The Anatomy of Dependence in Multivariate Ordered Choice

Tatiana Komarova and William Matcham*

June 25, 2026

Abstract

When individuals make simultaneous decisions across multiple ordered dimensions, standard multivariate ordered choice models impose narrow bracketing: each dimension is decided as if the others did not exist. We develop a general rectangular structure model capturing broad bracketing while nesting the standard model. The framework introduces two layers of dependence, one through the decision-threshold structure and the other through the correlation of latent utilities. We provide microfoundations, prove identification, and discuss estimation. We showcase the model in applications to parental investment in children's education, cryptocurrency familiarity and expectations, and health insurance, where the latter separates moral hazard from adverse selection.

1 Introduction

How do people make simultaneous decisions across multiple ordered dimensions? How does a worker choose a health insurance plan and adjust healthcare utilization? How do parents allocate time and money to their children's education? In these cases, each of the two decisions is shaped by the other, and the economic content of that interaction—such as complementarity and substitutability, or moral hazard and adverse selection—is precisely what researchers wish to recover.

The standard approach extends univariate ordered response models, dimension by dimension. Each dimension has its own latent utility and its own thresholds, and these thresholds are assumed to be functionally independent across dimensions. This is what we call the *lattice model*: its threshold intersections form a lattice in the latent space, and it corresponds to an agent who *narrowly brackets*, making each decision as if the others did not exist, thereby prioritizing simpler choice rules over general joint utility optimization.[1] All dependence between decisions enters only through the correlation of latent utilities.[2] This oversimplified decision-making process is illustrated in the left panel of Figure 1.

*Komarova: University of Cambridge (tk670@cam.ac.uk); Matcham: Royal Holloway University of London (william.matcham@rhul.ac.uk). This paper was previously circulated as "Multivariate ordered discrete response models with two layers of dependence." We are grateful to participants at Yale, Harvard-MIT, Erasmus University Rotterdam, LSE-STICERD seminars, and EC^2 2023 conference for helpful comments.

[1] Functionally independent thresholds are compatible with additively separable joint utility, as shown in Section 3. They are not, however, implied by general joint utility specifications. The lattice model should therefore be understood not as ruling out joint utility maximization altogether, but as sharply restricting the cross-dimensional interactions that joint utility can generate in the induced decision rules.

[2] In an auction context, the assumption of functionally independent decision thresholds 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, Komarova and Schiraldi (2023) and Gentry et al. (2019) for more detail.

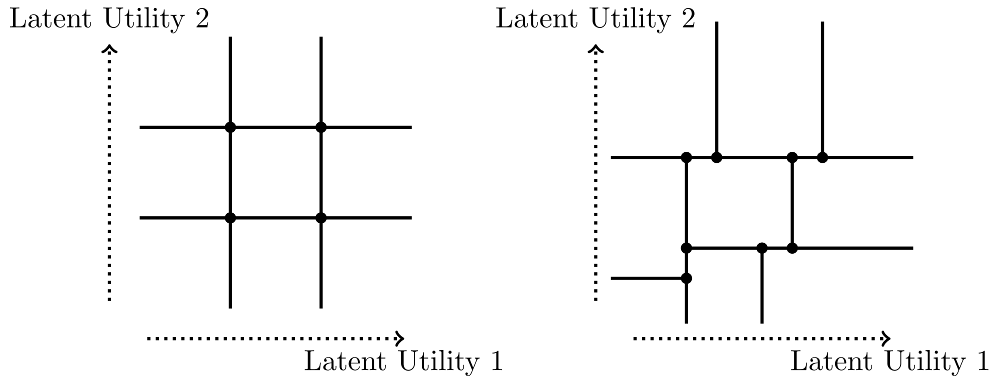


Figure 1: Models with a lattice (left) and a general rectangular structure (right)

This paper introduces a richer class of models we call *general rectangular structures*. The key innovation is allowing each dimension's decision threshold to depend on the outcome indices of all other dimensions. The right panel of Figure 1 illustrates the models we propose. The latent space remains partitioned into rectangular regions, and a clean mapping from latent utilities to observed discrete outcomes is preserved, but the boundaries of those rectangles can shift as a function of where the agent stands in other dimensions. This captures *broad bracketing* since the agent evaluates their options jointly, and their decision rule in one dimension adjusts to reflect the constraints and opportunities in others. The lattice model is the special case in which no such adjustment occurs. The flexibility of our new, more comprehensive, model enables us to capture more complex economic behavior than lattice models.

Allowing thresholds to shift across dimensions introduces a logical consistency requirement that is absent from lattice models. We formalize this problem as *coherency* and define it as the requirement that the rectangular regions partition the latent space. We characterize it through a set of local conditions on adjacent thresholds that capture choices in an immediate neighborhood. Our notion of coherency is the econometric counterpart of internal consistency in decision-making, and it constrains the threshold structure in a way that can be directly enforced in the estimation.

The separation of two layers of dependence is the central economic payoff of the framework. In any multivariate ordered choice setting, observed covariation between outcomes reflects two conceptually distinct forces. One is the *threshold layer*, through which one dimension's decision rule responds to the other, and the other one is the *unobservables layer*, through which latent shocks are correlated conditional on observables. In a lattice model, both forces are absorbed into a single dependence layer in correlation, which becomes a mixture of the two and is therefore interpretable as neither. Our framework assigns each force to a separate object, enabling the researcher to ask whether dependence between outcomes reflects a structural behavioral link or correlation of unobservables. For instance, the threshold decision rule may exhibit substitutability, while the dependence structure of the unobservables in the latent utilities may reflect complementarities. Furthermore, within the threshold decision structure itself, patterns of complementarity or substitutability can vary across different regions of the latent utility space. Another advantage of general rectangular structure models relative to lattice models is that they allow partial effects on the conditional probability of exceeding a given level to change sign across the domain, capturing more nuanced and flexible behavioral patterns. They also allow indirect effects of covariates from other processes on the conditional choice probability for a given process. 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.

Some might view the use of sequential timing as a natural response to the limitations of lattice models. This approach involves deciding one dimension first, with the other's thresholds then free to depend on that first-stage outcome. However, restricting thresholds to a strict sequential order blocks certain feedback channels by construction: for instance, if the insurance decision rule is modeled as functionally independent of utilization outcomes, individuals cannot select insurance based on their anticipated behavioral response. Since the ordering of decisions is challenging to observe, such assumptions are far from innocuous. Our framework nests sequential timing as a testable special case rather than a maintained assumption.

From a complementary conceptual perspective, the paper provides two microfoundations for general rectangular structures. The first introduces synergies and crowding-out effects directly into joint utility, while the second shows how general rectangular models arise from discretizing a continuous joint utility maximization problem.

Econometrically, we develop a semiparametric setting for our general rectangular structure models and formalize how the models generate richer behavior than lattice specifications, especially through flexible partial effects and cross-process covariate effects. We establish identification sequentially, beginning with parameters associated with exclusive covariates and eventually turning to the more involved identification of thresholds. The argument relies on independence between unobservables and covariates, as well as eventually on the presence of at least one relevant exclusive covariate in each process. We also discuss semiparametric estimation, building on Coppejans (2007) and imposing coherency through equality restrictions on thresholds.

We also study a parametric version in which unobservables are multivariate normal. A bivariate numerical exercise illustrates identification under weaker conditions than in the semiparametric case. We then discuss maximum likelihood estimation under the same coherency constraints and present Monte Carlo evidence comparing the proposed general rectangular structure (also called "non-lattice") probit estimator with standard bivariate ordered probit estimators, the latter of which effectively estimates a misspecified lattice model.

We demonstrate the practical importance of our general rectangular structures and the separation of two dependence layers in three applications. In U.S. health insurance markets, we draw on the Medical Expenditure Panel Survey and revisit the classic question of separating moral hazard from adverse selection. A dominant method since Chiappori and Salanie (2000) is to estimate a bivariate probit model and interpret the sign of the latent correlation. As Cohen and Siegelman (2010) emphasize, a positive correlation conflates adverse selection with moral hazard. Separating these two forces has required either quasi-experimental variation or strong structural assumptions. Our general rectangular structure resolves the identification problem and disentangles moral hazard from selection (adverse or advantageous) differently 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. We find that (i) the asymmetric information is mostly confined to moral hazard, (ii) adverse selection is minimal, and (iii) the general rectangular structure estimates are inconsistent with selection on moral hazard. These three findings are similar to recent papers in the literature that rely on context-specific identification strategies or fully fledged structural economic models, neither of which we require. Finally, we also estimate non-zero partial effects of the partner's insurance status on own utilization, which the lattice model rules out.

Our second application examines the relationship between cryptocurrency familiarity and optimism. It shows how the non-lattice approach reveals differences in how individuals form and express opinions about bitcoin's value, obscured by the lattice model. We also consider a sequential model in which cryptocurrency familiarity is determined first with thresholds functionally independent of optimism, then with thresholds for optimism allowed to depend on familiarity, and compare its findings and content with the general rectangular structure model. This application illustrates that the framework matters for index coefficient estimates: the sign of the male coefficient in the bitcoin optimism equation reverses between the lattice and non-lattice specifications, because the lattice coefficient conflates the structural effect with threshold misattribution.

The third application revisits parental investment in children's human capital using the PSID Child Development Supplement. The literature, following Del Boca, Flinn and Wiswall (2014), typically assumes complementarity between financial and time investments in child quality production. Our model allows the data to reveal the sign of this relationship at each layer separately. At the threshold layer, higher home cognitive stimulation lowers the financial spending threshold, confirming production complementarity in parental decision making. At the unobservables layer, however, the correlation between the two is negative, reflecting a labor-supply trade-off. More precisely, households in which parents work more accumulate money but lose time. The lattice model captures only the net effect and returns a moderate positive correlation, obscuring both mechanisms.

The paper proceeds as follows. After reviewing related work and situating our contribution in the literature, Section 2 introduces general rectangular structures and coherency. Section 3 provides microfoundations. Sections 4 and 5 develop identification and estimation for the semiparametric and parametric cases, respectively. Section 6 presents Monte Carlo evidence, Section 7 presents three empirical applications, and Section 8 concludes.

Our contributions and literature review

Our paper contributes to the literature on the economic content and econometrics of ordered choice models. The literature on *univariate* ordered response models is extensive. Cameron and Heckman (1998), Heckman, Lalonde and Smith (1999), and Carneiro, Hansen and Heckman (2003) develop foundational frameworks that ground thresholds in economic decision-making. Cunha, Heckman and Navarro (2007) generalizes this by allowing thresholds to depend on both observables and unobservables, capturing dynamic settings such as schooling decisions. On the semiparametric side, Lewbel (2000) establishes identification of binary, ordered, and multinomial choice models via a continuous special regressor that is conditionally independent of the error. Lewbel (2003) extends this approach to ordered models with random or sequentially determined thresholds.[3] Reduced-form treatments of thresholds as psychological cut-points or cost-benefit barriers without optimization foundations are surveyed in Greene and Hensher (2010) and Boes and Winkelmann (2006). Anderson (1984) pursues a related approach in which thresholds relate to category proportions rather than absolute utility levels.[4] Bhat and Pulugurta (1998) derives thresholds from a range-based utility specification, and Apesteguia and Ballester (2023) proposes type-ordered random utility models in which ordered choices arise from heterogeneous preference types without restrictive distributional assumptions.

Our paper moves this body of work from univariate to *multivariate* ordered decisions. The key departure is that thresholds in each dimension depend on the realized discrete outcomes of the other dimensions. This requires a joint model of all endogenous processes and introduces a flexibility in the threshold structure unavailable in any existing framework. Our contribution to the microfoundations of ordered choice (see Section 3) speaks to the strand of univariate models focused on utility optimization, as we show that general rectangular structures arise naturally both from joint utility specifications with complementarities/substitutabilities, and discretizing a continuous joint maximization problem.

The economic content of threshold interdependence connects our work to the literature on *choice*

[3] Lewbel (2014) provides a unified treatment of the special regressor method.

[4] In the Anderson (1984) model, ordinal categories are not tied to a single latent variable with fixed thresholds.

bracketing. The question of whether agents evaluate options jointly or dimension by dimension has several interpretations. For example, it is *sequential vs. simultaneous choice* in Simonson and Winer (1992), *narrow vs. broad decision frames* in Kahneman and Lovallo (1993), *local vs. overall value functions* in Heyman (1996), and *isolated vs. distributed choice* in Herrnstein and Prelec (1991). This literature is largely theoretical and experimental (Tversky and Kahneman, 1981; Read, Loewenstein and Rabin, 1999; Thaler, 1999; Rabin and Weizsäcker, 2009; Lian, 2020; Camara, 2021; Zhang, 2021), with only a few descriptive or structural empirical studies (Camerer et al., 1997; Thakral and Tô, 2021).[5] What has been missing is a structural econometric model that simultaneously nests both behaviors and allows the data to reveal which is operative. Our framework provides this. The lattice model is the precise econometric representation of narrow bracketing, whereas the general rectangular structure is its broad-bracketing generalization. The hypothesis of narrow bracketing becomes a testable restriction, which can be implemented via likelihood ratio tests.

A tangentially related literature is the discrete choice framework with *strategic interactions*, in which one player's outcomes depend on the others' actions (Tamer, 2003; Berry and Reiss, 2007; Ciliberto and Tamer, 2009; Honoré and De Paula, 2010; Chesher and Rosen, 2017, 2020; Aradillas-López and Rosen, 2022). A central challenge in that literature is that best-response correspondences can yield incoherent or incomplete outcome mappings. As we detail in the following section, by *coherency* we mean logical consistency in decision-making ensuring that rectangular regions representing different discrete responses do not overlap and together cover the entire latent space $\mathbb{R}^D$.[6] Our setting is conceptually distinct from models with strategic interactions, as we have a *single* agent choosing across multiple dimensions. Consequently, this decision problem is free from the strategic-interaction sources of incoherency studied in that literature and is internally consistent by construction. But since general rectangular structures do not guarantee coherency, explicit threshold constraints for coherency will be required and enforced in estimation, and characterizing these constraints is one of our theoretical contributions.

2 Model with a general rectangular structure

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

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

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

$$R_{j_1,\ldots,j_D} = \bigtimes_{d=1}^{D} \left(\alpha^{(d)}_{j_1,\ldots,j_{d-1},j_d-1,j_{d+1},\ldots,j_D}, \alpha^{(d)}_{j_1,\ldots,j_{d-1},j_d,j_{d+1},\ldots,j_D} \right], \tag{1}$$

with thresholds $\alpha^{(d)}_{j_1,\ldots,j_{d-1},j_d,j_{d+1},\ldots,j_D}$ increasing in j_d for fixed other indices and normalized at the boundary as

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

[5] Tversky and Kahneman (1981) provides a classic example of narrow bracketing in experimental settings.

[6] For more on coherency, see Heckman (1978), Lewbel (2007) and Tamer (2003). Tamer (2003) distinguishes 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.

The threshold intersections in Definition 1 do not *necessarily* form a lattice in $\mathbb{R}^D$. Instead, they reflect functionally interdependent decision rules, akin to *broad bracketing* in behavioral economics.[7]

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

$$(Y^{c_1},\ldots,Y^{c_D}) = (y^{(1)}_{j_1},\ldots,y^{(D)}_{j_D}) \quad \iff \quad Y^{*c_d} \in \left(\alpha^{(d)}_{j_d-1}, \alpha^{(d)}_{j_d}\right] \quad \forall d, \ \textit{with}$$

$$\alpha^{(d)}_{j_d} = +\infty \textit{ for } j_d = M_d, \quad \alpha^{(d)}_{j_d} = -\infty \textit{ for } j_d = 0.$$

Since the general rectangular structure model allows for the lattice form, the synonym "non-lattice" should not be viewed as a ruling out of the possibility of a lattice structure but instead a model that does not need to have a lattice structure.

In Definition 2, thresholds are functionally independent across dimensions, forming a lattice in $\mathbb{R}^D$ at their intersections. Lattice models describe decisions made dimension by dimension and will misspecify a decision maker with more complex decision rules.

Thus, in our setting, the distinction between broad and narrow bracketing is fully captured by the functional interdependence or independence of decision rules, which is determined by the thresholds. Both lattice and non-lattice models permit correlated decisions (conditional on observables) through latent processes, but the latter distinguishes 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 $R_{j_1,\ldots,j_D}$ is exhaustive or mutually exclusive. As a result, the condition that each latent profile maps to exactly one observed response (which is associated with logical consistency in decision-making) and to which we refer as *coherency* is not guaranteed by Definition 1. 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 coherency requires constraints on the thresholds. Lattice models, by contrast, are coherent by construction.

