

# Multivariate ordered discrete response models with two layers of dependence\*

Tatiana Komarova<sup>†</sup> William Matcham<sup>‡</sup>

November 5, 2025

## Abstract

We develop a class of multivariate ordered discrete response models featuring general rectangular structures, which allow for functionally interdependent thresholds across dimensions, extending beyond traditional (lattice) models that assume threshold independence. The new models incorporate two layers of dependence: one arising from the interdependence of decision rules (capturing broad bracketing behaviors) and another from the correlation of latent utilities conditional on observables. We provide microfoundations, explore semiparametric and parametric specifications, and establish identification conditions under logical consistency in decision-making. An empirical application to health insurance markets demonstrates the advantages of this new framework, showing how it disentangles moral hazard (captured via threshold dependence) from adverse selection (isolated in unobservable correlations), offering insights into behavioral responses obscured by lattice models.

**Keywords:** Ordered response, multiple dimensions, broad bracketing, narrow bracketing, coherency, insurance, moral hazard, adverse selection

**JEL Classification:** C14, C31, C35, D9

---

\*We are grateful to participants at Yale University, Harvard-MIT, Erasmus University Rotterdam, and LSE-STICERD seminars, *EC<sup>2</sup>* 2023 conference for helpful comments.

<sup>†</sup>Faculty of Economics, University of Cambridge. tk670@cam.ac.uk.

<sup>‡</sup>Department of Economics, Royal Holloway University of London. William.Matcham@rhul.ac.uk.

# 1 Introduction

This paper examines ordered discrete response models in which individuals make simultaneous decisions across multiple categorical dimensions, each with a meaningful ordering. A natural way to extend univariate ordered response models to these multivariate contexts – and the one adopted by most empirical work – is to define latent utilities and threshold decision rules for each dimension. However, such extensions, while intuitive, often oversimplify the decision-making process by assuming complete functional independence of threshold decision structures across dimensions, as illustrated in the left panel in Figure 1.<sup>1</sup> From a behavioral economics perspective, these models align with agents exhibiting *narrow bracketing*, prioritizing simpler choice rules over joint utility optimization. The structure of these models reveals that threshold intersections across dimensions create a lattice in multidimensional space, once again illustrated in the left panel in Figure 1, prompting us to term them *lattice models*.<sup>2</sup>

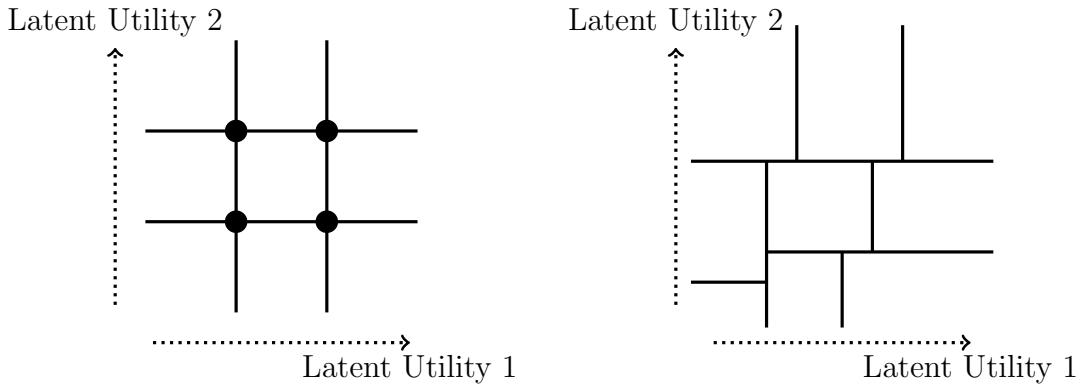


FIGURE 1: Models with a lattice (left) and a non-lattice (right) structure

We introduce and explore a comprehensive class of models for selecting ordered categories across multiple dimensions. These models retain familiar features of ordered response models, such as (a) reliance on latent utilities for each dimension, as in lattice models, and (b) threshold-based decision rules. However, the models innovate by allowing thresholds across dimensions to be functionally interdependent. This flexibility enables our models to capture more complex and nuanced economic behavior compared to lattice models.

Drawing on behavioral economics, our framework accommodates broad bracketing as a general

<sup>1</sup>In an auction context, this assumption is akin to suggesting that a firm bidding in two simultaneous auctions for complementary objects would employ functionally independent equilibrium strategies in each, a notion contradicted by auction theory literature. See discussion in [Gentry et al. \(2023\)](#) and [Gentry et al. \(2019\)](#) for more detail.

<sup>2</sup>This is the term we use ourselves for this model, it is not a commonly accepted terminology

case, while encompassing narrow bracketing as a special case, since lattice models are nested within our broader class. Like lattice models, we focus on a single economic agent making simultaneous decisions across multiple dimensions. In other words, different dimensions will not represent different economic agents interacting strategically. This does not mean that our setting is completely irrelevant to game-theoretical contexts. For instance, drawing on the auction analogy from above, our decision structure determined ex-ante can represent a single bidder's equilibrium strategies across multiple auctions.

The right panel in Figure 1 illustrates the models we propose. We refer to them as models with *general rectangular structures*. Sometimes to distinguish them from lattice models and to signify the fact that intersections of thresholds across dimensions no longer form a lattice, we may refer to them as *non-lattice* models.<sup>3</sup>

Models with general rectangular structures pose both theoretical and practical challenges. On the theoretical side, understanding economic behavior of agents making decisions and estimating latent utility parameters or their joint dependence requires disentangling two distinct elements: the functional dependence of decision thresholds and the interdependence of latent utilities across dimensions, conditional on observables. On the practical side, one must impose restrictions on the thresholds to ensure the internal coherence of the decision structure (a concept we formalize later) and estimate a greater number of parameters from the data.

While these challenges are nontrivial, they come with significant rewards, as models with general rectangular structures allow us to more accurately uncover the true decision structure and identify the underlying economic primitives.

Specifically, general rectangular structures come with two key advantages over existing multivariate ordered choice models. First, non-lattice models permit richer forms of interaction across dimensions by allowing two distinct layers of complementarity or substitutability. For instance, the threshold decision rule might display substitutability, while the dependence structure of unobservables in the latent utilities might reflect complementarities. Furthermore, within the threshold decision structure itself, patterns of complementarity or substitutability can vary across different regions of the latent utility space. Our application to health insurance exploits this feature of two distinct layers of complementarity/substitutability to disentangle moral hazard from advantageous/adverse selection.

---

<sup>3</sup>When using the term non-lattice, one has to keep in mind that lattice models are a special case of such models.

Second, lattice models force the sign of any partial effect on the conditional probability of exceeding a given level in one dimension to be constant across the domain (for which we will provide a concrete example in the text). By contrast, models with general rectangular structures allow these partial effects to change sign, capturing more nuanced and flexible behavioral patterns. Lattice models also restrict the conditional probability of exceeding a given level in one dimension (given all covariates across processes) from depending on covariates that do not belong to its own latent process. Models with general rectangular structures relax this restriction, allowing indirect effects from other covariates. For instance, a price subsidy meant to encourage health insurance enrollment can indirectly influence the probability of exceeding a given level in the pension plan dimension.

After reviewing the related work and situating our contribution within the broader literature, we begin our analysis in Section 3. There, we formally define models with general rectangular structures (which we sometimes refer to as non-lattice models), lattice models<sup>4</sup>, and develop our concept of coherency. In Section 4 we provide two microfoundations for general rectangular structures. The first includes explicit synergies/crowding out effects in the joint utility specification. The second microfounds general rectangular models as the outcome from discretizing a continuous joint utility maximization problem.

Section 5 develops a semiparametric specification of multivariate ordered response models with general rectangular structures and examines their properties in detail. In particular, Section 5.1 highlights how these models capture richer economic behavior than lattice models, focusing on the more flexible patterns of partial effects discussed above. Section 5.2 then turns to identification. We proceed step by step, starting from the parameters associated with exclusive covariates and ending with the identification of thresholds, which is more involved. The analysis assumes independence between unobservables and covariates, and that each process includes at least one exclusive covariate with a meaningful effect. Section 5.3 discusses possible estimation approaches for semiparametric models with general rectangular structures. The main method builds on Coppejans (2007) subject to coherency constraints expressed as equalities involving thresholds, though we do not provide formal inference results.

Section 6 focuses on the parametric case, where the joint distribution of unobservables is assumed to follow a multivariate normal distribution. We illustrate identification under much less

---

<sup>4</sup>For clarity, we emphasize that our non-lattice models encompass lattice models as a special case, even though the terminology might suggest otherwise.

stringent assumptions than in the semiparametric case through a numerical exercise using a bivariate model and discuss estimation via maximum likelihood, subject to the same coherency constraints.

Section 7 presents Monte Carlo simulations assessing the performance of the proposed non-lattice probit estimator under normal errors, compared with the standard bivariate ordered probit estimators which effectively estimates a mis-specified lattice model.

Section 8 contains applications. The first explores the relationship between cryptocurrency familiarity and optimism. It shows how the non-lattice approach helps uncover differences in how individuals form and express opinions about bitcoin’s value obscured by the lattice model. The second application concerns insurance markets, where we show how the non-lattice model can disentangle moral hazard from selection (adverse or advantageous) by allowing functionally dependent thresholds that capture moral hazard in a coherent and data-driven way leaving selection to be fully captured by correlation of unobservables.

Section 9 concludes and the online supplement contains more details on coherency and proofs of all formal results.

## 2 Our contributions and literature review

Our paper contributes to the literature on the economic foundations of ordered choice models by extending the analysis from *univariate* to *multivariate* decision problems. We study an agent making several ordered choices whose decisions interact through functionally dependent fixed thresholds across dimensions. Most existing work focuses on univariate models. For example, Cunha et al. (2007) develop a “generalized ordered choice” model with thresholds that depend on observables and unobservables, showing how this framework captures a wide range of economic settings, including dynamic ones such as schooling decisions. Earlier contributions include Cameron and Heckman (1998), Heckman et al. (1999), Carneiro et al. (2003), and Lewbel (2003), who study ordered models with random or sequentially determined thresholds.

In contrast, our model allows for complex interactions across multiple ordered dimensions while maintaining fixed thresholds. Here, thresholds depend on the realizations of other endogenous variables rather than regressors or unobservables, requiring a joint model of all endogenous processes. This extension offers a more flexible structure on thresholds than the one implied by

fixed thresholds and univariate stochastic thresholds determined by regressors and errors.

From a more foundational point of view, two main approaches have emerged in the literature on univariate threshold-based ordered response models. One treats thresholds as reduced-form tools that aid estimation but have limited behavioral interpretation (e.g., [Greene and Hensher \(2010\)](#) and [Boes and Winkelmann \(2006\)](#)). For example, [Greene and Hensher \(2010\)](#) notes that thresholds may capture psychological attitudes with “bunched” cut points suggesting strong preferences, and dispersed ones reflecting indifference. In economic contexts like schooling or job satisfaction, thresholds are often viewed as cost-benefit barriers, though this link is typically conceptual rather than derived from optimization. [Anderson \(1984\)](#) extends this idea with “stereotype” ordered regressions, where thresholds relate to category proportions rather than absolute utility levels.<sup>5</sup>

A second strand grounds thresholds in explicit optimization problems, offering clearer microfoundations. For instance, [Bhat and Pulugurta \(1998\)](#) derive thresholds from a range-based utility model, while [Apesteguia and Ballester \(2023\)](#) propose type-ordered random utility models, where ordered choices arise from heterogeneous preference types without restrictive distributional assumptions. Structural approaches such as [Cunha et al. \(2007\)](#) also belong to this line of work, though none extend to multivariate settings.

Our paper takes a first step toward microfoundations for multivariate ordered response models with general rectangular structures. Section 4 presents two approaches. The first, illustrated with a bivariate example, interprets higher discrete responses as bundles of lower outcomes that may be complements or substitutes, depending on latent utilities and thresholds. The second characterizes marginal utilities along each dimension and shows that a rectangular structure naturally arises as the optimal discrete response satisfying discrete analogues of first-order conditions. Together, these provide a foundation for understanding multivariate ordered decision-making.

Another related line of research concerns *choice bracketing* also referred to as *sequential vs. simultaneous choice* ([Simonson and Winer, 1992](#)), *narrow vs. broad decision frames* ([Kahneman and Lovallo, 1993](#)), *local vs. overall value functions* ([Heyman, 1996](#)), and *isolated vs. distributed choice* ([Herrnstein and Prelec, 1991](#)). This literature is largely theoretical and experimental ([Tversky and Kahneman, 1981; Read et al., 1999; Thaler, 1999; Rabin and Weizsäcker, 2009; Lian, 2020; Camara, 2021; Zhang, 2021](#)), with only a few descriptive or structural empirical studies ([Camerer,](#)

---

<sup>5</sup>In the [Anderson \(1984\)](#) model, ordinal categories are not tied to a single latent variable with fixed thresholds.

Babcock, Loewenstein, and Thaler, 1997; Thakral and Tô, 2021).<sup>6</sup> To date, econometric work has not explicitly modeled narrow versus broad bracketing behavior. Our framework of general rectangular structures offers a natural way to do so with lattice models corresponding to *narrow bracketing* and the broader rectangular structure capturing *broad bracketing* decisions. This setup also enables formal testing for broad bracketing by examining whether thresholds in the latent space conform to a lattice structure.

A tangentially related literature is the discrete choice framework with *strategic interactions*, where outcomes for one player depend on the actions of others (Tamer, 2003; Berry and Reiss, 2007; Ciliberto and Tamer, 2009; Honore and De Paula, 2010; Chesher and Rosen, 2017, 2020; Aradillas-López and Rosen, 2022). In these models, each agent represents a distinct dimension, and best responses can lead to incoherent or incomplete outcomes. In contrast, our paper focuses on a single economic agent making decisions along multiple dimensions. For such an agent, the decision problem is internally consistent by construction, and therefore models with general rectangular structures are *coherent*. As we detail in the following section, by *coherency* we mean logical consistency in decision-making that ensures that rectangular regions representing different discrete responses do not overlap and together cover the entire latent space  $\mathbb{R}^D$ .<sup>7</sup>

### 3 Model with a general rectangular structure

We formally define general rectangular structures and lattice multivariate ordered response models for an agent making decisions across  $D \geq 2$  dimensions. These models map a  $D$ -variate latent continuous metric  $(Y^{*c_1}, \dots, Y^{*c_D})$  to a discrete metric  $(Y^{c_1}, \dots, Y^{c_D})$ , with ordered responses in dimension  $d$  denoted as  $y_j^{(d)}$ ,  $j = 1, \dots, M_d$ , and satisfying  $y_1^{(d)} < \dots < y_{M_d}^{(d)}$ .

**Definition 1 (General rectangular structure model)** *A model has a general rectangular structure (or sometimes we refer to it as a non-lattice model) if*

$$(Y^{c_1}, \dots, Y^{c_D}) = (y_{j_1}^{(1)}, \dots, y_{j_D}^{(D)}) \iff (Y^{*c_1}, \dots, Y^{*c_D}) \in R_{j_1, \dots, j_D}, \text{ where}$$

---

<sup>6</sup>Tversky and Kahneman (1981) provides a classic example of narrow bracketing in experimental settings.

<sup>7</sup>For more on coherency, see Heckman (1978) and Tamer (2003). Tamer (2003) distinguish between incoherency and incompleteness in games with strategic interactions, a distinction followed by later studies. In our setting, we use the term “coherency” to refer more generally to overall logical consistency.

$$R_{j_1, \dots, j_D} = \bigtimes_{d=1}^D \left( \alpha_{j_1, \dots, j_{d-1}, \textcolor{red}{j_d-1}, j_{d+1}, \dots, j_D}^{(d)}, \alpha_{j_1, \dots, j_{d-1}, \textcolor{red}{j_d}, j_{d+1}, \dots, j_D}^{(d)} \right], \quad (1)$$

with thresholds  $\alpha_{j_1, \dots, j_{d-1}, \textcolor{red}{j_d}, j_{d+1}, \dots, j_D}^{(d)}$  increasing in  $j_d$  for given other indices and normalized at the boundary as

$$\alpha_{j_1, \dots, j_d, \dots, j_D}^{(d)} = +\infty \text{ for } j_d = M_d, \quad \alpha_{j_1, \dots, j_d, \dots, j_D}^{(d)} = -\infty \text{ for } j_d = 0.$$

Threshold intersections in Definition 1 do not necessarily form a lattice in  $\mathbb{R}^D$ , reflecting functionally interdependent decision rules, akin to *broad bracketing* in behavioral economics.

**Definition 2 (Lattice model)** A lattice model is a special case of a general rectangular structure (non-lattice) model in which each threshold  $\alpha_{j_1, \dots, j_d, \dots, j_D}^{(d)}$  depends only on the index in its own dimension:

$$(Y^{c_1}, \dots, Y^{c_D}) = (y_{j_1}^{(1)}, \dots, y_{j_D}^{(D)}) \iff Y^{*c_d} \in \left( \alpha_{j_d-1}^{(d)}, \alpha_{j_d}^{(d)} \right] \quad \forall d, \text{ with}$$

$$\alpha_{j_d}^{(d)} = +\infty \text{ for } j_d = M_d, \quad \alpha_{j_d}^{(d)} = -\infty \text{ for } j_d = 0.$$

Here, thresholds are functionally independent across dimensions, forming a lattice in  $\mathbb{R}^D$  in their intersections. Lattice models correspond to a decision maker who *narrowly brackets*, since decisions can now be seen as made dimension-by-dimension, as opposed to jointly. Lattice models will misspecify a decision maker who broadly brackets.

Thus, the distinction between broad and narrow bracketing is fully captured by functional interdependence or independence of decision rules, determined by the thresholds. Both lattice and non-lattice models permit correlated decisions through latent processes, but the latter distinguish correlation in unobservables from interdependent decision rules.

**Coherency** The flexibility of general rectangular structure models is achieved by allowing thresholds for each dimension to depend on the full vector of response indices. But this comes at a cost, as Definition 1 does not guarantee that the division of latent space into regions  $R_{j_1, \dots, j_D}$  is to be exhaustive or mutually exclusive. As a result, the condition that each latent profile maps to exactly one observed response (which is to us is associated with logical consistency in decision-making) and to which we refer to as *coherency* is not satisfied by Definition 1 construction.

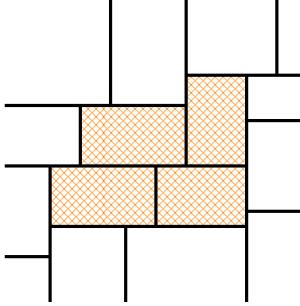


FIGURE 2: Intuition for a non-lattice model being coherent when  $D = 2$

Since the model aims to describe the behavior of a logically consistent decision-maker, it should always satisfy coherency, especially when we take it to the data. In general rectangular structure models ensuring this requires explicit constraints on the thresholds. Lattice models, by contrast, are coherent by construction.

We now examine coherency in a bivariate general rectangular structure ordered response model, providing a formal condition on thresholds for the model to be coherent. The characterization of coherency for  $D > 2$  is more involved and is given in the online supplement.

**Proposition 1 (Coherency for  $D = 2$ )** *Consider a bivariate general rectangular structure ordered response model, defined by a set of thresholds  $\left\{ \alpha_{j_1,j_2}^{(1)}, \alpha_{j_1,j_2}^{(2)} \right\}_{j_1=1,j_2=1}^{M_1-1,M_2-1}$ .*

*Given thresholds normalizations at the boundary, the model is coherent – i.e., the latent space is partitioned into mutually exclusive and exhaustive rectangular regions  $R_{j_1,j_2} = (\alpha_{j_1-1,j_2}^{(1)}, \alpha_{j_1,j_2}^{(1)}) \times (\alpha_{j_1,j_2-1}^{(2)}, \alpha_{j_1,j_2}^{(2)})$  each corresponding to a unique observed outcome – if and only if, for all  $(j_1, j_2)$ ,*

$$\left( \alpha_{j_1+1,j_2}^{(1)} - \alpha_{j_1,j_2}^{(1)} \right) \cdot \left( \alpha_{j_1,j_2+1}^{(2)} - \alpha_{j_1,j_2}^{(2)} \right) = 0. \quad (2)$$

In other words, for the model to be coherent, the thresholds must satisfy a local condition for each  $2 \times 2$  block of adjacent cells. Specifically, within each block, at least one of the dimensions must have constant thresholds across that block. This requirement prevents ambiguity in decision-making by ensuring that when an agent faces a choice within a  $2 \times 2$  block, they make their decision sequentially: first along one dimension, where the thresholds remain fixed, and then along the other dimension. An illustration of a local problem and the coherency requirement is given in Figure 2. Thus, in every local decision problem one dimension is leading and the leading dimension may be different across different parts of the domain (e.g., when one considers health insurance levels vs retirement contribution level, it may very well be the case that for lower levels

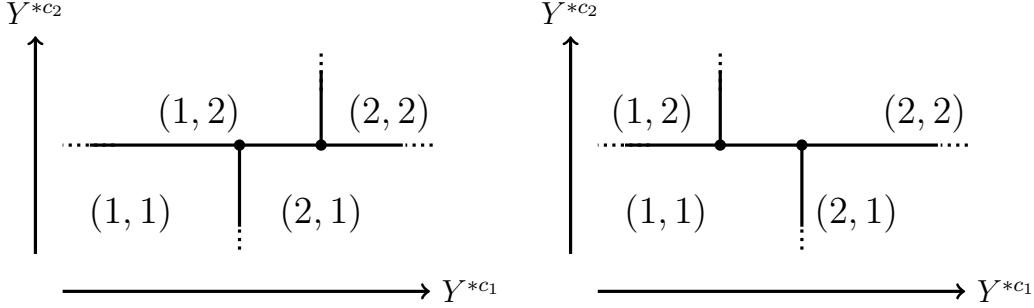


FIGURE 3: Illustration in the bivariate case with two discrete options in each dimension

the insurance decision is the leading one whereas for higher levels of both the leading decision is the retirement contributions as at those levels long-run financial planning may be of more relevance).

## 4 Microfoundations

There are several ways to approach a general rectangular structure model from a microeconomic foundations perspective. We propose two such approaches.

The first approach directly models complementarities and substitutabilities in the joint utility across different pairings of options. We illustrate how it can be done in a simple bivariate model with two discrete options (1 and 2) in each dimension. The left panel in Figure 3 shows substitutability in the decision structure reflected in the larger threshold in the second dimension when  $Y^{c_1} = 1$ . It shows that choosing a higher level in dimension 1 makes it harder to choose a higher level in the other, often due to resource constraints. The right panel in Figure 3 shows complementarity as choosing a higher level in dimension 1 facilitates a higher level in the other dimension.

Consider the following utilities across four pairs of discrete choices: for constant  $v < 0$ ,

$$\begin{aligned} U(1, 1) &= 0, & U(2, 1) &= Y^{*c_1}, & U(1, 2) &= Y^{*c_2}, \\ U(2, 2) &= D(Y^{*c_1} + Y^{*c_2} - v) + (1 - D)(Y^{*c_1} + Y^{*c_2}), \end{aligned} \quad (3)$$

where  $D = \mathbf{1}(Y^{*c_1} > 0)\mathbf{1}(Y^{*c_2} > v)$ . Then the  $\text{argmax}_{j_1, j_2} U(j_1, j_2)$  is (i) (1,1) when  $Y^{*c_1}, Y^{*c_2} \leq 0$ ; (ii) (1, 2) when  $Y^{*c_1} \leq 0, Y^{*c_2} > 0$ ; (iii) (2,1) when  $Y^{*c_1} > 0, Y^{*c_2} \leq v$ ; and (iv) (2,2) when  $Y^{*c_1} > 0, Y^{*c_2} > v$ . The lower threshold  $v < 0$  facilitates choosing  $Y^{*c_2} = 2$  when  $Y^{*c_1} = 2$ ,

suggesting complementarity (as in the right panel in Figure 3). The utility boost  $-v > 0$  in  $U(2, 2)$  when  $D = 1$  acts like a synergy term, incentivizing (2,2) over (2,1) or (1,2) when propensities are sufficient. This reflects scenarios where choosing one high option reduces the marginal cost of the other (e.g., economies of scale, subsidies).

To obtain substitutes in the decision structure (as in the left panel in Figure 3), consider  $w > 0$  and replace  $U(2, 2)$  in (3) with the following definition:

$$U(2, 2) = D(Y^{*c_1} + Y^{*c_2}) + (1 - D)(Y^{*c_1} + Y^{*c_2} - w),$$

with  $D$  defined in the same way as before. The higher threshold for  $Y^{*c_2}$  when  $Y^{c_1} = 2$  indicates that choosing  $Y^{c_1} = 2$  raises the level needed for  $Y^{c_2} = 2$ , reflecting substitutability. This captures resource competition (e.g., budget, time) where pursuing one high choice increases the cost of the other. The penalty  $-w < 0$  when  $D = 0$  reinforces the trade-off. An analogous construct could be employed for any number of ordered choices in each dimension.

