y)]/\|x - y\|$ for any x, y such that $\rho(x, y) = t$. By Lemma 9 and Lemma 10, the function f is well defined. To see that the image of f must be a compact interval in $\mathbb{R}$, take any lottery $x \neq \hat{x}$ with $U(x) = U(\hat{x})$. Then we have $U(\hat{x} + t(x - \hat{x})) - U(\bar{y}) = U(\hat{x}) - U(\bar{y})$ for all $t > 0$ and $\|\hat{x} + t(x - \hat{x}) - \bar{y}\| \geq t\|x - \hat{x}\| - \|\hat{x} - \bar{y}\|$ which goes to infinity when t goes to infinity. Hence $[U(\hat{x} + t(x - \hat{x})) - U(\bar{y})]/\|\hat{x} + t(x - \hat{x}) - \bar{y}\|$ goes to zero when t goes to infinity. Thus the image of f is the interval $[-T, T]$, where $T = [U(\hat{x}) - U(\bar{y})]/\|\hat{x} - \bar{y}\|$. By Lemma 9 f is strictly increasing and has an inverse. Repeating the argument in the proof of Lemma 10 shows f is continuous. Leting $F = f^{-1}$ be the continuous inverse of f, it follows that $(U, \|\cdot\|, F)$ is a linear differentiation representation for ρ. □

Proof of Theorem 2

We start the proof with a useful result about elliptical distributions:

Lemma 11. *Let $\beta = \mu + \Lambda\varepsilon$ be elliptical and F the marginal cdf of the spherical ε. Then, for every x and y,*

$$\mathbb{P}\{\beta' x \geq \beta' y\} = \begin{cases} \mathbf{1}_{\{\mu' x \geq \mu' y\}}, & \text{if } \Lambda'(x-y) = 0 \\ F\left(\frac{\mu'(x-y)}{\sqrt{(x-y)'\Lambda\Lambda'(x-y)}}\right), & \text{if } \Lambda'(x-y) \neq 0. \end{cases}$$

Proof. Since $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_k)$ has a spherical distribution, for any $a \in \mathbb{R}^k$ the random variable $\varepsilon' a$ has the same distribution as the random variable $\|a\|\varepsilon_1$ (see, for example, Theorem 2.4 in Fang et al. (1990)). Hence for any $x \neq y$, the random variable $\varepsilon'\Lambda'(y-x)$ has the same distribution as $\|\Lambda'(x-y)\|\varepsilon_1$ and

$$\begin{aligned} \mathbb{P}\{\beta' x \geq \beta' y\} &= \mathbb{P}\{(\mu + \Lambda\varepsilon)'(x-y) \geq 0\} \\ &= \mathbb{P}\{\varepsilon'\Lambda'(y-x) \leq \mu'(x-y)\} \\ &= \mathbb{P}\{\|\Lambda'(x-y)\|\varepsilon_1 \leq \mu'(x-y)\}. \end{aligned}$$

When $\Lambda'(x-y) = 0$ we have $\|\Lambda'(x-y)\| = 0$ and therefore

$$\mathbb{P}\{\beta' x \geq \beta' y\} = \mathbb{P}\{0 \leq \mu'(x-y)\} = \mathbf{1}_{\{\mu' x \geq \mu' y\}}.$$

However, when $\Lambda'(x-y) \neq 0$ we have $\|\Lambda'(x-y)\| > 0$ and

$$\mathbb{P}\{\beta'x \geq \beta'y\} = \mathbb{P}\left\{\varepsilon_1 \leq \frac{\mu'(x-y)}{\|\Lambda'(x-y)\|}\right\} = F\left(\frac{\mu'(x-y)}{\|\Lambda'(x-y)\|}\right)$$

as desired. □

Lemma 11 applies to all elliptical distributions. Since the elliptical coefficients model is a random coefficients model (3), by definition it must satisfy requirement (4). In particular, Lemma 11 shows the elliptical coefficients model restricts the distribution of $\beta = \mu + \Lambda\varepsilon$ to satisfy the regularity condition that $\mu'(x-y) \neq 0$ whenever $\Lambda'(x-y) = 0$.

To prove necessity, suppose ρ is an elliptical coefficients model. In particular, since ρ is a random coefficients model, it must be continuous, linear, monotone and extreme by Theorem 3 in Gul and Pesendorfer (2006). Using Lemma 11, it is also straightforward to verify that ρ must satisfy the remaining binary choice postulates.