We examine coherency in a bivariate general rectangular structure ordered response model and provide a formal condition on thresholds for the model to be coherent. The characterization of coherency for $D > 2$ is more involved and is presented 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 given by $\left\{\alpha^{(1)}_{j_1,j_2}\right\}^{M_1-1,M_2}_{j_1=1,j_2=1} \bigcup \left\{\alpha^{(2)}_{j_1,j_2}\right\}^{M_1,M_2-1}_{j_1=1,j_2=1}$. Given threshold normalizations at the boundary, the model is coherent – i.e., the latent space is partitioned into mutually exclusive and exhaustive rectangular regions $R_{j_1,j_2} = (\alpha^{(1)}_{j_1-1,j_2}, \alpha^{(1)}_{j_1,j_2}] \times (\alpha^{(2)}_{j_1,j_2-1}, \alpha^{(2)}_{j_1,j_2}]$ each corresponding to a unique observed outcome – if and only if, for all (j_1, j_2), $j_1 = 1,\ldots,M_1-1$, $j_2 = 1,\ldots,M_2-1$,*

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

[7] Definition 1 outlines the intended threshold decision rules. However, for this mapping to be valid for every latent draw, the defined regions must be exhaustive and mutually exclusive. Definition 1 alone does not guarantee this division and we discuss this below under "Coherency".

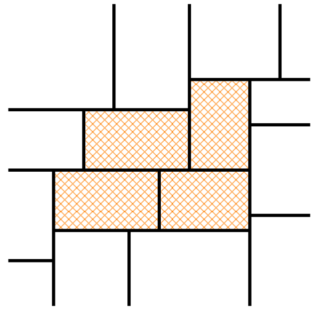


Figure 2: Intuition for a non-lattice model being coherent in the bivariate case

In other words, for the model to be coherent, the thresholds must satisfy a local condition for each 2×2 block of adjacent cells. Specifically, within each block, at least one of the dimensions must have constant thresholds across that block. One economic interpretation of that is when an agent faces a choice within a 2×2 block, they make their decision sequentially: first along one dimension, where the thresholds remain fixed, and then along the other. 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 well be the case that for lower levels 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).

3 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 a larger threshold in the second dimension when $Y^{c_1} = 2$. It shows that choosing a higher level in dimension 1 makes it harder to choose a higher level in the other dimension, for example, 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 through a lower threshold.

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

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

where $D = \mathbf{1}(Y^{*c_1} > 0)\mathbf{1}(Y^{*c_2} > v)$. Then $\text{argmax}_{j_1,j_2} U(j_1, j_2)$ is (i) (1,1) when $Y^{*c_1}, Y^{*c_2} \leqslant 0$; (ii) $(1,2)$ when $Y^{*c_1} \leqslant 0, Y^{*c_2} > 0$; (iii) $(2,1)$ when $Y^{*c_1} > 0, Y^{*c_2} \leqslant v$; 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 of 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 sufficiently high. This reflects scenarios where choosing one high option reduces the marginal cost of the other, such as economies of scale, shared fixed cost, learning spillovers, subsidies, or other mechanisms through which one high-level choice raises the marginal value of the other.

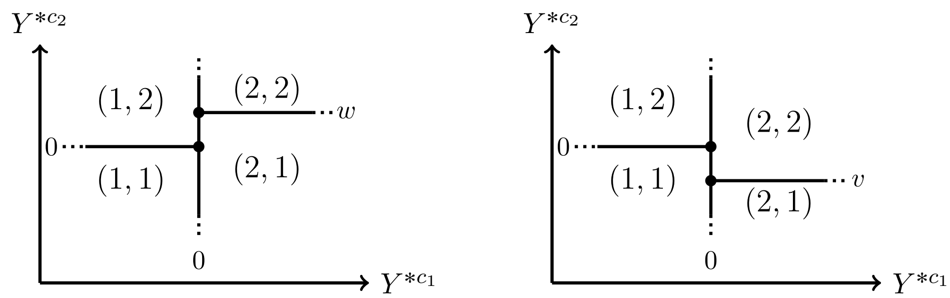


Figure 3: Illustration of first microfoundation in the bivariate case with two discrete options in each dimension

To obtain substitutes in the decision structure (as in the left panel of Figure 3), consider $w > 0$ and define $D = \mathbf{1}(Y^{*c_1} > 0)\mathbf{1}(Y^{*c_2} \leqslant w)$, $U(1,1) = 0$, $U(1,2) = Y^{*c_2}$, $U(2,1) = Y^{*c_1} + wD$, $U(2,2) = Y^{*c_1} + Y^{*c_2}$. This utility specification delivers the desired rectangular decision rule, with a unique argmax away from threshold boundaries. When $Y^{*c_1} > 0$ and $Y^{*c_2} < w$, $D = 1$ and $U(2,1) - U(2,2) = w - Y^{*c_2} > 0$, so $(2,1)$ is preferred to $(2,2)$. When $Y^{*c_2} > w$, $D = 0$ and $U(2,2)$ strictly exceeds both $U(2,1)$ and $U(1,2)$ because $Y^{*c_2} > 0$ and $Y^{*c_1} > 0$. The remaining regions follow analogously. The term wD is a specialization premium. It captures substitutability: when the high-level choice in the first dimension is attractive, but the second propensity remains below w, the agent receives an additional payoff for choosing the high option in the first dimension alone rather than choosing both high options jointly. Economically, this reflects competition for a common resource such as time, effort, attention, or budget. Once Y^{*c_2} exceeds w, the value of the second high-level choice dominates the specialization premium, and the joint alternative $(2,2)$ becomes optimal. Analogous constructs 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 other, similar, factors. If one had a smooth function $U(y^{(1)}, y^{(2)})$ of continuous responses $(y^{(1)}, y^{(2)})$ then the global maximum $(\overline{y}^{(1)}, \overline{y}^{(2)})$ would have necessarily satisfied $\frac{\partial U(\overline{y}^{(1)},\overline{y}^{(2)})}{\partial y^{(1)}} = 0$, $\frac{\partial U(\overline{y}^{(1)},\overline{y}^{(2)})}{\partial y^{(2)}} = 0$. With the discrete grid $(y^{(1)}_{j_1}, y^{(2)}_{j_2})$ 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, $(y^{(1)}_{j_1}, y^{(2)}_{j_2})$ is the maximizer of U on the grid only if

$$U(y^{(1)}_{j_1}, y^{(2)}_{j_2}) - U(y^{(1)}_{j_1-1}, y^{(2)}_{j_2}) > 0, \quad U(y^{(1)}_{j_1+1}, y^{(2)}_{j_2}) - U(y^{(1)}_{j_1}, y^{(2)}_{j_2}) \leqslant 0, \tag{4}$$

$$U(y^{(1)}_{j_1}, y^{(2)}_{j_2}) - U(y^{(1)}_{j_1}, y^{(2)}_{j_2-1}) > 0, \quad U(y^{(1)}_{j_1}, y^{(2)}_{j_2+1}) - U(y^{(1)}_{j_1}, y^{(2)}_{j_2}) \leqslant 0 \tag{5}$$

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

$$U(y^{(1)}_{j_1}, y^{(2)}_{j_2}) - U(y^{(1)}_{j_1-1}, y^{(2)}_{j_2}) = Y^{*c_1} - \alpha^{(1)}_{j_1-1,j_2}.$$

$$U(y^{(1)}_{j_1}, y^{(2)}_{j_2}) - U(y^{(1)}_{j_1}, y^{(2)}_{j_2-1}) = Y^{*c_2} - \alpha^{(2)}_{j_1,j_2-1}.$$

The displayed equations should be understood as assumptions on local discrete marginal utilities sufficient to generate the rectangular decision rule. Assuming coherency, 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 $(Y^{*c_1}, Y^{*c_2}) \in R_{j_1,j_2}$.

4 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:

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

where x_d is a row vector of observable covariates, β_d is a k_d-dimensional column vector of unknown parameters, and ε_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 ε_d capturing random shocks to the latent utility. The shocks $\varepsilon_1, \ldots, \varepsilon_D$ may be dependent, allowing latent processes Y^{*c_d} to be correlated conditional on covariates. This linear index, which is standard in discrete choice models, supports the 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: β_d, $d = 1, ..., D$, the joint c.d.f of $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_D)'$, and the thresholds. We start by imposing Assumption 1, relating to the independence of ε from index covariates:

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

In what follows, we make use of the following notation:

Notation 1 *For any $d = 1, \ldots, D$, let κ_d denote either the $\leqslant$ or $>$ sign. Denote*

$$F_{\kappa_1,\kappa_2,\ldots,\kappa_D}(t_1, ..., t_D) = P\left(\bigcap_{d=1}^{D} (\varepsilon_d \, \kappa_d \, t_d)\right)$$

for any $\kappa_d \in \{\leqslant, >\}$, $d = 1, .., D$. Functions $F_{\leqslant,\ldots,\leqslant}$ and $F_{>,\ldots,>}$ are the joint c.d.f. and the joint survival function of $\boldsymbol{\varepsilon}$, respectively, and for simplicity we will interchangeably denote them as F and $\overline{F}$, respectively. All the other cases correspond to hybrid forms of c.d.f.s and survival functions. Similar notations apply to subsets of dimensions.

Also denote $F_{d,\kappa_d}(t_d) = P(\varepsilon_d \, \kappa_d \, t_d)$ for $\kappa_d \in \{\leqslant, >\}$. Thus, $F_{d,\leqslant}$ is the marginal c.d.f. and $F_{d,>}$ is the marginal survival function of ε_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 ε_d from x_d for each $d = 1, \ldots, D$. However, as we argue there, the identification of the joint distribution of $\boldsymbol{\varepsilon}$ *does* rely on Assumption 1 even in lattice models. Thus, when one of the objectives is to identify the joint distribution of errors in our general rectangular structure models (motivation for this is discussed later), our Assumption 1 is no more restrictive than what is required in lattice models.

4.1 Advantages of non-lattice models over lattice models

Before turning to identification in the described semiparametric contexts, we formalize the advantages of general rectangular structure models over lattice models for capturing interdependencies among 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 conditional on observables (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. We now describe two advantages of general rectangular models relating to richer partial effects.

Advantage 1. Cross-partial effects: Partial effects in one dimension can depend on covariates exclusive to other processes.

Under Assumption 1, models with general rectangular structures allow $P(Y^{c_d} \geqslant y_j^{(d)}|x)$, which is the likelihood that an individual selects at least a certain level of commitment in one decision area, given all relevant personal and contextual factors, to depend on covariates exclusive to other processes (i.e., $x_h, h \neq d$), enabling analysis of partial effects $\frac{\partial P(Y^{c_d} \geqslant y_j^{(d)}|x)}{\partial x_{h,m}}$. This effect is *indirect* as 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).

For example, consider high-school students choosing academic effort, measured by study hours, and extracurricular participation, measured by involvement in sports or clubs. These choices are interdependent at the level of the decision structure, but the direction of that interdependence is ambiguous a priori as high academic effort may limit time for extracurriculars, while extensive extracurricular involvement may encourage greater academic effort for college applications. Our model allows the data to determine the form and strength of this interdependence. Some covariates may be dimension-specific, such as parental education or tutoring access for academic effort, and school resources or peer involvement for extracurriculars. Others, such as socioeconomic status or school quality, may be shared by both processes. A non-lattice model allows covariates entering only the extracurricular process to affect the probability of high academic effort through the indirect channel, which is informative for resource allocation policies. Furthermore, a lattice model rules out such indirect effects and would therefore miss it.

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

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

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 structure, $\frac{\partial P(Y^{c_1} \leqslant 1|x)}{\partial x_{2,m}} = 0$, 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, assuming F is partially differentiable and letting $e_2 = \alpha^{(2)} - x_2\beta_2$, we have

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

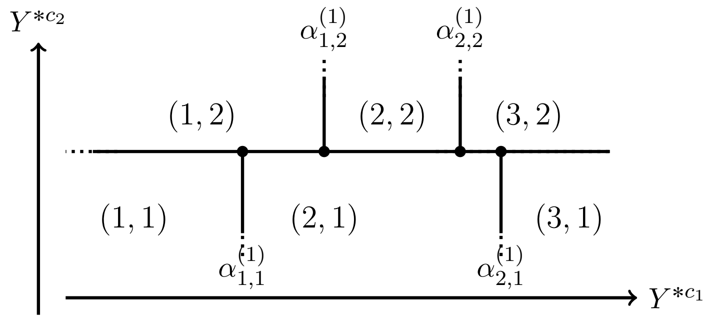


Figure 4: Example of a 3×2 bivariate model

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

Advantage 2. 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 $\frac{\partial P(Y^{c_1} \leqslant 1|x)}{\partial x_{2,m}}$ above, we can see that this partial effect will be positive with a positive probability (and non-negative a.e.) if and only if $\beta_{2,m}(\alpha^{(1)}_{1,2} - \alpha^{(1)}_{1,1}) > 0$. In a completely analogous way, we can consider $\frac{\partial P(Y^{c_1} \leqslant 2|x)}{\partial x_{2,m}}$ and obtain that this partial effect will be positive with a positive probability (and non-negative a.e.) if and only if $\beta_{2,m}(\alpha^{(1)}_{2,2} - \alpha^{(1)}_{2,1}) > 0$. Since $\alpha^{(1)}_{2,2} - \alpha^{(1)}_{2,1} < 0$ and $\alpha^{(1)}_{1,2} - \alpha^{(1)}_{1,1} > 0$, partial effects $\frac{\partial P(Y^{c_1} \leqslant 1|x)}{\partial x_{2,m}}$ and $\frac{\partial P(Y^{c_1} \leqslant 2|x)}{\partial x_{2,m}}$ 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 $\frac{\partial P(Y^{c_d} \leqslant y^{(d)}_j|x)}{\partial x_{d,m}}$ will have the same sign in a non-lattice model for any $y^{(d)}_j$, $j = 1, \ldots, M_d - 1$. 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,

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

where f_d denotes the p.d.f. of ε_d.

As an example, consider a special case of independent ε_1 and ε_2. Then

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

If $\beta_{2,m_2}(\alpha^{(1)}_{j,2} - \alpha^{(1)}_{j,1})$ 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 β_{2,m_2}.

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

4.2 Identification in the semiparametric model

Now we proceed to establishing identification in semiparametric models with non-lattice structures. The knowledge of index parameters and both layers of dependence, which is embedded in 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. For example, the double-layer dependence structure will determine whether bundled interventions reinforce or crowd out each other.

The approach in Komarova and Matcham (2025) for the identification of index parameters and threshold differences in lattice models relies on the ability to isolate different dimensions and consider one dimension at a time. This will not work in non-lattice models as we cannot isolate different dimensions. For example, from Figure 4 one can see that $P\left(Y^{(1)} \leqslant y_j^{(1)}|x\right)$ for $j = 1, 2$, cannot be expressed just in terms of the index $x_1\beta_1$ and the marginal c.d.f. $F_{1,\leqslant}$. General rectangular structures therefore require a different approach to identification. Intuitively, the identification of parameters β_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 general rectangular models become similarly demanding on the data, which is worthy of note.

To prove identification, we introduce some definitions and notations:

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

Notation 2 *For each $d = 1, \ldots, D$, 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 covariates within x_d.

Notation 3 *Due to the presence of shared covariates among processes, the vector $x = (x_1, ..., x_D)$ 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).*

We establish identification through the following four steps:

- 1st step: identification of parameters corresponding to exclusive covariates in each process (Theorem 1).
- 2nd step: identification of parameters corresponding to shared covariates (Theorem 2).
- 3rd step: identification of the joint c.d.f. of ε (Theorem 3).
- 4th step: identification of the thresholds α (Theorem 4).

Theorem 1 gives sufficient conditions for the identification of $\beta_{d;1;L_d}$, $d = 1, \ldots, D$, 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 and the model has a coherent general rectangular structure. Suppose that the following conditions are satisfied:*

For each $d = 1, \ldots, D$:

(a) $L_d \geqslant 1$;

(b) The coefficient $\beta_{d,1}$ corresponding to $x_{d,1}$ in $x_d\beta_d$ is equal to 1;

(c) There exists $j_d = 1, \ldots, M_d - 1$, such that $P(S_d(j_d)) > 0$, where

$$S_d(j_d) = \left\{ x : 0 < P(Y^{c_d} \leqslant y^{(d)}_{j_d}|x) < P(Y^{c_d} \leqslant y^{(d)}_{j_d+1}|x) \right\}.$$

In addition,

(c.i) $S_d(j_d)$ contains a Cartesian product $(\underline{x}_{d,1}, \overline{x}_{d,1}) \times S_{d,-1}(j_d)$, where $\overline{x}_{d,1} > \underline{x}_{d,1}$ and $S_{d,-1}(j_d) \subset \mathbb{R}^{K-1}$, such that $P((\underline{x}_{d,1}, \overline{x}_{d,1}) \times S_{d,-1}(j_d)) > 0$ (the order of covariates in this Cartesian product coincides with the order of covariates in x, i.e., the components of this product are understood to map to their corresponding coordinates in x)

(c.ii) $S_{d,-1}(j_d)$ has full affine dimension in $\mathbb{R}^{K-1}$.

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

The proof of all theorems in the paper are found in Appendix B.

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. Condition (b) further normalizes this coefficient 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 $d = 1, \ldots, D$, 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 take them to be two different consecutive values $y^{(d)}_{j_d}$ and $y^{(d)}_{j_d+1}$ (more generally, they need not be consecutive).

Our next result is on the identification of index parameters 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 $P(\cap_{d=1}^{D}(Y^{c_d}\, \kappa_d\, y^{(d)}_{j_d})|x)$, where $\kappa_d \in \{\leqslant, >\}$. 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, 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 $(j_1, ..., j_D)$ *such that the intersection term* $S = \cap_{d=1}^{D} S_d(j_d)$ *has positive probability measure and full affine dimension* K*;*

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

Then β_d*,* $d = 1, \ldots, D$*, are identified.*

Condition (b) in 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 their 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 β_d, including $\beta_{d,1:L_d}$ 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 third result is on the identification of the joint distribution of $\boldsymbol{\varepsilon}$. It is enough to identify one function $F_{\kappa_1 \ldots \kappa_D}$ 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.