The second approach to providing microeconomic foundations for general rectangular structure models is to view discrete options as the result of discretizing an underlying continuous space, whether due to survey design, categorical reasoning, or similar factors. If one had a smooth function  $U(y^{(1)}, y^{(2)})$  of continuous responses  $(y^{(1)}, y^{(2)})$  then the global maximum  $(\bar{y}^{(1)}, \bar{y}^{(2)})$  would have necessarily satisfied  $\frac{\partial U(\bar{y}^{(1)}, \bar{y}^{(2)})}{\partial y^{(1)}} = 0$ ,  $\frac{\partial U(\bar{y}^{(1)}, \bar{y}^{(2)})}{\partial y^{(2)}} = 0$ . With the discrete grid  $(y_{j_1}^{(1)}, y_{j_2}^{(2)})$  to find a maximum on the grid we have to ensure the function value does not increase when moving to any neighboring grid points in each coordinate: that is,  $(y_{j_1}^{(1)}, y_{j_2}^{(2)})$  is the maximizer of  $U$  on the grid only if

$$U(y_{j_1}^{(1)}, y_{j_2}^{(2)}) - U(y_{j_1-1}^{(1)}, y_{j_2}^{(2)}) > 0, \quad U(y_{j_1+1}^{(1)}, y_{j_2}^{(2)}) - U(y_{j_1}^{(1)}, y_{j_2}^{(2)}) \leq 0, \quad (4)$$

$$U(y_{j_1}^{(1)}, y_{j_2}^{(2)}) - U(y_{j_1}^{(1)}, y_{j_2-1}^{(2)}) > 0, \quad U(y_{j_1}^{(1)}, y_{j_2+1}^{(2)}) - U(y_{j_1}^{(1)}, y_{j_2}^{(2)}) \leq 0 \quad (5)$$

A general rectangular structure model is obtained when for any  $j_1 > 1$  and any  $j_2 > 1$ ,

$$\begin{aligned} U(y_{j_1}^{(1)}, y_{j_2}^{(2)}) - U(y_{j_1-1}^{(1)}, y_{j_2}^{(2)}) &= Y^{*c_1} - \alpha_{j_1-1, j_2}^{(1)}. \\ U(y_{j_1}^{(1)}, y_{j_2}^{(2)}) - U(y_{j_1}^{(1)}, y_{j_2-1}^{(2)}) &= Y^{*c_2} - \alpha_{j_1, j_2-1}^{(2)}. \end{aligned}$$

These specifications of the utility differences (analogues of marginal utilities) ensure that for any realization  $(Y^{*c_1}, Y^{*c_2})$  of latent processes, there is only one pair  $(j_1, j_2)$  that satisfies (4)-(5).

Namely, this is the pair  $(j_1, j_2)$  such that  $(Y^{*c_1}, Y^{*c_2}) \in R_{j_1, j_2}$ .

## 5 Semiparametric specification, partial effects, identification, estimation

To take our model to the data, we adopt the standard approach in the discrete response literature, specifying each  $d$ -th continuous latent process as a linear index:

$$Y^{*c_d} = x_d \beta_d + \varepsilon_d, \quad d = 1, \dots, D, \tag{6}$$

where  $x_d$  is a row vector of observable covariates,  $\beta_d$  is a column vector of unknown parameters, and  $\varepsilon_d$  is an unobservable error term. This structure interprets  $x_d \beta_d$  as the systematic component of an agent's latent propensity to choose an ordered category in dimension  $d$ , with  $\varepsilon_d$  capturing random shocks to the latent utility. The errors  $\varepsilon_1, \dots, \varepsilon_D$  may be dependent, allowing latent processes  $Y^{*c_d}$  to be correlated conditional on covariates. This linear index, standard in discrete choice models, supports estimation of threshold-based interdependence in general rectangular structure models while maintaining parsimony. While more general functions of  $x_d$  could be used without compromising identifiability, the linear form offers practical simplicity with minimal loss of flexibility.

Given the complexity of the model, particularly with regard to the two-layer dependence structure, we should be prepared for fairly stringent requirements on the data to ensure identification of the following objects of interest :  $\beta_d$ ,  $d = 1, \dots, D$ , the joint c.d.f of  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_D)'$ , and the thresholds. We start by imposing Assumption 1.

**Assumption 1**  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_D)'$  is independent of  $x = (x_1, \dots, x_D)$  and has a convex support in  $\mathbf{R}^D$ .

**Notation 1** For any  $d = 1, \dots, D$ , let  $\kappa_d$  denote either  $\leq$  or  $>$  sign. Denote

$$F_{\kappa_1, \kappa_2, \dots, \kappa_D}(t_1, \dots, t_D) = P(\cap_{d=1}^D (\varepsilon_d \kappa_d t_d))$$

for any  $\kappa_d \in \{\leq, >\}$ ,  $d = 1, \dots, D$ . Functions  $F_{\leq, \dots, \leq}$  and  $F_{>, \dots, >}$  are the joint c.d.f. and the joint survival function of  $\varepsilon$ , respectively, and for simplicity we will interchangeably denote them as

$F$  and  $\bar{F}$ , respectively. All the other cases correspond to hybrid forms of c.d.f.s and survival functions. Similar notation apply to subsets of dimensions.

Also denote  $F_{d,\kappa_d}(t_d) = P(\varepsilon_d \kappa_d t_d)$  for  $\kappa_d \in \{\leq, >\}$ . Thus,  $F_{d,\leq}$  is the marginal c.d.f. and  $F_{d,>}$  is the marginal survival function of  $\varepsilon_d$ .

In our related paper, [Komarova and Matcham \(2025\)](#), we outline the increasing degree of restrictions required to ensure identification in semiparametric lattice models. There, we discuss why identification of parameters and thresholds can rely on a weaker version of Assumption 1 – namely, the independence of  $\varepsilon_d$  from  $x_d$  for each  $d = 1, \dots, D$ . However, as we argue there, the identification of the joint distribution of  $\varepsilon$  does rely on Assumption 1 even in lattice models. Thus, when one of the objectives is to identify the joint distribution of errors (motivation for this is discussed later) then our Assumption 1 is no more restrictive than what is required in lattice models.

## 5.1 Advantages of non-lattice models over lattice models

Next, we describe how general rectangular structure models offer significant advantages over lattice models in capturing complex interdependencies between discrete choices. Even in the simplest possible case of a bivariate model with 2 discrete choices in each dimension we can talk about at least four different empirical structures. These arise from the interactions between complementarity or substitutability in decision thresholds (the first layer) and complementarity or substitutability in unobservables (the second layer), the latter captured through positive or negative dependence that reflects whether shocks reinforce or offset one another.

In many applications, substitutability/complementarity relationships at each layer may be unknown *a priori* and need to be identified from the available data.

**Cross-partial effects: Partial effects in one dimension can depend on covariates exclusive to other processes** In another aspect, under Assumption 1, models with general rectangular structures allow the likelihood  $P(Y^{c_d} \geq y_j^{(d)} | x)$  that an individual selects at least a certain level of commitment in one decision area, given all relevant personal and contextual factors, depend on covariates exclusive to other processes (e.g.,  $x_h, h \neq d$ ), enabling analysis of partial effects  $\frac{\partial P(Y^{c_d} \geq y_j^{(d)} | x)}{\partial x_{h,m}}$ . This effect is *indirect* – covariates exclusive to processes in other

dimensions affect the probabilities of decision in a given dimension *indirectly* through their influence on latent utilities in other dimensions. In lattice models under Assumption 1 such partial effects are absent (that is, zero).

Consider, for example, high school students choosing academic effort ( $Y^{c_1}$ : low = 1, medium = 2, high = 3) obtained from hours of study and extracurricular activity participation ( $Y^{c_2}$ : low = 1, moderate = 2, high = 3) e.g. expressed through involvement in sports or clubs. These choices are interdependent at the layer of the decision structure with the direction of that interdependence unknown a priori. Indeed, high academic effort may limit time for extracurriculars, but at the same time extensive extracurricular involvement may encourage academic effort for college applications. Covariates exclusive to  $Y^{c_1}$  could be parental education and access to tutors, covariates exclusive to  $Y^{c_2}$  could be school resources and peer involvement. Covariates shared by both  $Y^{c_1}$  and  $Y^{c_2}$  could be socioeconomic status and school quality. A non-lattice model would e.g. allow school resources (exclusive to  $Y^{c_2}$ ) affect the probability of high academic effort. It may show for example that better school resources (such as sports facilities) increase the probability of high academic effort by motivating students to balance high academic effort for college admissions. This would be informative for resource allocation policies. A lattice model would miss that effect.

To illustrate this theoretically, take the model in Figure 4 and note that

$$\begin{aligned} P(Y^{c_1} \leq 1|x) &= P(Y^{c_1} \leq 1, Y^{c_2} = 1|x) + P(Y^{c_1} \leq 1, Y^{c_2} = 2|x) \\ &= F\left(\alpha_{1,1}^{(1)} - x_1\beta_1, \alpha^{(2)} - x_2\beta_2\right) + F_1\left(\alpha_{1,2}^{(1)} - x_1\beta_1\right) - F\left(\alpha_{1,2}^{(1)} - x_1\beta_1, \alpha^{(2)} - x_2\beta_2\right) \end{aligned}$$

With the lattice structure ( $\alpha_{1,1}^{(1)} = \alpha_{1,2}^{(1)}$ ), only the middle term in this representation is left. Therefore, with the lattice stricture,  $\frac{\partial P(Y^{c_1} \leq 1|x)}{\partial x_{2,m}} = 0$ , where  $x_{2,m}$  is a covariate exclusive to the process  $Y^{c_2}$ . Thus, cross-partial effects are not possible in the lattice model.

With the non-lattice structure,

$$\frac{\partial P(Y^{c_1} \leq 1|x)}{\partial x_{2,m}} = -\beta_{2,m} \frac{\partial F\left(\alpha_{1,1}^{(1)} - x_1\beta_1, \alpha^{(2)} - x_2\beta_2\right)}{\partial e_2} + \beta_{2,m} \frac{\partial F\left(\alpha_{1,2}^{(1)} - x_1\beta_1, \alpha^{(2)} - x_2\beta_2\right)}{\partial e_2}$$

is not necessarily 0 because of  $\alpha_{1,1}^{(1)} \neq \alpha_{1,2}^{(1)}$ . In contrast, under Assumption 1, lattice models restrict  $P(Y^{c_d} \geq y_j^{(d)}|x) = P(Y^{c_d} \geq y_j^{(d)}|x_d)$ , ignoring cross-process effects.

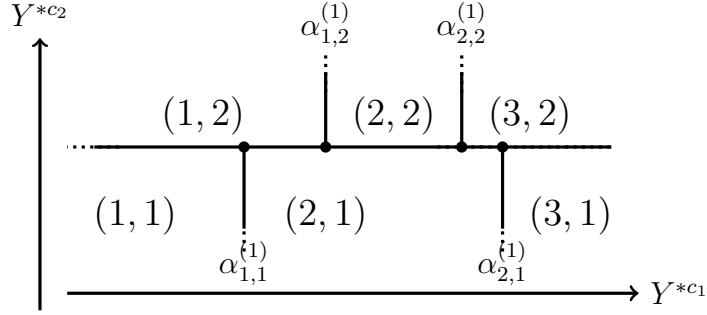


FIGURE 4: Example of a  $3 \times 2$  bivariate model

**Partial and cross-partial effects can vary in sign across the domain.** Let us first illustrate this property for cross-partial effects. Using the model in Figure 4 and the formula for  $\frac{\partial P(Y^{c_1} \leq 1|x)}{\partial x_{2,m}}$  above, we can see that this partial effect is guaranteed to be positive with a positive probability (and non-negative a.e.) iff  $\beta_{2,m}(\alpha_{1,2}^{(1)} - \alpha_{1,1}^{(1)}) > 0$ . In a completely analogous way, we can consider  $\frac{\partial P(Y^{c_1} \leq 2|x)}{\partial x_{2,m}}$  and obtain this partial effect is guaranteed to be positive with a positive probability (and non-negative a.e.) iff  $\beta_{2,m}(\alpha_{2,2}^{(1)} - \alpha_{2,1}^{(1)}) > 0$ . Since  $\alpha_{2,2}^{(1)} - \alpha_{2,1}^{(1)} < 0$  and  $\alpha_{1,2}^{(1)} - \alpha_{1,1}^{(1)} > 0$ , partial effects  $\frac{\partial P(Y^{c_1} \leq 1|x)}{\partial x_{2,m}}$  and  $\frac{\partial P(Y^{c_1} \leq 2|x)}{\partial x_{2,m}}$  will have different signs. In contrast, in lattice models the sign of such cross-partial effect would remain consistent across the whole domain,

Let us now look at the partial effects with respect to own covariates. If  $x_{d,m}$  is exclusive to  $Y^{*c_d}$ , then the own partial effect  $\frac{\partial P(Y^{c_d} \leq y_j^{(d)}|x)}{\partial x_{d,m}}$  will have the same sign in a non-lattice model for any  $y_j^{(d)}$ ,  $j = 1, \dots, M_d$ . The situation is different if  $x_{d,m}$  is shared with another latent process. Using our model in Figure 4, suppose  $x_{1,m} = x_{2,m_2}$  and obtain that in this case,

$$\begin{aligned} \frac{\partial P(Y^{c_1} \leq y_j^{(1)}|x)}{\partial x_{1,m}} &= -\beta_{1,m} \frac{\partial F(\alpha_{j,1}^{(1)} - x_1\beta_1, \alpha^{(2)} - x_2\beta_2)}{\partial e_1} - \beta_{2,m_2} \frac{\partial F(\alpha_{j,1}^{(1)} - x_1\beta_1, \alpha^{(2)} - x_2\beta_2)}{\partial e_2} \\ &\quad - \beta_{1,m} f_1(\alpha_{j,2}^{(1)} - x_1\beta_1) + \beta_{1,m} \frac{\partial F(\alpha_{j,2}^{(1)} - x_1\beta_1, \alpha^{(2)} - x_2\beta_2)}{\partial e_1} + \beta_{2,m_2} \frac{\partial F(\alpha_{j,2}^{(1)} - x_1\beta_1, \alpha^{(2)} - x_2\beta_2)}{\partial e_2}. \end{aligned}$$

As an example, consider a special case of independent  $\varepsilon_1$  and  $\varepsilon_2$ . Then

$$\begin{aligned} \frac{\partial P(Y^{c_1} \leq y_j^{(1)}|x)}{\partial x_{1,m}} &= \beta_{2,m_2} f_2(\alpha^{(2)} - x_2\beta_2)(F_1(\alpha_{j,2}^{(1)} - x_1\beta_1) - F_1(\alpha_{j,1}^{(1)} - x_1\beta_1)) \\ &\quad - \beta_{1,m}(f_1(\alpha_{j,1}^{(1)} - x_1\beta_1)F_2(\alpha^{(2)} - x_2\beta_2) \\ &\quad + f_1(\alpha_{j,2}^{(1)} - x_1\beta_1)(1 - F_2(\alpha^{(2)} - x_2\beta_2))). \end{aligned}$$

If  $\beta_{2,m_2}(\alpha_{j,2}^{(1)} - \alpha_{j,1}^{(1)})$  and  $\beta_{1,m}$  have the same sign, then we have the difference of either two

positive or two negative terms. Each of them can potentially dominate the other depending on the values of indices (and, hence,  $x_1$  and  $x_2$ ), thresholds, as well as  $\beta_{1,m}$  and  $\beta_{2,m_2}$ .

The feature of own partial effects with respect to a shared covariate or cross-partial facts not having the same sign across the domain complicates the identification analysis.

## 5.2 Identification in the semiparametric model

Next, we establish identification in semiparametric models with non-lattice structures. The knowledge of index parameters and both layers of dependence – through threshold parameters and the joint c.d.f. – is central to counterfactual analysis and policy design involving joint outcomes such as household decisions on, say, healthcare and education investments. The double-layer dependence structure will e.g. determine whether bundled interventions reinforce or crowd out each other.

The approach used in lattice models in [Komarova and Matcham \(2025\)](#) for the identification of index parameters and threshold differences relies on the ability to isolate different dimensions and consider one dimension at a time. This is not going to work here as we cannot isolate different dimensions. E.g., from Figure 4 one can see that  $P(Y^{(1)} \leq y_j^{(1)} | x)$  for  $j = 1, 2$ , cannot be expressed just in terms of the index  $x_1\beta_1$  and the marginal c.d.f.  $F_{1,\leq}$ . General rectangular structure cases therefore require a different approach to identification. Intuitively, the identification of parameters  $\beta_d$  and the threshold structure in these models should be more demanding on the data compared to lattice models, especially given an unknown dependence structure of unobservables. This is indeed the case until we get to the stage of identifying the joint c.d.f. of unobservables. At that stage, as follows from Theorem 3 here and Theorem 4 in [Komarova and Matcham \(2025\)](#) both lattice and non-lattice models become similarly demanding on the data, which is an interesting result.

After introducing a helpful definition and notations, we proceed to prove identification in several steps as follows:

1st step: identification of parameters corresponding to exclusive covariates in each process (Theorem 1).

2nd step: identification of parameters corresponding to non-exclusive covariates (Theorem 2).

3rd step: identification of the joint c.d.f. (Theorem 3).

4th step: identification of the thresholds (Theorem 4).

**Definition 3** Covariate  $x_{d,i}$  is exclusive to process  $d$  if  $x_{d,i} | x_{-d}$  has a non-degenerate distribution almost everywhere for  $x_{-d} \equiv (x_1, \dots, x_{d-1}, x_{d+1}, \dots, x_D)$ .

**Notation 2** For each  $d = 1, \dots, D$ , let  $x_{d,1:L_d}$  denote the subvector of  $x_d$  that consists of all the covariates in  $x_d$  that are exclusive to the process  $Y^{*c_d}$ .

Intuitively, an exclusive covariate is one that contains information unique to the  $d$ -th process  $Y^{*c_d}$  and cannot be perfectly predicted from the covariates associated with other processes. It is without a loss of generality that these exclusive covariates are arranged to be the first few covariates within  $x_d$ .

**Notation 3** Due to the presence of shared covariates among processes, the vector  $x = (x_1, \dots, x_D)$  may contain several identical variables. Its effective dimension will count each of these shared covariates only once. Let us denote the effective dimension of  $x$  as  $K$ . In other words,  $K$  is the number of exclusive covariates across all processes plus the number of covariates shared by at least two processes (and counted only once in this effective dimension).

Theorem 1 gives sufficient conditions for the identification of  $\beta_{d,1:L_d}$ ,  $d = 1, \dots, D$ , which are the parameters corresponding to the exclusive covariates in each process.

**Theorem 1** Consider a  $D$ -variate ordered discrete response model with the index structure (6). Suppose Assumption 1 holds for each  $d = 1, \dots, D$ , and the model has a coherent general rectangular structure. Suppose that the following conditions are satisfied:

- (a)  $L_d \geq 1$  for each  $d = 1, \dots, D$ .
- (b) The coefficient  $\beta_{d,1}$  corresponding to  $x_{d,1}$  in  $x_d \beta_d$  is 1,  $d = 1, \dots, D$ .
- (c) For each  $d = 1, \dots, D$ , there exists  $j_d = 1, \dots, M_d - 1$ , such that  $P(S_d(j_d)) > 0$ ,

$$\text{where } S_d(j_d) = \left\{ x : 0 < P(Y^{c_d} \leq y_{j_d}^{(d)} | x) < P(Y^{c_d} \leq y_{j_d+1}^{(d)} | x) \right\}.$$

In addition,  $S_d(j_d)$  contains a Cartesian product  $(\underline{x}_{d,1}, \bar{x}_{d,1}) \times S_{d,-1}(j_d)$ , where  $\bar{x}_{d,1} > \underline{x}_{d,1}$  and  $S_{d,-1}(j_d) \subset \mathbf{R}^{K-1}$ , such that  $P((\underline{x}_{d,1}, \bar{x}_{d,1}) \times S_{d,-1}(j_d)) > 0$  (the order of covariates in this Cartesian product coincides with the order of covariates in  $x$ ) and  $S_{d,-1}(j_d)$  is not contained in any proper linear subspace of  $\mathbf{R}^{K-1}$ .

Then parameters  $\beta_{d,1:L_d}$ ,  $d = 1, \dots, D$ , corresponding to the exclusive covariates in each process are identified.

Condition (a) states that each process has at least one exclusive covariate, and condition (b) effectively states that the first (and potentially only) exclusive covariate in process  $Y^{*c_d}$  has a non-zero coefficient; it further normalizes it to 1 (alternatively, could be normalized to  $-1$  if the impact is negative). Normalization restrictions like these are standard in semiparametric models where parameter vectors generally can only be identified up to scale. These normalizations can be different across  $d$  (some normalized to 1, some to  $-1$ ). Condition (c) is a version of the rank condition and, intuitively, requires that for  $d = 1, \dots, D$ , there is some continuous variation in at least one exclusive covariate in  $x_d$ , conditional on other covariates, at least in that part of domain that gives non-trivial (and, thus, informative) probabilities of choice with respect to dimension  $d$ . One of the requirements is that  $Y^{c_d}$  takes at least two different values with positive probabilities. In condition (c) for simplicity we took them to be two different consecutive values  $y_{j_d}^{(d)}$  and  $y_{j_d+1}^{(d)}$  (more generally, they don't need to be consecutive).

Our next result is on the identification of those parameters components that correspond to shared regressors. It is given in Theorem 2 and relies on strengthening conditions on exclusive covariates to have a large enough support. The result leverages the multidimensional nature of the problem and the ability to consider probabilities  $P(\cap_{d=1}^D (Y^{c_d} \kappa_d y_{j_d}^{(d)}) | x)$ , where  $\kappa_d \in \{\leq, >\}$ . In bivariate models the regions inside these probabilities are easy to visualize in the latent space as constructed using rectangles starting from one “corner” of partitioning structure. Note that large (or large enough) support assumptions are common in the semiparametric literature and, in particular, in semiparametric univariate ordered response models (see e.g. Manski (1985, 1988); Horowitz (2010); Lewbel (2000, 2003)).

**Theorem 2** Suppose all the conditions of Theorem 1 hold. Also assume that:

- (a) there is a collection of indices  $(j_1, \dots, j_d)$  such that the intersection  $S = \cap_{d=1}^D S_d(j_d)$  has positive probability measure and full affine dimension  $K$ ;

(b) for each  $d$  either  $\underline{x}_{d,1}$  is small enough to guarantee that  $\alpha_{j_1, \dots, j_D}^{(d)} - \underline{x}_{d,1} - x_{d,-1}\beta_{d,-1}$  is at the upper support point of the  $\varepsilon_d$  distribution, or  $\bar{x}_{d,1}$  is large enough to guarantee that  $\alpha_{j_1, \dots, j_D}^{(d)} - \bar{x}_{d,1} - x_{d,-1}\beta_{d,-1}$  is at the lower support point of the  $\varepsilon_d$  distribution for  $x_{d,-1} \in S_{d,-1}(j_d)$ .

Then  $\beta_d$ ,  $d = 1, \dots, D$ , are identified.

Condition (b) Theorem 2 can be reformulated in terms of observed choice probabilities (and, thus, verified in practice) where one would need to check that some of them can attain one of its natural bounds (either 0 or 1) through the variation in an exclusive covariates with other covariates taking values in some subset of a positive measure. Note that Theorem 2 does not rely on the result of Theorem 1 as its proof does not use the fact that all parameters for exclusive covariates have been identified and establishes the identification of the whole vector  $\beta_d$ , including  $\beta_{d,1:L_1}$  independently of Theorem 1. It is instructive though to have Theorem 1 as a separate result to emphasize that the identification of exclusive covariates' parameters requires weaker conditions.

Our next result is on the identification of the joint distribution of  $\varepsilon$ . It is enough to identify one function  $F_{\kappa_1 \dots \kappa_D}$  to fully characterize this distribution. We can identify the distribution from one of the “corners” in our partitioning that gives us enough variation in probabilities.

**Theorem 3** Suppose all the conditions of Theorem 2 hold for a collection of indices  $(j_1, \dots, j_D)$  such that  $j_d = 1$  or  $j_d + 1 = M_d$  for each  $d$ .

Additionally, in condition (b) of Theorem 2 suppose that for each  $d = 1, \dots, D$  both of the following conditions hold:  $\underline{x}_{d,1}$  is small enough to guarantee that  $\alpha_{j_1, \dots, j_D}^{(d)} - \underline{x}_{d,1} - x_{d,-1}\beta_{d,-1}$  is at the upper support point of the  $\varepsilon_d$  distribution, and  $\bar{x}_{d,1}$  is large enough to guarantee that  $\alpha_{j_1, \dots, j_D}^{(d)} - \bar{x}_{d,1} - x_{d,-1}\beta_{d,-1}$  is at the lower support point of the  $\varepsilon_d$  distribution for  $x_{d,-1} \in S_{d,-1}(j_d)$  (this is in contrast with either/or required in Theorem 2).

Then, under the normalization  $F_{d,\leq}(e_{0d}) = c_{0d}$  for each marginal c.d.f.  $F_{d,\leq}$  for some known  $e_{0d}$  and  $c_{0d} \in (0, 1)$ ,  $d = 1, \dots, D$ , the distribution of  $\varepsilon$  is identified.

Conditions on covariates in Theorem 3 first identify marginal distributions up to a shift and then, coupled with the normalization restrictions, fully identify them. In addition, threshold

parameters of the “corner” of the partitioning structure in the latent space implicitly specified in the formulation of Theorem 3 (through the values of indices  $j_d$ ,  $d = 1, \dots, D$ ) are identified. Then the observed probabilities of that “corner” region together with the knowledge of thresholds defining it identify the joint distribution of  $\varepsilon$ .