For sufficiency, our Theorem 1 implies the binary choice restriction of ρ admits a linear differentiation model representation (2). In addition, Theorem 3 in Gul and Pesendorfer (2006) implies ρ is a random coefficients model, that is, there exists a random vector β satisfying (3) and (4). We now prove β has the required elliptical distribution.

The assumption that ρ is full means $\rho(x, y) = 1$ for some x, y. For ease of notation we prove the case where $\rho(x, y) = 1$ in the direction $x - y = (0, \ldots, 0, 1)$. The general case follows easily by using an orthogonal rotation in $\mathbb{R}^n$. Let β be a random vector, and let (u, Σ, F) be the linear differentiation model parameters such that

$$\mathbb{P}\{\beta'(x-y) \geq 0\} = \rho(x, y) = F\left(\frac{u'(x-y)}{\sqrt{(x-y)'\Sigma(x-y)}}\right)$$

for every $x \neq y$.

Since $x - y = (0, \ldots, 0, 1)$ implies $\mathbb{P}\{\beta'(x-y) > 0\} = \rho(x, y) = 1$ we have $\beta_n > 0$ almost surely. Dividing every entry β_i by $\beta_n > 0$, if necessary, guarantees that the last coordinate of $\beta = (\beta_1, \ldots, \beta_{n-1}, 1)$ is constant. For the same reason, it must be that $u_n > 0$ and, by Proposition 2, we can assume $u_n = 1$ without loss of generality.

We obtain a new representation with parameters $u, \hat{\Sigma}, \hat{F}$ by taking the limit $C \to 0$ for the vertical differentiation scaling constant in Proposition 2. The matrix $\hat{\Sigma} = (\Sigma - uu'/T^2)$ induces a semi-norm in $\mathbb{R}^n$, in which every vector in the direction of maximum choice probability has length zero. We let

$$\hat{F}(t) = F\left(t/\sqrt{1+(t/T)^2}\right)$$

for each $t \in \mathbb{R}$ and, in addition, set $\hat{F}(-\infty) = F(-T)$ and $\hat{F}(+\infty) = F(T)$. Since ρ is full we have $F(T) = 1$ and hence $\hat{F}$ is a strictly increasing and continuous cumulative distribution function defined on the extended real line $[-\infty, \infty]$. The parameters $u, \hat{\Sigma}, \hat{F}$ yield a representation

$$\rho(x,y) = \hat{F}\left(\frac{u'(x-y)}{\sqrt{(x-y)'\hat{\Sigma}(x-y)}}\right)$$

for every $x \neq y$, since a zero denominator occurs if and only if $x - y$ is a multiple of the maximum choice probability direction $\Sigma^{-1}u$. In such cases, we have a non-zero numerator divided by zero, yielding a ratio of $\pm\infty$ which by construction belongs to the extended domain of $\hat{F}$.

Proposition 1 shows the maximum choice probability direction $\Sigma^{-1}u$ is unique, so $(0, \ldots, 0, 1)$ is a positive multiple of $\Sigma^{-1}u$. This implies the last column and row of Σ is given by a multiple of u. Moreover, the matrix $\hat{\Sigma}$ must have zeroes on the last row and column, and can be written as

$$\hat{\Sigma} = \begin{bmatrix} M & 0 \\ 0 & 0 \end{bmatrix}$$

where M is an $(n-1) \times (n-1)$ positive definite matrix.

For each vector $v = (v_1, \ldots, v_{n-1}, v_n)$ in $\mathbb{R}^n$ we write $v_{-n} = (v_1, \ldots, v_{n-1}) \in \mathbb{R}^{n-1}$. For each x, y with $\|x-y\|_{\hat{\Sigma}} = \sqrt{(x-y)'\hat{\Sigma}(x-y)} \neq 0$ we have

$$\begin{aligned}
\rho(x,y) &= \mathbb{P}\{\beta'(x-y) \geq 0\} \\
&= \mathbb{P}\{(\beta - u + u)'_{-n}(x-y)_{-n} + x_n - y_n > 0\} \\
&= \mathbb{P}\left\{\frac{u'_{-n}(x-y)_{-n}}{\|x-y\|_{\hat{\Sigma}}} + \frac{x_n - y_n}{\|x-y\|_{\hat{\Sigma}}} > (\beta - u)'_{-n} M^{-\frac{1}{2}} \frac{M^{\frac{1}{2}}(y-x)_{-n}}{\|x-y\|_{\hat{\Sigma}}}\right\}.
\end{aligned}$$