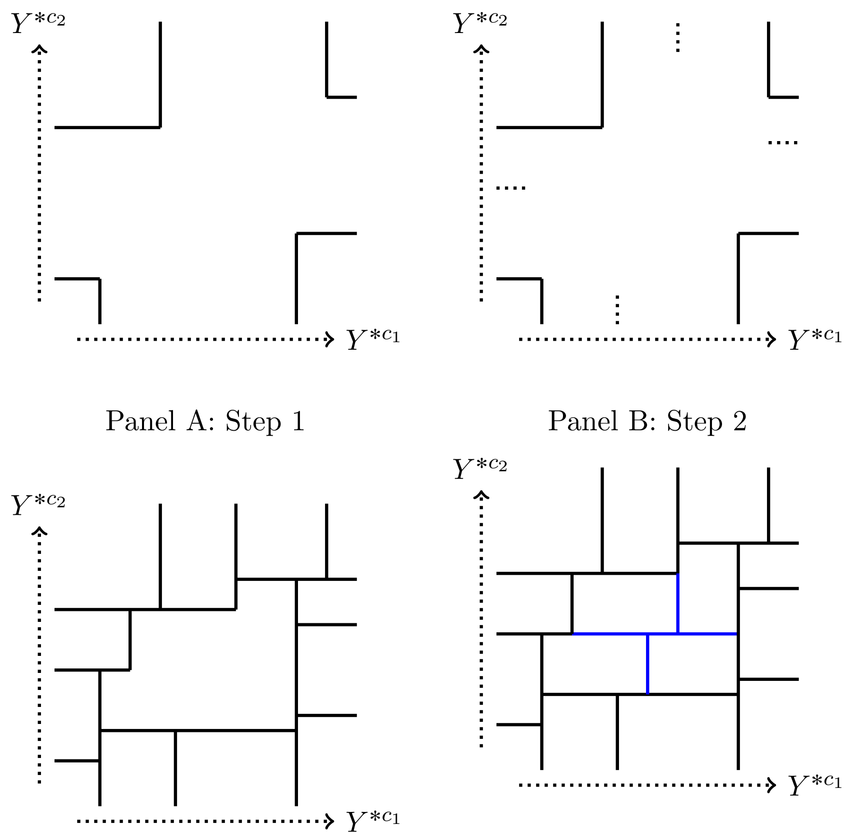
Theorem 3 *Suppose all the conditions of Theorem 2 hold for a collection of indices* $(j_1, ..., j_D)$ *such that* $j_d = 1$ *or* $j_d + 1 = M_d$ *for each* d*.*

Additionally, in condition (b) of Theorem 2, suppose that for each $d = 1, ..., D$ ***both*** *of the following conditions hold:* $\underline{x}_{d,1}$ *is small enough to guarantee that* $\alpha^{(d)}_{j_1,...,j_D} - \underline{x}_{d,1} - x_{d,-1}\beta_{d,-1}$ *is at the upper support point of the* ε_d *distribution,* ***and*** $\overline{x}_{d,1}$ *is large enough to guarantee that* $\alpha^{(d)}_{j_1,...,j_D} - \overline{x}_{d,1} - x_{d,-1}\beta_{d,-1}$ *is at the lower support point of the* ε_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* $F_{d,\leqslant}(e_{0d}) = c_{0d}$ *for each marginal c.d.f.* $F_{d,\leqslant}$ *for some known* e_{0d} *and* $c_{0d} \in (0, 1)$*,* $d = 1, \ldots, D$*, the distribution of* ε *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, $d = 1, \ldots, D$,) are identified. Then the observed probabilities of that "corner" region together with the knowledge of thresholds defining it identify the joint distribution of $\boldsymbol{\varepsilon}$.

Conditions strengthen across Theorems 1–3, reflecting the increasing demand on data as we move from exclusive-covariate parameters to shared parameters and then to the joint distribution of unobservables and finally to the thresholds.

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 $F_{\kappa_1,\ldots,\kappa_D}$. Theorem 4 gives a formal identification result for the thresholds. It strengthens previous conditions by essentially requiring that for any rectangle $R_{j_1,...,j_D}$ 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).



Panel A: Step 1 Panel B: Step 2

Panel C: Step 3 Panel D: Step 4

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

Theorem 4 *Suppose all the conditions of Theorem 3 hold for any collection of indices* $(j_1, \ldots, j_D)$ *with* $j_d \in \{1, M_d - 1\}$*,* $d = 1, \ldots, D$*. Then all the thresholds* $\alpha^{(d)}_{q_1,\ldots,q_{d-1},q_d,q_{d+1},\ldots,q_D}$ *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 $R_{q_1,\ldots,q_D}$ 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 $d = 1, .., D$, the thresholds $a^{(d)}_{q_1,\ldots,q_{d-1},q_d,q_{d+1},\ldots q_D}$ 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 $M_d - 2$. 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 $d = 1, .., D$ and identifies thresholds $\alpha^{(d)}_{q_1,\ldots,q_{d-1},q_d,q_{d+1},\ldots q_D}$ for when q_d remains fixed at its value 1 or $M_d - 1$ whereas q_h, $h \neq d$, 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" (in blue), 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 $x_d\beta_d$, $d = 1, \ldots, D$, and the knowledge of $F_{\kappa_1,\ldots,\kappa_D}$ for a suitable $\kappa_1, \ldots, \kappa_D$ (Theorem 3 implies the knowledge of $F_{\kappa_1,\ldots,\kappa_D}$ for any $\kappa_1, \ldots, \kappa_D$). Which $F_{\kappa_1,\ldots,\kappa_D}$ 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 $\boldsymbol{\varepsilon}$ 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 $\boldsymbol{\varepsilon}$ 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.

4.3 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 scale function like in Chen and Khan, 2003) do not extend to multivariate models with general rectangular structures.[8] 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 (2024). 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 $P(Y^{(1)} = y^{(1)}_{j_1}, Y^{(2)} = y^{(2)}_{j_2} \mid x_1, x_2) = P(Y^{(1)} = y^{(1)}_1, Y^{(2)} = y^{(2)}_1 \mid \tilde{x}_1, \tilde{x}_2)$ 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 shared covariates and all other dimensions are similar, while MRC would exploit the stochastic dominance, which is behind the result in Theorem 1.

We found that the only existing approach that can be extended to non-lattice models and which permits the estimation of all the index parameters as well as all the thresholds and the unknown 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, we describe how to 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 prob-

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

ability of the bivariate latent process falling into the rectangle R_{j_1,j_2}:

$$\ell^{(i)}_{j_1,j_2} = \sum_{\ell_1=0}^{1}\sum_{\ell_2=0}^{1}(-1)^{\ell_1+\ell_2}\widehat{F}\left(a^{(1)}_{j_1-\ell_1,j_2} - x^{(i)}_1 b_1, a^{(2)}_{j_1,j_2-\ell_2} - x^{(i)}_2 b_2\right)$$

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

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

where θ 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 $\mathcal{R}_{1;s_1,S_1}(e_1; q_1) \cdot \mathcal{R}_{2;s_2,S_2}(e_2; q_2)$, $s_1 = 1, \ldots, S_1$, $s_2 = 1, \ldots, S_2$, 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

$$\begin{aligned} h_{s_1 s_2} &\leqslant h_{s_1+1,s_2}, \quad s_1 = 1, \ldots, S_1 - 1,\ s_2 = 1, \ldots, S_2, \\ h_{s_1 s_2} &\leqslant h_{s_1,s_2+1}, \quad s_2 = 1, \ldots, S_2 - 1,\ s_1 = 1, \ldots, S_1; \end{aligned}$$

(b) c.d.f. *bounds* are maintained by $0 \leqslant h_{s_1 s_2} \leqslant 1$ for all s_1, s_2 and (c) the supermodularity (2-increasing) property is enforced through

$$h_{s_1+1,s_2+1} - h_{s_1+1,s_2} - h_{s_1,s_2+1} + h_{s_1,s_2} \geqslant 0$$

for $s_1 = 1, \ldots, S_1 - 1$ and $s_2 = 1, \ldots, S_2 - 1$.[9] 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, for example,

$$-\lambda_N \left(\alpha^{(1)}_{j_1+1,j_2} - \alpha^{(1)}_{j_1,j_2}\right)^2 \cdot \left(\alpha^{(2)}_{j_1,j_2+1} - \alpha^{(2)}_{j_1,j_2}\right)^2, \tag{7}$$

for a large $\lambda_N > 0$.

The distribution theory in Coppejans (2007) generalizes as well with modifications required to make it applicable in multivariate non-lattice setting: regularity conditions and conditions on the growth of the B-spline bases would, however, need to be adjusted.

[9] For more details on shape constraints in tensor-product B-splines, see Bhattacharya and Komarova (2026).

5 Parametric specification

Our semiparametric framework offers an approach to achieving identification results despite model complexity and offers an estimation method that generalizes Coppejans (2007). In practice, researchers may opt for a rich parametric family for the joint distribution of unobserved $\boldsymbol{\varepsilon}$, in part for computational convenience. This choice eliminates the need for nonparametric modeling of the joint distribution (which, as described above, can be done with B-splines) and typically reduces the data requirements for identifying all unknown parameters. These parametric families must be sufficiently flexible to accommodate various negative and positive dependencies 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 the logistic distribution.

5.1 Identification

Although parametric versions of the model will be identified under the conditions outlined in Section 4.2, intuitively, 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.

Adopting parametric assumptions in our general rectangular structure setting poses 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 instead illustrate parametric identification without exclusive covariates through a numerical identification exercise.

Consider the bivariate case $D = 2$. Let $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2)'$ be jointly normal with mean $(0,0)'$, unit variances, and correlation ρ.[10] 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 $\beta_1, \beta_2, \alpha^{(1)}_{1,1}, \alpha^{(2)}_{1,1}, \rho$ can be studied using $T \geqslant k_1 + k_2 + 3$ distinct covariate values $x^{(i)} = (x^{(i)}_1, x^{(i)}_2)'$ by writing a system of T equations in the $k_1 + k_2 + 3$ unknowns: for $i = 1, \ldots, T$

$$\underbrace{P(Y^{c_1} = y^{(1)}_1, Y^{c_2} = y^{(2)}_1 | x^{(i)})}_{p_{\text{obs}}(x^{(i)})} = \Phi_2(\alpha^{(1)}_{1,1} - x^{(i)}_1 \beta_1, \alpha^{(2)}_{1,1} - x^{(i)}_2 \beta_2; \rho), \tag{8}$$

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 point, consider the case of each latent process having a single covariate shared between the two processes. Then the vector $\theta = (\beta_1, \beta_2, \alpha^{(1)}_{1,1}, \alpha^{(2)}_{1,1}, \rho)$ 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

[10] As usual, we normalize means and variances because shift and scale changes yield observationally equivalent parameter vectors.

the objective function for a candidate parameter θ as

$$Q_T(\theta) = \sum_{i=1}^{T} \Big(p_{\text{obs}}(x^{(i)}) - \Phi_2\big(\alpha_{1,1}^{(1)} - x_1^{(i)}\beta_1,\ \alpha_{1,1}^{(2)} - x_2^{(i)}\beta_2;\ \rho\big)\Big)^2 .$$

We compute $p_{\text{obs}}(x_i)$ from the known θ_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 θ_0, we reparametrize ρ as $z = \operatorname{atanh}(\rho)$ (so $z \in \mathbb{R}$ and $\rho = \tanh(z) \in [-1, 1]$), and then consider hyperspheres in this transformed parameter space. For a given radius r we sample 5,000 directions on the sphere of Euclidean radius r about θ_0; for each sampled point θ, 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 of the minimum objective Q_T (log scale) against radius r for both sampling schemes (left: fixed-direction; right: random-sphere). These plots show 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 ($Q_T(\theta)$ increases in T for all radii r). The pronounced jitter visible in the random-sphere plot reflects sampling variability across radii.

5.2 Estimation

To outline an estimation approach in the parametric case, we continue with $D = 2$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2)'$ being jointly normal with zero mean, unit variances, and correlation ρ. Given a random sample $\left\{(y^{(1)(i)}, y^{(2)(i)}, x_1^{(i)}, x_2^{(i)})\right\}_{i=1}^{N}$ and collecting β_1, β_2, ρ and all the thresholds in α in the parameter vector θ, consider the log-likelihood function

$$\begin{aligned}
\mathcal{L}(\theta) &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j_1=1}^{M_1}\sum_{j_2=1}^{M_2} 1\left[(y^{(1)(i)}, y^{(2)(i)}) = (y_{j_1}^{(1)}, y_{j_2}^{(2)})\right]\log(\ell_{j_1,j_2}^{(i)}(\theta)) \\
&= \frac{1}{N}\sum_{i=1}^{N}\log(\ell^{(i)}(\theta)),
\end{aligned}$$

$$\text{with} \quad \ell_{j_1,j_2}^{(i)} = \sum_{t_1=0}^{1}\sum_{t_2=0}^{1}(-1)^{t_1+t_2}\Phi_2\left(\alpha_{j_1-t_1,j_2}^{(1)} - x_1^{(i)}\beta_1, \alpha_{j_1,j_2-t_2}^{(2)} - x_2^{(i)}\beta_2; \rho\right).$$

Analogously to the semiparametric model case, this log-likelihood needs to be maximized subject to linear inequality constraints that describe ordering of thresholds in each dimension, normalization constraints on the thresholds, and non-linear equality constraints $q(\theta) = 0$ 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 coherency constraints on thresholds. Assuming the true parameters lie at a regular point of the constraint set (such that the Jacobian of the active coherency constraints has full row rank), the coherency constraints are differentiable and we can invoke typical MLE regularity conditions (e.g., Newey and McFadden (1994)) to obtain the asymptotic result $\sqrt{N}(\hat{\theta}-\theta_0) \xrightarrow{d} \mathcal{N}(0, V)$, where $V = J^{-1} - J^{-1}Q'(QJ^{-1}Q')^{-1}QJ^{-1}$ for terms $Q = \dfrac{\partial q(\theta_0)}{\partial \theta'}$ and $J = \mathbb{E}\left[\frac{\partial \log(\ell^{(i)}(\theta_0))}{\partial \theta}\frac{\partial \log(\ell^{(i)}(\theta_0))}{\partial \theta'}\right]$.

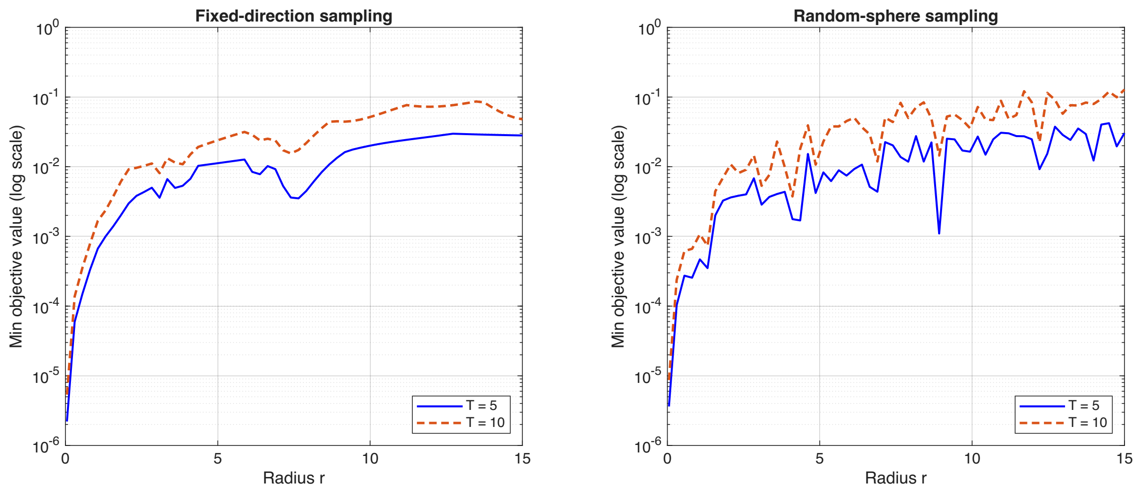


Figure 6: Numerical illustration of identification for the bivariate-normal model. The vertical axis (log scale) reports the minimum squared objective Q_T found on a sphere of radius r around the true parameter vector; the horizontal axis is r. Left: fixed-direction sampling. Right: random-sphere sampling.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

The natural plug-in sample-analogue estimator of V provides a consistent estimator for the variance-covariance matrix.