Note that conditions in Theorems 1-3 are increasingly more restrictive. This is not surprising as, first, when thinking of Theorem 2 vs Theorem 1 conditions, it is intuitive that the identification of the parameters corresponding to shared covariates is harder due to multiple effects occurring together when this covariate varies. When comparing Theorem 3 with Theorem 2 conditions, we see that in the former conditions are more restrictive in requiring that there is sufficient variation in covariates at the boundary of the partitioning of the latent space (namely, in one of the “corners”).

Our final result is on the identification of threshold parameters. This result allows us to find out whether decision-making is consistent with broad bracketing or narrow bracketing. Identification comes from variation in covariates and consideration of probabilities of various rectangular regions, which can be expressed in terms of  $F_{\kappa_1, \dots, \kappa_D}$ . Theorem 4 gives a formal identification result for the thresholds. It strengthens previous conditions by essentially requiring that for any rectangle  $R_{j_1, \dots, j_D}$  there is a positive mass of  $x$  that delivers a strictly positive choice probability for this rectangle (in contrast, Theorem 3 only required that to apply in one of the “corner” regions).

**Theorem 4** *Suppose all the conditions of Theorem 3 hold for any collection of indices  $(j_1, \dots, j_D)$  with  $j_d \in \{1, M_d - 1\}$ ,  $d = 1, \dots, D$ . Then all the thresholds  $\alpha_{q_1, \dots, q_{d-1}, q_d, q_{d+1}, \dots, q_D}^{(d)}$  are identified.*

We prove the identification of thresholds in Theorem 4 sequentially in a manner somewhat consistent with solving a puzzle and it is best illustrated in the bivariate case in Figure 5. In Stage 1, thresholds are identified that define “corner” regions  $R_{q_1, \dots, q_D}$  with  $q_d \in \{1, M_{d-1}\}$ . The result of this step is given in Panel A. In Stage 2, border regions are considered and for any  $d = 1, \dots, D$ , the thresholds  $a^{(d)q_1, \dots, q_{d-1}, q_d, q_{d+1}, \dots, q_D}$  are identified when  $q_h$ ,  $h \neq d$ , remains fixed at its value 1 or  $M_h - 1$  whereas  $q_d$  varies from 2 to  $M_h - 2$ . In the bivariate case the thresholds identified after Stage 2 are in Panel B in Figure 5 as dotted lines (dotted because their length is not known). Stage 3 continues to consider border regions for each  $d = 1, \dots, D$  and identifies thresholds  $a_{q_1, \dots, q_{d-1}, q_d, q_{d+1}, \dots, q_D}^{(d)}$  for when  $q_d$  remains fixed at its value 1 or  $M_d - 1$  whereas  $q_h$ ,

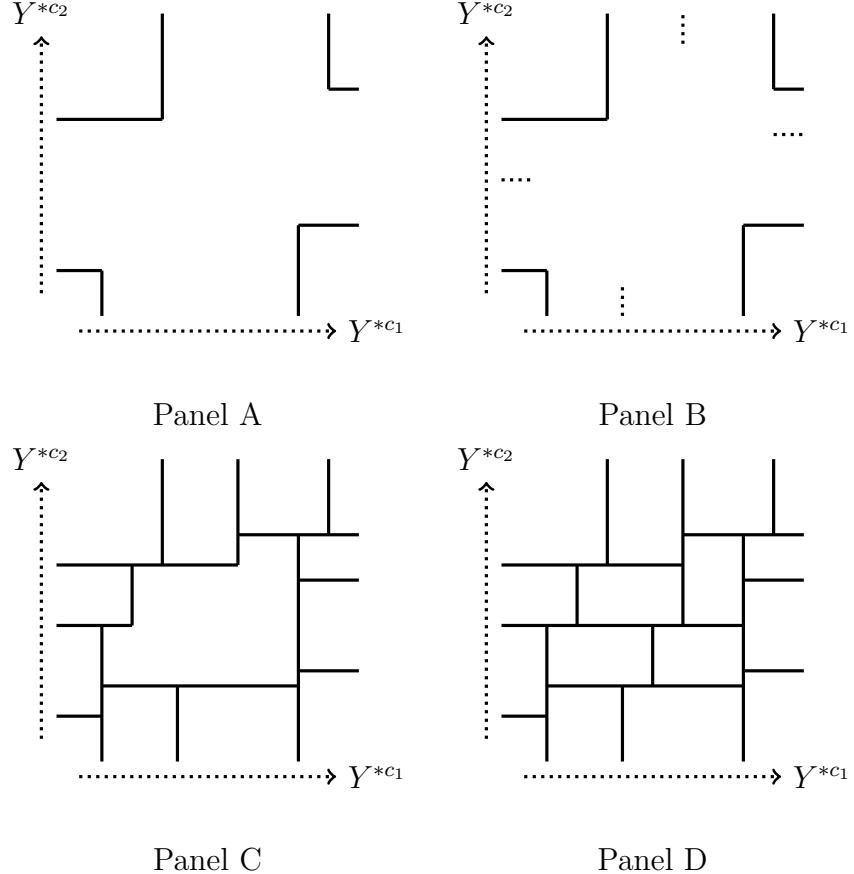


FIGURE 5: Stages of threshold system identification in the bivariate case.

$h \neq 2$ , vary from 2 to  $M_h - 2$ . In the bivariate case the thresholds identified after Stage 3 are in Panel C in Figure 5 (the thresholds from Stage 2 are now in solid lines at their lengths are known). Stage 4 identifies all the thresholds “in the middle” proceeding sequentially from the rectangular regions close to the border further into the depth of partitioning. Each stage builds on the results of the previous stage. Importantly, the sequential nature of the threshold identification process ensures that at each phase there are at most  $D$  unknown thresholds (out of the overall  $2D$  thresholds forming a rectangle of interest) that need to be identified. Identification of yet unknown thresholds is obtained through a variation in  $D$  indices  $x_d\beta_d$ ,  $d = 1, \dots, D$ , and the knowledge of  $F_{\kappa_1, \dots, \kappa_D}$  for a suitable  $\kappa_1, \dots, \kappa_D$  (Theorem 3 implies the knowledge of  $F_{\kappa_1, \dots, \kappa_D}$  for any  $\kappa_1, \dots, \kappa_D$ ). Which  $F_{\kappa_1, \dots, \kappa_D}$  is suitable for the identification task of a particular rectangle depends on which thresholds forming this rectangle are already known and which still need to be identified (up to  $D$  of these).

Note that the only assumption we make about the support of  $\varepsilon$  is that it is convex. This support can be bounded, partially bounded (that is, bounded in some directions but not others), or

extend over the entire  $\mathbb{R}^D$ . The geometry of this support is closely related to what we require from the support of the exclusive covariates. For example, if the support of  $\varepsilon$  is bounded, then it is sufficient for the exclusive covariates to vary only within a finite range as well. Regardless of its specific form, our identification argument remains flexible as it can rely on areas near the finite boundary of the support, or on directions in which the support is unbounded, as long as those directions form a sufficiently large cone.

### 5.3 Estimation in a semiparametric model

The majority of existing estimation approaches for univariate semiparametric ordered response models (either under full stochastic independence of the unobservable from covariates or under a slightly more general formulation with a multiplicative scedastic function like in [Chen and Khan \(2003\)](#)) do not extend to multivariate models with general rectangular structures.<sup>8</sup> For example, in the two-stage approach of [Klein and Sherman \(2002\)](#), which first estimates the index parameter using kernel density estimates of the conditional choice probabilities and then identifies the threshold parameters through shift restrictions, adapting the shift restrictions to non-lattice settings proves challenging. The same limitation applies to [Liu and Yu \(2019\)](#). Similarly, the approaches proposed in [Lewbel \(2000, 2003\)](#) do not generalize to non-lattice frameworks. Moreover, the strategy of [Chen and Khan \(2003\)](#) cannot be directly applied to models with general rectangular structures to estimate all finite-dimensional parameters of interest. In such settings, equality of joint probabilities  $P(Y^{(1)} = y_{j_1}^{(1)}, Y^{(2)} = y_{j_2}^{(2)} | x_1, x_2) = P(Y^{(1)} = y_1^{(1)}, Y^{(2)} = y_1^{(2)} | \tilde{x}_1, \tilde{x}_2)$  only implies that  $\tilde{x}_d\beta_d = x_d\beta_d$  if and only if  $x_{-d} = \tilde{x}_{-d}$ , for  $d = 1, 2$ . Hence, their method may only be suitable for estimating parameters associated with exclusive covariates.

If we were interested in the estimation of parameters on exclusive covariates, we could proceed in many ways. We could combine pairwise differences ([Honoré and Powell, 2005](#)) with maximum rank correlation (MRC) estimation ([Han, 1987](#)) or any other method suitable for single-index models to estimate  $\beta_{d,1:L_d}$  for exclusive regressors. Pairwise differencing would restrict attention to comparisons where non-exclusive covariates and all other dimensions are similar, while MRC would exploit the stochastic dominance which is behind the result in [Theorem 1](#).

We find that the only existing approach that can be extended to non-lattice model and which permits the estimation of all the index parameters as well as all the thresholds and the unknown

---

<sup>8</sup>Most of these methods, however, can be extended to multivariate lattice models, as discussed in [Komarova and Matcham \(2025\)](#).

joint distribution of unobservables is [Coppejans \(2007\)](#), which, analogously to us, relies on the assumption of independence of the unobservable from covariates as well as enough variation in covariates. Focusing on the bivariate case for clarity, let us describe how we can extend [Coppejans \(2007\)](#) to our setting.

Consider a random sample  $\left\{ (y^{(1)(i)}, y^{(2)(i)}, x_1^{(i)}, x_2^{(i)}) \right\}_{i=1}^N$ . First, create an estimate of the probability of the bivariate latent process falling into the rectangle  $R_{j_1, j_2}$ :

$$\ell_{j_1, j_2}^{(i)} = \sum_{\ell_1=0}^1 \sum_{\ell_2=0}^1 (-1)^{\ell_1 + \ell_2} \widehat{F} \left( a_{j_1 - \ell_1, j_2}^{(1)} - x_1^{(i)} b_1, a_{j_1, j_2 - \ell_2}^{(2)} - x_2^{(i)} b_2 \right)$$

[Coppejans \(2007\)](#) deals with the univariate  $\widehat{F}$  and models it using a quadratic  $B$ -spline whose coefficients are estimated jointly with the index and threshold parameters. In the multivariate case, we can model  $\widehat{F}$  using tensor-product  $B$ -splines and estimate their coefficients jointly with thresholds and index parameters. The estimation proceeds by

$$\max_{\theta, \widehat{F}} \mathcal{L}(\theta) = \frac{1}{N} \sum_{i=1}^N \sum_{j_1=1}^{M_1} \sum_{j_2=1}^{M_2} \mathbf{1} \left[ (y^{(1)(i)}, y^{(2)(i)}) = (y_{j_1}^{(1)}, y_{j_2}^{(2)}) \right] \log(\ell_{j_1, j_2}^{(i)}),$$

where  $\theta$  combines all the index and threshold parameters. In the bivariate setting, the basis functions in tensor-product representation for  $\widehat{F}$  consist of  $S_1 \cdot S_2$  products  $\mathcal{R}_{1;s_1, S_1}(e_1; q_1) \cdot \mathcal{R}_{2;s_2, S_2}(e_2; q_2)$ ,  $s_1 = 1, \dots, S_1$ ,  $s_2 = 1, \dots, S_2$ , of univariate  $B$ -splines evaluated at specific  $(e_1, e_2)$ . Here  $q_d$  is the degree in dimension  $d = 1, 2$ . There is a system of knots in each dimension which is not explicitly incorporated in our notation. The full tensor-product  $B$ -spline for  $\widehat{F}(e_1, e_2)$  is a linear combination of these basis functions:

$$\sum_{s_1=1}^{S_1} \sum_{s_2=1}^{S_2} h_{s_1 s_2} \mathcal{R}_{1;s_1, S_1}(e_1; q_1) \mathcal{R}_{2;s_2, S_2}(e_2; q_2),$$

with coefficients  $\{h_{s_1 s_2}\}$  constrained to ensure valid c.d.f. properties. Specifically, (a) *monotonicity* in each dimension is enforced by

$$h_{s_1 s_2} \leq h_{s_1 + 1, s_2}, \quad s_1 = 1, \dots, S_1 - 1, \quad s_2 = 1, \dots, S_2,$$

$$h_{s_1 s_2} \leq h_{s_1, s_2 + 1}, \quad s_2 = 1, \dots, S_2 - 1, \quad s_1 = 1, \dots, S_1;$$

(b) c.d.f. *bounds* are maintained by  $0 \leq h_{s_1 s_2} \leq 1$  for all  $s_1, s_2$ .<sup>9</sup> Linear equality constraints on some  $h_{s_1 s_2}$  can also impose normalization restrictions on marginal distributions of unobservables. Coherency requires additional constraints on thresholds. In the bivariate case, these can be imposed via a penalty term added to the objective: e.g. in the form of

$$-\lambda_N \left( \alpha_{j_1+1, j_2}^{(1)} - \alpha_{j_1, j_2}^{(1)} \right)^2 \cdot \left( \alpha_{j_1, j_2+1}^{(2)} - \alpha_{j_1, j_2}^{(2)} \right)^2, \quad (7)$$

for a large  $\lambda_N > 0$ .

The distribution theory in [Coppejans \(2007\)](#) generalize as well with some obvious modifications required to make it applicable in multivariate non-lattice setting: regularity conditions and conditions on the growth of the *B*-spline based would need to be adjusted.

## 6 Parametric specification

The semiparametric framework offers a solid approach for achieving identification results despite model complexity, while also providing an estimation method that generalizes the approach of [Coppejans \(2007\)](#). However, in practice, researchers may opt for a parametric family for the joint distribution of unobserved  $\varepsilon$  due to computational convenience. This choice eliminates the need for nonparametric estimation of the joint distribution and typically reduces the data requirements for identifying all unknown parameters. These parametric families must, however, remain flexible to accommodate correlations among unobservables in the latent processes. Following established traditions in statistics and econometrics, natural choices for the joint distribution of unobservables are the multivariate normal distribution and a multivariate extension of a logistic distribution.

Even though parametric versions of the model will be identified under the conditions outlined in Section 5.2, intuitively, much weaker conditions ensuring sufficient variation (potentially discrete and/or exclusive covariates) should suffice for identification in the parametric case. A useful analogy is the comparison between a univariate single-index model with an unknown monotone link function and the probit model. The probit model requires only a finite number of points satisfying a rank condition, whereas the single-index model demands richer variation, often guaranteed by the presence of a continuous covariate, which is not required in the probit model.

---

<sup>9</sup>For more details on shape constraints in tensor-product B-splines, see [Bhattacharya and Komarova \(2024\)](#).

Adopting parametric assumptions within our general rectangular structure setting creates the following theoretical challenge for identification. While they may exist in theory, deriving clear-cut weaker conditions that do not involve exclusive or continuous covariates and are sufficient to identify all model parameters is complex. This difficulty mirrors the lack of straightforward identification conditions in multinomial probit or logit models that allow unknown correlations among different latent processes. We illustrate parametric identification without exclusive covariates through a numerical identification exercise.

Consider the case  $D = 2$ . Let  $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2)'$  be jointly normal with mean  $(0, 0)'$ , unit variances, and correlation  $\rho$ .<sup>10</sup> Denote its bivariate normal cumulative distribution function by  $\Phi_2(\cdot, \cdot; \rho)$ . Focusing on one of the “corner” regions (for concreteness, the south-west quadrant), identification of the  $k_1 + k_2 + 3$  parameters  $\beta_1, \beta_2, \rho, \alpha_{11}^{(1)}, \alpha_{11}^{(2)}$  can be studied using  $T \geq k_1 + k_2 + 3$  distinct covariate values  $x_i = (x_{i1}, x_{i2})'$  by writing a system of  $T$  equations in the  $k_1 + k_2 + 3$  unknowns: for  $i = 1, \dots, T$

$$\underbrace{P(Y^{(c_1)} = y_1^{(1)}, Y^{(c_2)} = y_1^{(2)} | x_i)}_{p_{\text{obs}}(x_i)} = \Phi_2(\alpha_{11}^{(1)} - x_{i1}' \beta_1, \alpha_{11}^{(2)} - x_{i2}' \beta_2; \rho), \quad (8)$$

where the left-hand side  $p_{\text{obs}}(x_i)$  is observable. Larger  $T$  or the presence of covariates exclusive to one of the processes can aid identification, but identification can proceed even when all covariates are shared, provided the  $x_i$  vary sufficiently.

To illustrate this, consider the case of each latent processes having a single covariate which is shared between the two processes. Then  $\theta = (\beta_1, \beta_2, \alpha_{11}^{(1)}, \alpha_{11}^{(2)}, \rho)$ , has five unknowns. We perform a numerical identifiability exercise around the true  $\theta_0 = (1.2, -0.8, 0.5, -0.2, 0.4)$ . Define the objective function for a candidate parameter  $\theta$  as

$$Q_T(\theta) = \sum_{i=1}^T \left( p_{\text{obs}}(x_i) - \Phi_2(\alpha_{11}^{(1)} - x_{i1}' \beta_1, \alpha_{11}^{(2)} - x_{i2}' \beta_2, \rho) \right)^2.$$

We compute  $p_{\text{obs}}(x_i)$  from the known  $\theta_0$ . The experiment is conducted for two designs: (i)  $T = 5$  design points drawn from the interval  $[-2, 2]$ ; and (ii)  $T = 10$  points obtained by adding five more draws from the same interval.

To explore the local geometry of  $Q_T$  around  $\theta_0$ , we reparametrize  $\rho$  as  $z = \text{atanh}(\rho)$  (so  $z \in \mathbb{R}$

---

<sup>10</sup>As usual, we normalize means and variances because shifts and scale changes yield observationally equivalent parameter vectors.

and  $\rho = \tanh(z) \in [0, 1]$ ), and then consider hyperspheres in this transformed parameter space. For a given radius  $r$  we sample 5000 directions on the sphere of Euclidean radius  $r$  about  $\theta_0$ ; for each sampled point  $\theta$  we evaluate  $Q_T(\theta)$  and record the *minimum* value found on that sphere. Two direction-sampling schemes are applied: *fixed-direction* sampling in which we draw random directions once and scale them to each radius  $r$ , and *random-sphere* sampling in which we draw new random directions independently for each radius  $r$ .

Figure 6 shows the resulting plots (log scale) of the minimum objective  $Q_T$  versus radius  $r$  for both sampling schemes (left: fixed-direction; right: random-sphere). These plots display how well the model discriminates the true parameter vector from alternatives at varying distances, and the extent to which this discrimination improves with the number of design points  $T$  (shown through the fact that  $Q_T(\theta)$  is increasing in  $T$  for all radii  $r$ ). The pronounced jitter visible in the random-sphere plot reflects sampling variability across radii.

**Estimation** To outline an estimation approach in the case of parametric assumptions on the distribution of unobservables, we continue with  $D = 2$  and  $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2)'$  being jointly normal with zero mean, unit variances, and correlation  $\rho$ . Given a random sample  $\left\{ (y^{(1)(i)}, y^{(2)(i)}, x_1^{(i)}, x_2^{(i)}) \right\}_{i=1}^N$  and collecting  $\beta_1, \beta_2, \rho$  and all the thresholds in  $\alpha$  in one parameter vector  $\theta$ , we construct the log-likelihood function

$$\begin{aligned} \mathcal{L}(\theta) &= \frac{1}{N} \sum_{i=1}^N \sum_{j_1=1}^{M_1} \sum_{j_2=1}^{M_2} \mathbb{1} \left[ (y^{(1)(i)}, y^{(2)(i)}) = (y_{j_1}^{(1)}, y_{j_2}^{(2)}) \right] \log(\ell_{j_1, j_2}^{(i)}(\theta)) = \frac{1}{N} \sum_{i=1}^N \log(\ell^{(i)}(\theta)), \\ \text{with } \ell_{j_1, j_2}^{(i)} &= \sum_{t_1=0}^1 \sum_{t_2=0}^1 (-1)^{t_1+t_2} \Phi_2 \left( \alpha_{j_1-t_1, j_2}^{(1)} - x_1^{(i)} \beta_1, \alpha_{j_1, j_2-t_2}^{(2)} - x_2^{(i)} \beta_2; \rho \right). \end{aligned}$$

Analogously to semiparametric model case, this log-likelihood needs to be maximized subject to a linear inequality constraints that describe ordering of thresholds in each dimensions, normalization constraints on the the thresholds, and non-linear equality constraints  $q(\alpha) = 0$  that collect coherency constraints (2) across all the local models.

The constrained maximum likelihood estimator (MLE)  $\hat{\theta}$  solves the optimization problem  $\max_{\theta} \mathcal{L}(\theta)$  subject to the described constraints on thresholds. The coherency constraints are differentiable and so under the typical MLE regularity conditions (e.g. Newey and McFadden (1994)), we have  $\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, V)$ , where  $V = BJB'$ , with  $J = \mathbb{E} \left[ \frac{\partial \log(\ell^{(i)}(\theta_0))}{\partial \theta} \frac{\partial \log(\ell^{(i)}(\theta_0))}{\partial \theta'} \right]$ ,

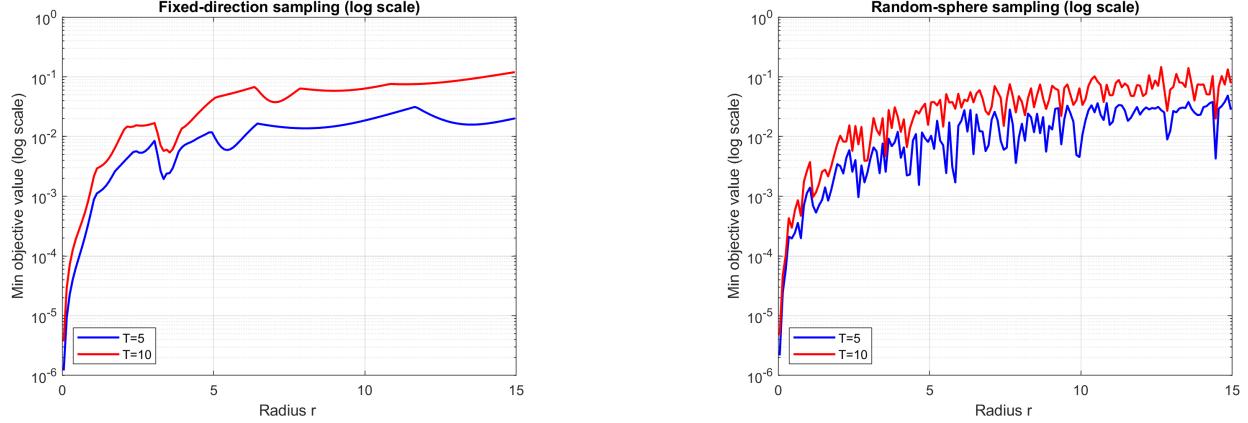


FIGURE 6: Numerical illustration of identification for the bivariate-normal model. The vertical axis (log scale) reports the minimum squared objective  $Q_T$  found on a sphere of radius  $r$  around the true parameter vector; the horizontal axis is  $r$ . Left: fixed-direction sampling. Right: random-sphere sampling. Curves are shown for  $T = 5$  and  $T = 10$  design points.

$B = J^{-1} - J^{-1}Q'(QJ^{-1}Q')^{-1}QJ^{-1}$ , and  $Q = \frac{\partial q(\theta_0)}{\partial \theta'}$ . The natural plug-in sample-analogue estimator of  $V$  provides a consistent estimator for the variance-covariance matrix.

## 7 Monte Carlo experiments

We now examine Monte Carlo simulations for the parametric case with normal errors, as outlined in Section 6. We compare the constrained maximum likelihood estimator described in Section 6 for models with general rectangular structures (we will refer to it from now as *non-lattice probit*) to the standard bivariate ordered probit estimator (that is, the estimator of a lattice model under normal errors). The baseline model is

$$Y^{*c_1} = x\beta_1 + w_1\gamma_1 + \varepsilon_1, \quad Y^{*c_2} = x\beta_2 + w_2\gamma_2 + \varepsilon_2,$$

where unobservables are independent of regressors and jointly normal with zero means and unit variances. This setup distinguishes exclusive and non-exclusive covariates. We explore scenarios with no exclusive covariates ( $\gamma_1 = \gamma_2 = 0$ ) and with an exclusive covariate in one latent process. The key findings are: (i) non-lattice model parameters may be estimated well without exclusive covariates, and (ii) using lattice instead of non-lattice models can yield inconsistent estimators for all parameters by ignoring broad bracketing in the decision process. Our simulations and applications show cases in which an expected positive correlation between unobservables is esti-

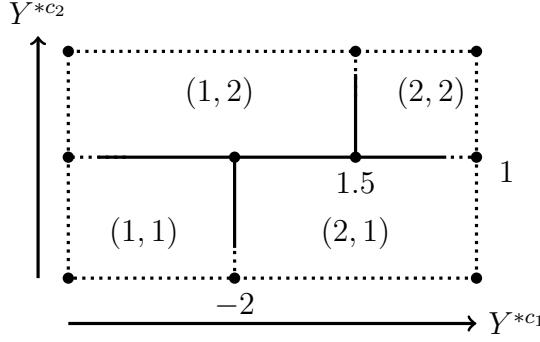


FIGURE 7: Latent variable space in Design 1

mated to be significantly negative under a lattice model. The inconsistency in index parameters depends on how well the lattice model approximates the true non-lattice model.