In addition, for any such x, y we have

$$\rho(x,y) = \hat{F}\left(\frac{u'(x-y)}{\|x-y\|_{\hat{\Sigma}}}\right) = \hat{F}\left(\frac{u'_{-n}(x-y)_{-n}}{\|x-y\|_{\hat{\Sigma}}} + \frac{x_n - y_n}{\|x-y\|_{\hat{\Sigma}}}\right).$$

Fix any arbitrary unitary vector v in $\mathbb{R}^{n-1}$ and any $t \in \mathbb{R}$. Setting $y_{-n} = M^{-1/2}v$, $y_n = 0$, $x_{-n} = 0$ and $x_n = t + u'_{-n}M^{-1/2}v$, the last two display equations yield

$$\mathbb{P}\{(\beta - u)'_{-n}M^{-1/2}v < t\} = \rho(x, y) = \hat{F}(t).$$

Since the unitary vector v and the constant t are arbitrary, this shows $M^{-1/2}(\beta - u)_{-n}$ has an unbounded spherical distribution with a marginal cdf $\hat{F}$ (see, for example, Theorem 2.5 in Fang et al. (1990)) and therefore β has the required elliptical distribution. $\square$

Proof of Proposition 3

Let the elliptical coefficient vector be $\beta = \mu + \Lambda\varepsilon$ for some $n \times (n-1)$ full rank (i.e. rank n-1) matrix with ϵ of dimension n-1. By Lemma 11 and condition (4), we have

$$\mathbb{P}\{\beta' x \geq \beta' y\} = F\left(\frac{\mu'(x-y)}{\sqrt{(x-y)'\Sigma(x-y)}}\right)$$

where F is the one-dimensional marginal cdf of ε and $\Sigma := \Lambda\Lambda'$. Similarly, any alternative elliptical coefficients vector $\hat{\beta} := \hat{\mu} + \hat{\Lambda}\hat{\varepsilon}$ generates binary choice probabilities that can be analogously parametrized by $\hat{\mu}, \hat{\Sigma}, \hat{F}$ where $\hat{\Sigma} := \hat{\Lambda}\hat{\Lambda}'$ and $\hat{F}$ is the uni-dimensional marginal of $\hat{\varepsilon}$.

Notice that, by definition, the indifference sets $\{x \in \mathbb{R}^n : \frac{1}{2} = \mathbb{P}\{\beta' x \geq \beta' y\}\}$ are hyperplanes of dimension $n-1$ and orthogonal to μ. Since $\hat{\beta}$ generates the same choices, we have $\mu = A\hat{\mu}$ for some constant $A > 0$.

Now since β and $\hat{\beta}$ represents the same binary choices, $\frac{\mu' w}{||w||_\Sigma} \geq \frac{\mu' v}{||v||_\Sigma}$ iff $\frac{\hat{\mu}' w}{||w||_{\hat{\Sigma}}} \geq \frac{\hat{\mu}' v}{||v||_{\hat{\Sigma}}}$ which is equivalent to $\frac{\mu' w}{||w||_\Sigma} \geq \frac{\mu' v}{||v||_\Sigma}$ iff $\frac{\mu' w}{||w||_{\hat{\Sigma}}} \geq \frac{\mu' v}{||v||_{\hat{\Sigma}}}$.

Observe that $||z||_\Sigma = 0$ iff $||z||_{\hat{\Sigma}} = 0$ because the direction of choice probability 1 (as in Lemma 11 with condition (4)) has to match in both elliptical coefficient models. Among such z's, since by condition (4) of elliptical coefficient models, $\mu' z \neq 0$, we can fix for the following analysis a z such that $\mu' z = \mu' \mu$. Then for any vector v, we have

$$||\frac{\mu' v}{\mu' \mu} z||_\Sigma = ||\frac{\mu' v}{\mu' \mu} z||_{\hat{\Sigma}} = 0 \text{ and } \mu' \frac{\mu' v}{\mu' \mu} z = \mu' v.$$

Now for all the w, v such that $\mu' w = \mu' v = \mu' z$, $\frac{\mu' w}{||w||_\Sigma} \geq \frac{\mu' v}{||v||_\Sigma}$ iff $\frac{\mu' w}{||w||_{\hat{\Sigma}}} \geq \frac{\mu' v}{||v||_{\hat{\Sigma}}}$. Therefore whenever $\mu' w = \mu' v = \mu' z$ we have $||w||_\Sigma \geq ||v||_\Sigma$ iff $||w||_{\hat{\Sigma}} \geq ||v||_{\hat{\Sigma}}$.