6 Monte Carlo experiments

We examine Monte Carlo simulations for the parametric case with bivariate normal errors, as outlined in Section 5. We compare the constrained MLE for general rectangular structure models (*non-lattice probit*) with the standard bivariate ordered probit, which estimates a misspecified lattice model. The baseline model is

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

where unobservables are independent of regressors and jointly normal with zero means and unit variances. We explore scenarios with no exclusive covariates ($\gamma_1 = \gamma_2 = 0$) and separately with an exclusive covariate in one latent process. Each simulation design uses 250 independent random samples of size $N = 5,000$, with the penalty term in (7) set to N. [11]

Two findings stand out. First, non-lattice model parameters are estimated with negligible bias, even in the absence of exclusive covariates. Second, imposing a lattice structure on a truly non-lattice data-generating process yields severely inconsistent estimators for *all* parameters, not merely the thresholds. The degree of inconsistency depends on how well lattice models approximate the true non-lattice. The most dramatic distortion concerns the correlation parameter ρ. For example, in Design 1, where the truth is $\rho = 0.33$, the lattice estimator consistently returns values near -0.93 to -0.97 across all covariate distributions (standard deviations of 0.01–0.02). This is a near-deterministic sign reversal of magnitude greater than one on a parameter bounded in $[-1, 1]$. The mechanism for this is intuitive. Namely, the true non-lattice structure exhibits substitutability

[11] Alternative N and λ_N yield similar results, but we favor a large sample size so that any issues with the lattice or non-lattice performance cannot be attributed to a small sample size.

among outcomes via the threshold layer. The lattice estimator, lacking any threshold-layer channel, must attribute this to ρ, forcing it toward large negative values. The result is that the sign of $\hat{\rho}$ under a lattice model cannot be taken as evidence of the sign of the true unobservables correlation. This has direct implications for any empirical study that uses bivariate ordered probit, bivariate probit, or correlated ordered logit to draw conclusions about adverse selection, complementarity, or the direction of latent covariation.

Design 1: 2×2 structure, no excluded covariates

First, we set $\gamma_1 = \gamma_2 = 0$, thereby removing w_1 and w_2. We also set $\beta_1 = 1$, $\beta_2 = 0.5$, $\rho = 0.33$, and use a 2×2 non-lattice structure with thresholds $\alpha^{(2)}_{1,1} = \alpha^{(2)}_{2,1} = 1$, $\alpha^{(1)}_{1,1} = -2$, and $\alpha^{(1)}_{1,2} = 1.5$. We consider three distributions for the common regressor x: uniform $[-5, 5]$ (Design 1A), discrete on 10 points $\{\pm 5, \pm 3.5, \pm 2.5, \pm 1.5, 0, 0.5\}$ with equal probabilities (Design 1B), and discrete on only three points $\{\pm 2.5, 0\}$ with equal probabilities (Design 1C).

Table 1 reports the across-simulation means and standard deviations for the estimated parameters. Across all sub-designs A-C, the non-lattice method estimates all parameters with minimal bias, even with the minimal variation in x as in Design 1C. The bivariate lattice ordered probit method performs especially poorly on β_2, ρ, and the thresholds. Because of the very limited variation in the common regressor, Design 1C exhibits higher standard deviations than designs 1A and 1B.

Design 2: 4×3 with one excluded covariate

We now increase the number of discrete values, M_d, in both dimensions. The discrete dependent variables Y^{c_1} and Y^{c_2} can take four and three values, respectively. This yields a 4×3 non-lattice structure, illustrated in Figure 7. The common covariate x follows a uniform $[-3, 3]$ distribution (though continuity is not necessary here). The covariate w_1 takes values $\{-2.5, -1.5, \pm 0.5\}$ with equal probability. We set $\gamma_2 = 0$, thus 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 2 reports the across-simulation means and standard deviations of the estimated index parameters and the correlation coefficient. Table 4 in the online supplement provides the estimated threshold values. The non-lattice bivariate ordered probit method estimates all the parameters with almost no bias. By contrast, the lattice bivariate ordered probit method estimates all 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 the parameters. Assuming a lattice structure makes estimating the correlation parameter ρ markedly difficult, with the method failing to estimate the correct sign for ρ, let alone a value close to the true one. The intuition is similar to that in Design 1.

7 Applications

7.1 Identifying moral hazard and adverse selection in insurance markets

Motivation. The dominant empirical framework for estimating asymmetric information in insurance markets, due to Chiappori and Salanie (2000), tests for adverse selection by estimating a bivariate probit and interpreting the latent correlation as evidence of asymmetric information. As Cohen and Siegelman (2010) emphasize, this correlation conflates adverse selection with moral hazard, since both forces contribute to the positive covariation between coverage and utilization. Separating them has required either quasi-experimental variation or strong structural assumptions (Handel,

Table 1: Simulation results Design 1

Parameter	Design 1A		Design 1B		Design 1C	
	Non-latt	**Latt**	**Non-latt**	**Latt**	**Non-latt**	**Latt**
$\beta_1 = 1$	1.00 (0.03)	0.77 (0.02)	1.00 (0.03)	0.75 (0.02)	1.00 (0.03)	0.84 (0.02)
$\beta_2 = 0.5$	0.50 (0.02)	0.01 (0.01)	0.50 (0.02)	0.04 (0.01)	0.50 (0.05)	-0.18 (0.01)
$\rho = 0.33$	0.33 (0.12)	-0.93 (0.02)	0.34 (0.12)	-0.94 (0.01)	0.33 (0.14)	-0.97 (0.01)
$\alpha_{1,1}^{(2)} = \alpha_{2,1}^{(2)}$=1	1.00 (0.04) 1.00 (0.04)	0.72 (0.04)	1.00 (0.04) 1.00 (0.04)	0.68 (0.03)	1.00 (0.05) 1.00 (0.05)	0.64 (0.03)
$\alpha_{1,1}^{(1)}$ = -2 $\alpha_{1,2}^{(1)} = 1.5$	-1.99 (0.07) 1.50 (0.08)	-0.42 (0.02)	-1.99 (0.07) 1.50 (0.08)	-0.40 (0.02)	-1.99 (0.12) 1.49 (0.15)	-0.53 (0.02)

Notes: This table reports the sample mean and sample standard deviations (in parentheses) of the estimates of the Design 1 parameters, over 250 samples. The "Non-latt" columns provide estimates from using the newly proposed non-lattice bivariate ordered probit model. The "Latt" columns assume a lattice structure.

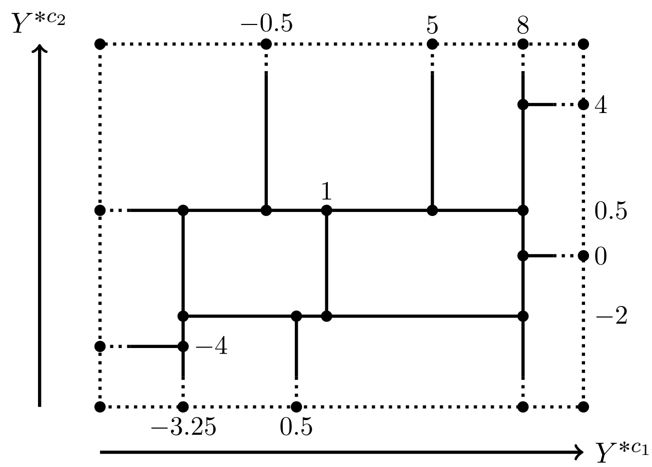


Figure 7: Latent variable space for two equations: Design 2 simulation

2013; Einav et al., 2013; Hackmann, Kolstad and Kowalski, 2015).[12] Our general rectangular structure resolves this limitation without quasi-experimental variation or structural assumptions, as moral hazard enters through coverage-dependent utilization thresholds, leaving the unobservables correlation to capture selection alone. Our Monte Carlo results further motivate this approach since we saw the lattice estimator's $\hat{\rho}$ being severely distorted when threshold-layer dependence is present, so studies interpreting the sign of a bivariate probit's latent correlation as evidence for or against adverse selection may be reading an artifact of model misspecification rather than a structural parameter. Moreover, our general rectangular structure model permits two-way dependence of thresholds, so not only can utilization threshold depend on insurance coverage, but insurance-choice thresholds can also 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.

Data and Model Specification. We apply the 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. The quality of the MEPS data has been

[12]Despite these advances, the method remains popular in practice, with recent applications including Matcham (2026) and Benetton and Buchak (2026).

Table 2: Simulation results for Design 2

Parameter	Non-latt	Latt
$\beta_1 = 1.5$	1.50 (0.04)	0.61 (0.01)
$\gamma_1 = -4$	-4.01 (0.09)	-2.51 (0.04)
$\beta_2 = 3$	3.00 (0.09)	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.

verified in Zuvekas and Olin (2009). Our sample includes approximately 60,000 individuals from 2005 to 2010, before the Affordable Care Act. Following the standard framework, we specify latent insurance and utilization equations as

$$Y^*_{\text{ins}} = x_{\text{ins}}\beta_{\text{ins}} + \varepsilon_{\text{ins}}, \quad Y^*_{\text{use}} = x_{\text{use}}\beta_{\text{use}} + \varepsilon_{\text{use}}. \tag{9}$$

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, and arthritis). An exclusive covariate for insurance is partner's job-provided coverage. In this and the two other empirical applications to follow, we assume a bivariate normal distribution for the vector of unobservables, with zero mean, unit variances, and an unknown correlation ρ.

The insurance outcome Y_{ins} equals 1 if the individual holds private health insurance in January and is 0 otherwise. Utilization Y_{use} is measured categorically (0 for no charges, 1 for below-median charges, 2 for above-median charges).[13]

Moral hazard is isolated by allowing utilization thresholds to depend on coverage status: $Y_{\text{ins}} = 1$ lowers the thresholds for utilization if individuals respond to coverage by consuming more care. Insurance-dimension thresholds may also vary with utilization status, accommodating "selection on moral hazard" (Einav et al., 2013), which is the possibility that individuals select coverage partly based on their anticipated behavioral response. We estimate the model subject only to coherency constraints on the thresholds, thus remaining agnostic on both questions and letting the data determine whether either threshold interdependence is present.[14] Once the behavioral effect is accounted for, the remaining correlation between ε_{ins} and ε_{use} can be interpreted as evidence of adverse or advantageous selection.

Estimated Threshold Structure. Figure 8 presents the threshold results for the non-lattice model. First, the model reveals moral hazard as utilization thresholds shift downward when $Y_{\text{ins}} = 1$, with the low-to-medium usage threshold falling from -0.02 (0.04) to -0.45 (0.05) and the medium-to-high threshold from 0.68 (0.04) to 0.46 (0.05), where here and in what follows, standard errors are in parentheses.

[13]There could be a concern that insured and uninsured individuals pay different prices, implying that charges may not reflect actual usage on an equivalent scale for the insured and uninsured. However, our utilization is constructed from MEPS total expenditures, defined as the sum of all payments received by the provider across patient out-of-pocket, private insurance, and public payers, rather than from billed charges. This is the variable the MEPS data programme designed precisely to be payer-neutral at the transaction level. Furthermore, it is also possible instead to use the number of billable events; results are similar and available on request.

[14]In this, and all empirical examples, we use a penalty of N^2, since the eventual sample size in the empirical applications are generally smaller than in the Monte Carlo simulations.

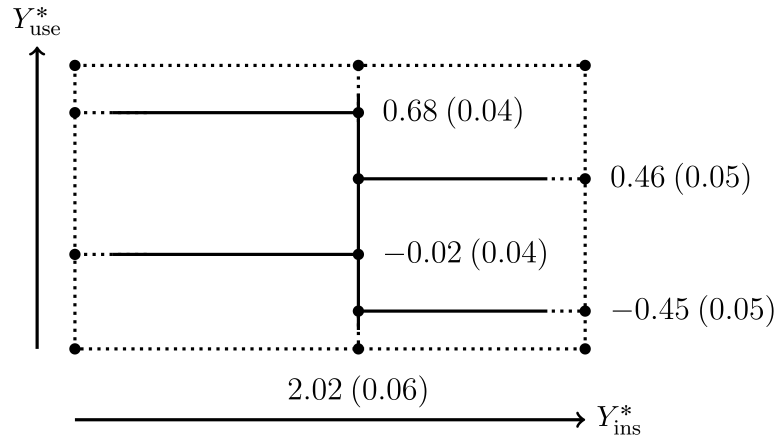


Figure 8: Estimated thresholds for insurance coverage (Y^*_{ins}) and healthcare utilization (Y^*_{use}) (standard errors in parentheses)

Second, there is no evidence of "selection on moral hazard" since insurance-coverage thresholds are estimated as invariant to utilization status. Crucially, *this is a finding of the estimation, not a maintained assumption.* A researcher who imposed sequential timing by assumption would obtain the same insurance threshold structure but have no way of knowing whether the data supported it.

Third, once moral hazard is absorbed into the threshold layer, the residual unobservables correlation $\hat{\rho}$ falls from 0.21 (0.01) in the lattice model to 0.04 (0.02) in the non-lattice model, suggesting minimal adverse selection, which is consistent with the broader literature that finds limited selection once moral hazard is accounted for (Einav et al., 2013; Handel, 2013).

Our approach 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.

Partial Effects. As we explained in Section 4.1, the non-lattice model allows for indirect effects of exclusive covariates on outcome variables in other dimensions. In the insurance case, the insurance status of one's partner may indirectly affect one's own utilization. We estimate these partial effects and examine their magnitude.

First, we calculate average partial effects, taking each individual in the dataset at their observed covariates and evaluating the impact on the probability of exceeding a certain value of utilization when the individual has a partner with insurance, and when they do not. Specifically, we calculate

$$APE_k = \frac{1}{N}\sum_{i=1}^{N}\left[P(Y_{\text{use}} \geqslant k | x_{i,\text{partner}=1}) - P(Y_{\text{use}} \geqslant k | x_{i,\text{partner}=0})\right]$$

for $k = 1$ (utilization but below the median) and $k = 2$ (utilization above the median). We estimate APE_1 as 0.055 (0.0034) and APE_2 as 0.034 (0.0043), implying that having a partner with insurance raises the probability of high utilization by three percentage points on average, and the probability of non-zero utilization by 5.5 percentage points on average.

Second, we calculate partial effects at a specific vector of covariates (typically the average) that is,

$$PEA_k = P(Y_{\text{use}} \geqslant k | x_{i,\text{partner}=1}, \bar{x}_{i,-\text{partner}}) - P(Y_{\text{use}} \geqslant k | x_{i,\text{partner}=0}, \bar{x}_{i,-\text{partner}})$$

For $\bar{x}_{i,-\text{partner}}$, we take numerical covariates (logged income, dividend payments, family size, hours worked, years of education, and age) at their means, use the modal values of the binary variables:

marital status (married), race (white), year (2009), metropolitan statistical area (yes), and region. Since the two other binary variables, male and pre-existing conditions, have means close to 0.5, we calculate PEA for the four possibilities. We estimate PEA_2 between 0.03 and 0.04 (0.01) across all four cases, whereas PEA_1 varies from 0.08 (0.01) for males without pre-existing conditions to 0.03 (0.005) for females with pre-existing conditions. Hence, the impact of partner's insurance on non-zero utilization depends on other covariates in the model.

7.2 Cryptocurrency Familiarity and Optimism

Motivation, Data and Model Specification. In the second application, we use data from the Survey of Consumer Payment Choice (SCPC) (Foster, Greene and Stavins, 2021) to study opinions on future movements in cryptocurrency prices.[15] Conducted annually by the Federal Reserve Banks of Atlanta, Boston, and San Francisco, the SCPC tracks U.S. consumers' adoption of payment methods, and has recently noted a shift toward online and mobile payments due to the COVID-19 pandemic. Our sample includes around 4,600 individuals, 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, Greene and Stavins (2021) for details.

We examine whether opinions about 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).[16] For Y^{c_2}, we use an ordered variable for bitcoin's expected value in one year (-1: decrease, 0: no change, 1: increase).

Estimated Threshold Structure and Coefficients. Figure 9 shows the estimated threshold structure for a non-lattice model. It reveals varied thresholds. For example, individuals with low familiarity expect no change (the blue dotted region for no change) at a lower threshold in the opinion dimension (i.e., lower $Y^{*Opinion}$ values yield a "no change" opinion compared to other familiarity groups). Conversely, those with high familiarity have closely spaced thresholds (the orange crosshatch region for no change), thereby defining a narrow "no change" region. For this group, most $Y^{*Opinion}$ values reflect strong opinions, either decreasing or increasing. These findings describe decision-making structures and not probabilistic outcomes.

Table 3 provides estimates of β and ρ. The correlation ρ ranges from 0.09 (lattice model, *s.e.* = 0.03) to 0.75 (non-lattice model, *s.e.* = 0.14), where the latter is large and reflects most plausibly a general disposition toward novel financial products that simultaneously raises both latent familiarity and latent optimism. Some index coefficient estimates differ by over 10% in magnitude. Notably, the coefficient on male changes sign, being negative in the lattice model and positive in the non-lattice model. The lattice model thus suggests males are more pessimistic about bitcoin's value as the effect in this model would lead to $P(Y^{Optimism} \geqslant j|x_{-male}, x_{male} = 1) - P(Y^{Optimism} \geqslant j|x_{-male}, x_{male} = 0)$ being 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, however, confirm that given the estimated thresholds the non-lattice model yields $P(Y^{Optimism} \geqslant j|x_{-male}, x_{male} = 1) - P(Y^{Optimism} \geqslant j|x_{-male}, x_{male} = 0)$ as positive for any level

[15] See also Benetton and Compiani (2024) and Kahn and Linares Zegarra (2016).