Each simulation design uses 250 independent random samples of size  $N = 5,000$ , with the penalty term in (7) set to  $N$ . Alternative  $N$  and  $\lambda_N$  values yield similar results.

### Design 1: $2 \times 2$ structure, no excluded regressors

We investigate parameter estimation in non-lattice probit models without exclusive covariates by setting  $\gamma_1 = \gamma_2 = 0$ , thereby removing  $w_1$  and  $w_2$ . We set  $\beta_1 = 1$ ,  $\beta_2 = 0.5$ ,  $\rho = 0.33$ , and use a  $2 \times 2$  non-lattice structure with thresholds  $\alpha_{11}^{(2)} = \alpha_{21}^{(2)} = 1$ ,  $\alpha_{11}^{(1)} = -2$ , and  $\alpha_{12}^{(1)} = 1.5$  (see Figure 7). We consider three distributions for the common regressor  $x$ : (Design 1A) uniform  $[-5, 5]$ , (Design 1B) discrete on 10 points  $\{\pm 5, \pm 3.5, \pm 2.5, \pm 1.5, 0, 0.5\}$  with equal probabilities, and (Design 1C) discrete on five points  $\{\pm 5, \pm 2.5, 0\}$  with equal probabilities.

Table 1 reports mean and standard deviation of parameter estimates for Design 1A. The non-lattice method estimates all parameters with minimal bias. The bivariate lattice ordered probit method estimates  $\beta_1$  and the first-dimension threshold well, but performs poorly for  $\beta_2$ ,  $\rho$ , and the second-dimension thresholds. Estimates of  $\rho$  are not very precise due to the absence of excluded regressors.

For Designs 1B and 1C, we estimate using only the non-lattice probit method. Results in Table 1 show that all parameters are estimated well, even with the minimal discrete variation in  $x$  in Design 1C. As expected, as a result of the limited variation in the common regressor, Design 1C exhibits higher standard deviations.

TABLE 1: Simulation results Design 1A

| Parameter                               | Truth | Non-lattice model          | Lattice model |
|---|-------|----------------------------|---------------|
| $\beta_1$                               | 1     | 1.00 (0.03)                | 0.77 (0.019)  |
| $\beta_2$                               | 0.5   | 0.50 (0.02)                | 0.00 (0.01)   |
| $\rho$                                  | 0.33  | 0.33 (0.12)                | -0.93 (0.02)  |
| $\alpha_{11}^{(2)} = \alpha_{21}^{(2)}$ | 1     | 1.00 (0.04)<br>1.00 (0.04) | 0.72 (0.04)   |
| $\alpha_{11}^{(1)}$                     | -2    | -1.99 (0.07)               | -0.42 (0.02)  |
| $\alpha_{12}^{(1)}$                     | 1.5   | 1.50 (0.08)                |               |

Notes: This table reports the sample mean and sample standard deviations (in parentheses) of the estimates of the Design 1A parameters, over 250 samples. The “Nonlattice model” column provides estimates from using the newly proposed nonlattice bivariate ordered probit model. The “Lattice model” column assumes a lattice structure, but estimates the two equations jointly.

TABLE 2: Simulation results Designs 1B and 1C

| Parameter                               | Truth | Design 1B                  | Design 1C                    |
|---|-------|----------------------------|------------------------------|
| $\beta_1$                               | 1     | 1.00 (0.03)                | 1.00 (0.03)                  |
| $\beta_2$                               | 0.5   | 0.50 (0.02)                | 0.50 (0.03)                  |
| $\rho$                                  | 0.33  | 0.34 (0.12)                | 0.32 (0.14)                  |
| $\alpha_{11}^{(2)} = \alpha_{21}^{(2)}$ | 1     | 1.00 (0.04)<br>1.00 (0.04) | 0.997 (0.058)<br>1.00 (0.05) |
| $\alpha_{11}^{(1)}$                     | -2    | -2.00 (0.07)               | -1.99 (0.12)                 |
| $\alpha_{12}^{(1)}$                     | 1.5   | 1.50 (0.08)                | 1.49 (0.14)                  |

Notes: This table reports the sample mean and sample standard deviations (in parentheses) of the estimates of the non-lattice model in Designs 1B and 1C, over 250 samples.

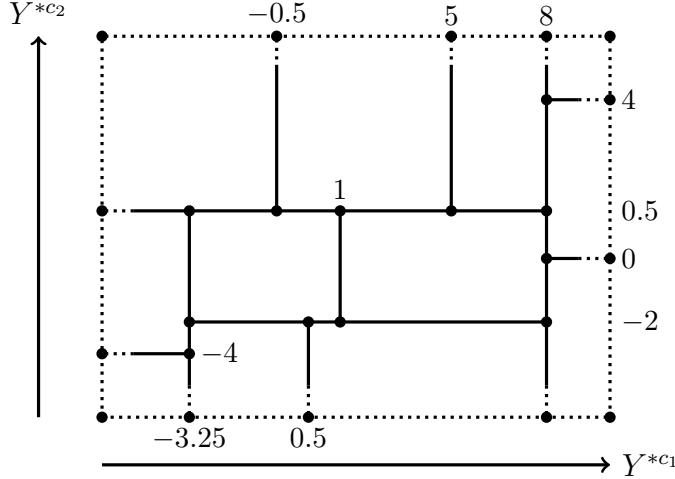


FIGURE 8: Latent variable space for two equations: Design 2

### Design 2: $4 \times 3$ with one excluded covariate

In the second simulation design, we extend the number of discrete values  $M_d$  in both dimensions. The discrete dependent variable  $Y^{c_1}$  can take four values and  $Y^{c_2}$  can take three values. This generates a  $4 \times 3$  non-lattice structure, illustrated in Figure 8. The common covariate  $x$  follows a uniform  $[-3, 3]$  distribution, though continuity here is not necessary. The covariate  $w_1$  is a discrete random variable taking values  $-2.5, -1.5, -0.5$  and  $0.5$  with equal probability. We set  $\gamma_2 = 0$  thus effectively removing  $w_2$  in the second equation. The parameter values are  $\beta_1 = 1.5, \gamma_1 = -4, \beta_2 = 3$  and  $\rho = 0.5$ .

Table 3 lists the across-simulation means and standard deviations of the index parameters and the correlation coefficient. Table 5 in the online supplement provides the values of the thresholds, together with their estimated means and standard deviations. The non-lattice bivariate ordered probit method estimates all the parameters with almost no bias. On the contrary, the lattice bivariate ordered probit method estimates the parameters with a relatively large bias. The mean squared errors in the non-lattice method are far lower than those in the lattice method for all of the parameters. Assuming a lattice structure makes estimating the correlation parameter  $\rho$  decidedly difficult, with the method failing to estimate the correct sign for  $\rho$ , let alone an approximately close value.

TABLE 3: Simulation results for Design 2

| Parameter  | Truth | Non-lattice model | Lattice model |
|------------|-------|-------------------|---------------|
| $\beta_1$  | 1.5   | 1.50 (0.04)       | 0.61 (0.01)   |
| $\gamma_1$ | -4    | -4.01 (0.09)      | -2.51 (0.04)  |
| $\beta_2$  | 3     | 2.99 (0.10)       | 1.64 (0.03)   |
| $\rho$     | 0.5   | 0.50 (0.06)       | -0.60 (0.03)  |

Notes: Sample means and sample standard deviations (in parentheses) of the estimates of the model parameters, over 250 repeated samples.

## 7.1 Cryptocurrency Familiarity and Optimism

In the first of two applications, we use data from the Survey of Consumer Payment Choice (SCPC) (Foster et al., 2021) to study opinions on future movements in cryptocurrency prices.<sup>11</sup> Conducted annually by the Federal Reserve Banks of Atlanta, Boston, and San Francisco, the SCPC tracks U.S. consumers' payment method adoption, recently noting a shift toward online and mobile payments due to the COVID-19 pandemic. Our sample includes 4,600 individuals (2015–2020), with data on demographics (e.g., income, age, gender, education), payment method use (e.g., credit cards, cryptocurrencies, mobile platforms like Google Pay), and perceptions of safety, convenience, and cost. Additional data cover fraud exposure, FICO score ranges, and household financial roles. See Foster et al. (2021) for details.

We examine whether opinions on bitcoin's future value are interdependent with cryptocurrency familiarity.  $Y^{c_1}$  is an ordered variable for familiarity with bitcoin (-1: not familiar, 0: slightly familiar, 1: somewhat familiar, 2: moderately/extremely familiar).<sup>12</sup> For  $Y^{c_2}$ , we use an ordered variable for bitcoin's expected value in one year (-1: decrease, 0: no change, 1: increase). We use a bivariate normal distribution specification for the vector of unobservables with zero mean and unit variances and an unknown correlation  $\rho$ .

## 8 Applications

Figure 9 shows the estimated threshold structure for a non-lattice model. It reveals varied thresholds. For example, individuals with low familiarity (blue dots) expect no value change at a lower threshold in the opinion dimension (i.e., lower  $Y^{Opinion}$  values yield a "no change" opinion com-

---

<sup>11</sup>See also Benetton and Compiani (2024) and Kahn and Linares Zegarra (2016).

<sup>12</sup>Moderate and extremely familiar are combined due to few respondents reporting extreme familiarity.

TABLE 4: Estimation coefficients: bitcoin familiarity and optimism

| Variable                        | O-Probit     | O-Probit     | Non-lattice  | Lattice      |
|---------------------------------|--------------|--------------|--------------|--------------|
| <i>Familiarity with Bitcoin</i> |              |              |              |              |
| LOW INCOME                      | -0.16 (0.06) |              | -0.11 (0.05) | -0.16 (0.06) |
| AGE                             | -0.02 (0.00) |              | -0.02 (0.00) | -0.01 (0.00) |
| MALE                            | 0.55 (0.06)  |              | 0.42 (0.06)  | 0.55 (0.06)  |
| LOW EDUCATION                   | -0.54 (0.09) |              | -0.40 (0.09) | -0.54 (0.09) |
| <i>Bitcoin “optimism”</i>       |              |              |              |              |
| LOW INCOME                      |              | 0.07 (0.06)  | -0.00 (0.06) | 0.07 (0.06)  |
| AGE                             |              | -0.01 (0.00) | -0.01 (0.00) | -0.01 (0.00) |
| MALE                            |              | -0.13 (0.05) | 0.02 (0.06)  | -0.13 (0.05) |
| LOW EDUCATION                   |              | 0.13 (0.07)  | -0.00 (0.08) | 0.13 (0.07)  |
| $\rho$                          | NA           | NA           | 0.84 (0.23)  | 0.03 (0.03)  |
| N                               | 1818         | 1818         | 1818         | 1818         |

Notes: Columns labeled “O-probit” provide estimates from univariate ordered probit models. The “Non-lattice” column provides estimates from using non-lattice bivariate ordered probit model. The “Lattice” column assumes a lattice structure, but estimates the two equations jointly.

pared to other familiarity groups). Conversely, those with high familiarity (red crosshatch) have closely spaced thresholds, defining a narrow ”no change” region. For this group, most  $Y^{\ast}Opinion$  values reflect strong opinions, either decreasing or increasing. These findings describe decision-making structures, not probabilistic outcomes.

Table 4 provides estimates of  $\beta$  and  $\rho$ . The correlation  $\rho$  ranges from 0.03 (lattice model) to 0.84 (non-lattice model), with coefficients differing by over 20% in magnitude. Notably, the coefficient on male changes sign – it is negative (and statistically significant at the 5% level) in the lattice model and positive (not statistically significant at the 5% level) in the non-lattice model. The lattice model thus suggests males are more pessimistic about bitcoin’s value as the effect as  $P(Y^{Optimism} \geq j|x_{-male}, x_{male} = 1) - P(Y^{Optimism} \geq j|x_{-male}, x_{male} = 0)$  is negative for any level  $j$  and any  $x_{-male}$ . As discussed earlier, a non-lattice model allows for changes in the signs of partial effects across the domain, therefore, the positive coefficient for males obtained there does not directly imply that this model suggests males are more optimistic for any level  $j$  and any  $x_{-male}$ . Additional post-estimation analysis we have conducted does confirm, however, that given the estimated thresholds the non-lattice model gives  $P(Y^{Optimism} \geq j|x_{-male}, x_{male} = 1) - P(Y^{Optimism} \geq j|x_{-male}, x_{male} = 0)$  as positive for any level  $j$  and any  $x_{-male}$  in the data, thus, giving a stable sign of this partial effect across the domain.<sup>13</sup>.

---

<sup>13</sup>A potentially different estimated threshold structure could have resulted in the switching of signs for this partial effect.

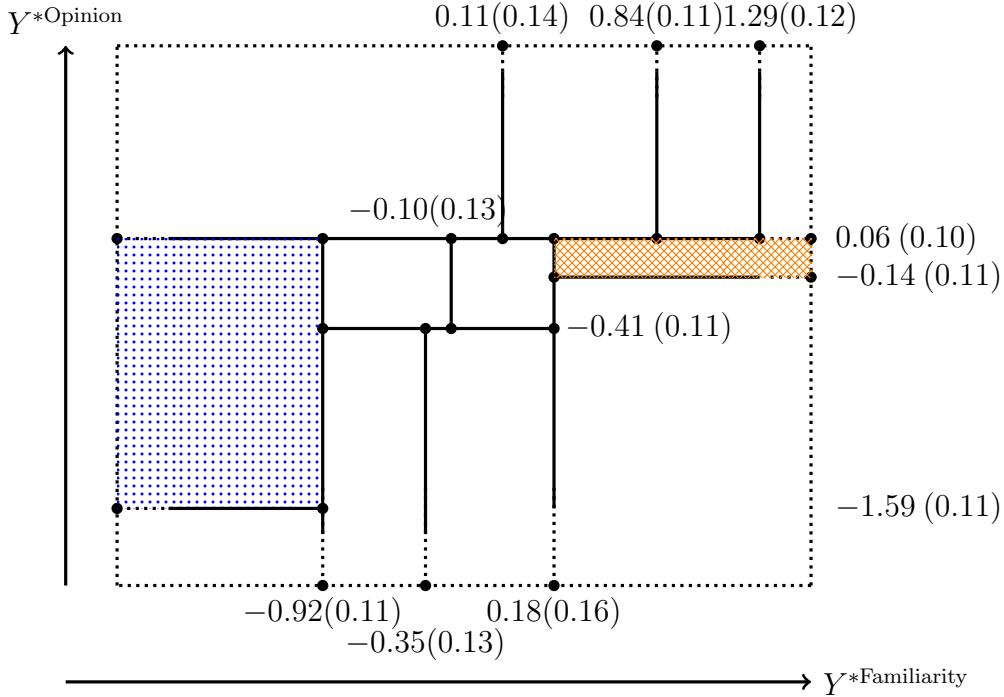


FIGURE 9: Estimates from cryptocurrency example when assuming a general rectangular structure model.

Another aspect illustrated by the thresholds in Figure 9 regarding the decision-making process is that individual decisions can be modeled sequentially using a binary decision tree, where each node represents a decision based on a single latent process. This structure can be termed a *hierarchical non-lattice model*, a specific subset of non-lattice models.<sup>14</sup> Hierarchical non-lattice models maintain coherence, as each node in the decision tree further refines the partitioning of the latent space. Figure 12 in the online supplement depicts the binary decision tree that outlines the estimated hierarchical decision-making process in this cryptocurrency application.

## 8.1 Identifying moral hazard and adverse selection in insurance markets

In the empirical analysis of asymmetric information in insurance markets, a highly influential framework was introduced by Chiappori and Salanie (2000), which proposed testing for adverse selection by estimating a *bivariate probit model* linking insurance coverage decisions and ex post risk realizations and taking a positive correlation between the latent errors of the insurance and

---

<sup>14</sup> Any model with a general rectangular structure can be represented by a decision tree, but typically, each node may involve multiple or all latent processes.

risk equations as evidence of asymmetric information. Subsequent research has emphasized the limitations of this framework, pointing out its inability to disentangle (*adverse*) *selection* from *moral hazard* because both forces contribute to the estimated positive correlation, but with different policy implications. As emphasized in [Cohen and Siegelman \(2010\)](#), “the disentanglement of adverse selection and moral hazard is probably the most significant and difficult challenge that empirical work on adverse selection in insurance markets faces.” Recent work has attempted to move beyond these limitations. On the theoretical side, a large literature has developed richer models of contract choice and risk response (e.g. [Einav et al., 2010](#), [Hendren, 2013](#)). Empirically, researchers have leveraged quasi-experimental variation or structural models to separately identify selection and moral hazard (e.g. [Handel, 2013](#), [Einav et al., 2013](#), [Hackmann et al., 2015](#), among many other). However, these approaches often require highly specific data environments or strong structural assumptions.

Our general rectangular structure framework provides an alternative empirical strategy. First, it allows utilization thresholds to vary with insurance status, thereby directly incorporating individuals’ behavioral (“moral hazard”) response to coverage into the econometric model. At the same time, it permits insurance-choice thresholds to depend on anticipated behavioral responses to insurance, accommodating what [Einav et al. \(2013\)](#) and others term “selection on moral hazard.” Of course, coherency still needs to be satisfied.

We apply this general rectangular structure to U.S. health insurance markets using the Medical Expenditure Panel Survey (MEPS), which offers nationally representative data on insurance coverage and healthcare utilization. Our sample includes approximately 60,000 individuals from 2005 to 2010, prior to the Affordable Care Act. Following the standard framework, we specify latent insurance and utilization equations as

$$Y_{\text{ins}}^* = x_{\text{ins}}\beta_{\text{ins}} + \varepsilon_{\text{ins}}, \quad Y_{\text{use}}^* = x_{\text{use}}\beta_{\text{use}} + \varepsilon_{\text{use}}. \quad (9)$$

Common covariates include demographics (logged income and its square, dividend payments, family size, logged hourly wage, age, education, gender, marital status, race, region, and year) and pre-existing conditions (diabetes, asthma, high blood pressure, high cholesterol, angina, heart attack, stroke, emphysema, arthritis). An excluded covariate for  $x_{\text{ins}}$  is partner’s job-provided coverage.

The insurance outcome  $Y_{\text{ins}} = 1$  if the individual holds private health insurance in January and

is 0 otherwise. Utilization  $Y_{\text{use}}$  is measured categorically (0 for no charges, 1 for below-median charges, 2 for above-median charges)

Moral hazard is isolated in the general rectangular structure by allowing utilization thresholds to depend on coverage. E.g.,  $Y_{\text{ins}} = 1$  lowers these thresholds if moral hazard is present. Insurance-dimension thresholds may potentially differ across utilization responses, provided coherency is maintained. One might argue, however, that due to the natural sequencing, where insurance coverage is selected first and utilization occurs second, the thresholds in the insurance dimension are invariant to utilization status. This restriction can be directly incorporated into the identification and estimation processes, automatically ensuring coherency while simplifying both identification and estimation. In this case, moral hazard is fully captured by differences in utilization thresholds across coverage levels. Once this behavioral effect is accounted for, the remaining correlation between  $\varepsilon_{\text{ins}}$  and  $\varepsilon_{\text{use}}$  can be interpreted as evidence of adverse or advantageous selection.

If, in contrast, the thresholds in the insurance-coverage dimension are permitted to vary with utilization status (in addition to utilization thresholds varying with insurance coverage), this accommodates the aforementioned “selection on moral hazard” where individuals may select coverage partly based on their anticipated behavioral (“moral hazard”) response to insurance. Then, the correlation between  $\varepsilon_{\text{ins}}$  and  $\varepsilon_{\text{use}}$  represents residual adverse selection, while moral hazard continues to be captured by lower utilization thresholds when  $Y_{\text{ins}} = 1$ .

We estimate the model with a general rectangular structure subject only to coherency constraints, thus potentially allowing for “selection on moral hazard” and letting the model and the data reveal to us in particular if that phenomenon exists or whether the choices can be considered to be sequential and consistent with the widely perceived natural timing of things.

Estimation results are presented in Figure 10. First, the estimated general rectangular structure model does reveal moral hazard through coverage-dependent thresholds. Utilization thresholds shift downward when  $Y_{\text{ins}} = 1$ : low-to-medium usage shifts from  $-0.02$  ( $0.04$ ) uninsured to  $-0.45$  ( $0.07$ ) insured, and medium-to-high from  $0.68$  ( $0.04$ ) to  $0.46$  ( $0.08$ ). Second, the results reveal no “selection on moral hazard” as the insurance coverage thresholds are estimated as invariant to utilization levels. The adverse selection is captured by the correlation coefficient  $\rho$  which drops from  $0.21$  ( $0.01$ ) in the lattice model to  $0.04$  ( $0.02$ ) in the non-lattice model, suggesting minimal adverse selection after accounting for moral hazard. Our findings align with a growing body of evidence from structural and experimental studies (e.g. [Einav et al., 2013](#), [Handel, 2013](#)), which generally find limited adverse selection once moral hazard is accounted for. Note that

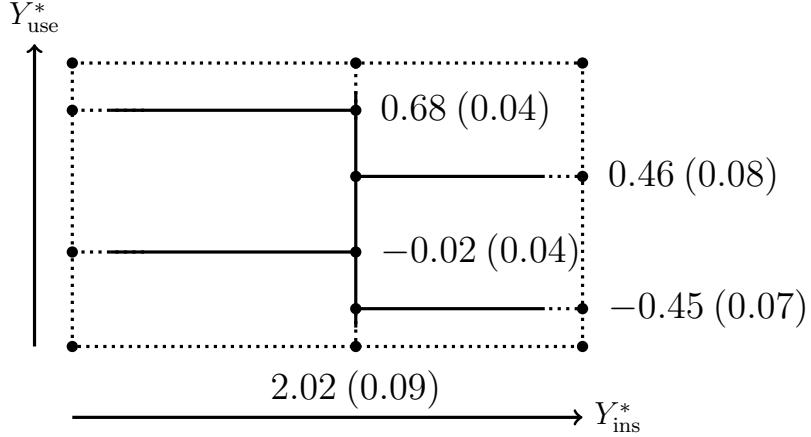


FIGURE 10: Estimated thresholds for insurance coverage ( $Y_{\text{ins}}^*$ ) and healthcare utilization ( $Y_{\text{use}}^*$ ), with standard errors in parentheses.

additionally the estimated structure allows for a cross-partial (indirect) effect of partner's job-provided coverage on the degree of utilization, even though this variable does not enter directly the latent process in the utilization dimension.

Thus, our general rectangular structure approach disentangles moral hazard from adverse selection while avoiding strong assumptions, using only observational data and flexible thresholds. It outperforms traditional lattice models and, because it can be applied with standard survey or administrative data, provides a powerful way to revisit much of the empirical literature built on lattice-based or ordered probit models.

## 9 Conclusion

This paper proposes a framework with general rectangular structures that extends the reach of traditional ordered response models, thereby providing new insights into economic decision-making processes. The framework captures interactions across dimensions in both decision rules and latent factors, with traditional lattice models as a special case. By formalizing coherency, deriving utility-based microfoundations, and proving identification, our framework enables realistic modeling of complex multidimensional choices. The approach reveals sign-changing partial effects, indirect covariate influences, and varying complementarity or substitutability, all of which are absent in lattice models. In doing so, it expands the econometric toolkit for studying multidimensional decisions and improves interpretability. As our empirical examples demonstrate, existing important economic contexts such as selection markets can be revisited (and new envi-

ronments can be traversed) with the models we have introduced.

## References

The numbers at the end of every reference link to the pages citing the reference.