Observe for all $v \in \mathbb{R}^n$, it holds that $||v||_\Sigma = ||v - \frac{\mu' v}{\mu' \mu} z + z||_\Sigma$, and $||v||_{\hat{\Sigma}} = ||v - \frac{\mu' v}{\mu' \mu} z +$

$z||_{\hat{\Sigma}}$, and also for all $v \in \mathbb{R}^n$, $\mu'(v - \frac{\mu' v}{\mu'\mu} z + z) = \mu' z$. Therefore we have $||w||_\Sigma \geq ||v||_\Sigma$ iff $||w||_{\hat{\Sigma}} \geq ||v||_{\hat{\Sigma}}$ for all $v, w \in \mathbb{R}^n$.

Now suppose for some v we have $||v||_\Sigma = B||v||_{\hat{\Sigma}} > 0$, and $||w||_\Sigma = C||w||_{\hat{\Sigma}} > 0$. We can find a λ so that $||v||_\Sigma = ||\lambda w||_\Sigma = C||\lambda w||_{\hat{\Sigma}}$. So $||v||_{\hat{\Sigma}} \geq ||\lambda w||_{\hat{\Sigma}} \Rightarrow ||v||_\Sigma / B \geq ||v||_\Sigma / C$. Also $||v||_{\hat{\Sigma}} \leq ||\lambda w||_{\hat{\Sigma}} \Rightarrow ||v||_\Sigma / B \leq ||v||_\Sigma / C$. Therefore $B = C$ and Σ and $\hat{\Sigma}$ are scalings of each other.

We have shown that $\hat{\mu} = A\mu, \hat{\Sigma} = B\Sigma$ for some $A, B > 0$. Since $\hat{\beta}$ and β represents the same choices, we must have $\hat{F}(tA/B) = F(t)$. In other words,

$$\hat{\beta} = A\mu + B\Lambda \frac{A}{B} \varepsilon = A\beta.$$

Theorem 2.5 in Fang et al. (1990) shows the uni-dimensional marginal cdf pins down the n-dimensional spherical distribution, as we wanted to show. $\square$

Proof of Proposition 4

To simplify notation, it suffices to show the statement for the two-attribute case. For any option $x = (x_1, x_2, p)$ the indirect utility is $x_1\beta_1 + x_2\beta_2 - p$. Consider the following pair of alternatives $y_a = (x_1 + 1, x_2, p + a)$ and $y_b = (x_1, x_2 + 1, p + b)$. It is easy to see that for this random coefficients model, we have

$$\begin{aligned}
\rho(x, \{x, y_a\}) &= \mathbb{P}\{\beta_1 < a\} \\
\rho(x, \{x, y_b\}) &= \mathbb{P}\{\beta_2 < b\} \\
\rho(x, \{x, y_a, y_b\}) &= \mathbb{P}\{\beta_1 < a \text{ and } \beta_2 < b\}
\end{aligned}$$

Factorable implies for any real values a, b we have

$$\rho(x, \{x, y_a, y_b\}) = \rho(x, \{x, y_a\}) \times \rho(x, \{x, y_b\})$$

which means for any real numbers a, b we have

$$\mathbb{P}\{\beta_1 < a \text{ and } \beta_2 < b\} = \mathbb{P}\{\beta_1 < a\} \times \mathbb{P}\{\beta_2 < b\}.$$

Therefore β_1 is independent of β_2. This implies Σ is diagonal (covariance is zero at off-diagonals). By Theorem 4.11 in Fang et al. (1990), β is Gaussian. $\square$

References

Apesteguia, Jose and Miguel A Ballester, "Monotone stochastic choice models: The case of risk and time preferences," *Journal of Political Economy*, 2018, *126* (1), 74–106.

Berry, Steven and Ariel Pakes, "The pure characteristics demand model," *International Economic Review*, 2007, *48* (4), 1193–1225.

_ , James Levinsohn, and Ariel Pakes, "Automobile prices in market equilibrium," *Econometrica: Journal of the Econometric Society*, 1995, pp. 841–890.

Chambers, Christopher P, Tugce Cuhadaroglu, and Yusufcan Masatlioglu, "Behavioral influence," *Journal of the European Economic Association*, 2023, *21* (1), 135–166.

Dean, Mark and Nathaniel Neligh, "Experimental tests of rational inattention," *Journal of Political Economy*, 2023, *131* (12), 000–000.

Dutta, Rohan, "Gradual pairwise comparison and stochastic choice," *Theoretical Economics*, 2020, *15* (4), 1335–1364.