[16] Moderate and extremely familiar are combined due to few respondents reporting extreme familiarity.

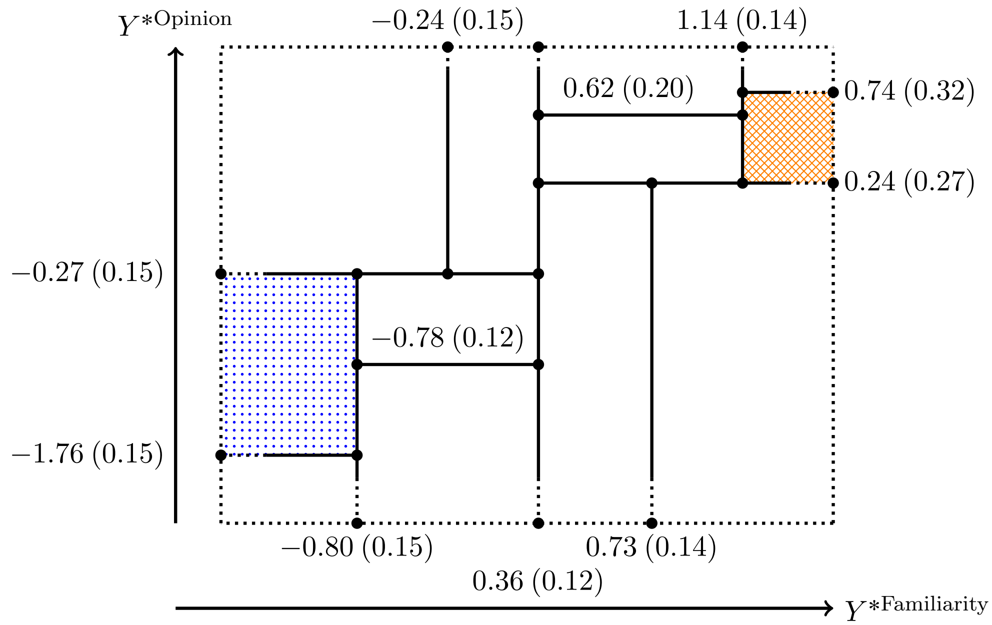


Figure 9: Estimates from cryptocurrency example when assuming a general rectangular structure model (standard errors in parentheses)

j and any x_{-male} in the data, thus, giving a stable sign of this partial effect across the domain.[17].

Another aspect illustrated by the thresholds in Figure 9 regarding the decision-making process is that individual decisions can be modeled 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.[18] Hierarchical non-lattice models maintain coherence, as each node in the decision tree further refines the partitioning of the latent space. Figure 13 in the online supplement depicts a binary decision tree outlining the estimated hierarchical decision-making process for this cryptocurrency application.

Sequential Timing Model. As an interim model between the lattice and the general rectangular structure, we estimate a sequential timing model in which familiarity is determined first, with thresholds functionally independent of optimism, while thresholds for optimism are then free to depend on familiarity. This is strictly narrower than the hierarchical non-lattice structure described above, as it imposes a lattice restriction in the familiarity dimension while leaving the optimism dimension unrestricted. The sequential model is coherent by construction.

The estimated index coefficients (Table 3) follow a monotone pattern with $\hat{\rho}_{\text{latt}} = 0.09 < \hat{\rho}_{\text{seq}} = 0.26 < \hat{\rho}_{\text{non-latt}} = 0.75$. This ordering is expected and the explanation is as follows. The lattice model misses this shifting threshold and infers from fewer people crossing into the highly optimistic category than a pure positive correlation would predict that $\hat{\rho}$ must be low. But the non-lattice model instead allows a high threshold into the highly optimistic category and, once that is accounted for, correctly interprets the residual covariation, given thresholds as a larger correlation in unobservables. The sequential model goes some way to fixing this but the fixed thresholds in

[17] A potentially different estimated threshold structure could have resulted in the switching of signs for this partial effect.

[18] 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.

Table 3: Estimation coefficients: bitcoin familiarity and optimism

Variable	Non-lattice	Lattice	Sequential
Familiarity with Bitcoin			
Low Income	0.02 (0.05)	0.01 (0.06)	0.01 (0.05)
Age	-0.02 (0.00)	-0.02 (0.00)	-0.02 (0.00)
Male	0.46 (0.06)	0.51 (0.06)	0.51 (0.05)
Low Education	-0.44 (0.09)	-0.48 (0.09)	-0.48 (0.08)
Bitcoin "optimism"			
Low Income	0.05 (0.06)	0.04 (0.06)	0.03 (0.05)
Age	-0.01 (0.00)	-0.01 (0.00)	-0.01 (0.00)
Male	0.12 (0.07)	-0.08 (0.05)	-0.02 (0.09)
Low Education	-0.10 (0.08)	0.08 (0.07)	0.02 (0.08)
ρ	0.75 (0.14)	0.09 (0.03)	0.26 (0.21)

Notes: The "Non-lattice" column provides estimates from using non-lattice bivariate ordered probit model. The "Lattice" column assumes a lattice structure. The "Sequential" column assumes that thresholds in the familiarity dimension are lattice in structure, but thresholds in the optimism dimension are free (see Appendix Figure 12).

familiarity dimension block the full increase in ρ as in the fully flexible model.

The general rectangular structure appears to reject sequential timing in the familiarity direction. Figure 9 shows that familiarity thresholds vary systematically across opinion outcomes. In particular, individuals who already hold strong opinions (either bullish or bearish about bitcoin) require a meaningfully different level of latent familiarity to self-report as, say, "moderately familiar" than do those who are agnostic about price movements. This two-way feedback is inconsistent with any sequential timing model. Economically, it suggests that stated familiarity and price beliefs are formed jointly, consistent with belief-consistent information acquisition, which means that people who are already bullish may selectively engage with the asset and thereby report higher familiarity, while those who are agnostic remain less engaged. The lattice and sequential timing models suppress this channel entirely.

The male coefficient reinforces this conclusion. In the lattice, the coefficient on male in the optimism equation is -0.08, in the sequential timing model it attenuates to -0.02, and then only in the general rectangular structure does the coefficient become positive ($+0.12$). The lattice misspecification of the familiarity thresholds contaminates the optimism coefficients, and the sequential timing restriction is insufficient to correct that.[19]

7.3 Parental investments in children

Motivation. Our third and final application considers parental investment in their children, which is a central topic in the economics of human capital. The recent empirical literature is framed around how families allocate their time and money towards their children's educational development. Recent notable work (e.g., Cunha and Heckman (2007) and Cunha, Heckman and Schennach, 2010) focuses on the dynamic technology that converts parental investments into the cognitive and non-cognitive

[19]One could implement a likelihood ratio test of the sequential timing restriction against the general rectangular structure to formally quantify this rejection.

skills of the children, stressing the importance of early investments.[20]

Our application uses a general rectangular structure framework to study the *interdependence* of families' decisions on (i) financial investments into their child's education, and (ii) the non-financial cognitive environment they create for their child, the latter of which relates to their time investment into their child's education. Structural studies on this topic, most notably Del Boca, Flinn and Wiswall (2014), estimate production technologies with parental time and money as inputs. Both parental utility and production technologies typically take the Cobb Douglas form, which imposes complementarity of time and monetary investments on the margin. The argument is that a dollar of education spending yields more when the child is already in a cognitively stimulating home environment, and hours of stimulating parental time yield more when schooling investments reinforce them. However, parents may view time and monetary investments as substitutable, or they may narrow bracket their time and money investments into their child's education, as a lattice model would imply.

Furthermore, in line with Beckerian tradition, there is a budgetary constraint between time investments and financial investments, arising from the labor supply decision of parents. Households in which parents work longer hours have more monetary resources available for education spending but less time and attention available for cognitive stimulation at home. A lattice model bundles all these forces into the correlation between unobservables. But a general rectangular structure model allows the data to inform on the complementarity/substitutability of parents' financial and time investments into their children's education through correlation in latent outcome unobservables and the threshold structure.

Data and Model Specification. To estimate a general rectangular structure model in this context, we use the Panel Study of Income Dynamics (PSID) Child Development Supplement (CDS).[21] The CDS sampled approximately 3,500 children of PSID household members, aged 0–12 in 1997. We pool the second and third waves of the survey, which took place in 2002 and 2007, as they contain all requisite variables. The model we consider is

$$Y^*_{\text{spend}} = x_{\text{spend}}\beta_{\text{spend}} + \varepsilon_{\text{spend}} \qquad Y^*_{\text{stim}} = x_{\text{stim}}\beta_{\text{stim}} + \varepsilon_{\text{stim}}$$

with Y_{spend} reflecting ordinal categories of financial investments into the child's education, and Y_{stim} as an ordinal variable reflecting the quality of the child's cognitive environment. More specifically, for illustration, Y_{spend} denotes annual education spending on the child, coded into four ordered categories split by quartile of educational spending. The quality of the child's cognitive home environment, Y_{stim} is measured by the HOME Cognitive Stimulation scale, coded into terciles: low, medium, and high. The HOME scale is constructed from interviewer-reported and parent-reported items on play with educational toys, stimulating activities, and parent-child reading and conversation.[22] It is a measure of parental time and environmental investment, not monetary investment. Education spending and HOME investment reflect the canonical inputs of money and time, respectively, in the child-quality production function of Del Boca, Flinn and Wiswall (2014).

The index vectors x_{spend} and x_{stim} share a common set of demographic and socioeconomic controls: log family income, family size, child age, child gender, head's race, head's years of schooling, and three region indicators. We include an indicator for if the child has been previously classified as gifted and talented (`ever-gifted`) in both equations, and an indicator for if the primary-caregiver works a non-standard-shift as an exclusive covariate in the cognitive environment equation. The

[20] A more foundational contribution is Becker and Tomes (1986), which describes the partial inheritance of endowments and resources across generations.

[21] See Tsouderou and Tuzel (2026) for another paper using the CDS data.

[22] The analysis sample restricts to children aged 10 or older so that the HOME-10+ scale is consistently applied.

argument is that, conditional on income, if the primary-caregiver works a non-standard shift, it will not affect financial investments into the child's education, but the nature of their job will render it more difficult to, for example, read to children and create a favorable cognitive environment for their child.[23]

Results and Interpretation. Figure 10 shows the full threshold structure in the latent space. At higher HOME levels, a lower latent index is required to enter the higher education-spending category, consistent with complementarity of the decisions between home inputs and financial investment in child production. The thresholds for the cognitive environment shift less in the same way. The estimate of the correlation coefficient is negative, $\widehat{\rho} = -0.27$ (0.12). This is not inconsistent with the complementarity displayed by the thresholds as the threshold layer captures the production complementarity emphasized by Del Boca, Flinn and Wiswall (2014), while the error correlation captures dependence in unobservables. Its negative sign is consistent with labor-supply constraints, whereby households with a higher latent propensity for monetary investment may have a lower latent propensity for time- and attention-intensive home investment, conditional on observed income and covariates.

These results illustrate why separating the two layers matters. By contrast, the lattice model estimates a positive correlation, $\widehat{\rho} = 0.17$ (0.02). We interpret this as a consequence of forcing threshold dependence and residual dependence into a single correlation coefficient, where the complementarity embedded in the threshold structure dominates the negative trade-off in unobservables, giving a positive estimate.

Finally, the index coefficients (reported in Table 5) differ across specifications in some places: in the education spending equation, the positive effect of child age is exacerbated in the non-lattice model, whereas the positive effects of family income, child gender (female), parental race (white), and having a gifted child are muted.

Our results have implications for the literature on investments in children's human capital. First, the complementarity assumption between financial and non-financial investment into a child's human capital as adopted in the literature is supported by our threshold structure. Second, a key finding of Del Boca, Flinn and Wiswall (2014) is that non-financial investments are more productive than financial ones, and cash subsidies to households with children have small impacts on child quality due to the relatively low impact of money investments on child outcomes. Our model does not speak to the productivity of each of the investment types. Instead, we point out that while cash subsidies themselves may have a muted effect, they may relax budget constraints. Furthermore, even if financial investments into education are less productive than non-financial ones, the complementarity between financial and non-financial investments can make financial subsidies foster non-financial investments, rather than lead to a substitution between them.

8 Conclusion

This paper develops a class of multivariate ordered discrete response models with general rectangular structures that substantially extends the reach of traditional lattice-based ordered choice models. The central insight is that the thresholds governing decisions in one dimension need not be functionally independent of outcomes in other dimensions. Allowing thresholds to shift as a function

[23] Two points are worth noting. First, a specification excluding `ever-gifted` from the HOME equation (so that each equation has an exclusive covariate) yields the same qualitative conclusions. Second, there may be a concern that parents working non-standard shifts may compensate for their lack of time by spending more on their child's education, invalidating the exclusion. However, because our measure of financial investment focuses mainly on direct schooling costs, it is unlikely to capture this the direct substitution of paid daytime care for parental time.

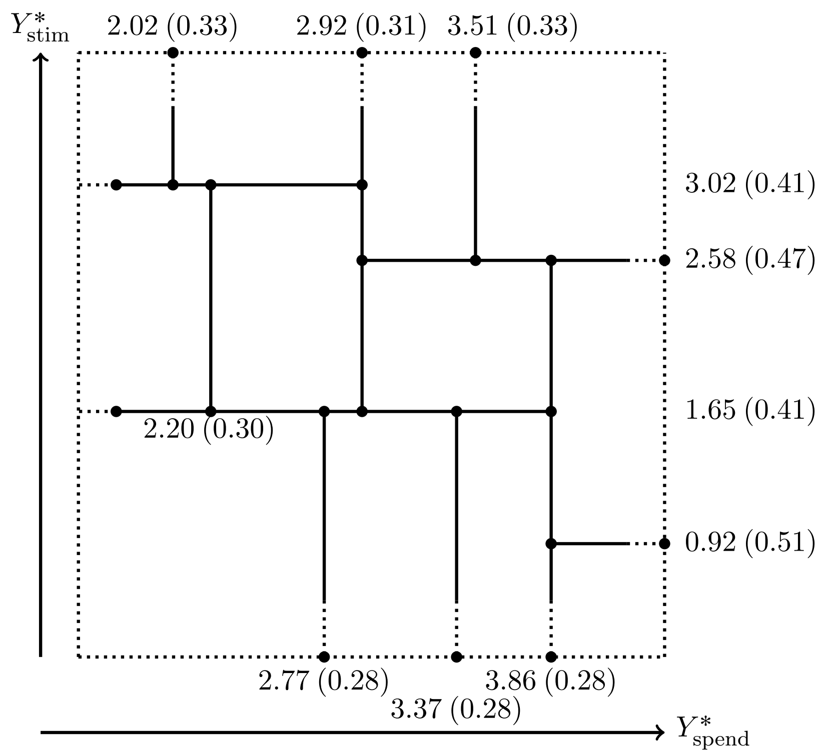


Figure 10: Threshold structure: parental investment example

of the full vector of outcome indices introduces a second layer of dependence through the decision-rule structure itself, which is conceptually distinct from the correlation of latent unobservables. Standard lattice models collapse these two layers into one, conflating behavioral interdependence (narrow/broad bracketing) with unobserved heterogeneity and thereby misattributing the sources of covariation across outcomes.

By formalizing coherency, deriving utility-based microfoundations, and proving identification, our framework enables realistic modeling of complex multidimensional choices. The approach can reveal sign-changing partial effects, indirect covariate influences, and varying complementarity or substitutability, all of which are absent in lattice models. In doing so, our contribution expands the econometric toolkit for studying multidimensional decisions and improves interpretability. As our empirical examples demonstrate, important economic contexts, such as selection markets, can be revisited, and new environments explored, with the models we have introduced.

Several directions for future research are natural. A dynamic extension, in which thresholds evolve as earlier outcomes resolve, would allow the framework to accommodate decision-making over time while still nesting the static case. Allowing for endogeneity of covariates (perhaps through a control-function approach adapted to the two-layer structure) would broaden the range of empirical settings in which the model is applicable. A third direction concerns scalability as the number of dimensions D grows. The number of free threshold parameters grows rapidly with D, and disciplining this expansion through structures in which thresholds depend only on a low-dimensional summary of the other outcome indices, or through shape constraints that encode anticipated complementarity or substitutability patterns, would make high-dimensional applications tractable.

There are also important current empirical settings where the methods developed here can be applied directly. Firms adopting artificial intelligence face simultaneous choices over the depth of integration into production, the intensity of workforce reskilling, and the scale of investment in complementary data infrastructure. These margins are functionally interdependent as AI integration yields larger productivity gains when the workforce has been trained to use it, while reskilling investments pay more when AI tools are embedded throughout production. However, these are also subject to budget and managerial attention constraints that work in the opposite direction. The general rectangular structure would let the threshold layer carry the technological complementarities, leaving the unobservable correlation to capture forces such as managerial type, organizational culture, and beliefs about the technology's trajectory.

More broadly, our framework makes a testable restriction for any lattice model. Wherever researchers have used bivariate ordered probit, bivariate probit, or correlated ordered logit models to study joint discrete decisions, the lattice assumption can now be potentially tested rather than maintained, and the two dependence layers can be estimated rather than conflated. We hope the framework opens new opportunities to revisit important questions in applied economics where the joint structure of ordered decisions carries the economic content of interest.