- ANDERSON, J. A. (1984): “Regression and Ordered Categorical Variables,” *Journal of the Royal Statistical Society: Series B (Methodological)*, 46, 1–30. [6](#)
- APESTEGUIA, J. AND M. A. BALLESTER (2023): “Random utility models with ordered types and domains,” *Journal of Economic Theory*, 208, 105605. [6](#)
- ARADILLAS-LÓPEZ, A. AND A. M. ROSEN (2022): “Inference in ordered response games with complete information,” *Journal of Econometrics*, 226, 451–476. [7](#)
- BENETTON, M. AND G. COMPIANI (2024): “Investors’ Beliefs and Cryptocurrency Prices,” *The Review of Asset Pricing Studies*, 14, 197–236. [31](#)
- BERRY, S. AND P. REISS (2007): “Empirical Models of Entry and Market Structure,” *Handbook of Industrial Organization*, 3, 1845–1886. [7](#)
- BHAT, C. R. AND V. PULUGURTA (1998): “A comparison of two alternative behavioral choice mechanisms for household auto ownership decisions,” *Transportation Research Part B: Methodological*, 32, 61–75. [6](#)
- BHATTACHARYA, D. AND T. KOMAROVA (2024): “Incorporating Social Welfare in Program Evaluation and Treatment Choice,” *The Review of Economics and Statistics*, 106, forthcoming; volume and pages to be determined. [24](#)
- BOES, S. AND R. WINKELMANN (2006): “Ordered Response Models,” in *Modern Econometric Analysis*, Springer, chap. 12, 167–181. [6](#)
- CAMARA, M. (2021): “Computationally Tractable Choice,” *Working Paper*. [6](#)
- CAMERER, C., L. BABCOCK, G. LOEWENSTEIN, AND R. THALER (1997): “Labor Supply of New York City Cabdrivers: One Day at a Time,” *The Quarterly Journal of Economics*, 112, 407–441. [6](#)

CAMERON, S. V. AND J. J. HECKMAN (1998): “Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males,” *Journal of Political Economy*, 106, 262–333. [5](#)

CARNEIRO, P., K. HANSEN, AND J. HECKMAN (2003): “Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice,” *International Economic Review*, 44. [5](#)

CHEN, S. AND S. KHAN (2003): “Rates of convergence for estimating regression coefficients in heteroskedastic discrete response models,” *Journal of Econometrics*, 117, 245–278. [22](#)

CHESHER, A. AND A. M. ROSEN (2017): “Generalized Instrumental Variable Models,” *Econometrica*, 85, 959–989. [7](#)

——— (2020): “Structural modeling of simultaneous discrete choice,” *Cemmap Working Paper*. [7](#)

CHIAPPORI, P. AND B. SALANIE (2000): “Testing for Asymmetric Information in Insurance Markets,” *Journal of Political Economy*, 108, 56–78. [33](#)

CILIBERTO, F. AND E. TAMER (2009): “Market Structure and Multiple Equilibria in Airline Markets,” *Econometrica*, 77, 1791–1828. [7](#)

COHEN, A. AND P. SIEGELMAN (2010): “Testing for Adverse Selection in Insurance Markets,” *The Journal of Risk and Insurance*, 77, 39–84. [34](#)

COPPEJANS, M. (2007): “On efficient estimation of the ordered response model,” *Journal of Econometrics*, 137, 577–614. [4](#), [23](#), [24](#)

CUNHA, F., J. J. HECKMAN, AND S. NAVARRO (2007): “The Identification and Economic Content of Ordered Choice Models with Stochastic Thresholds,” *International Economic Review*, 48, 1273–1309. [5](#), [6](#)

EINAV, L., A. FINKELSTEIN, AND J. LEVIN (2010): “Beyond Testing: Empirical Models of Insurance Markets,” *Annual Review of Economics*, 2, 311–336. [34](#)

EINAV, L., A. FINKELSTEIN, S. P. RYAN, P. SCHRIMPF, AND M. R. CULLEN (2013): “Selection on Moral Hazard in Health Insurance,” *American Economic Review*, 103, 178–219. [34](#), [35](#)

FOSTER, K., C. GREENE, AND J. STAVINS (2021): “The 2020 Survey of Consumer Payment Choice: Summary Results,” *Federal Reserve Bank of Atlanta Research Data Reports*. [31](#)

GENTRY, M., T. KOMAROVA, AND P. SCHIRALDI (2023): “Preferences and Performance in Simultaneous First-Price Auctions: A Structural Analysis,” *The Review of Economic Studies*, 90, 852–878. [2](#)

GENTRY, M., T. KOMAROVA, P. SCHIRALDI, AND W. SHIN (2019): “On Monotone Strategy Equilibria in Simultaneous Auctions for Complementary Goods,” *Journal of Mathematical Economics*, 85, 68–81. [2](#)

GREENE, W. H. AND D. A. HENSHER (2010): *Modeling Ordered Choices: A Primer*, Cambridge University Press. [6](#)

HACKMANN, M. B., J. T. KOLSTAD, AND A. E. KOWALSKI (2015): “Adverse Selection and an Individual Mandate: When Theory Meets Practice,” *American Economic Review*, 105, 1030–1066. [34](#)

HAN, A. K. (1987): “Non-parametric analysis of a generalized regression model: The maximum rank correlation estimator,” *Journal of Econometrics*, 35, 303–316. [22](#)

HANDEL, B. R. (2013): “Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts,” *American Economic Review*, 103, 2643–2682. [34](#), [35](#)

HECKMAN, J. J. (1978): “Dummy Endogenous Variables in a Simultaneous Equation System,” *Econometrica*, 46, 931–959. [7](#)

HECKMAN, J. J., R. J. LALONDE, AND J. A. SMITH (1999): “The Economics and Econometrics of Active Labor Market Programs,” *Handbook of Labor Economics*, 3, Part A, 1865–2097. [5](#)

HENDREN, N. (2013): “Private Information and Insurance Rejections,” *Econometrica*, 81, 1713–1762. [34](#)

HERRNSTEIN, R. J. AND D. PRELEC (1991): “Melioration: A Theory of Distributed Choice,” *The Journal of Economic Perspectives*, 5, 137–156. [6](#)

HEYMAN, G. M. (1996): “Resolving the contradictions of addiction,” *Behavioral and Brain Sciences*, 19, 561–574. [6](#)

HONORE, B. E. AND A. DE PAULA (2010): “Interdependent Durations,” *The Review of Economic Studies*, 77, 1138–1163. [7](#)

HONORÉ, B. E. AND J. L. POWELL (2005): *Pairwise Difference Estimators for Nonlinear Models*, Cambridge University Press, 520–553. [22](#)

HOROWITZ, J. L. (2010): *Semiparametric and Nonparametric Methods in Econometrics*, Springer Series in Statistics, Springer-Verlag. [18](#)

KAHN, C. AND J. M. LINARES ZEGARRA (2016): “Identity Theft and Consumer Payment Choice: Does Security Really Matter?” *Journal of Financial Services Research*, 50, 121–159. [31](#)

KAHNEMAN, D. AND D. LOVALLO (1993): “Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking,” *Management Science*, 39, 17–31. [6](#)

KLEIN, R. W. AND R. P. SHERMAN (2002): “Shift Restrictions and Semiparametric Estimation in Ordered Response Models,” *Econometrica*, 70, 663–691. [22](#)

KOMAROVA, T. AND W. MATCHAM (2025): “Multivariate Ordered Discrete Response Models with Lattice Structures,” Working paper. [13](#), [16](#), [22](#)

LEWBEL, A. (2000): “Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables,” *Journal of Econometrics*, 97, 145–6177. [18](#), [22](#)

——— (2003): “Ordered Response Threshold Estimation,” *Unpublished Working Paper*. [5](#), [18](#), [22](#)

LIAN, C. (2020): “A Theory of Narrow Thinking,” *The Review of Economic Studies*, 88, 2344–2374. [6](#)

LIU, R. AND Z. YU (2019): “Simple Semiparametric Estimation of Ordered Response Models: with an Application to the Interdependence Duration Models,” *Tsukuba Economics Working Papers*. [22](#)

MANSKI, C. F. (1985): “Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator,” *Journal of Econometrics*, 27, 313–333. [18](#)

——— (1988): “Identification of Binary Response Models,” *Journal of the American Statistical Association*, 83, 729–738. [18](#)

NEWEY, W. K. AND D. MCFADDEN (1994): “Chapter 36: Large sample estimation and hypothesis testing,” *Handbook of Econometrics*, 4, 2111–2245. [26](#)

RABIN, M. AND G. WEIZSÄCKER (2009): “Narrow Bracketing and Dominated Choices,” *American Economic Review*, 99, 1508–43. [6](#)

READ, D., G. LOEWENSTEIN, AND M. RABIN (1999): “Choice Bracketing,” *Journal of Risk and Uncertainty*, 19, 171–197. [6](#)

SIMONSON, I. AND R. S. WINER (1992): “The Influence of Purchase Quantity and Display Format on Consumer Preference for Variety,” *Journal of Consumer Research*, 19, 133–138. [6](#)

TAMER, E. (2003): “Incomplete Simultaneous Discrete Response Model with Multiple Equilibria,” *The Review of Economic Studies*, 70, 147–165. [7](#)

THAKRAL, N. AND L. T. TÔ (2021): “Daily Labor Supply and Adaptive Reference Points,” *American Economic Review*, 111, 2417–43. [7](#)

THALER, R. H. (1999): “Mental accounting matters,” *Journal of Behavioral Decision Making*, 12, 183–206. [6](#)

TVERSKY, A. AND D. KAHNEMAN (1981): “The Framing of Decisions and the Psychology of Choice,” *Science*, 211, 453–458. [6](#), [7](#)

ZHANG, M. (2021): “A Theory of Choice Bracketing under Risk,” *Working Paper*. [6](#)

# Online Supplement

## 1 Appendix A: Coherency

Coherency plays a critical role in both our identification and estimation routines, as well as in the conceptual framing of the model. Specifically, we interpret the model as representing the behavior of a single decision maker, for whom incoherent (i.e., logically inconsistent) choices would be implausible.

Although the literature on strategic interaction often distinguishes between incoherency and incompleteness, we subsume both under the broader notion of incoherency. From a technical standpoint, we define a model with the given set of thresholds in the latent space as *coherent* (or *coherent in the latent space*) if the rectangles  $R_{j_1, \dots, j_D}$  defined in (1) form a partition of  $\mathbb{R}^D$ ; that is, they are mutually exclusive and collectively exhaustive over  $\mathbb{R}^D$ .

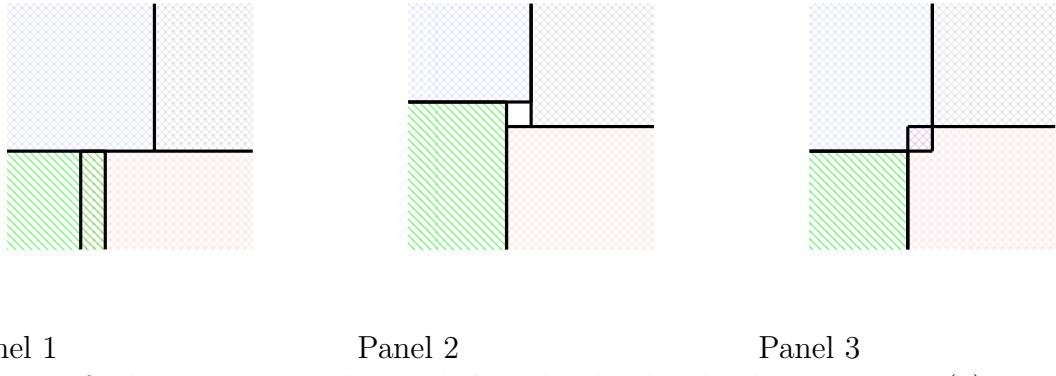
An equivalent way to characterize this notion of coherence is to ask: under what conditions on the latent thresholds does the observed coherency in choice probabilities and reflected in the fact that  $\sum_{d=1}^D \sum_{j_d=1}^{M_d} P(Y = (y_{j_1}^{(1)}, \dots, y_{j_D}^{(D)}) | x) = 1$  translate into coherence in the latent space? These conditions must be generic; that is, they should not depend on the distributional assumptions regarding observables or unobservables.

### 1.1 Bivariate case

We begin by examining the bivariate case,  $D = 2$ . The main result for this case is presented in Proposition 1, which states that a bivariate model with a general rectangular structure is generically coherent in the latent space if and only if it is locally hierarchical in every local configuration. Specifically, this involves examining each of four outcomes  $(y_{j_1}^{(1)}, y_{j_2}^{(2)})$ ,  $(y_{j_1+1}^{(1)}, y_{j_2}^{(2)})$ ,  $(y_{j_1}^{(1)}, y_{j_2+1}^{(2)})$  and  $(y_{j_1+1}^{(1)}, y_{j_2+1}^{(2)})$ , where we consider incremental moves from a given point  $(j_1, j_2)$  along one or both dimensions. The model must satisfy local hierarchies across all such configurations to ensure overall coherency.

*Proof of Proposition 1. Sufficiency.* First, observe that an incremental move from  $(j_1, j_2)$  in only one dimension – either to  $(j_1 + 1, j_2)$  or to  $(j_1, j_2 + 1)$  – cannot by itself generate incoherency. This is because the definition of the rectangles  $R_{j_1, j_2}$  in equation (1) ensures continuity at the

FIGURE 11: Potential violations of coherency.



Notes: Violation of coherency in Panel 1 is ruled out by the thresholds structure in (1). Violations of coherency in Panels 2 and 3 are not immediately ruled out by (1).

thresholds along single-dimensional moves. Specifically, in dimension 1, the rectangle  $R_{j_1, j_2}$  ends at the threshold  $\alpha_{j_1, j_2}^{(1)}$ , which simultaneously serves as the lower bound for the adjacent rectangle  $R_{j_1+1, j_2}$ . In other words, incoherency patterns such as the one illustrated in Panel 1 of Figure 11 are ruled out by construction. A similar argument holds in dimension 2: the rectangle  $R_{j_1, j_2}$  terminates at threshold  $\alpha_{j_1, j_2}^{(2)}$ , which also acts as the starting threshold for  $R_{j_1, j_2+1}$ . Thus, the design of the threshold structure inherently prevents discontinuities along single-dimensional moves.

Thus, the only case requiring careful attention is a two-dimensional move from  $(j_1, j_2)$  to  $(j_1 + 1, j_2 + 1)$ . In such moves, coherency violations can arise, as illustrated in Panels 2 and 3 of Figure 11. Condition (2) ensures proper alignment of the rectangles  $R_{j_1, j_2}$  and  $R_{j_1+1, j_2+1}$  along their shared boundary. It guarantees that there is no gap between the rectangles and no overlap in their interiors.

*Necessity.* Consider a general coherent model. Suppose it fails to be locally hierarchical in the sense that the threshold condition stated in Proposition 1 is violated for some local configuration  $\{(j_1 + \ell_1, j_2 + \ell_2)\}_{\ell_1, \ell_2 \in \{0, 1\}}$ . In such a case, the violation necessarily leads to either a gap (as shown in Panel 2) or an interior overlap (as shown in Panel 3) in Figure 11.

Since we are considering a generic model and looking for conditions in terms of thresholds alone, we can assume that the vector of latent utilities  $(Y^{*c_1}, Y^{*c_2})$ , conditional on  $x$  (and for a set of  $x$  with positive measure), has a strictly positive probability of falling into any rectangle with a non-empty interior. This implies that either the gap (Panel 2) or the overlap (Panel 3) will have a strictly positive probability mass conditional on  $x$ .

Consequently, in the case of a gap, we would observe:

$$\sum_{\ell_1=0}^1 \sum_{\ell_2=0}^1 P \left( Y = (y_{j_1+\ell_1}^{(1)}, y_{j_2+\ell_2}^{(2)}) \mid x, Y \in \bigcup_{\ell_1=0}^1 \bigcup_{\ell_2=0}^1 \{(y_{j_1+\ell_1}^{(1)}, y_{j_2+\ell_2}^{(2)})\} \right) < 1,$$

while in the case of an overlap with non-empty interior, we would have:

$$\sum_{\ell_1=0}^1 \sum_{\ell_2=0}^1 P \left( Y = (y_{j_1+\ell_1}^{(1)}, y_{j_2+\ell_2}^{(2)}) \mid x, Y \in \bigcup_{\ell_1=0}^1 \bigcup_{\ell_2=0}^1 \{(y_{j_1+\ell_1}^{(1)}, y_{j_2+\ell_2}^{(2)})\} \right) > 1.$$

Either case contradicts the finite additivity of the probability measure, namely:

$$P(Y \in A \mid x) = \sum_{(j_1, j_2): (y_{j_1}^{(1)}, y_{j_2}^{(2)}) \in A} P(Y = (y_{j_1}^{(1)}, y_{j_2}^{(2)}) \mid x). \blacksquare$$

The core principles underlying Proposition 1 can be naturally extended for any  $D > 2$ .

## 1.2 D Greater Than 2

For  $D > 2$ , we again examine each local configuration of the form  $\{(j_1 + \ell_1, \dots, j_D + \ell_D)\}_{\ell_d \in \{0,1\}, d=1,\dots,D}$ . Coherency within every such local configuration ensures global coherency of the model. We proceed incrementally in deriving coherency conditions: starting with moves in one dimension, then in pairs of dimensions, and so on, up to moves in all  $D$  dimensions.

For concreteness, consider the case  $D = 3$ . As in the bivariate case, any single-dimensional move within a local configuration (i.e., from  $(t_1, t_2, t_3)$  to a neighboring point along one axis) cannot induce incoherency. This follows from the definition of the rectangles  $R_{t_1, t_2, t_3}$  in equation (1), which ensures continuity at threshold boundaries along each coordinate axis.

The next step is to consider moves that involve changes along two dimensions. For example, transitions such as  $(j_1 + 1, j_2 + 1, j_3)$  to  $(j_1, j_2 + 1, j_3 + 1)$ , among others, must also preserve coherency. To verify this, we project the local configuration onto the relevant two-dimensional subspace, holding the third coordinate fixed. For each such projection, we apply the bivariate condition from Proposition 1. This yields conditions

$$(\alpha_{j_1, j_2, t_3}^{(1)} - \alpha_{j_1, j_2+1, t_3}^{(1)})(\alpha_{j_1+1, j_2, t_3}^{(2)} - \alpha_{j_1, j_2, t_3}^{(2)}) = 0, \quad t_3 \in \{j_3, j_3 + 1\},$$

$$(\alpha_{j_1, t_2, j_3}^{(1)} - \alpha_{j_1, t_2, j_3+1}^{(1)}) (\alpha_{j_1+1, t_2, j_3}^{(3)} - \alpha_{j_1, t_2, j_3}^{(3)}) = 0, \quad t_2 \in \{j_2, j_2 + 1\},$$

$$(\alpha_{t_1, j_2, j_3}^{(2)} - \alpha_{t_1, j_2, j_3+1}^{(2)}) (\alpha_{t_1, j_2+1, j_3}^{(3)} - \alpha_{t_1, j_2, j_3}^{(3)}) = 0, \quad t_1 \in \{j_1, j_1 + 1\}.$$

This must hold for all pairs of dimensions and all fixed values of the remaining coordinate.

Finally, we must consider transitions involving simultaneous changes in all three dimensions. These correspond to movements between opposite orthants within the local configuration. To prevent incoherency, the thresholds where these orthants “meet” must align in at least one dimension. Define the pair of opposite orthants as those corresponding to

$$\mathbf{j} = (j_1 + \ell_1, j_2 + \ell_2, j_3 + \ell_3), \quad \text{and} \quad \mathbf{j}' = (j_1 + 1 - \ell_1, j_2 + 1 - \ell_2, j_3 + 1 - \ell_3).$$

for  $\boldsymbol{\ell} = (\ell_1, \ell_2, \ell_3)' \in \{0, 1\}^3$ . Then the coherency condition for such pairs is  $\prod_{d=1}^3 (\alpha_{\mathbf{j}}^{(d)} - \alpha_{\mathbf{j}'_{(d)}}^{(d)}) = 0$ , where  $\mathbf{j}'_{(d)}$  denotes the index vector obtained from  $\mathbf{j}$  by replacing the  $d$ -th coordinate with that of  $\mathbf{j}'$ :

$$\mathbf{j}'_{(d)} = (j_1 + \ell_1, \dots, j_{d-1} + \ell_{d-1}, j_d + 1 - \ell_d, j_{d+1} + \ell_{d+1}, \dots, j_3 + \ell_3).$$

This holds for all 8 choices of  $\boldsymbol{\ell}$  (or equivalently, all 4 diagonal pairs up to symmetry).

This condition ensures that, for each pair of opposite orthants within a local configuration, the corresponding threshold surfaces meet along at least one boundary, thereby ruling out both gaps and overlaps in the latent space.

For general  $D > 3$ , the approach proceeds inductively, leveraging coherency conditions established for lower dimensions. As in prior cases, single-dimensional moves within any local configuration preserve coherency due to the continuity enforced by the rectangle definitions in equation (1) along each axis. Coherency for moves involving changes in any  $k$  dimensions ( $2 \leq k < D$ ) is ensured by projecting the local configuration onto the corresponding  $k$ -dimensional subspace (fixing the remaining  $D - k$  coordinates) and applying the coherency conditions derived for dimension  $k$  (e.g., Proposition 1 for  $k = 2$ , or the trivariate conditions for  $k = 3$ ). Finally, for simultaneous changes across all  $D$  dimensions, which correspond to transitions between opposite corners of the  $D$ -dimensional hyperrectangle, the coherency condition requires that the threshold hypersurfaces meet along at least one boundary, with the formulation analogous to the case of  $D = 3$ . This gives  $2^{D-1}$  conditions.

This inductive framework ensures global coherency for arbitrary  $D > 2$ .

## 2 Appendix B: Proofs

*Proof of Theorem 1.* Without a loss of generality, take  $d = 1$ . The proof proceeds in the following way. First, we establish an auxiliary result that  $P(Y^{c_1} \leq y_{j_1}^{(1)} | x)$  is non-increasing in the index  $x_1\beta_1$  with other indices fixed and is strictly decreasing for  $x$  in  $S_d(j_1)$  that satisfies condition (c) of the theorem. Second, having established that strict monotonicity, we then use techniques in the spirit of single-index identification approaches by varying  $x_{1,1}$  to establish identification.

From the model definition,

$$\begin{aligned} P(Y^{c_1} \leq y_{j_1}^{(1)} | x) &= \sum_{\tilde{j}=1}^{j_1} \sum_{j_2=1}^{M_2} \dots \sum_{j_D=1}^{M_D} P((Y^{*c_1}, \dots, Y^{*c_D}) \in R_{\tilde{j}, j_2, \dots, j_D} | x) \\ &= \sum_{\tilde{j}=1}^{j_1} \sum_{j_2=1}^{M_2} \dots \sum_{j_D=1}^{M_D} P((x_1\beta_1 + \varepsilon_1, \dots, x_D\beta_D + \varepsilon_D) \in R_{\tilde{j}, j_2, \dots, j_D} | x), \end{aligned}$$

$j_1 = 1, \dots, M_1$ . Let us show that this probability is non-increasing in  $x_1\beta_1$  when other indices  $x_\ell\beta_\ell$ ,  $\ell \neq 1$ , remain fixed.

For any  $j_1 = 1, \dots, M_1$ , the partitioning structure in the decision rule guarantees that

$$\cup_{\tilde{j}=1}^{j_1} \cup_{j_2=1}^{M_2} \dots \cup_{j_D=1}^{M_D} R_{\tilde{j}, j_2, \dots, j_D} = \cup_{j_2=1}^{M_2} \dots \cup_{j_D=1}^{M_D} R_{j_1, j_2, \dots, j_D}^*, \quad \text{where}$$

$$R_{j_1, j_2, \dots, j_D}^* = (-\infty, \alpha_{j_1, j_2, \dots, j_D}^{(1)}] \times_{d=2}^D (\alpha_{j_1, j_2, \dots, j_{d-1}, j_d-1, j_{d+1}, \dots, j_D}^{(d)}, \alpha_{j_1, j_2, \dots, j_{d-1}, j_d, j_{d+1}, \dots, j_D}^{(d)}].$$

In turn, this gives

$$\begin{aligned} P(Y^{c_1} \leq y_{j_1}^{(1)} | x) &= \sum_{j_2=1}^{M_2} \dots \sum_{j_D=1}^{M_D} \left( \bar{F}(-\infty, \alpha_{j_1, j_2, \dots, j_D}^{(2)} - x_2\beta_2, \dots, \alpha_{j_1, j_2, \dots, j_D}^{(D)} - x_D\beta_D) \right. \\ &\quad \left. + F(\alpha_{j_1, j_2, \dots, j_D}^{(1)} - x_1\beta_1, \alpha_{j_1, j_2, \dots, j_D}^{(2)} - x_2\beta_2, \dots, \alpha_{j_1, j_2, \dots, j_D}^{(D)} - x_D\beta_D) - 1 \right), \end{aligned}$$

where  $F$  and  $\bar{F}$ , as stated in Notation 1, denote the joint c.d.f. and survival functions of  $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_D)'$ , respectively. By coordinate-wise monotonicity of  $F$ ,  $P(Y^{c_1} \leq y_{j_1}^{(1)} | x)$  is non-increasing in  $x_1\beta_1$  when other indices  $x_\ell\beta_\ell$ ,  $\ell \neq 1$ , remain fixed. Note that condition (c) of the theorem guarantees that  $\alpha_{j_1, j_2, \dots, j_D}^{(1)} - x_1\beta_1$  is in the interior of the support of  $\varepsilon_1$  for  $x \in S_1(j_1)$ . Using the fact that the support of  $\varepsilon_1$  is convex (implied by convexity of the support of  $\boldsymbol{\varepsilon}$ ), we then conclude that  $P(Y^{c_1} \leq y_{j_1}^{(1)} | x)$  is strictly decreasing in  $x_1\beta_1$  when other indices remain

fixed and  $x \in S_1(j_1)$ ,

Let us now take two vectors  $b = (b'_1, \dots, b'_D)', \beta = (\beta'_1, \dots, \beta'_D)' \in \mathbb{R}^{\sum_{d=1}^D k_d}$  that satisfy normalization condition (b) of the theorem and suppose that both are consistent with the observed conditional probabilities of choice. If  $L_1 = 1$ , then the result of the theorem is already established for  $d = 1$ . Suppose  $L_1 > 1$  and  $b_{1,2:L_1} \neq \beta_{1,2:L_1}$ . Then from the condition on the probability of  $(\underline{x}_{1,1}, \bar{x}_{1,1}) \times S_{1,-1}(j_1)$  as well as the full affine dimension of  $S_1(j_1)$  in condition (c) we conclude that  $x_{1,2:L_1}\beta_{1,2:L_1} \neq x_{1,2:L_1}b_{1,2:L_1}$  for a positive measure of  $x_{1,2:L_1}$  that belong to the projection of  $S_{1,-1}(j_1)$  on the last  $L_1 - 1$  components (that is, those corresponding to  $x_{1,2:L_1}$ ). Without a loss of generality, suppose that for a positive measure of such  $x_{1,2:L_1}$  we have

$$x_{1,2:L_1}\beta_{1,2:L_1} > x_{1,2:L_1}b_{1,2:L_1}. \quad (10)$$

Now fix any  $x_{1,2:L_1}$  that satisfies (10). Then for any  $\tilde{x}_{1,1} \in (\underline{x}_{1,1}, \bar{x}_{1,1})$ , we have  $\tilde{x}_{1,1} + x_{1,2:L_1}\beta_{1,2:L_1} > \tilde{x}_{1,1} + x_{1,2:L_1}b_{1,2:L_1}$ , and for given  $x_{1,2:L_1}$  we can find  $\tilde{\tilde{x}}_{1,1} \in (\underline{x}_{1,1}, \bar{x}_{1,1})$  such that

$$\tilde{x}_{1,1} + x_{1,2:L_1}\beta_{1,2:L_1} \stackrel{(a)}{>} \tilde{\tilde{x}}_{1,1} + x_{1,2:L_1}\beta_{1,2:L_1} \stackrel{(b)}{>} \tilde{x}_{1,1} + x_{1,2:L_1}b_{1,2:k_1}. \quad (11)$$

Because of  $x_{1,1}$  being exclusive for  $Y^{*c_1}$ , when we vary  $x_{1,1}$ , the values of  $x_2, \dots, x_D$  remain exactly the same. This means that in the expression for  $P(Y^{c_1} \leq y_j^{(1)} | x_1, \dots, x_D)$ , the values of indices  $x_\ell\beta_\ell, x_\ell b_\ell, \ell \neq 1$ , remain exactly the same. This means that by varying  $x_{1,1}$ , we can equivalently express the ordering of  $P(Y^{c_1} \leq y_j^{(1)} | x_1, \dots, x_D)$  with the reverse ordering of the first argument in the first index.