Echenique, Federico and Kota Saito, "General luce model," *Economic Theory*, 2019, *68* (4), 811–826.

Einav, Liran, Amy Finkelstein, Stephen P Ryan, Paul Schrimpf, and Mark R Cullen, "Selection on moral hazard in health insurance," *American Economic Review*, 2013, *103* (1), 178–219.

Enke, Benjamin and Thomas Graeber, "Cognitive uncertainty," *The Quarterly Journal of Economics*, 2023, *138* (4), 2021–2067.

_ , _ , and Ryan Oprea, "Complexity and Time," Technical Report, National Bureau of Economic Research 2023.

Fang, Kai Wang, Samuel Kotz, and Kai Wang Ng, *Symmetric Multivariate and Related Distributions*, CRC Press, 1990.

Faro, José Heleno, "The Luce model with replicas," *Journal of Economic Theory*, 2023, *208*, 105596.

Ford, "Ford Dealer Training: 1989 Ford Taurus vs 1989 Honda Accord," John C. [Taurus91] 1989.

Fudenberg, Drew, Ryota Iijima, and Tomasz Strzalecki, "Stochastic Choice and Revealed Perturbed Utility," *Econometrica*, November 2015, *83* (6), 2371–2409.

Gentzkow, Matthew, "Valuing new goods in a model with complementarity: Online newspapers," *American Economic Review*, 2007, *97* (3), 713–744.

Gul, Faruk and Wolfgang Pesendorfer, "Random Expected Utility," *Econometrica*, January 2006, *74* (1), 121–146.

_ , Paulo Natenzon, and Wolfgang Pesendorfer, "Random Choice as Behavioral Optimization," *Econometrica*, September 2014, *82* (5), 1873–1912.

Gurari, N.I. and Y.I. Sozonov, "On normed spaces which have no bias of the unit sphere," *Math. Notes*, 1970, *7*, 187–189.

Halff, Henry M, "Choice theories for differentially comparable alternatives," *Journal of Mathematical Psychology*, 1976, *14* (3), 244–246.

Hausman, Jerry A. and David A. Wise, "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica*, 1978, *46* (2), pp. 403–426.

He, Junnan, "Bayesian contextual choices under imperfect perception of attributes," *Management Science*, 2023.

_ and Paulo Natenzon, "Moderate Utility," *American Economic Review: Insights*, 2023, *forthcoming.*

Hebert, Benjamin and Michael Woodford, "Neighborhood-based information costs," *American Economic Review*, 2021, *111* (10), 3225–3255.

Horan, Sean, "Stochastic semi-orders," *Journal of Economic Theory*, 2021, *192*, 105171.

Lancaster, Kelvin, *Variety, equity, and efficiency*, Columbia University Press, 1979.

Lu, Jay and Kota Saito, "Mixed logit and pure characteristics models," *Working Paper*, 2022.

Luce, R. Duncan, *Individual Choice Behavior: a Theoretical Analysis*, Wiley New York, 1959.

Matejka, Filip and Alisdair McKay, "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *The American Economic Review*, 2015, *105* (1), 272–298.

McFadden, Daniel and Kenneth Train, "Mixed MNL models for discrete response," *Journal of applied Econometrics*, 2000, *15* (5), 447–470.

Natenzon, Paulo, "Random Choice and Learning," *Journal of Political Economy*, February 2019, *1* (127), 419–457.

Nevo, Aviv, "A practitioner's guide to estimation of random-coefficients logit models of demand," *Journal of economics & management strategy*, 2000, *9* (4), 513–548.

_ , "Measuring market power in the ready-to-eat cereal industry," *Econometrica*, 2001, *69* (2), 307–342.

Oprea, Ryan, "Simplicity Equivalents," *Working Paper*, 2022.

Pomatto, Luciano, Philipp Strack, and Omer Tamuz, "The cost of information: The case of constant marginal costs," *American Economic Review*, 2023, *113* (5), 1360–1393.

Puri, Indira, "Simplicity and Risk," *Working Paper*, September 2023.

Sims, Christopher A, "Implications of Rational Inattention," *Journal of Monetary Economics*, 2003, *50* (3), 665–690.

Thompson, Anthony C, *Minkowski geometry*, Cambridge University Press, 1996.

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

Tversky, Amos and J. Edward Russo, "Substitutability and similarity in binary choices," *Journal of Mathematical Psychology*, 1969, *6* (1), 1–12.

Woodford, Michael, "Modeling imprecision in perception, valuation, and choice," *Annual Review of Economics*, 2020, *12*, 579–601.