References

Anderson, J. A. 1984. "Regression and Ordered Categorical Variables." *Journal of the Royal Statistical Society: Series B (Methodological)*, 46(1): 1–30. 4

Apesteguia, Jose, and Miguel A. Ballester. 2023. "Random utility models with ordered types and domains." *Journal of Economic Theory*, 208: 105605. 4

Aradillas-López, Andrés, and Adam M. Rosen. 2022. "Inference in ordered response games with complete information." *Journal of Econometrics*, 226(2): 451–476. 5

Becker, Gary S., and Nigel Tomes. 1986. "Human Capital and the Rise and Fall of Families." *Journal of Labor Economics*, 4(3): S1–S39. 28

Benetton, Matteo, and Giovanni Compiani. 2024. "Investors' Beliefs and Cryptocurrency Prices." *The Review of Asset Pricing Studies*, 14(2): 197–236. 25

Benetton, Matteo, and Greg Buchak. 2026. "Revolving Credit to SMEs: The Role of Business Credit Cards." *SSRN Working Paper*.
`http://dx.doi.org/10.2139/ssrn.4997456`. 22

Berry, Steven, and Peter Reiss. 2007. "Empirical Models of Entry and Market Structure." *Handbook of Industrial Organization*, 3: 1845–1886. 5

Bhat, Chandra R., and Vamsi Pulugurta. 1998. "A comparison of two alternative behavioral choice mechanisms for household auto ownership decisions." *Transportation Research Part B: Methodological*, 32(1): 61–75. 4

Bhattacharya, Debopam, and Tatiana Komarova. 2026. "Incorporating Social Welfare in Program-Evaluation and Treatment Choice." *The Review of Economics and Statistics*, 108(2): 311–326. 17

Boes, Stefan, and Rainer Winkelmann. 2006. "Ordered Response Models." In *Modern Econometric Analysis.* Chapter 12, 167–181. Springer. 4

Camara, Modibo. 2021. "Computationally Tractable Choice." *Working Paper.* 5

Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler. 1997. "Labor Supply of New York City Cabdrivers: One Day at a Time." *The Quarterly Journal of Economics*, 112(2): 407–441. 5

Cameron, Stephen V., and James J. Heckman. 1998. "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males." *Journal of Political Economy*, 106(2): 262–333. 4

Carneiro, Pedro, Karsten Hansen, and James 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. 4

Chen, Songnian, and Shakeeb Khan. 2003. "Rates of convergence for estimating regression coefficients in heteroskedastic discrete response models." *Journal of Econometrics*, 117: 245–278. 16

Chesher, Andrew, and Adam M. Rosen. 2017. "Generalized Instrumental Variable Models." *Econometrica*, 85(3): 959–989. 5

Chesher, Andrew, and Adam M. Rosen. 2020. "Structural modeling of simultaneous discrete choice." *Cemmap Working Paper.* 5

Chiappori, Pierre-Andre, and Bernard Salanie. 2000. "Testing for Asymmetric Information in Insurance Markets." *Journal of Political Economy*, 108(1): 56–78. 3, 21

Ciliberto, Federico, and Elie Tamer. 2009. "Market Structure and Multiple Equilibria in Airline Markets." *Econometrica*, 77(6): 1791–1828. 5

Cohen, Alma, and Peter Siegelman. 2010. "Testing for Adverse Selection in Insurance Markets." *The Journal of Risk and Insurance*, 77(1): 39–84. 3, 21

Coppejans, Mark. 2007. "On efficient estimation of the ordered response model." *Journal of Econometrics*, 137(2): 577–614. 3, 16, 17, 18

Cunha, Flavio, and James Heckman. 2007. "The Technology of Skill Formation." *American Economic Review*, 97(2): 31–47. 27

Cunha, Flavio, James J. Heckman, and Salvador Navarro. 2007. "The Identification and Economic Content of Ordered Choice Models with Stochastic Thresholds." *International Economic Review*, 48(4): 1273–1309. 4

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica*, 78(3): 883–931. 27

Del Boca, Daniela, Christopher Flinn, and Matthew Wiswall. 2014. "Household Choices and Child Development." *The Review of Economic Studies*, 81(1): 137–185. 4, 28, 29

Einav, Liran, Amy Finkelstein, Stephen P. Ryan, Paul Schrimpf, and Mark R. Cullen. 2013. "Selection on Moral Hazard in Health Insurance." *American Economic Review*, 103(1): 178–219. 22, 23, 24

Foster, Kevin, Claire Greene, and Joanna Stavins. 2021. "The 2020 Survey of Consumer Payment Choice: Summary Results." *Federal Reserve Bank of Atlanta Research Data Reports.* 25

Gentry, Matthew, Tatiana Komarova, and Pasquale Schiraldi. 2023. "Preferences and Performance in Simultaneous First-Price Auctions: A Structural Analysis." *The Review of Economic Studies*, 90(2): 852–878. 1

Gentry, Matthew, Tatiana Komarova, Pasquale Schiraldi, and Wiroy Shin. 2019. "On Monotone Strategy Equilibria in Simultaneous Auctions for Complementary Goods." *Journal of Mathematical Economics*, 85: 68–81. 1

Greene, William H., and David A. Hensher. 2010. *Modeling Ordered Choices: A Primer.* Cambridge University Press. 4

Hackmann, Martin B., Jonathan T. Kolstad, and Amanda E. Kowalski. 2015. "Adverse Selection and an Individual Mandate: When Theory Meets Practice." *American Economic Review*, 105(3): 1030–1066. 22

Han, Aaron K. 1987. "Non-parametric analysis of a generalized regression model: The maximum rank correlation estimator." *Journal of Econometrics*, 35(2-3): 303–316. 16

Handel, Benjamin R. 2013. "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts." *American Economic Review*, 103(7): 2643–2682. 21, 24

Heckman, James J. 1978. "Dummy Endogenous Variables in a Simultaneous Equation System." *Econometrica*, 46(4): 931–959. 5

Heckman, James J., Robert J. Lalonde, and Jeffrey A. Smith. 1999. "The Economics and Econometrics of Active Labor Market Programs." *Handbook of Labor Economics*, 3, Part A: 1865–2097. 4

Herrnstein, Richard J., and Dražen Prelec. 1991. "Melioration: A Theory of Distributed Choice." *The Journal of Economic Perspectives*, 5(3): 137–156. 5

Heyman, Gene M. 1996. "Resolving the contradictions of addiction." *Behavioral and Brain Sciences*, 19(4): 561–574. 5

Honoré, Bo E., and Aureo De Paula. 2010. "Interdependent Durations." *The Review of Economic Studies*, 77(3): 1138–1163. 5

Honoré, Bo E., and James L. Powell. 2005. "Pairwise Difference Estimators for Nonlinear Models." *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, 520–553. Cambridge University Press. 16

Horowitz, Joel L. 2010. *Semiparametric and Nonparametric Methods in Econometrics.* Springer Series in Statistics, Springer-Verlag. 13

Kahn, Charles, and Jose Manuel Linares Zegarra. 2016. "Identity Theft and Consumer Payment Choice: Does Security Really Matter?" *Journal of Financial Services Research*, 50(1): 121–159. 25

Kahneman, Daniel, and Dan Lovallo. 1993. "Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking." *Management Science*, 39(1): 17–31. 5

Klein, Roger W., and Robert P. Sherman. 2002. "Shift Restrictions and Semiparametric Estimation in Ordered Response Models." *Econometrica*, 70(2): 663–691. 16

Komarova, Tatiana, and William Matcham. 2025. "Multivariate Ordered Discrete Response Models with Lattice Structures." Working paper. 9, 12, 16

Lewbel, Arthur. 2000. "Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables." *Journal of Econometrics*, 97: 145–6177. 4, 13, 16

Lewbel, Arthur. 2003. "Ordered Response Threshold Estimation." *Unpublished Working Paper*. 4, 13, 16

Lewbel, Arthur. 2007. "Coherency and Completeness of Structural Models Containing a Dummy Endogenous Variable." *International Economic Review*, 48(4): 1379–1392. 5

Lewbel, Arthur. 2014. "An Overview of the Special Regressor Method." In *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics.* , ed. Jeffrey S. Racine, Liangjun Su and Aman Ullah. Oxford University Press. 4

Lian, Chen. 2020. "A Theory of Narrow Thinking." *The Review of Economic Studies*, 88(5): 2344–2374. 5

Liu, Ruixuan, and Zhengfei Yu. 2024. "Simple Semiparametric Estimation of Ordered Response Models." *Econometric Theory*, 40(1): 1–36. 16

Manski, Charles F. 1985. "Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator." *Journal of Econometrics*, 27(3): 313–333. 13

Manski, Charles F. 1988. "Identification of Binary Response Models." *Journal of the American Statistical Association*, 83(403): 729–738. 13

Matcham, William. 2026. "Risk-Based Borrowing Limits in Credit Card Markets." *SSRN Working Paper.*
`http://dx.doi.org/10.2139/ssrn.4926974`. 22

Newey, Whitney K., and Daniel McFadden. 1994. "Chapter 36: Large sample estimation and hypothesis testing." *Handbook of Econometrics*, 4: 2111–2245. 19

Rabin, Matthew, and Georg Weizsäcker. 2009. "Narrow Bracketing and Dominated Choices." *American Economic Review*, 99(4): 1508–43. 5

Read, Daniel, George Loewenstein, and Matthew Rabin. 1999. "Choice Bracketing." *Journal of Risk and Uncertainty*, 19(1/3): 171–197. 5

Simonson, Itamar, and Russell S. Winer. 1992. "The Influence of Purchase Quantity and Display Format on Consumer Preference for Variety." *Journal of Consumer Research*, 19(1): 133–138. 5

Tamer, Elie. 2003. "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria." *The Review of Economic Studies*, 70(1): 147–165. 5

Thakral, Neil, and Linh T. Tô. 2021. "Daily Labor Supply and Adaptive Reference Points." *American Economic Review*, 111(8): 2417–43. 5

Thaler, Richard H. 1999. "Mental accounting matters." *Journal of Behavioral Decision Making*, 12(3): 183–206. 5

Tsouderou, Athena, and Selale Tuzel. 2026. "The Paradox of Opportunity: Socioeconomic Origins of Financial Gender Gaps." University of Miami Business School University of Miami Business School Research Paper No. 4860259. 28

Tversky, Amos, and Daniel Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." *Science*, 211(4481): 453–458. 5

Zhang, Mu. 2021. "A Theory of Choice Bracketing under Risk." *Working Paper*. 5

Zuvekas, Samuel H., and Gary L. Olin. 2009. "Validating Household Reports of Health Care Use in the Medical Expenditure Panel Survey." *Health Services Research*, 44(5p1): 1679–1700. 23

Online Supplement

A 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 $R_{j_1,\ldots,j_D}$ 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 $\sum_{j_1=1}^{M_1} \cdots \sum_{j_D=1}^{M_D} P\left(Y = \left(y_{j_1}^{(1)}, \ldots, y_{j_D}^{(D)}\right) \mid x\right) = 1$ 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.

A.1 Bivariate case

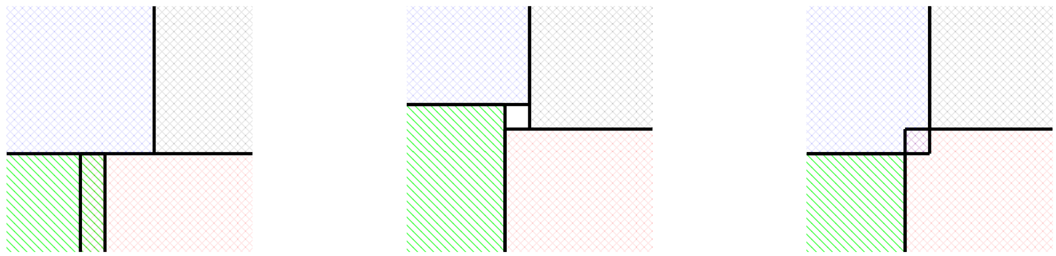
We begin by examining the bivariate case, $D = 2$. The main result for this case is presented in Proposition 1, which states that a bivariate model with a general rectangular structure is generically coherent in the latent space if and only if it is locally hierarchical in every local configuration. 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 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.

Figure 11: Potential violations of coherency.



Panel 1 Panel 2 Panel 3

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).

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:

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

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

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

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

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

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

A.2 D Greater Than 2

For $D > 2$, we again examine each local configuration of the form $\{(j_1+\ell_1, \ldots, j_D+\ell_D)\}_{\ell_d \in \{0,1\},\, d=1,\ldots,D}$. 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 second coordinate fixed. For each such projection, we apply the bivariate condition from Proposition 1. This yields conditions

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

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

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

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 $\prod_{d=1}^{3}(\alpha^{(d)}_{\mathbf{j}} - \alpha^{(d)}_{\mathbf{j}'_{(d)}}) = 0$, where $\mathbf{j}'_{(d)}$ denotes the index vector obtained from $\mathbf{j}$ by replacing the d-th coordinate with that of $\mathbf{j}'$:

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

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 ($2 \leqslant k < D$) 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 $D > 2$.

B Appendix B: Proofs

Proof of Theorem 1. Without loss of generality, take $d=1$. The proof proceeds in the following way. First, we establish an auxiliary result that $P(Y^{c_1} \leqslant y^{(1)}_{j_1} \mid x)$ 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,

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

$j_1 = 1, \dots, M_1$. 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

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

$$
R^*_{j,j_2,\dots,j_D} = (-\infty, \alpha^{(1)}_{j,j_2,\dots,j_D}] \times_{d=2}^{D} (\alpha^{(d)}_{j,j_2,\dots,j_{d-1},j_d-1,j_{d+1},\dots,j_D}, \alpha^{(d)}_{j,j_2,\dots,j_{d-1},j_d,j_{d+1},\dots,j_D}].
$$

In turn, this gives

$$
\begin{aligned}
P(Y^{c_1} \leqslant y^{(1)}_{j_1} \,|\, x) = \sum_{j_2=1}^{M_2} \dots \sum_{j_D=1}^{M_D} \Big(&\overline{F}\left(-\infty, \alpha^{(2)}_{j,j_2-1,\dots,j_D} - x_2\beta_2, \dots, \alpha^{(D)}_{j,j_2,\dots,j_D-1} - x_D\beta_D\right) \\
&+ F\left(\alpha^{(1)}_{j_1,j_2,\dots,j_D} - x_1\beta_1, \alpha^{(2)}_{j_1,j_2,\dots,j_D} - x_2\beta_2, \dots, \alpha^{(D)}_{j_1,j_2,\dots,j_D} - x_D\beta_D\right) - 1\Big),
\end{aligned}
$$

where F and $\overline{F}$, 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, $P(Y^{c_1} \leqslant y^{(1)}_{j_1} \,|\, x)$ 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 $\alpha^{(1)}_{j_1,j_2,\dots,j_D} - x_1\beta_1$ is in the interior of the support of ε_1 for $x \in S_1(j_1)$. Using the fact that the support of ε_1 is convex (implied by convexity of the support of $\boldsymbol{\varepsilon}$), we then conclude that $P(Y^{c_1} \leqslant y^{(1)}_{j_1} \,|\, x)$ 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 $(\underline{x}_{1,1}, \overline{x}_{1,1}) \times S_{1,-1}(j_1)$ 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

$$
x_{1,2:L_1}\beta_{1,2:L_1} > x_{1,2:L_1}b_{1,2:L_1}. \tag{10}
$$

Now fix any $x_{1,2:L_1}$ that satisfies (10). Then for any $\tilde{x}_{1,1} \in (\underline{x}_{1,1}, \overline{x}_{1,1})$, 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 $\tilde{\tilde{x}}_{1,1} \in (\underline{x}_{1,1}, \overline{x}_{1,1})$ such that

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

Because of $x_{1,1}$ being exclusive for Y^{*c_1}, when we vary $x_{1,1}$, the values of x_2, ..., x_D remain exactly the same. This means that in the expression for $P(Y^{c_1} \leqslant y^{(1)}_j \,|\, x_1, \dots, x_D)$, 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 $P(Y^{c_1} \leqslant y_j^{(1)} \,|\, x_1, \ldots, x_D)$ with the reverse ordering of the first argument in the first index.