Therefore, (a) in (11) implies that

$$P(Y^{c_1} \leq y_j^{(1)} | (\tilde{x}_{1,1}, x_{1,2:L_1}, x_2, \dots, x_D)) < P(Y^{c_1} \leq y_j^{(1)} | (\tilde{\tilde{x}}_{1,1}, x_{1,2:L_1}, x_2, \dots, x_D)).$$

Since we supposed that both  $\beta$  and  $b$  can generate observable choice probabilities of choice, then (b) in (11) implies that

$$P(Y^{c_1} \leq y_j^{(1)} | (\tilde{\tilde{x}}_{1,1}, x_{1,2:L_1}, x_2, \dots, x_D)) < P(Y^{c_1} \leq y_j^{(1)} | (\tilde{x}_{1,1}, x_{1,2:k_1}, x_2, \dots, x_D)).$$

Combining the last two inequalities results in an obvious contradiction

$$P\left(Y^{c_1} \leq y_j^{(1)} \mid (\tilde{x}_{1,1}, x_{1,2:L_1}, x_2, \dots, x_D)\right) < P\left(Y^{c_1} \leq y_j^{(1)} \mid (\tilde{x}_{1,1}, x_{1,2:L_1}, x_2, \dots, x_D)\right),$$

and from our discussion it is clear that this contradiction is obtained for a positive measure of  $(\tilde{x}_{1,1}, x_{1,2:L_1}, x_2, \dots, x_D)$ . Thus, we cannot have  $\beta_{1,2:L_1} \neq b_{1,2:L_1}$ , and, therefore,  $\beta_{1,2:L_1}$  is identified relative to any  $b_{1,2:L_1} \neq \beta_{1,2:L_1}$ . We can do this for any  $d$ .  $\square$

**Proof of Theorem 2.** Fix  $d$ . If in condition (b) of the theorem we have  $\underline{x}_{d,1}$  is small enough then we take  $\kappa_d$  to be  $\leq$ . If in that condition  $\bar{x}_{d,1}$  is large enough, we take  $\kappa_d$  to be  $>$ . Analyze now  $P(\cap_{d=1}^D (Y^{c_d} \kappa_d y_{j_d}^{(d)} \mid x))$  for any  $x$  in the intersection indicated in condition (a). First of all, that condition implies that this probability is strictly between 0 and 1. Second,

$$P(\cap_{d=1}^D (Y^{c_d} \kappa_d y_{j_d}^{(d)} \mid x)) = \sum_{d=1}^D \sum_{\tilde{j}_d \kappa_d j_d} P\left((x_1 \beta_1 + \varepsilon_1, \dots, x_D \beta_D + \varepsilon_D) \in R_{\tilde{j}_1, \dots, \tilde{j}_D} \mid x\right)$$

Focus e.g. on  $d = 1$  and for any  $d \geq 2$  take  $x_{d,1} \rightarrow \underline{x}_{d,1}$  if  $\kappa_d$  is  $\leq$  and take  $x_{d,1} \rightarrow \bar{x}_{d,1}$  if  $\kappa_d$  is  $>$ . Condition (a) guarantees that this limit can be taken within the intersection indicated in that condition. By condition (b), in such a limit of  $P(\cap_{d=1}^D (Y^{c_d} \kappa_d y_{j_d}^{(d)} \mid x))$  we obtain a function that no longer depends on indices  $x_d \beta_d$ ,  $d \neq 1$ , and is strictly monotone with respect to  $x_1 \beta_1$  for  $x_1$  from the projection of  $S_1(j_1)$  on the first  $k_1$  components.

For instance, if all the relevant for this limit boundaries  $\underline{x}_{d,1}$ ,  $\bar{x}_{d,1}$  are infinite (that is,  $-\infty$ ,  $\infty$ , respectively), then we obtain

$$P(\cap_{d=1}^D (Y^{c_d} \kappa_d y_{j_d}^{(d)} \mid x)) \rightarrow F_{1,\kappa_1}\left(\alpha_{j_1, m_2, \dots, m_D}^{(1)} - x_1 \beta_1\right),$$

where  $m_d = M_d$  is  $\kappa_d$  is  $>$  and  $m_d = 1$  if  $\kappa_d$  is  $\leq$ , for  $d \neq 1$ . Condition (a) of the theorem as well as the fact that  $P((Y^{*c_1}, \dots, Y^{*c_D}) \in R_{\tilde{j}_1, \tilde{j}_2, \dots, \tilde{j}_D} \mid x) \rightarrow 0$  for  $\tilde{j}_d \kappa_d j_d$ ,  $d \geq 2$ , and  $(\tilde{j}_2, \dots, \tilde{j}_D) \neq (m_2, \dots, m_D)$ , guarantee that  $\alpha_{j_1, m_2, \dots, m_D}^{(1)} - x_1 \beta_1$  is in the interior of the support of  $\varepsilon_1$ . Therefore, the limit is strictly monotone on the projection of  $S_1(j_1)$  on the first  $k_1$  components (corresponding to vector  $x_1$ ). It will be strictly increasing if  $\kappa_1$  is  $>$  and strictly decreasing if  $\kappa_1$  is  $\leq$ .

If some (or all) of the relevant boundaries  $\underline{x}_{d,1}$ ,  $\bar{x}_{d,1}$  are finite, then the limit of  $P((Y^{*c_1}, \dots, Y^{*c_D}) \in R_{\tilde{j}_1, \tilde{j}_2, \dots, \tilde{j}_D} \mid x)$  has a more complex form and can involve several thresholds. However it will still remain the case the overall limit will not depend on any indices except for  $x_1 \beta_1$  and will be

strictly monotone in  $x_1\beta_1$  on the projection of  $S_1(j_1)$  on the first  $k_1$  components.

Then, using the single-index approach analogous to the one we used in Theorem 1 we can establish the identification of the full vector  $\beta_1$ . This can be done for any  $\beta_d$ .  $\square$

**Proof of Theorem 3.** If in the condition of the theorem  $j_d = 1$ , we take  $\kappa_d$  to be  $\leq$  and if  $j_d = M_d - 1$ , then we take  $\kappa_d$  to be  $>$ . We start by showing identification of all  $\alpha_{j_1, j_2, \dots, j_D}^{(d)}$  shaping the “corner” rectangular region corresponding to

$$P(\cap_{h=1}^D \left( Y^{*c_h} \kappa_h \alpha_{j_1, j_2, \dots, j_D}^{(h)} \right) | x) = P(\cap_{h=1}^D \left( Y^{c_h} \kappa_h y_{j_h}^{(h)} \right) | x).$$

Indeed, these probabilities are observed. Fix  $d$ . Just like in the proof of Theorem 2, for any  $h \neq d$  take  $x_{h,1} \rightarrow \underline{x}_{h,1}$  if  $\kappa_h$  is  $\leq$  and take  $x_{h,1} \rightarrow \bar{x}_{h,1}$  if  $\kappa_h$  is  $>$ . By doing this, in the limit of  $x_{h,1}$ ,  $h \neq d$ , we identify  $F_{d,\kappa_d}(\alpha_{j_1, j_2, \dots, j_D}^{(d)} - x_d \beta_d)$ .

Now, using the support condition on  $x_{d,1}$ , we obtain that  $\alpha_{j_1, j_2, \dots, j_D}^{(d)} - x_d \beta_d$  goes through the whole support of  $\varepsilon_d$ . Using the normalization on  $F_{d,\kappa_d}$  we find  $x_{0d}$  such that  $F_{d,\kappa_d}(\alpha_{j_1, j_2, \dots, j_D}^{(d)} - x_{0d} \beta_d) = c_{0d}$  if  $\kappa_d$  is  $\leq$ , or  $x_{0d}$  such that  $1 - F_{d,\kappa_d}(\alpha_{j_1, j_2, \dots, j_D}^{(d)} - x_{0d} \beta_d) = c_{0d}$  if  $\kappa_d$  is  $>$ . From this we can identify  $\alpha_{j_1, j_2, \dots, j_D}^{(d)}$  as  $\alpha_{j_1, j_2, \dots, j_D}^{(d)} = e_{0d} + x_{0d} \beta_d$  ( $e_{0d}$  and  $\beta_d$  are known).

Combining the knowledge of  $\alpha_{j_1, j_2, \dots, j_D}^{(d)}$  for any  $d = 1, \dots, D$ , with the exclusiveness of some covariates in each index and support conditions on  $x_{d,1}$ ,  $d + 1, \dots, D$ , we can identify the joint distribution of  $\varepsilon$  as

$$P \left( \cap_{h=1}^D (Y^{*c_h} \kappa_h \alpha_{j_1, \dots, j_D}^{(h)}) | x \right) = F_{\kappa_1, \kappa_2, \dots, \kappa_D}(\alpha_{j_1, \dots, j_D}^{(1)} - x_1 \beta_1, \dots, \alpha_{j_1, \dots, j_D}^{(D)} - x_D \beta_D)$$

as the vector  $(\alpha_{j_1, \dots, j_D}^{(1)} - x_1 \beta_1, \dots, \alpha_{j_1, \dots, j_D}^{(D)} - x_D \beta_D)$  is known and can be taken to be any value on the support of  $\varepsilon$ .  $\square$

**Proof of Theorem 4.**

**Stage 1.** Pick any “corner” rectangular region  $R_{j_1, \dots, j_D}$ ,  $j_d \in \{1, M_d\}$  for each  $d = 1, \dots, D$ . It is described by  $D$  unknown thresholds  $\alpha_{j_1, \dots, q_{d-1}, r_d(j_d), j_{d+1}, \dots, j_D}^{(d)}$ , where

$$r_d(j_d) = \begin{cases} 1, & \text{if } j_d = 1, \\ M_d - 1, & \text{if } j_d = M_d. \end{cases}$$

Our goal is to identify them. Once again, it is convenient to associate a certain direction for the distribution of  $\varepsilon$  with this “corner”. If  $j_d = 1$  we take  $\kappa_d$  to be  $\leq$ , and if  $j_d = M_d$  we take  $\kappa_d$  to be  $>$ ,

In Theorem 3 we only considered one “corner” region associated with one particular direction  $(\kappa_1, \dots, \kappa_D)$  and identified the  $D$  thresholds that shape it (other  $D$  thresholds are either at  $\infty$  or  $-\infty$  depending on the location of the “corner”). Conditions of this Theorem 4 imply that we can now consider any  $(\kappa_1, \dots, \kappa_D)$  with its associated “corner” region  $n$  and then apply the machinery of the Theorem 3 to identify its unknown thresholds.

Thus, this stage identifies thresholds shaping all the “corner” rectangular regions.

**Stage 2.** In this stage we continue to consider rectangular regions near the border. In Stage 1 for each “corner” region we considered the  $D$  known thresholds were fixed at  $\infty$  or  $-\infty$ . Now we will have only  $D - 1$  known thresholds fixed at  $\infty$  or  $-\infty$ . At least one known threshold will be finite and known from the previous stage.

Namely, fix  $d$  and consider a border rectangular region  $R_{j_1, \dots, j_{d-1}, q_d, j_{d+1}, \dots, j_D}$  where  $j_h \in \{1, M_h\}$  for  $h \neq d$  and  $q_d = 2, \dots, M_d - 1$ . It is described by  $D + 1$  thresholds  $\alpha_{j_1, \dots, j_{d-1}, q_d-1, j_{d+1}, \dots, j_D}^{(d)}$ ,  $\alpha_{j_1, \dots, j_{d-1}, q_d, j_{d+1}, \dots, j_D}^{(d)}$  (in dimension  $d$ ) and  $\alpha_{j_1, \dots, j_{d-1}, q_d, j_{d+1}, \dots, j_D}^{(h)}$ ,  $h \neq d$ . For  $q_d = 2$  and  $q_d = M_d - 1$  we only have  $D$  unknown thresholds since  $\alpha_{j_1, \dots, j_{d-1}, 1, j_{d+1}, \dots, j_D}^{(d)}$  and  $\alpha_{j_1, \dots, j_{d-1}, M_d-1, j_{d+1}, \dots, j_D}^{(d)}$  have been identified in Stage 1. The idea is then to indeed proceed sequentially from, say,  $q_d = 2$  in an increasing manner.

Within this stage, note that we can identify thresholds  $\alpha_{j_1, \dots, j_{d-1}, q_d, j_{d+1}, \dots, j_D}^{(d)}$ , for  $q_d = 2, \dots, M_d - 2$  in the following way. Choosing, once again,  $\kappa_h$  to be  $\leq$  if  $j_h = 1$  and to be  $>$  if  $j_h = M_h$ ,  $h \neq d$ , obtain that

$$\begin{aligned} P \left( Y^{c_d} = y_{q_d}^{(d)}, \cap_{h \neq d} \left( Y^{c_h} = y_{j_h}^{(h)} \right) | x \right) &= P \left( \alpha_{j_1, \dots, j_{d-1}, q_d-1, j_{d+1}, \dots, j_D}^{(d)} < Y^{*c_d} \right. \\ &\leq \left. \alpha_{j_1, \dots, j_{d-1}, q_d, j_{d+1}, \dots, j_D}^{(d)}, \cap_{h \neq d} \left( Y^{*c_h} \leq \kappa_h \leq \alpha_{j_1, \dots, j_{d-1}, q_d, j_{d+1}, \dots, j_D}^{(h)} \right) \right). \end{aligned}$$

By taking  $x_{h,1} \rightarrow \underline{x}_{h,1}$  if  $\kappa_h$  is  $\leq$  or  $x_{h,1} \rightarrow \bar{x}_{h,1}$  if  $\kappa_h$  is  $>$  for all  $h \neq d$ , we identify

$$\lim_{x_{h,1} \rightarrow h \neq d} P(Y^{c_d} = y_{q_d}^{(d)} \cap_{h \neq d} (Y^{c_h} = y_{j_h}^{(h)}) | x) = \\ F_{d,\leq}(\alpha_{j_1, \dots, j_{d-1}, q_d, j_{d+1}, \dots, j_D}^{(d)} - x_d \beta_d) - F_{d,\leq}(\alpha_{j_1, \dots, j_{d-1}, q_d-1, j_{d+1}, \dots, j_D}^{(d)} - x_d \beta_d).$$

It is clear that from the knowledge of  $F_{d,\leq}$  (Theorem 3) and  $\alpha_{j_1, \dots, j_{d-1}, 1, j_{d+1}, \dots, j_D}^{(d)}$  we can identify from this limit  $\alpha_{j_1, \dots, j_{d-1}, 2, j_{d+1}, \dots, j_D}^{(d)}$  simply by choosing  $x_d$  such that  $F_{d,\leq}(\alpha_{j_1, \dots, j_{d-1}, q_d, j_{d+1}, \dots, j_D}^{(d)} - x_d \beta_d) \in (0, 1)$ . Using the same arguments, we can identify  $\alpha_{j_1, \dots, j_{d-1}, 3, j_{d+1}, \dots, j_D}^{(d)}$ , etc. Proceeding sequentially, we will establish identification of any such  $\alpha_{j_1, \dots, j_{d-1}, q_d, j_{d+1}, \dots, j_D}^{(d)}$ .

Thus, in each rectangular region  $R_{j_1, \dots, j_{d-1}, q_d, j_{d+1}, \dots, j_D}$  under consideration in this Stage 2 we now only have  $D - 1$  unknown thresholds  $\alpha_{j_1, \dots, j_{d-1}, q_d, j_{d+1}, \dots, j_D}^{(h)}$ ,  $h \neq d$ .

Now we also fix one  $h$  such that  $h \neq d$  and show the threshold  $\alpha_{j_1, \dots, j_{d-1}, q_d, j_{d+1}, \dots, j_D}^{(h)}$  is identified. For this, we consider  $P(Y^{c_d} = y_{q_d}^{(d)} \cap_{h \neq d} (Y^{c_h} = y_{j_h}^{(h)}) | x)$  again (as above) but now take to the limit  $x_{\tilde{h},1}$  for  $\tilde{h} \neq h$ ,  $\tilde{h} \neq d$ . Namely, for such  $\tilde{h}$  take  $x_{\tilde{h},1} \rightarrow \underline{x}_{\tilde{h},1}$  if  $\kappa_{\tilde{h}}$  is  $\leq$  and take  $x_{\tilde{h},1} \rightarrow \bar{x}_{\tilde{h},1}$  if  $\kappa_{\tilde{h}}$  is  $>$ . In such a limit we identify

$$F_{d,h;\leq,\kappa_h} \left( \alpha_{j_1, \dots, j_{d-1}, q_d, j_{d+1}, \dots, j_D}^{(d)} - x_d \beta_d, \alpha_{j_1, \dots, j_{d-1}, q_d, j_{d+1}, \dots, j_D}^{(h)} - x_h \beta_h \right) \\ - F_{d,h;\leq,\kappa_h} \left( \alpha_{j_1, \dots, j_{d-1}, q_d-1, j_{d+1}, \dots, j_D}^{(d)} - x_d \beta_d, \alpha_{j_1, \dots, j_{d-1}, q_d, j_{d+1}, \dots, j_D}^{(h)} - x_h \beta_h \right), \quad (12)$$

where  $F_{d,h;\leq,\kappa_h}(a_1, a_2) \equiv P(\varepsilon_d \leq a_1, \varepsilon_h \leq a_2)$ . Function  $F_{d,h;\leq,\kappa_h}$  is identified as an implication of Theorem 3. Thus, in the known limit (12) there is only one unknown threshold  $\alpha_{j_1, \dots, j_{d-1}, q_d, j_{d+1}, \dots, j_D}^{(h)}$ .

Denote  $a_1 = \alpha_{j_1, \dots, j_{d-1}, q_d-1, j_{d+1}, \dots, j_D}^{(d)} - x_d \beta_d$ ,  $a_2 = \alpha_{j_1, \dots, j_{d-1}, q_d, j_{d+1}, \dots, j_D}^{(h)} - x_h \beta_h$ ,  $\Delta a_1 = \alpha_{j_1, \dots, j_{d-1}, q_d, j_{d+1}, \dots, j_D}^{(d)} - \alpha_{j_1, \dots, j_{d-1}, q_d-1, j_{d+1}, \dots, j_D}^{(d)}$ . Since the function  $F_{d,h;\leq,\kappa_h}(a_1 + \Delta a_1, a_2) - F_{d,h;\leq,\kappa_h}(a_1, a_2)$  is strictly monotone in  $a_2$  as long as  $(a_1 + \Delta a_1, a_2)$  and  $(a_1 + \Delta a_1, a_2)$  remain in the support of  $(\varepsilon_d, \varepsilon_h)$  (namely, it is strictly increasing if  $\kappa_h$  is  $\leq$  and strictly decreasing if  $\kappa_h$  is  $>$ ), then we the right choice of  $x_d$ ,  $x_h$  we can identify  $\alpha_{j_1, \dots, j_{d-1}, q_d, j_{d+1}, \dots, j_D}^{(h)}$ .

**Stage 3.** In this stage we continue to consider rectangular regions near the border. Compared with Stages 1 and 2, now we will have only  $D - 2$  known thresholds fixed at  $\infty$  or  $-\infty$ . At least two other thresholds will be known and finite from previous stages.

Consider a rectangular border region  $R_{j_1, \dots, j_{d_1-1}, q_{d_1}, j_{d_1+1}, \dots, j_{d_2-1}, q_{d_2}, j_{d_2+1}, \dots, j_D}$  for some  $d_1, d_2$  such

that  $d_1 < d_2$  and where  $j_h \in \{1, M_h\}$  for  $h \neq d_1, d_2$  and  $q_{d_j} = 2, \dots, M_{d_j} - 1$ ,  $j = 1, 2$ . In this border region we allow discrete processes in dimensions  $d_1$  and  $d_2$  to take any of their possible values whereas in all the other dimensions they take their boundary values. without a loss of generality we will take  $d_1 = 1$  and  $d_2 = 2$ .

Region  $R_{q_1, q_2, j_3, \dots, j_D}$ , where  $j_h \in \{1, M_h\}$  for  $h \neq d_1, d_2 \geq 3$  is described by  $D + 2$  thresholds  $\alpha_{q_1-1, q_2, j_3, \dots, j_D}^{(1)}, \alpha_{q_1, q_2, j_3, \dots, j_D}^{(1)}, \alpha_{q_1, q_2-1, j_3, \dots, j_D}^{(2)}, \alpha_{q_1, q_2, j_3, \dots, j_D}^{(2)}$  (in dimensions 1 and 2) and  $\alpha_{q_1, q_2, j_3, \dots, j_D}^{(h)}$ ,  $h \geq 3$  (other dimensions). We proceed sequentially – first, taking  $q_1 \in \{2, M_1 - 2\}$ ,  $q_2 \in \{2, M_2 - 2\}$  and then changing them one unit at a time. The direction in which we proceed (from a high to low index, or the other way around) though may depend on the properties of the support  $\mathcal{E}_{12}$  of  $(\varepsilon_1, \varepsilon_2)'$ .

**Subcase 1** If  $\mathcal{E}_{12}$  is not bounded in any directions, it means that it is  $\mathbf{R}^2$  and in the condition of the theorem we necessarily have  $\underline{x}_{h,1} = -\infty$  and  $\bar{x}_{h,1} = \infty$  for  $h = 1, 2$ . Then it does not matter in which direction we proceed. E.g., we can first consider  $q_1 = 2, q_2 = 2$ . Using results of Stage 2, we find that we only deal with  $D$  unknown thresholds  $\alpha_{2,2,j_3,\dots,j_D}^{(1)}, \alpha_{2,2,j_3,\dots,j_D}^{(2)}$  and  $\alpha_{2,2,j_3,\dots,j_{h-1},r_h(j_h),j_{h+1},\dots,j_D}^{(h)}, h \geq 3$ . Notice that at this stage, due to coherency requirements, it may be strictly fewer than  $D$  of these thresholds unknown. However, all  $D$  may potentially be unknown and that is why we need to develop a general identification strategy.

Consider the observed probability  $P(Y^{c_1} = y_2^{(1)}, Y^{c_d} = y_2^{(2)}, \cap_{h \geq 3} (Y^{c_h} = y_{j_h}^{(h)}) | x)$  and find its limit when  $x_{h,1} \rightarrow$  for  $h \geq 3$ . Just as before, we take  $x_{h,1} \rightarrow \underline{x}_{h,1}$  if  $j_h = 1$  and take  $x_{h,1} \rightarrow \bar{x}_{h,1}$  if  $j_h = M_h$ . In this limit we identify

$$\sum_{\ell_1=0}^1 \sum_{\ell_2=0}^1 (-1)^{\ell_1+\ell_2} F_{1,2;\leq} \left( \alpha_{1+\ell_1, 2, j_3, \dots, j_D}^{(1)} - x_1 \beta_1, \alpha_{2, 1+\ell_2, j_3, \dots, j_D}^{(2)} - x_2 \beta_2 \right) \in (0, 1).$$

Let us now take  $x_{1,1} \rightarrow -\infty$ . Then the known limit becomes

$$F_{2,\leq} \underbrace{(\alpha_{2,2,j_3,\dots,j_D}^{(2)} - x_2 \beta_2)}_{z_2} - F_{2,\leq} \underbrace{(\alpha_{2,1,j_3,\dots,j_D}^{(2)} - x_2 \beta_2)}_{z_1} \in (0, 1).$$

Using the knowledge of  $F_{2,\leq}$  and the strict monotonicity of the obtained probability with respect to  $z_2$  with  $z_1$  fixed we identify  $\alpha_{2,2,j_3,\dots,j_D}^{(2)}$ . Analogously we can identify  $\alpha_{2,2,j_3,\dots,j_D}^{(1)}$  by keeping  $x_1$  fixed and taking  $x_{2,1} \rightarrow -\infty$ . Once  $\alpha_{2,2,j_3,\dots,j_D}^{(1)}, \alpha_{2,2,j_3,\dots,j_D}^{(2)}$  are identified, we can identify  $\alpha_{2,2,r_3(j_3),\dots,j_D}^{(3)}$  by taking in our observed probability  $x_{h,1} \rightarrow$  in the manner described above but

now only for  $h \geq 4$ . In the limit we identify

$$\sum_{\ell_1=0}^1 \sum_{\ell_2=0}^1 \sum_{\ell_3=0}^1 (-1)^{\ell_1+\ell_2+\ell_3} F_{1,2,3; \leq \leq \leq} \left( \alpha_{1+\ell_1, 2, 1, j_4, \dots, j_D}^{(1)} - x_1 \beta_1, \alpha_{2, 1+\ell_2, 1, j_4, \dots, j_D}^{(2)} - x_2 \beta_2, \alpha_{2, 2, \ell_3, j_4, \dots, j_D}^{(3)} - x_3 \beta_3 \right) \in (0, 1)$$

if  $j_3 = 1$ , and identify

$$\sum_{\ell_1=0}^1 \sum_{\ell_2=0}^1 \sum_{\ell_3=0}^1 (-1)^{\ell_1+\ell_2+\ell_3} F_{1,2,3; \leq \leq >} \left( \alpha_{1+\ell_1, 2, M_3, j_4, \dots, j_D}^{(1)} - x_1 \beta_1, \alpha_{2, 1+\ell_2, M_3, j_4, \dots, j_D}^{(2)} - x_2 \beta_2, \alpha_{2, 2, M_3-\ell_3, j_4, \dots, j_D}^{(3)} - x_3 \beta_3 \right) \in (0, 1)$$

if  $j_3 = M_3$ . No matter which of these situations we have, we use the knowledge of  $F_{1,2,3; \leq \leq \kappa_3}$ , the knowledge of 5 out of 6 thresholds in them and the strict monotonicity of that limit with respect to the unknown threshold ( $\alpha_{2,2,\ell_3,j_4,\dots,j_D}^{(3)}$  in the first and  $\alpha_{2,2,M_3-\ell_3,j_4,\dots,j_D}^{(3)}$  in the second situation) to identify this threshold. Analogously we can identify  $\alpha_{2,2,j_3,\dots,j_{h-1},r_h(j_h),j_{h+1},\dots,j_D}^{(h)}$  for any  $h \geq 4$ . Thus, all the thresholds of  $R_{2,2,j_3,\dots,j_D}$  are identified.

Building on this result, we will next look at the rectangles  $R_{3,2,j_3,\dots,j_D}$  and  $R_{2,3,j_3,\dots,j_D}$  (one index change at a time) and identify their thresholds in a similar manner. Then we look at  $R_{3,3,j_3,\dots,j_D}$ ,  $R_{4,2,j_3,\dots,j_D}$ ,  $R_{2,4,j_3,\dots,j_D}$ , etc. until we identify thresholds of all the rectangles considered in this stage.

**Subcase 2** Second, consider the case when  $\mathcal{E}_{12}$  is bounded in some directions. Recall that it is convex and has non-empty interior in  $\mathbb{R}^2$  by Assumption 1. The idea is to use points near the finite boundary to establish the identification of threshold. The machinery of the identification procedure depends on some properties of points at the finite boundary. To describe it, let us define the four quadrants in  $\mathbf{R}^2$  originating from  $(0, 0)'$  as

$$\mathcal{O}_{s_1 s_2} = \{(s_1 \lambda_1, s_2 \lambda_2) : \lambda_i \geq 0, i = 1, 2\} \quad \text{for } s_1, s_2 \in \{+, -\}.$$

E.g.,  $\mathcal{O}_{-+}$  e.g. contains all bivariate vectors with the first non-positive and the second non-negative coordinate.  $\bar{s}$  will denote  $-$  if  $s$  is  $+$ , and will denote  $+$  if  $s$  is  $-$ .

For every point  $a = (a_1, a_2)'$  at the finite boundary the interior of at least one quadrant  $a + \mathcal{O}_{s_1 s_2}$

for some  $(s_1, s_2)$  does not intersect  $\mathcal{E}_{12}$  (if the interiors of all four such quadrants intersected with  $\mathcal{E}_{12}$ , because of convexity of  $\mathcal{E}_{12}$  it would contradict the fact that  $a$  is at the boundary). At the same time, there are points  $a$  at the finite boundary for which two consecutive quadrants – either  $a + \mathcal{O}_{s_1 s_2}$  and  $a + \mathcal{O}_{\bar{s}_1 s_2}$ , or  $a + \mathcal{O}_{s_1 s_2}$  and  $a + \mathcal{O}_{s_1 \bar{s}_2}$  for some  $(s_1, s_2)$  intersect  $\mathcal{E}_{12}$  in their interior (if it were not the case for all  $a$  at the finite boundary, then this would contradict the fact that  $\mathcal{E}_{12}$  has a non-empty interior in  $\mathbb{R}^2$ ).

Once we found  $(s_1, s_2)$  such that  $a + \mathcal{O}_{s_1 s_2}$  intersects  $\mathcal{E}_{12}$  in its interior and  $a + \mathcal{O}_{\bar{s}_1 \bar{s}_2}$  does not intersect  $\mathcal{E}_{12}$  in its interior, the direction of the proof depends which of the remaining quadrants  $a + \mathcal{O}_{\bar{s}_1 s_2}$  or  $a + \mathcal{O}_{s_1 \bar{s}_2}$  intersects  $\mathcal{E}_{12}$  in its interior (it may be both or just one of them). Suppose  $a + \mathcal{O}_{\bar{s}_1 s_2}$  intersects  $\mathcal{E}_{12}$  in its interior. When  $\bar{s}_1 = -, s_2 = +$ , we proceed from “north-west” corner by starting with  $q_1 = 2, q_2 = M_2 - 1$  and gradually increasing  $q_1$  and gradually decreasing  $q_2$ . When  $\bar{s}_1 = +, s_2 = +$ , we proceed from “north-east” corner by starting with  $q_1 = M_1 - 1, q_2 = M_2 - 1$  and gradually decreasing both. When  $\bar{s}_1 = +, s_2 = -$ , we proceed from “south-east” corner by starting with  $q_1 = M_1 - 1, q_2 = 2$  and gradually decreasing  $q_1$  and increasing  $q_2$ . Finally, when  $\bar{s}_1 = -, s_2 = -$ , we proceed from “south-west” corner by starting with  $q_1 = 2, q_2 = 2$  and gradually increasing both.

For concreteness, in our proof suppose the interior of  $a + \mathcal{O}_{++}$  does not overlap with  $\mathcal{E}_{12}$  whereas the interiors of  $a + \mathcal{O}_{-+}$  do  $a + \mathcal{O}_{--}$  do. We proceed in our identification from the “north-west corner” by starting with  $q_1 = 2, q_2 = M_2 - 1$ . Consider the observed probability  $P(Y^{c_1} = y_2^{(1)}, Y^{c_2} = y_{M_2-1}^{(2)}, \cap_{h \geq 3}(Y^{c_h} = y_{j_h}^{(h)}) | x)$  and find its limit when  $x_{h,1}$  converges for  $h \geq 3$  in a way described earlier (take  $x_{h,1} \rightarrow \underline{x}_{h,1}$  if  $j_h = 1$  and take  $x_{h,1} \rightarrow \bar{x}_{h,1}$  if  $j_h = M_h$ ). In this limit we identify

$$Q(x_1, x_2) = \sum_{\ell_1=0}^1 \sum_{\ell_2=0}^1 (-1)^{\ell_1+\ell_2} F_{1,2;\leq>}(\alpha_{1+\ell_1, M_2-1, j_3, \dots, j_D}^{(1)} - x_1 \beta_1, \alpha_{2, M_2-1-\ell_2, j_3, \dots, j_D}^{(2)} - x_2 \beta_2) \in (0, 1) \quad (13)$$

for all  $(x_1, x_2)$  in  $S_1(2) \cap S_2(M_2 - 1)$ . The limit has two unknown thresholds:  $\alpha_{2, M_2-1, j_3, \dots, j_D}^{(1)}$  and  $\alpha_{2, M_2-2, j_3, \dots, j_D}^{(2)}$ . Suppose there is another set of parameters  $(\tilde{\alpha}_{2, M_2-1, j_3, \dots, j_D}^{(1)}, \tilde{\alpha}_{2, M_2-2, j_3, \dots, j_D}^{(2)})$  different from  $\alpha_{2, M_2-1, j_3, \dots, j_D}^{(1)}$  and  $\alpha_{2, M_2-2, j_3, \dots, j_D}^{(2)}$  and that generate the same observed probabilities (13) for all  $(x_1, x_2)$  in  $S_1(2) \cap S_2(M_2 - 1)$ . Denote

$$\Delta_1 = \alpha_{2, M_2-1, j_3, \dots, j_D}^{(1)} - \alpha_{1, M_2-1, j_3, \dots, j_D}^{(1)} > 0, \quad \Delta_2 = \alpha_{2, M_2-1, j_3, \dots, j_D}^{(2)} - \alpha_{2, M_2-2, j_3, \dots, j_D}^{(2)} > 0,$$

$$\delta_1 = \tilde{\alpha}_{2, M_2-1, j_3, \dots, j_D}^{(1)} - \alpha_{1, M_2-1, j_3, \dots, j_D}^{(1)} > 0, \quad \delta_2 = \alpha_{2, M_2-1, j_3, \dots, j_D}^{(2)} - \tilde{\alpha}_{2, M_2-2, j_3, \dots, j_D}^{(2)} > 0$$

for the two sets of thresholds. The observational equivalence in terms of probabilities implies that  $(\Delta_1 - \delta_1)(\Delta_2 - \delta_2) < 0$  as it would be easy to obtain a contradiction otherwise from the properties of  $F_{1,2;\leq>}$ . Now define  $z_1 = a_1 - \min\{\Delta_1, \delta_1\}$ ,  $z_2 = a_2 + \min\{\Delta_2, \delta_2\}$  and choose  $x_1$  and  $x_2$  such that  $z_1 = \alpha_{1,M_2-1,j_3,\dots,j_D}^{(1)} - x_1\beta_1$ ,  $z_2 = \alpha_{2,M_2-1,j_3,\dots,j_D}^{(2)} - x_2\beta_2$ . Define rectangles

$$R_\Delta = [z_1, z_1 + \Delta_1] \times [z_2 - \Delta_2, z_2], \quad R_\delta = [z_1, z_1 + \delta_1] \times [z_2 - \delta_2, z_2].$$

From the property  $(\Delta_1 - \delta_1)(\Delta_2 - \delta_2) < 0$ , we can show that  $R_\Delta \cap R_\delta = [z_1, a_1] \times [a_2, z_2]$ . From the properties of orthants  $a + \mathcal{O}_{s_1 s_2}$  supposed earlier we conclude that the interior of  $R_\Delta \cap R_\delta$  overlaps with  $\mathcal{E}_{12}$  thus showing that for the chosen  $(x_1, x_2)$  the probability  $Q(x_1, x_2)$  computed in (13) is strictly positive (by our supposition, both sets of thresholds produce the same  $Q(x_1, x_2)$ ). Then we can equivalently represent  $Q(x_1, x_2)$  in the following two ways:

$$\begin{aligned} Q(x_1, x_2) &= Pr_{(\varepsilon_1, \varepsilon_2)}((\varepsilon_1, \varepsilon_2)' \in R_\Delta \cap R_\delta) + Pr_{(\varepsilon_1, \varepsilon_2)}((\varepsilon_1, \varepsilon_2)' \in R_\Delta \setminus R_\delta) \\ Q(x_1, x_2) &= Pr_{(\varepsilon_1, \varepsilon_2)}((\varepsilon_1, \varepsilon_2)' \in R_\Delta \cap R_\delta) + Pr_{(\varepsilon_1, \varepsilon_2)}((\varepsilon_1, \varepsilon_2)' \in R_\delta \setminus R_\Delta). \end{aligned}$$

However, this gives us a contradiction since from the properties of quadrants  $a + \mathcal{O}_{s_1 s_2}$  supposed in the beginning and the convexity of  $\mathcal{E}_{12}$  we have one of  $Pr_{(\varepsilon_1, \varepsilon_2)}((\varepsilon_1, \varepsilon_2)' \in R_\Delta \setminus R_\delta)$  and  $Pr_{(\varepsilon_1, \varepsilon_2)}((\varepsilon_1, \varepsilon_2)' \in R_\delta \setminus R_\Delta)$  is 0 whereas the other one is strictly positive. E.g., if  $\Delta_1 < \delta_1$ , then we must have  $\Delta_2 > \delta_2$  (as required by  $(\Delta_1 - \delta_1)(\Delta_2 - \delta_2) < 0$ ) and then  $Pr_{(\varepsilon_1, \varepsilon_2)}((\varepsilon_1, \varepsilon_2)' \in R_\Delta \setminus R_\delta) > 0$ ,  $Pr_{(\varepsilon_1, \varepsilon_2)}((\varepsilon_1, \varepsilon_2)' \in R_\delta \setminus R_\Delta) = 0$ . Note that if we now vary  $(x_1, x_2)$  within a small enough neighborhood, we will continue to obtain contradictions for  $Q(\tilde{x}_1, \tilde{x}_2)$  for covariate values  $(\tilde{x}_1, \tilde{x}_2)$  within that neighborhood. Thus, the contradiction will in fact be obtained on a positive mass set of  $(x_1, x_2)$ . This contradiction means that the set of two thresholds we are looking for is unique.

Once thresholds  $\alpha_{2,M_2-1,j_3,\dots,j_D}^{(1)}$ ,  $\tilde{\alpha}_{2,M_2-2,j_3,\dots,j_D}^{(2)}$  are identified, we can proceed analogously to Subcase 1 to identify  $\alpha_{2,M_2-1,j_3,\dots,j_{h-1},r(j_h),j_{h+1},\dots,j_D}^{(h)}$  for any  $h \geq 3$ . Thus, all the thresholds of  $R_{2,M_2-1,j_3,\dots,j_D}$  are identified.

Building on this result, we will next look at the rectangles  $R_{3,M_2-1,j_3,\dots,j_D}$  and  $R_{2,M_2-2,j_3,\dots,j_D}$  (one index change at a time) and identify their thresholds in a similar manner. Then we look at  $R_{3,M_2-2,j_3,\dots,j_D}$ ,  $R_{4,M_2-1,j_3,\dots,j_D}$ ,  $R_{2,M_2-3,j_3,\dots,j_D}$ , etc. until we identify thresholds of all the rectangles considered in this stage.

**Stage 4.** In this stage we build on the results of previous stages and consider rectangular border

regions where we allow discrete processes in three dimensions  $d_1, d_2, d_3$  to take any of their possible values whereas in all the other dimensions they take their boundary values. To analyze the identification of threshold considered in this stage, without a loss of generality we can take  $d_1 = 1, d_2 = 2, d_3 = 3$ .

When considering rectangles  $R_{q_1, q_2, q_3, j_3, \dots, j_D}$ , where  $j_h \in \{1, M_h\}$  for  $h \geq 4$ , the main idea is to start building the identification (the knowledge of relevant) of thresholds gradually, first, e.g. by taking with  $q_1 = 2, q_2 = M_2 - 1, q_3 = M_3 - 1$  which guarantees that at every step at most  $D$  thresholds are unknown. At every steps, we will have  $D - 3$  known thresholds fixed at  $\infty$  or  $-\infty$  and three other known thresholds be finite and identified from previous stages and steps.

The way in which one proceeds gradually depends on the properties of the support  $\mathcal{E}_{123}$  of  $(\varepsilon_1, \varepsilon_2, \varepsilon_3)'$ . Previously, in Step 3, we considered quadrants in  $\mathbf{R}^2$ . Now, we have to consider eight orthants in  $\mathbf{R}^3$  originating from  $(0, 0, 0)'$ :

$$\mathcal{O}_{s_1 s_2 s_3} = \{(s_1 \lambda_1, s_2 \lambda_2, s_3 \lambda_3) : \lambda_i \geq 0, i = 1, 2, 3\} \quad \text{for } s_1, s_2, s_3 \in \{+, -\}.$$

**Subcase 1** If there is a point  $\mathbf{e} = (e_1, e_2, e_3)' \in \mathcal{E}_{123}$  such that  $e + \mathcal{O}_{s_1 s_2 s_3}$  is fully contained in  $\mathcal{E}_{123}$ , then we can proceed with the identification of the thresholds from the “corner” with  $q_d = M_d - 1$  and taking  $\kappa_d$  as  $>$  if  $s_d = -$ , and with  $q_d = 2$  if and taking  $\kappa_d$  as  $\leq$  if  $s_d = +$ . The indices then change gradually by one further step in their respective directions.

For concreteness, suppose  $e + \mathcal{O}_{+, +, -}$  is fully contained in  $\mathcal{E}_{123}$ . By the condition of the theorem we necessarily have  $\underline{x}_{1,1} = -\infty, \underline{x}_{2,1} = -\infty, \bar{x}_{3,1} = \infty$ . Then we first consider  $q_1 = 2, q_2 = 2, q_3 = M_3 - 1$  ( $\kappa_1$  and  $\kappa_2$  are then  $>$  and  $\kappa_3$  is  $\leq$ ). Using results of Stage 3, we find that we only deal with  $D$  unknown thresholds  $\alpha_{2,2,M_3-1,j_4,\dots,j_D}^{(1)}, \alpha_{2,2,M_3-1,j_4,\dots,j_D}^{(2)}, \alpha_{2,2,M_3-2,j_4,\dots,j_D}^{(3)}$  and  $\alpha_{2,2,M_3-1,j_4,\dots,j_{h-1},r_h(j_h),j_{h+1},\dots,j_D}^{(h)}$ ,  $h \geq 4$ . Notice that at this stage, due to coherency requirements, it may be strictly fewer than  $D$  of these thresholds unknown. However, all  $D$  may potentially be unknown and that is why we need to develop a general identification strategy.

Consider the observed  $P(Y^{c_1} = y_2^{(1)}, Y^{c_d} = y_2^{(2)}, Y^{c_3} = y_{M_3-1}^{(2)} \cap_{h \geq 4} (Y^{c_h} = y_{j_h}^{(h)}) | x)$  and find its limit when  $x_{h,1} \rightarrow$  for  $h \geq 4$ . Just as before, we take  $x_{h,1} \rightarrow \underline{x}_{h,1}$  if  $j_h = 1$  and take  $x_{h,1} \rightarrow \bar{x}_{h,1}$

if  $j_h = M_h$ . In this limit we identify

$$\sum_{\ell_1=0}^1 \sum_{\ell_2=0}^1 (-1)^{\ell_1+\ell_2+\ell_3} F_{1,2,3;\leq,\leq,>} \left( \alpha_{1+\ell_1,2,M_3-1,j_4,\dots,j_D}^{(1)} - x_1 \beta_1, \right. \\ \left. \alpha_{2,1+\ell_2,M_3-1,j_4,\dots,j_D}^{(2)} - x_2 \beta_2, \alpha_{2,2,M_3-1-\ell_3,j_4,\dots,j_D}^{(3)} - x_3 \beta_3 \right) \in (0, 1).$$

Let us now take  $x_{1,1} \rightarrow -\infty$ ,  $x_{2,1} \rightarrow -\infty$ . Then the known limit becomes

$$F_{3,>} \underbrace{(\alpha_{2,2,M_3-2,j_4,\dots,j_D}^{(3)} - x_3 \beta_3)}_{z_2} - F_{3,>} \underbrace{(\alpha_{2,2,M_3-1,j_4,\dots,j_D}^{(3)} - x_3 \beta_3)}_{z_1} \in (0, 1).$$

Using the knowledge of  $F_{3,>}$  and the strict monotonicity of the obtained probability with respect to  $z_2$  when  $z_1$  is known (recall that  $\alpha_{2,2,M_3-1,j_4,\dots,j_D}^{(3)}$  is known from Stage 3) we identify  $\alpha_{2,2,M_3-2,j_4,\dots,j_D}^{(3)}$ .

Analogously, when taking  $x_{1,1} \rightarrow -\infty$ ,  $x_{3,1} \rightarrow \infty$ , the known limit becomes

$$F_{2,\leq} \underbrace{(\alpha_{2,2,M_3-1,j_4,\dots,j_D}^{(2)} - x_2 \beta_2)}_{z_2} - F_{2,\leq} \underbrace{(\alpha_{2,1,M_3-1,j_4,\dots,j_D}^{(2)} - x_2 \beta_2)}_{z_1} \in (0, 1).$$

We can identify  $\alpha_{2,2,M_3-1,\dots,j_D}^{(2)}$  using the knowledge of  $F_{2,\leq}$  and the strict monotonicity of the obtained probability with respect to  $z_2$  when  $z_1$  is known (recall that  $\alpha_{2,1,M_3-1,j_4,\dots,j_D}^{(2)}$  is known from Stage 3). Analogously we identify  $\alpha_{2,2,M_3-1,j_4,\dots,j_D}^{(1)}$ .

Once  $\alpha_{2,2,M_3-1,j_4,\dots,j_D}^{(1)}$ ,  $\alpha_{2,2,M_3-1,j_4,\dots,j_D}^{(2)}$  and  $\alpha_{2,2,M_3-2,j_4,\dots,j_D}^{(3)}$  are identified, we can identify  $\alpha_{2,2,M_3-1,j_4,\dots,j_{h-1},r_h(j_h),j_{h+1},\dots,j_D}^{(h)}$ ,  $h \geq 4$ , by taking in our observed probability  $x_{h,1} \rightarrow$  in the manner described above but now only for  $h \geq 4$ . E.g. for  $h = 4$  we identify

$$\sum_{\ell_1=0}^1 \sum_{\ell_2=0}^1 \sum_{\ell_3=0}^1 (-1)^{\ell_1+\ell_2+\ell_3+\ell_4} F_{1,2,3,4;\leq,\leq,>,\leq} \left( \alpha_{1+\ell_1,2,M_3-1,1,\mathbf{j}_{5:D}}^{(1)} - x_1 \beta_1, \right. \\ \left. \alpha_{2,1+\ell_2,M_3-1,1,\mathbf{j}_{5:D}}^{(2)} - x_2 \beta_2, \alpha_{2,2,M_3-1,1,\mathbf{j}_{5:D}}^{(3)} - x_3 \beta_3, \alpha_{2,2,M_3-1,\ell_4,\mathbf{j}_{5:D}}^{(4)} - x_4 \beta_4 \right) \in (0, 1)$$

if  $j_4 = 1$  (here  $\mathbf{j}_{5:D} \equiv (j_1, \dots, j_D)$ ), and identify

$$\sum_{\ell_1=0}^1 \sum_{\ell_2=0}^1 \sum_{\ell_3=0}^1 (-1)^{\ell_1+\ell_2+\ell_3+\ell_4} F_{1,2,3,4;\leq,\leq,>,>} \left( \alpha_{1+\ell_1,2,M_3-1,M_3,\mathbf{j}_{5:D}}^{(1)} - x_1 \beta_1, \right. \\ \left. \alpha_{2,1+\ell_2,M_3-1,M_3,\mathbf{j}_{5:D}}^{(2)} - x_2 \beta_2, \alpha_{2,2,M_3-1,M_3,\mathbf{j}_{5:D}}^{(3)} - x_3 \beta_3, \alpha_{2,2,M_3-1,\ell_4,\mathbf{j}_{5:D}}^{(4)} - x_4 \beta_4 \right) \in (0, 1)$$

if  $j_4 = M_4$ . No matter which of these situations we have, we use the knowledge of  $F_{1,2,3,4;\leq\leq>\kappa_4}$ , the knowledge of 7 out of 8 thresholds in them and the strict monotonicity of that limit with respect to the unknown threshold ( $\alpha_{2,2,M_3-1,1,\dots,j_D}^{(4)}$  in the first and  $\alpha_{2,2,M_3-1,M_4-1,j_4\dots,j_D}^{(4)}$  in the second situation) to identify this threshold. Proceeding analogously with other  $h \geq 4$ , all the thresholds of  $R_{2,2,M_3-1,j_4,\dots,j_D}$  are identified.

Building on this result, we will next look at the rectangles  $R_{3,2,M_3-1,j_4,\dots,j_D}$  and  $R_{2,3,M_3-1,j_4,\dots,j_D}$  and  $R_{2,2,M_3-2,j_4,\dots,j_D}$  (one index change at a time in the respective direction) and identify their thresholds in a similar manner and so on until we identify thresholds of all the rectangles considered in this stage.