Therefore, (a) in (11) implies that

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

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

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

Combining the last two inequalities results in an obvious contradiction

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

and from our discussion it is clear that this contradiction is obtained for a positive measure of $(\tilde{x}_{1,1}, x_{1,2:L_1}, x_2, \ldots, x_D)$. 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. ▫

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 κ_d to be $\leqslant$. If in that condition $\overline{x}_{d,1}$ is large enough, we take κ_d to be $>$. Analyze now $P(\cap_{d=1}^D (Y^{c_d}\, \kappa_d\, y_{j_d}^{(d)} \,|\, x)$ 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,

$$P(\cap_{d=1}^D (Y^{c_d}\, \kappa_d\, y_{j_d}^{(d)}) \,|\, x) = \sum_{\tilde{j}_1\, \kappa_1\, j_1} \ldots \sum_{\tilde{j}_D\, \kappa_D\, j_D} P\left((x_1\beta_1 + \varepsilon_1, \ldots, x_D\beta_D + \varepsilon_D) \in R_{\tilde{j}_1,\ldots,\tilde{j}_D} | x\right)$$

Focus e.g. on $d = 1$ and for any $d \geqslant 2$ take $x_{d,1} \to \underline{x}_{d,1}$ if κ_d is $\leqslant$ and take $x_{d,1} \to \overline{x}_{d,1}$ if κ_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 $P(\cap_{d=1}^D (Y^{c_d}\, \kappa_d\, y_{j_d}^{(d)}) \,|\, x)$ 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 $\underline{x}_{d,1}$, $\overline{x}_{d,1}$ are infinite (that is, $-\infty$, ∞, respectively), then we obtain

$$P(\cap_{d=1}^D (Y^{c_d}\, \kappa_d\, y_{j_d}^{(d)}) \,|\, x) \to F_{1,\kappa_1}\left(\alpha^{(1)}_{j_1, m_2 \ldots, m_D} - x_1\beta_1\right),$$

where $m_d = M_d$ if κ_d is $>$ and $m_d = 1$ if κ_d is $\leqslant$, for $d \neq 1$. Condition (a) of the theorem as well as the fact that $P((Y^{*c_1}, ..., Y^{*c_D} \in R_{\tilde{j}_1,\tilde{j}_2,..,\tilde{j}_D}|x) \to 0$ for $\tilde{j}_d\, \kappa_d\, j_d$, $d \geqslant 2$, and $(\tilde{j}_2, .., \tilde{j}_D) \neq (m_2, ..., m_D)$, guarantee that $\alpha^{(1)}_{j_1,m_2\ldots,m_D} - x_1\beta_1$ is in the interior of the support of ε_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 κ_1 is $>$ and strictly decreasing if κ_1 is $\leqslant$.

If some (or all) of the relevant boundaries $\underline{x}_{d,1}$, $\overline{x}_{d,1}$ are finite, then the limit of $P((Y^{*c_1}, ..., Y^{*c_D}) \in R_{\tilde{j}_1,\tilde{j}_2,..,\tilde{j}_D}|x)$ 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 β_1. This can be done for any β_d. ▫

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

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

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

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

Combining the knowledge of $\alpha^{(d)}_{j_1,j_2,\dots,j_D}$ 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 $\boldsymbol{\varepsilon}$ as

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

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

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 $\alpha^{(d)}_{j_1,\dots,q_{d-1},r_d(j_d),j_{d+1},\dots,j_D}$, 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 $\boldsymbol{\varepsilon}$ with this "corner". If $j_d = 1$ we take κ_d to be $\leqslant$, and if $j_d = M_d$ we take κ_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 ∞ 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 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 ∞ or $-\infty$. Now we will have only $D-1$ known thresholds fixed at ∞ 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 $\alpha^{(d)}_{j_1,\dots,j_{d-1},q_d-1,j_{d+1},\dots,j_D}$, $\alpha^{(d)}_{j_1,\dots,j_{d-1},q_d,j_{d+1},\dots,j_D}$ (in dimension d) and $\alpha^{(h)}_{j_1,\dots,j_{d-1},q_d,j_{d+1},\dots,j_D}$, $h \neq d$. For $q_d = 2$ and $q_d = M_d - 1$ we only have D unknown thresholds since $\alpha^{(d)}_{j_1,\dots,j_{d-1}1,j_{d+1},\dots,j_D}$ and $\alpha^{(d)}_{j_1,\dots,j_{d-1},M_d-1,j_{d+1},\dots,j_D}$ 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 $\alpha^{(d)}_{j_1,\ldots,j_{d-1},q_d,j_{d+1},\ldots,j_D}$, for $q_d = 2,\ldots,M_d-2$ in the following way. Choosing, once again, κ_h to be $\leqslant$ if $j_h = 1$ and to be $>$ if $j_h = M_h$, $h \neq d$, obtain that

$$P\left(Y^{c_d} = y^{(d)}_{q_d},\ \cap_{h\neq d}\left(Y^{c_h} = y^{(h)}_{j_h}\right) \mid x\right) = P\Bigg(\alpha^{(d)}_{j_1,\ldots,j_{d-1},q_d-1,j_{d+1},\ldots,j_D} < Y^{*c_d} \leqslant \alpha^{(d)}_{j_1,\ldots,j_{d-1},q_d,j_{d+1},\ldots,j_D},\ \cap_{h\neq d}\left(Y^{*c_h}\ \ \kappa_h\ \ \alpha^{(h)}_{j_1,\ldots,j_{d-1},q_d,j_{d+1},\ldots,j_D}\right)\Bigg).$$

By taking $x_{h,1} \to \underline{x}_{h,1}$ if κ_h is $\leqslant$ or $x_{h,1} \to \overline{x}_{h,1}$ if κ_h is $>$ for all $h \neq d$, we identify

$$\lim_{x_{h,1}\to,h\neq d} P(Y^{c_d} = y^{(d)}_{q_d}\ \cap_{h\neq d}\left(Y^{c_h} = y^{(h)}_{j_h}\right) \mid x) = F_{d,\leqslant}\big(\alpha^{(d)}_{j_1,\ldots,j_{d-1},q_d,j_{d+1},\ldots,j_D} - x_d\beta_d\big) - F_{d,\leqslant}\big(\alpha^{(d)}_{j_1,\ldots,j_{d-1},q_d-1,j_{d+1},\ldots,j_D} - x_d\beta_d\big).$$

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

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

Now we also fix one h such that $h \neq d$ and show the threshold $\alpha^{(h)}_{j_1,\ldots,j_{d-1},q_d,j_{d+1},\ldots,j_D}$ is identified. For this, we consider $P(Y^{c_d} = y^{(d)}_{q_d}\ \cap_{h\neq d}\left(Y^{c_h} = y^{(h)}_{j_h}\right) \mid x)$ 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 $x_{\tilde{h},1} \to \underline{x}_{\tilde{h},1}$ if $\kappa_{\tilde{h}}$ is $\leqslant$ and take $x_{\tilde{h},1} \to \overline{x}_{\tilde{h},1}$ if $\kappa_{\tilde{h}}$ is $>$. In such a limit we identify

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

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

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

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 ∞ or $-\infty$. At least two other thresholds will be known and finite from previous stages.

Consider a rectangular border region $R_{j_1,\ldots,j_{d_1-1},q_{d_1},j_{d_1+1},\ldots,j_{d_2-1},q_{d_2},j_{d_2+1},\ldots,j_D}$ 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 $q_{d_j} = 2,\ldots,M_{d_j}-1$, $j = 1,2$. 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 $R_{q_1,q_2.j_3,\ldots,j_D}$, where $j_h \in \{1, M_h\}$ for $h \neq d_1, d_2 \geqslant 3$ is described by $D+2$ thresholds $\alpha^{(1)}_{q_1-1,q_2,j_3,\ldots,j_D}$, $\alpha^{(1)}_{q_1,q_2,j_3,\ldots,j_D}$, $\alpha^{(2)}_{q_1,q_2-1,j_3,\ldots,j_D}$, $\alpha^{(2)}_{q_1,q_2,j_3,\ldots,j_D}$ (in dimensions 1 and 2) and $\alpha^{(h)}_{q_1,q_2,j_3,\ldots,j_D}$, $h \geqslant 3$ (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 direction, it means that it is $\mathbb{R}^2$ and in the condition of the theorem we necessarily have $\underline{x}_{h,1} = -\infty$ and $\overline{x}_{h,1} = \infty$ 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 $\alpha^{(1)}_{2,2,j_3,\ldots,j_D}$, $\alpha^{(2)}_{2,2,j_3,\ldots,j_D}$ and $\alpha^{(h)}_{2,2,j_3,\ldots,j_{h-1},r_h(j_h),j_{h+1},\ldots,j_D}$, $h \geqslant 3$. 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 $P\left(Y^{c_1} = y^{(1)}_2, Y^{c_2} = y^{(2)}_2, \cap_{h\geqslant 3}\left(Y^{c_h} = y^{(h)}_{j_h}\right) \mid x\right)$ and find its limit $h \geqslant 3$. Just as before, we take $x_{h,1} \to \underline{x}_{h,1}$ if $j_h = 1$ and take $x_{h,1} \to \overline{x}_{h,1}$ if $j_h = M_h$. In this limit we identify

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

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

$$F_{2,\leqslant}(\underbrace{\alpha^{(2)}_{2,2,j_3,\ldots,j_D} - x_2\beta_2}_{z_2}) - F_{2,\leqslant}(\underbrace{\alpha^{(2)}_{2,1,j_3,\ldots,j_D} - x_2\beta_2}_{z_1}) \in (0,1).$$

Using the knowledge of $F_{2,\leqslant}$ and the strict monotonicity of the obtained probability with respect to z_2 with z_1 fixed we identify $\alpha^{(2)}_{2,2,j_3,\ldots,j_D}$. Analogously we can identify $\alpha^{(1)}_{2,2,j_3,\ldots,j_D}$ by keeping x_1 fixed and taking $x_{2,1} \to -\infty$. Once $\alpha^{(1)}_{2,2,j_3,\ldots,j_D}$, $\alpha^{(2)}_{2,2,j_3,\ldots,j_D}$ are identified, we can identify $\alpha^{(3)}_{2,2,r_3(j_3),\ldots,j_D}$ by taking in our observed probability $x_{h,1} \to$ in the manner described above but now only for $h \geqslant 4$. In the limit we identify

$$\sum_{\ell_1=0}^{1}\sum_{\ell_2=0}^{1}\sum_{\ell_3=0}^{1}(-1)^{\ell_1+\ell_2+\ell_3}F_{1,2,3;\leqslant\leqslant\leqslant}\Big(\alpha^{(1)}_{1+\ell_1,2,1,j_4,\ldots,j_D} - x_1\beta_1,$$
$$\alpha^{(2)}_{2,1+\ell_2,1,j_4,\ldots,j_D} - x_2\beta_2, \alpha^{(3)}_{2,2,\ell_3,j_4\ldots,j_D} - x_3\beta_3\Big) \in (0,1)$$

if $j_3 = 1$, and identify

$$\sum_{\ell_1=0}^{1}\sum_{\ell_2=0}^{1}\sum_{\ell_3=0}^{1}(-1)^{\ell_1+\ell_2+\ell_3}F_{1,2,3;\leqslant\leqslant>}\Big(\alpha^{(1)}_{1+\ell_1,2,M_3,j_4,\ldots,j_D} - x_1\beta_1,$$
$$\alpha^{(2)}_{2,1+\ell_2,M_3,j_4,\ldots,j_D} - x_2\beta_2, \alpha^{(3)}_{2,2,M_3-\ell_3,j_4\ldots,j_D} - x_3\beta_3\Big) \in (0,1)$$

if $j_3 = M_3$. No matter which of these situations we have, we use the knowledge of $F_{1,2,3;\leqslant\leqslant\kappa_3}$, the knowledge of 5 out of 6 thresholds in them and the strict monotoncity of that limit with respect to the unknown threshold ($\alpha^{(3)}_{2,2,\ell_3,j_4\dots,j_D}$ in the first and $\alpha^{(3)}_{2,2,M_3-\ell_3,j_4\dots,j_D}$ in the second situation) to identify this threshold. Analogously we can identify $\alpha^{(h)}_{2,2,j_3,\dots j_{h-1},r_h(j_h),j_{h+1},\dots,j_D}$ for any $h \geqslant 4$. 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 $\mathbb{R}^2$ originating from $(0,0)'$ as

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

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 $P\left(Y^{c_1} = y^{(1)}_2, Y^{c_2} = y^{(2)}_{M_2-1}, \cap_{h\geqslant 3}(Y^{c_h} = y^{(h)}_{j_h}\right) \mid x)$ and find its limit when $x_{h,1}$ converges for $h \geqslant 3$ in a way described earlier (take $x_{h,1} \to \underline{x}_{h,1}$ if $j_h = 1$ and take $x_{h,1} \to \overline{x}_{h,1}$ if $j_h = M_h$). In this limit we identify

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

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

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

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

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 $F_{1,2;\leqslant>}$). 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 $z_1 = \alpha^{(1)}_{1,M_2-1,j_3,\ldots,j_D} - x_1\beta_1$, $z_2 = \alpha^{(2)}_{2,M_2-1,j_3,\ldots,j_D} - x_2\beta_2$. 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:

$$Q(x_1, x_2) = Pr_{(\varepsilon_1,\varepsilon_2)}((\varepsilon_1, \varepsilon_2)' \in R_\Delta \cap R_\delta) + Pr_{(\varepsilon_1,\varepsilon_2)}((\varepsilon_1, \varepsilon_2)' \in R_\Delta \backslash R_\delta)$$
$$Q(x_1, x_2) = Pr_{(\varepsilon_1,\varepsilon_2)}((\varepsilon_1, \varepsilon_2)' \in R_\Delta \cap R_\delta) + Pr_{(\varepsilon_1,\varepsilon_2)}((\varepsilon_1, \varepsilon_2)' \in R_\delta \backslash R_\Delta).$$

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 $Pr_{(\varepsilon_1,\varepsilon_2)}((\varepsilon_1, \varepsilon_2)' \in R_\Delta \backslash R_\delta)$ and $Pr_{(\varepsilon_1,\varepsilon_2)}((\varepsilon_1, \varepsilon_2)' \in R_\delta \backslash R_\Delta)$ 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 $Pr_{(\varepsilon_1,\varepsilon_2)}((\varepsilon_1, \varepsilon_2)' \in R_\Delta \backslash R_\delta) > 0$, $Pr_{(\varepsilon_1,\varepsilon_2)}((\varepsilon_1, \varepsilon_2)' \in R_\delta \backslash R_\Delta) = 0$. 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 $\alpha^{(1)}_{2,M_2-1,j_3,\ldots,j_D}$, $\tilde{\alpha}^{(2)}_{2,M_2-2,j_3,\ldots,j_D}$ are identified, we can proceed analogously to Subcase 1 to identify $\alpha^{(h)}_{2,M_2-1,j_3,\ldots j_{h-1},r(j_h),j_{h+1},\ldots,j_D}$ for any $h \geqslant 3$. Thus, all the thresholds of $R_{2,M_2-1,j_3,\ldots,j_D}$ are identified.

Building on this result, we will next look at the rectangles $R_{3,M_2-1,j_3,\ldots,j_D}$ and $R_{2,M_2-2,j_3,\ldots,j_D}$ (one index change at a time) and identify their thresholds in a similar manner. Then we look at $R_{3,M_2-2,j_3,\ldots,j_D}$, $R_{4,M_2-1,j_3,\ldots,j_D}$, $R_{2,M_2-3,j_3,\ldots,j_D}$, 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 $R_{q_1,q_2,q_3,j_3,\ldots,j_D}$, where $j_h \in \{1, M_h\}$ for $h \geqslant 4$, the main idea is to start building the identification (the knowledge of the relevant) of thresholds gradually, first, e.g. by taking $q_1 = 2$, $q_2 = M_2 - 1$, $q_3 = M_3 - 1$ which guarantees that at every step at most D thresholds are unknown. At every step, we will have $D - 3$ known thresholds fixed at ∞ 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 $\mathbb{R}^2$. Now, we have to consider eight orthants in $\mathbb{R}^3$ originating from $(0,0,0)'$:

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

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 κ_d as $>$ if $s_d = -$, and with $q_d = 2$ and κ_d as $\leqslant$ 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 $\underline{x}_{1,1} = -\infty$, $\underline{x}_{2,1} = -\infty$, $\overline{x}_{3,1} = \infty$. Then we first consider $q_1 = 2$, $q_2 = 2$, $q_3 = M_3 - 1$ (κ_1 and κ_2 are then $\leqslant$ and κ_3 is $>$). Using results of Stage 3, we find that we only deal with D unknown thresholds $\alpha^{(1)}_{2,2,M_3-1,j_4,\ldots,j_D}$, $\alpha^{(2)}_{2,2,M_3-1,j_4,\ldots,j_D}$, $\alpha^{(3)}_{2,2,M_3-2,j_4,\ldots,j_D}$ and $\alpha^{(h)}_{2,2,M_3-1,j_4,\ldots,j_{h-1},r_h(j_h),j_{h+1},\ldots,j_D}$, $h \geqslant 4$. 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 $P\left(Y^{c_1} = y^{(1)}_2, Y^{c_2} = y^{(2)}_2, Y^{c_3} = y^{(3)}_{M_3-1}. \cap_{h\geqslant 4}\left(Y^{c_h} = y^{(h)}_{j_h}\right) \mid x\right)$ and find its limit when $x_{h,1} \to$ for $h \geqslant 4$. Just as before, we take $x_{h,1} \to \underline{x}_{h,1}$ if $j_h = 1$ and take $x_{h,1} \to \overline{x}_{h,1}$ if $j_h = M_h$. In this limit we identify

$$\sum_{\ell_1=0}^{1}\sum_{\ell_2=0}^{1}\sum_{\ell_3=0}^{1}(-1)^{\ell_1+\ell_2+\ell_3} F_{1,2,3;\leqslant,\leqslant,>}\Big(\alpha^{(1)}_{1+\ell_1,2,M_3-1,j4,\ldots,j_D} - x_1\beta_1,$$
$$\alpha^{(2)}_{2,1+\ell_2,M_3-1,j_4,\ldots,j_D} - x_2\beta_2,\, \alpha^{(3)}_{2,2,M_3-1-\ell_3,j_4,\ldots,j_D} - x_3\beta_3\Big) \in (0,1).$$

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

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

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 $\alpha^{(3)}_{2,2,M_3-1,j_4,\ldots,j_D}$ is known from Stage 3) we identify $\alpha^{(3)}_{2,2,M_3-2,j_4,\ldots,j_D}$.