**Subcase 2** Suppose there is no point  $\mathbf{e} = (e_1, e_2, e_3)' \in \mathcal{E}_{123}$  and no orthant  $\mathcal{O}_{s_1 s_2 s_3}$  such that  $\mathbf{e} + \mathcal{O}_{s_1 s_2 s_3}$  is fully contained in  $\mathcal{E}_{123}$ . Then the convexity and non-empty interior properties of  $\mathcal{E}_{123}$  guarantee that there is point  $\mathbf{e} \in \partial\mathcal{E}_{123}$  at the finite boundary of  $\mathcal{E}_{123}$  such that at least two adjacent orthants  $\mathbf{e} + \mathcal{O}_{s_1 s_2 s_3}$  (orthants  $\mathcal{O}_{s_1 s_2 s_3}$  and  $\mathcal{O}_{\tau_1 \tau_2 \tau_3}$  are adjacent if they have a common face – thus, at least two of signs are the same) do not intersect the interior of  $\mathcal{E}_{123}$  and four orthants  $\mathbf{e} + \mathcal{O}_{s_1, s_2, s_3}$  with a consistent sign in one dimension intersect  $\mathcal{E}_{123}$  in their interior.

The exact nature of these orthants will determine the direction of the proof (from which “corner” we start and which  $\kappa_d$ ,  $d = 1, 2, 3$ , we use in the proof). After this is decided, we consider the first 3-dimensional rectangle and assume there are two sets of thresholds. To derive a contradiction, we construct two 3-dimensional rectangles – with one determined by the first set of thresholds and the other determined by the second set of thresholds – near  $\mathbf{e}$ , and show their symmetric differences have mismatched masses under the distribution of  $(\varepsilon_1, \varepsilon_2, \varepsilon_3)'$  (e.g. one zero in an empty orthant, one positive in an intersecting orthant).

For concreteness, assume the consistent sign is in dimension 3 with  $s_3 = -$ , so all four orthants  $\mathbf{e} + \mathcal{O}_{s_1 s_2 -}$  (for  $s_1, s_2 \in \{+, -\}$ ) intersect the interior  $\text{int}(\mathcal{E}_{123})$  of  $\mathcal{E}_{123}$  (positive mass in any small rectangle with a vertex at  $\mathcal{E}_{123}$  and located in  $\cup_{s_1, s_2} (\mathbf{e} + \mathcal{O}_{s_1 s_2 -})$  then, by convexity).

First, consider the case when  $e_3$  in  $\mathbf{e}$  provides a global maximum value of  $\mathcal{E}_{123}$ . This implies that the interiors of four orthants  $\mathbf{e} + \mathcal{O}_{s_1 s_2 +}$  do not intersect  $\mathcal{E}_{123}$ . In this case we proceed in the decreasing order in dimension 3 (thus, choosing  $q_3 = M_3 - 1$  and  $\kappa_3$  as  $>$ ). Directions in other two dimensions can be any. For concreteness, let us take them to be increasing – thus, choose  $q_1 = 2$ ,  $q_2 = 2$  and  $\kappa_1, \kappa_2$  as  $\leq$ ).

Consider observed  $P\left(Y^{c_1} = y_2^{(1)}, Y^{c_2} = y_2^{(2)}, Y^{c_3} = y_{M_3-1}^{(3)}, \cap_{h \geq 4} (Y^{c_h} = y_{j_h}^{(h)}) \mid x\right)$  and find its limit when  $x_{h,1} \rightarrow$  for  $h \geq 4$  in a way described earlier (take  $x_{h,1} \rightarrow \underline{x}_{h,1}$  if  $j_h = 1$  and take  $x_{h,1} \rightarrow \bar{x}_{h,1}$  if  $j_h = M_h$ ). Denote  $\mathbf{j} \equiv (j_4, \dots, j_D)$ . In the described limit we identify

$$Q(x_1, x_2, x_3) = \sum_{\ell_1=0}^1 \sum_{\ell_2=0}^1 (-1)^{\ell_1+\ell_2+\ell_3} F_{1,2,3; \leq \leq \geq} (\alpha_{1+\ell_1, 2, M_3-1, \mathbf{j}}^{(1)} - x_1 \beta_1, \alpha_{2, 1+\ell_2, M_3-1, \mathbf{j}}^{(2)} - x_2 \beta_2, \alpha_{2, 2, M_3-1-\ell_3, \mathbf{j}}^{(3)} - x_3 \beta_3) \in (0, 1) \quad (14)$$

for all  $(x_1, x_2, x_3)$  in  $S_1(2) \cap S_2(2) \cap S_3(M_3-1)$ . The limit has three unknown thresholds:  $\alpha_{2,2,M_3-1,\mathbf{j}}^{(d)}$ ,  $d = 1, 2$ ,  $\alpha_{2,2,M_3-2,\mathbf{j}}^{(3)}$ . Suppose there is different set of such three threshold parameters – we can use the same notation but with tildes for this alternative threshold set – that generate the same observed probabilities (14) for all  $(x_1, x_2, x_3)$  in  $S_1(2) \cap S_2(2) \cap S_3(M_3-1)$ . Denote

$$\begin{aligned} \Delta_1 &= \alpha_{2,2,M_3-1,\mathbf{j}}^{(1)} - \alpha_{1,2,M_3-1,\mathbf{j}}^{(1)}, \quad \Delta_2 = \alpha_{2,2,M_3-1,\mathbf{j}}^{(2)} - \alpha_{2,1,M_3-1,\mathbf{j}}^{(2)}, \quad \Delta_3 = \alpha_{2,2,M_3-1,\mathbf{j}}^{(3)} - \alpha_{2,2,M_3-2,\mathbf{j}}^{(3)}, \\ \delta_1 &= \tilde{\alpha}_{2,2,M_3-1,\mathbf{j}}^{(1)} - \alpha_{1,2,M_3-1,\mathbf{j}}^{(1)}, \quad \delta_2 = \tilde{\alpha}_{2,2,M_3-1,\mathbf{j}}^{(2)} - \alpha_{2,1,M_3-1,\mathbf{j}}^{(2)}, \quad \delta_3 = \tilde{\alpha}_{2,2,M_3-1,\mathbf{j}}^{(3)} - \alpha_{2,2,M_3-2,\mathbf{j}}^{(3)}, \end{aligned}$$

for the two sets of thresholds. Clearly,  $\Delta_k > 0$ ,  $\delta_k > 0$ ,  $k = 1, 2, 3$ . The observational equivalence in terms of probabilities implies that  $\exists d_1, d_2 \in \{1, 2, 3\}$  such that  $\Delta_{d_1} > \delta_{d_1}$ ,  $\Delta_{d_2} < \delta_{d_2}$  as otherwise would mean that two sets of thresholds are ordered in the coordinate-wise sense and then it would be easy to obtain a contradiction otherwise from the properties of  $F_{1,2,3; \leq \leq \geq}$ . We can take any  $\Delta_d$  and  $\delta_d$ ,  $d = 1, 2, 3$ , be different (otherwise we would revert to earlier stages and obtain a contradiction from results there).

Define  $z_1 = e_1 + \min\{\delta_1, \Delta_1\}$ ,  $z_2 = e_2 - \min\{\delta_2, \Delta_2\}$ ,  $z_3 = e_3 - \min\{\delta_3, \Delta_3\}$  and choose  $x_1, x_2, x_3$  such that  $z_1 = \alpha_{1,2,M_3-1,\mathbf{j}}^{(1)} - x_1 \beta_1$ ,  $z_2 = \alpha_{2,1,M_3-1,\mathbf{j}}^{(2)} - x_2 \beta_2$ ,  $z_3 = \alpha_{2,2,M_3-2,\mathbf{j}}^{(3)} - x_3 \beta_3$  (all the thresholds used here are known at this identification stage). Define rectangles

$$T_v = [z_1, z_1 + v_1] \times [z_2, z_2 + v_2] \times [z_3, z_3 - v_3], \quad v \in \{\Delta, \delta\}. \quad (15)$$

Note that  $T_\Delta \cap T_\delta = [z_1, e_1] \times [z_2, e_2] \times [z_3, e_3] \in \mathbf{e} + \mathcal{O}_{--+}$ . Suppose  $\Delta_3 > \delta_3$ . Then the intersection of  $T_\Delta \setminus T_\delta$  with any neighborhood of  $\mathbf{e}$  as well as the intersection with the interior of  $\mathcal{O}_{---}$  is non-empty and these intersections have interiors in  $\mathbb{R}^3$ . Hence,  $Pr_{(\varepsilon_1, \varepsilon_2, \varepsilon_3)}((\varepsilon_1, \varepsilon_2, \varepsilon_3)' \in T_\Delta \setminus T_\delta) > 0$ . At the same time the interior of  $T_\Delta \setminus T_\delta$  is full contained in  $\cup_{s_1, s_2} \mathcal{O}_{s_1 s_2 +}$ . Hence  $Pr_{(\varepsilon_1, \varepsilon_2, \varepsilon_3)}((\varepsilon_1, \varepsilon_2, \varepsilon_3)' \in T_\delta \setminus T_\Delta) = 0$ , which gives us a contradiction. This contradiction can be obtained on a positive measure of  $(x_1, x_2, x_3)$  by moving around the boundary point  $\mathbf{e}$ . This contradiction eliminates a possibility of two different sets of thresholds that can generate observed

probabilities of choice.

Continuing with the case of all four orthants  $\mathbf{e} + O_{s_1 s_2 -}$  (for  $s_1, s_2 \in \{+, -\}$ ) intersecting  $\text{int}(\mathcal{E}_{123})$ , now consider the case when  $e_3$  in  $\mathbf{e}$  is not a global maximum of  $\mathcal{E}_{123}$  in dimension 3. Then by convexity of  $\mathcal{E}_{123}$  there will be two orthants among  $\mathbf{e} + O_{s_1 s_2 +}$  that have 3-dimensional intersections with  $\text{int}(\mathcal{E}_{123})$  (and of, course, by convexity of  $\mathcal{E}_{123}$  and by the fact that  $\mathbf{e}$  is at a finite boundary the other two such orthants will have no intersection with  $\text{int}(\mathcal{E}_{123})$ ). Suppose  $\mathbf{e} + O_{--+}$  and  $\mathbf{e} + O_{+-+}$  have no intersection with  $\text{int}(\mathcal{E}_{123})$  whereas  $\mathbf{e} + O_{-++}$  and  $\mathbf{e} + O_{+++}$  have 3-dimensional intersections with  $\text{int}(\mathcal{E}_{123})$ . The nature of these orthants determines that the proof proceeds in the increasing order in dimension 3 (thus, choosing  $q_3 = 2$  and  $\kappa_3$  as  $\leq$ ) and in the increasing order in dimension 2 (thus, choosing  $q_2 = 2$  and  $\kappa_2$  as  $\leq$ ). As for dimension 1, we can proceed in any direction by taking  $q_1$  to be either 2 or  $M_1 - 1$ , so let's e.g. choose  $q_1 = 2$  and  $\kappa_1$  to be  $\leq$  and, thus, proceed in the increasing direction too in this dimension. .

Consider observed

$$P \left( Y^{c_1} = y_2^{(1)}, Y^{c_2} = y_2^{(2)}, Y^{c_3} = y_2^{(3)}, \cap_{h \geq 4} (Y^{c_h} = y_{j_h}^{(h)}) \mid x \right)$$

and find its limit when  $x_{h,1} \rightarrow$  for  $h \geq 4$  in a way described earlier (take  $x_{h,1} \rightarrow \underline{x}_{h,1}$  if  $j_h = 1$  and take  $x_{h,1} \rightarrow \bar{x}_{h,1}$  if  $j_h = M_h$ ). In this limit we identify

$$\begin{aligned} Q(x_1, x_2, x_3) = \sum_{\ell_1=0}^1 \sum_{\ell_2=0}^1 (-1)^{\ell_1+\ell_2+\ell_3} F_{1,2,3; \leq \leq \leq} (\alpha_{1+\ell_1, 2, 2, \mathbf{j}}^{(1)} - x_1 \beta_1, \\ \alpha_{2, 1+\ell_2, 2, \mathbf{j}}^{(2)} - x_2 \beta_2, \alpha_{2, 2, 1+\ell_3, \mathbf{j}}^{(3)} - x_3 \beta_3) \in (0, 1) \end{aligned} \quad (16)$$

for all  $(x_1, x_2, x_3)$  in  $S_1(2) \cap S_2(2) \cap S_3(2)$ . The limit has three unknown thresholds:  $\alpha_{2,2,2,\mathbf{j}}^{(d)}$ ,  $d = 1, 2, 3$ . Suppose there is different set of such three threshold parameters – we can use the same notation but with tildes for this alternative threshold set – that generate the same observed probabilities (16) for all  $(x_1, x_2, x_3)$  in  $S_1(2) \cap S_2(2) \cap S_3(2)$ . Denote

$$\begin{aligned} \Delta_1 &= \alpha_{2,2,2,\mathbf{j}}^{(1)} - \alpha_{1,2,2,\mathbf{j}}^{(1)}, & \Delta_2 &= \alpha_{2,2,2,\mathbf{j}}^{(2)} - \alpha_{2,1,2,\mathbf{j}}^{(2)}, & \Delta_3 &= \alpha_{2,2,2,\mathbf{j}}^{(3)} - \alpha_{2,2,1,\mathbf{j}}^{(3)}, \\ \delta_1 &= \tilde{\alpha}_{2,2,2,\mathbf{j}}^{(1)} - \alpha_{1,2,2,\mathbf{j}}^{(1)}, & \delta_2 &= \tilde{\alpha}_{2,2,2,\mathbf{j}}^{(2)} - \alpha_{2,1,2,\mathbf{j}}^{(2)}, & \delta_3 &= \tilde{\alpha}_{2,2,2,\mathbf{j}}^{(3)} - \alpha_{2,2,1,\mathbf{j}}^{(3)}, \end{aligned}$$

for the two sets of thresholds. Clearly,  $\Delta_k > 0$ ,  $\delta_k > 0$ ,  $k = 1, 2, 3$ . Similar to before, the observational equivalence in terms of probabilities implies that  $\Delta_{d_1} > \delta_{d_1}$ ,  $\Delta_{d_2} < \delta_{d_2}$  for some  $\exists d_1, d_2 \in \{1, 2, 3\}$  and we can take any  $\Delta_d$  and  $\delta_d$ ,  $d = 1, 2, 3$ , be different.

**Sub-case 1A.**  $\Delta_1 > \delta_1, \Delta_2 > \delta_2, \Delta_3 < \delta_3$ . Define  $z_1 = e_1 - \delta_1, z_2 = e_2 - \delta_2, z_3 = e_3 - \Delta_3$  and choose  $x_1, x_2, x_3$  such that  $z_1 = \alpha_{1,2,2,\mathbf{j}}^{(1)} - x_1\beta_1, z_2 = \alpha_{2,1,2,\mathbf{j}}^{(2)} - x_2\beta_2, z_3 = \alpha_{2,2,1,\mathbf{j}}^{(3)} - x_3\beta_3$  (note that all the thresholds used here are known at this identification stage). Define

$$T_v = [z_1, z_1 + v_1] \times [z_2, z_2 + v_2] \times [z_3, z_3 + v_3], \quad v \in \{\Delta, \delta\}. \quad (17)$$

Note that  $T_\Delta \cap T_\delta = [z_1, e_1] \times [z_2, e_2] \times [z_3, e_3] \in \mathbf{e} + \mathcal{O}_{---}$ . Note that  $T_\delta \setminus T_\Delta = [z_1, e_1] \times [z_2, e_2] \times (e_3, e_3 + \delta_3 - \Delta_3]$  is in the closure of  $\mathcal{O}_{--+}$  and, thus, has probability 0. At the same time,  $T_\Delta \setminus T_\delta \in \mathcal{O}_{+--} \cup \mathcal{O}_{-+-} \cup \mathcal{O}_{++-}$  and its intersection with any neighborhood of  $\mathbf{e} = (e_1, e_2, e_3)'$  has a non-empty 3-dimensional interior. By the properties of  $\mathcal{E}_{123}$  and its boundary point  $\mathbf{e}$ , this implies that  $Pr_{(\varepsilon_1, \varepsilon_2, \varepsilon_3)}((\varepsilon_1, \varepsilon_2, \varepsilon_3)' \in T_\Delta \setminus T_\delta) > 0$ , which gives us a contradiction in this sub-case. This contradiction can be obtained on a positive measure of  $(x_1, x_2, x_3)$  by moving around the boundary point  $\mathbf{e}$ .

**Sub-case 1B:**  $\Delta_1 > \delta_1, \Delta_2 < \delta_2, \Delta_3 > \delta_3$ , and **Sub-case 1C:**  $\Delta_1 > \delta_1, \Delta_2 < \delta_2, \Delta_3 < \delta_3$ . Define  $z_1 = e_1 - \delta_1, z_2 = e_2 - \Delta_2, z_3 = e_3$ . Choose  $x_1, x_2, x_3$  such that  $z_1 = \alpha_{1,2,2,\mathbf{j}}^{(1)} - x_1\beta_1, z_2 = \alpha_{2,1,2,\mathbf{j}}^{(2)} - x_2\beta_2, z_3 = \alpha_{2,2,1,\mathbf{j}}^{(3)} - x_3\beta_3$  (note that all the thresholds used here are known at this identification stage). Define rectangles  $T_\Delta, T_\delta$  as in (15). Then  $T_\delta \setminus T_\Delta$  is in  $\mathcal{O}_{-++}$  and its intersection with any neighborhood of  $\mathbf{e}$  in  $\mathcal{O}_{-++}$  has a non-empty interior in  $\mathbb{R}^3$ . This implies that its probability is strictly positive. At the same time,  $T_\Delta \setminus T_\delta$  is in the closure of  $\mathcal{O}_{+--}$  and has probability zero. This gives a contradiction with a positive probability since we can vary  $(z_1, z_2, z_3)$  (and, hence,  $(x_1, x_2, x_3)$ ) slightly and get a contradiction through the discontinuity of probabilities of  $T_\delta \setminus T_\Delta$  and  $T_\Delta \setminus T_\delta$  then no orthants  $\mathcal{O}_{s_1 s_2 +}$  in its interior intersects  $\mathcal{E}_{123}$ .

*Sub-case 1D.*  $\Delta_1 < \delta_1, \Delta_2 > \delta_2, \Delta_3 < \delta_3$ . This is completely analogous to Sub-case 1B with the roles of  $\Delta$ 's and  $\delta$ 's reversed.

*Sub-case 1E.*  $\Delta_1 < \delta_1, \Delta_2 < \delta_2, \Delta_3 > \delta_3$ . This is completely analogous to Sub-case 1A with the roles of  $\Delta$ 's and  $\delta$ 's reversed.

*Sub-case 1F.*  $\Delta_1 < \delta_1, \Delta_2 > \delta_2, \Delta_3 > \delta_3$ . This is completely analogous to Sub-case 1C with the roles of  $\Delta$ 's and  $\delta$ 's reversed.

Contradictions obtained in each Sub-case mean that the set of two thresholds we are looking for is unique. Hence, thresholds  $\alpha_{2,2,2,\mathbf{j}}^{(d)}, d = 1, 2, 3$  are identified. After that we can proceed analogously to Subcase 1 to identify  $\alpha_{2,2,2,j_4, \dots, j_{h-1}, r(j_h), j_{h+1}, \dots, j_D}^{(h)}$  for any  $h \geq 4$ . Thus, all the thresholds of  $R_{2,2,2,\mathbf{j}}$

are identified. Building on this result, we will next look at the rectangles  $R_{3,2,2,j}$ ,  $R_{2,3,2,j}$  and  $R_{2,2,3,j}$  (one index change at a time) and identify their thresholds in a similar manner. Then we look at  $R_{3,3,2,j}$ ,  $R_{3,2,3,j}$ ,  $R_{2,3,3,j}$ , etc. until we identify thresholds of all the rectangles considered in this stage.

**Stages 5 to  $D+1$ ,** Stage  $m$  here would deal with the case when  $D - (m - 1)$  out of  $D$  discrete responses are fixed at their boundary values but the rest can take any values. Identification would proceed sequentially analogously to Stages 3 and 4. ■

### 3 Appendix C: Additional simulation results for Design 2

In Table 5, we provide the results for the thresholds in simulation Design 2. It is evident how poorly the lattice model does in this case, relative to essentially no bias in the non-lattice model.

### 4 Appendix D: Additional illustration for the cryptocurrency application

TABLE 5: Simulation results: Design 2 thresholds

| Parameter           | Truth | Non-lattice model | Lattice model |
|---------------------|-------|-------------------|---------------|
| $\alpha_{11}^{(1)}$ | -3.25 | -3.27 (0.12)      |               |
| $\alpha_{12}^{(1)}$ |       | -3.24 (0.12)      | -1.48 (0.04)  |
| $\alpha_{13}^{(1)}$ | -0.5  | -0.50 (0.07)      |               |
| $\alpha_{21}^{(1)}$ | 0.5   | 0.51 (0.09)       |               |
| $\alpha_{22}^{(1)}$ | 1     | 0.97 (0.14)       | 1.59 (0.04)   |
| $\alpha_{23}^{(1)}$ | 5     | 5.02 (0.13)       |               |
| $\alpha_{31}^{(1)}$ |       | 8.03 (0.19)       |               |
| $\alpha_{32}^{(1)}$ | 8     | 8.03 (0.19)       | 5.12 (0.09)   |
| $\alpha_{33}^{(1)}$ |       | 8.03 (0.19)       |               |
| $\alpha_{11}^{(2)}$ | -4    | -3.94 (0.32)      |               |
| $\alpha_{21}^{(2)}$ | -2    | -2.04 (0.16)      | -1.10 (0.04)  |
| $\alpha_{31}^{(2)}$ |       | -1.99 (0.09)      |               |
| $\alpha_{41}^{(2)}$ | 0     | -0.01 (0.09)      |               |
| $\alpha_{12}^{(2)}$ |       | 0.50 (0.05)       |               |
| $\alpha_{22}^{(2)}$ | 0.5   | 0.50 (0.05)       | 0.90 (0.04)   |
| $\alpha_{32}^{(2)}$ |       | 0.50 (0.05)       |               |
| $\alpha_{42}^{(2)}$ | 4     | 3.99 (0.17)       |               |

Notes: Sample means and sample standard deviations (in parentheses) of the estimates of the Design 2 threshold parameters, over 250 repeated samples.

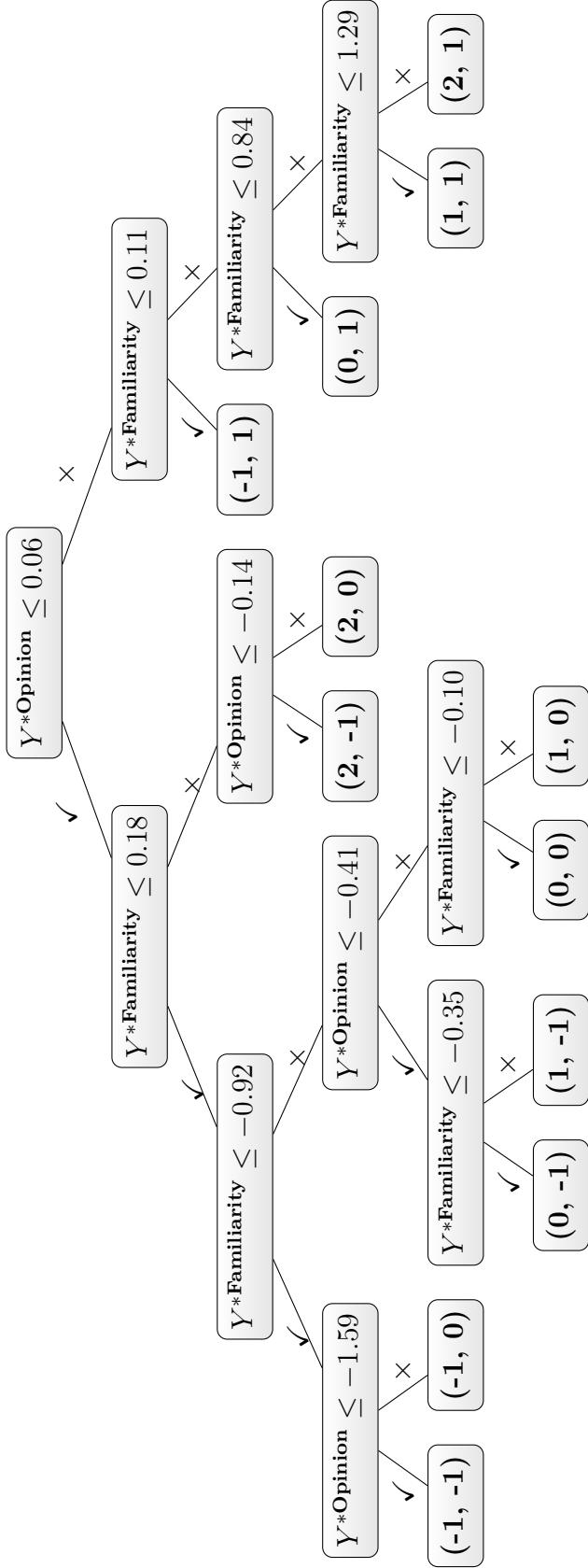


FIGURE 12: Binary decision tree describing the sequential (hierarchical) decision process