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

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

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

Once $\alpha^{(1)}_{2,2,M_3-1,j_4,\ldots,j_D}$, $\alpha^{(2)}_{2,2,M_3-1,j_4,\ldots,j_D}$ and $\alpha^{(3)}_{2,2,M_3-2,j_4,\ldots,j_D}$ are identified, we can identify $\alpha^{(h)}_{2,2,M_3-1,j_4,\ldots,j_{h-1},r_h(j}$ $h \geqslant 4$, by taking in our observed probability $x_{h,1} \to$ in the manner described above but now only for $h \geqslant 4$. E.g. for $h = 4$ we identify

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

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

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

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

Building on this result, we will next look at the rectangles $R_{3,2,M_3-1,j_4,\ldots,j_D}$ and $R_{2,3,M_3-1,j_4,\ldots,j_D}$ and $R_{2,2,M_3-2,j_4,\ldots,j_D}$ (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 $\mathcal{O}_{s_1s_2s_3}$ such that $\mathbf{e} + \mathcal{O}_{s_1s_2s_3}$ 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 $\mathbf{e} + O_{s_1s_2s_3}$ (orthants $O_{s_1s_2s_3}$ and $O_{\tau_1\tau_2\tau_3}$ 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 $\mathbf{e} + O_{s1,s2,s3}$ 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 κ_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 $\mathbf{e} + O_{s_1s_2-}$ (for $s_1, s_2 \in \{+, -\}$) intersect the interior $\mathrm{int}(\mathcal{E}_{123})$ of $\mathcal{E}_{123}$ (positive mass in any small rectangle with a vertex at $\mathcal{E}_{123}$ and located in $\cup_{s_1,s_2}(\mathbf{e} + O_{s_1s_2-})$ 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 $e + O_{s_1s_2+}$ 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 κ_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 κ_1, κ_2 as $\leqslant$).

Consider observed $P\left(Y^{c_1} = y^{(1)}_2, Y^{c_2} = y^{(2)}_2, Y^{c_3} = y^{(3)}_{M_3-1}, \cap_{h\geqslant 4}(Y^{c_h} = y^{(h)}_{j_h}\right) | \, x)$ and find its limit when $x_{h,1} \to$ for $h \geqslant 4$ in a way described earlier (take $x_{h,1} \to \underline{x}_{h,1}$ if $j_h = 1$ and take $x_{h,1} \to \overline{x}_{h,1}$ if $j_h = M_h$). Denote $\mathbf{j} \equiv (j_4, \ldots, j_D)$. In the described limit we identify

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

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: $\alpha^{(d)}_{2,2,M_3-1,\mathbf{j}}$, $d = 1, 2$, $\alpha^{(3)}_{2,2,M_3-2,\mathbf{j}}$. 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

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

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 $F_{1,2,3;\leqslant\leqslant>}$). We can take any Δ_d and δ_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 $z_1 = \alpha^{(1)}_{1,2,M_3-1,\mathbf{j}} - x_1\beta_1$, $z_2 = \alpha^{(2)}_{2,1,M_3-1,\mathbf{j}} - x_2\beta_2$, $z_3 = \alpha^{(3)}_{2,2,M_3-2,\mathbf{j}} - x_3\beta_3$ (all the thresholds used here are known at this identification stage). Define rectangles

$$T_\upsilon = [z_1, z_1 + \upsilon_1] \times [z_2, z_2 + \upsilon_2] \times [z_3 - \upsilon, z_3], \quad \upsilon \in \{\Delta, \delta\}. \tag{15}$$

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 $T_\Delta \backslash T_\delta$ 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, $Pr_{(\varepsilon_1,\varepsilon_2,\varepsilon_3)}((\varepsilon_1, \varepsilon_2, , \varepsilon_3)' \in T_\Delta \backslash T_\delta) > 0$. At the same time the interior of $T_\Delta \backslash T_\delta$ is fully contained in $\cup_{s_1,s_2} \mathcal{O}_{s_1 s_2 +}$. Hence $Pr_{(\varepsilon_1,\varepsilon_2,\varepsilon_3)}((\varepsilon_1, \varepsilon_2, \varepsilon_3)' \in T_\delta \backslash T_\Delta) = 0$, 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 κ_3 as $\leqslant$) and in the increasing order in dimension 2 (thus, choosing $q_2 = 2$ and κ_2 as $\leqslant$). 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 κ_1 to be $\leqslant$ and, thus, proceed in the increasing direction too in this dimension. .

Consider observed

$$P\left(Y^{c_1} = y^{(1)}_2, Y^{c_2} = y^{(2)}_2, Y^{c_3} = y^{(3)}_2, \cap_{h \geqslant 4} (Y^{c_h} = y^{(h)}_{j_h}\right) | \, x)$$

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

$$Q(x_1, x_2, x_3) = \sum_{\ell_1=0}^{1} \sum_{\ell_2=0}^{1} \sum_{\ell_3=0}^{1} (-1)^{\ell_1+\ell_2+\ell_3} F_{1,2,3;\leqslant\leqslant\leqslant}(\alpha^{(1)}_{1+\ell_1,2,2,\mathbf{j}} - x_1\beta_1,$$
$$\alpha^{(2)}_{2,1+\ell_2,2,\mathbf{j}} - x_2\beta_2, \alpha^{(3)}_{2,2,1+\ell_3,\mathbf{j}} - x_3\beta_3) \in (0, 1) \tag{16}$$

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: $\alpha^{(d)}_{2,2,2,\mathbf{j}}$, $d = 1, 2, 3$. 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

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

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 Δ_d and δ_d, $d = 1, 2, 3$, be different.

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

$$T_\upsilon = [z_1, z_1 + \upsilon_1] \times [z_2, z_2 + \upsilon_2] \times [z_3, z_3 + \upsilon_3], \quad \upsilon \in \{\Delta, \delta\}. \tag{17}$$

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 $T_\delta \backslash T_\Delta = [z_1, e_1] \times [z_2, e_2] \times (e_3, e_3 + \delta_3 - \Delta_3]$ is in the closure of $\mathcal{O}_{--+}$ and, thus, has probability zero. At the same time, $T_\Delta \backslash T_\delta \in \mathcal{O}_{+--} \cup \mathcal{O}_{-+-} \cup \mathcal{O}_{++-}$ 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 $Pr_{(\varepsilon_1,\varepsilon_2,\varepsilon_3)}((\varepsilon_1, \varepsilon_2, , \varepsilon_3)' \in T_\Delta \backslash T_\delta) > 0$, 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: $\underline{\Delta_1 > \delta_1,\ \Delta_2 < \delta_2,\ \Delta_3 > \delta_3}$**, and Sub-case 1C:** $\underline{\Delta_1 > \delta_1,\ \Delta_2 < \delta_2,\ \Delta_3 < \delta_3}$**.** 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 $z_1 = \alpha^{(1)}_{1,2,2,\mathbf{j}} - x_1\beta_1$, $z_2 = \alpha^{(2)}_{2,1,2,\mathbf{j}} - x_2\beta_2$, $z_3 = \alpha^{(3)}_{2,2,1,\mathbf{j}} - x_3\beta_3$ (note that all the thresholds used here are known at this identification stage). Define rectangles T_Δ, T_δ as in (15). Then $T_\delta \backslash T_\Delta$ 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, $T_\Delta \backslash T_\delta$ 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 $T_\delta \backslash T_\Delta$ and $T_\Delta \backslash T_\delta$ then no orthants $\mathcal{O}_{s_1 s_2 +}$ in its interior intersects $\mathcal{E}_{123}$.

Sub-case 1D. $\underline{\Delta_1 < \delta_1,\ \Delta_2 > \delta_2,\ \Delta_3 < \delta_3.}$ This is completely analogous to Sub-case 1B with the roles of Δ's and δ's reversed.

Sub-case 1E. $\underline{\Delta_1 < \delta_1,\ \Delta_2 < \delta_2,\ \Delta_3 > \delta_3.}$ This is completely analogous to Sub-case 1A with the roles of Δ's and δ's reversed.

Sub-case 1F. $\underline{\Delta_1 < \delta_1,\ \Delta_2 > \delta_2,\ \Delta_3 > \delta_3.}$ This is completely analogous to Sub-case 1C with the roles of Δ's and δ's reversed.

Contradictions obtained in each Sub-case mean that the set of two thresholds we are looking for is unique. Hence, thresholds $\alpha^{(d)}_{2,2,2,\mathbf{j}}$, $d = 1, 2, 3$ are identified. After that we can proceed analogously to Subcase 1 to identify $\alpha^{(h)}_{2,2,2,j_4,\ldots j_{h-1},r(j_h),j_{h+1},\ldots,j_D}$ for any $h \geqslant 4$. Thus, all the thresholds of $R_{2,2,2,\mathbf{j}}$ are identified. Building on this result, we will next look at the rectangles $R_{3,2,2,\mathbf{j}}$, $R_{2,3,2,\mathbf{j}}$ and $R_{2,2,3,\mathbf{j}}$ (one index change at a time) and identify their thresholds in a similar manner. Then we look at $R_{3,3,2,\mathbf{j}}$, $R_{3,2,3,\mathbf{j}}$, $R_{2,3,3,\mathbf{j}}$, etc. until we identify thresholds of all the rectangles considered in this stage.

Table 4: Simulation results: Design 2 thresholds

Parameter	**Truth**	**Non-lattice model**	**Lattice model**
$\alpha^{(1)}_{1,1}$	-3.25	-3.26 (0.11)	
$\alpha^{(1)}_{1,2}$	-3.25	-3.25 (0.10)	-1.48 (0.04)
$\alpha^{(1)}_{1,3}$	-0.5	-0.50 (0.07)	
$\alpha^{(1)}_{2,1}$	0.5	0.51 (0.09)	
$\alpha^{(1)}_{2,2}$	1	0.99 (0.12)	1.59 (0.04)
$\alpha^{(1)}_{2,3}$	5	5.02 (0.14)	
$\alpha^{(1)}_{3,1}$	8	8.03 (0.19)	
$\alpha^{(1)}_{3,2}$	8	8.03 (0.19)	5.12 (0.09)
$\alpha^{(1)}_{3,3}$	8	8.03 (0.19)	
$\alpha^{(2)}_{1,1}$	-4	-3.98 (0.21)	
$\alpha^{(2)}_{2,1}$	2	-2.02 (0.11)	-1.10 (0.04)
$\alpha^{(2)}_{3,1}$	2	-1.99 (0.08)	
$\alpha^{(2)}_{4,1}$	0	-0.01 (0.09)	
$\alpha^{(2)}_{1,2}$	0.5	0.50 (0.05)	
$\alpha^{(2)}_{2,2}$	0.5	0.50 (0.05)	0.90 (0.04)
$\alpha^{(2)}_{3,2}$	0.5	0.50 (0.05)	
$\alpha^{(2)}_{4,2}$	4	4.00 (0.15)	

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

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. ■

C Appendix C: Additional simulation results for Design 2

In Table 4, 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.

D Appendix D: Additional Exhibits for the applications

D.1 Cryptocurrency

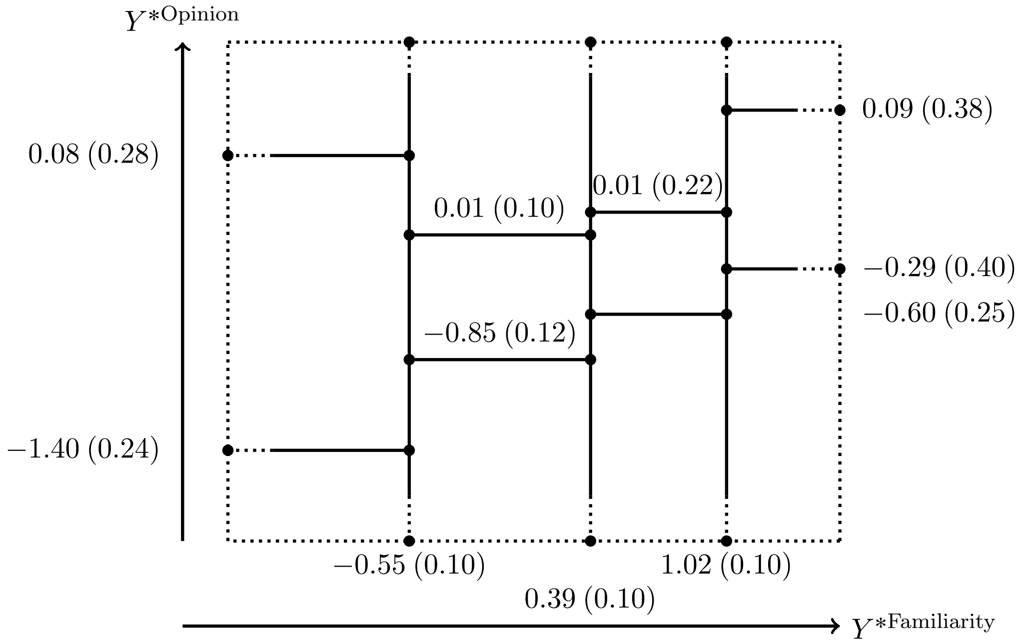


Figure 12: Estimates from cryptocurrency example when assuming a sequential model.

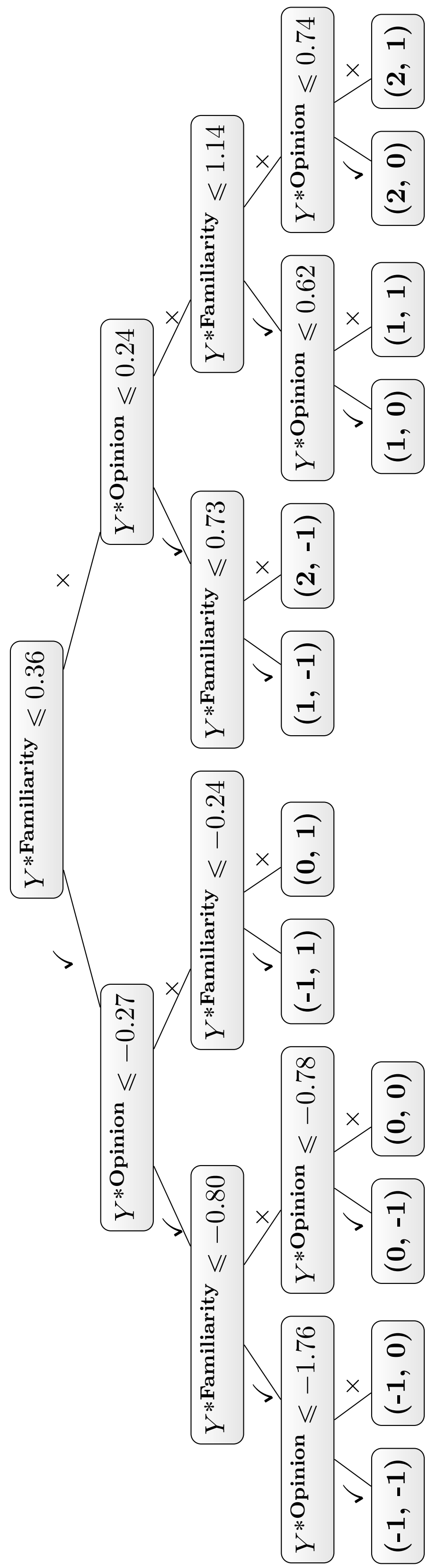


Figure 13: Binary decision tree describing the hierarchical decision process for general rectangular model

D.2 Parental Investment in Children's Education

Table 5: Estimation coefficients: parental investments in children

Variable	**Non-lattice**	**Lattice**
Education Spending (Y_1)		
Log family income	0.21 (0.03)	0.24 (0.03)
Family size	−0.05 (0.02)	−0.04 (0.02)
Child age	0.04 (0.01)	0.02 (0.01)
Child female	0.08 (0.04)	0.12 (0.04)
Head white	0.06 (0.05)	0.10 (0.05)
Head education yrs	0.02 (0.01)	0.04 (0.01)
Ever gifted	0.03 (0.05)	0.10 (0.04)
HOME Cognitive Stimulation (Y_2)		
Log family income	0.15 (0.03)	0.18 (0.03)
Family size	0.05 (0.02)	0.04 (0.02)
Child age	−0.10 (0.01)	−0.09 (0.01)
Child female	0.16 (0.04)	0.18 (0.04)
Head white	0.23 (0.05)	0.25 (0.05)
Head education yrs	0.11 (0.01)	0.11 (0.01)
Ever gifted	0.41 (0.05)	0.42 (0.05)
Non-standard shift	−0.10 (0.07)	−0.12 (0.06)

Notes: The "Non-lattice" column provides estimates from the non-lattice bivariate ordered probit model. The "Lattice" column assumes a lattice structure, estimated jointly. Both specifications include region and wave fixed effects (not reported). Standard errors in parentheses.