Counterfactuals in Discrete Choice Models Under Flexible Assumptions on Random Utility *

PRELIMINARY

Jonas Lieber† Pietro Tebaldi‡ Alexander Torgovitsky§

June 4, 2026

Abstract

Recent research has shown how to characterize sharp identified sets for nonparametric discrete choice models. The methods work by enumerating all sets of the latent component of random utility (the latent valuations) that are relevant for characterizing counterfactuals and rationalizing the data. We show how to solve this enumeration problem at a larger scale by exploiting connections to mechanism design and graph theory. The computational gains afforded by this method allow us to explore sensitivity to the distributional assumptions which are common in parametric models. We also show how to produce faster-to-compute outer sets by using coarser partitions of latent valuations. In simulations, we find that these outer sets are highly informative and often close to the sharp set. The results are illustrated using simulations calibrated to data on hospital choice in Austin, Texas.

*We thank Steve Berry, Christoph Breunig, Joachim Freyberger, Matt Gentzkow, Sukjin Han, Phil Haile, Alessandro Iaria, Simon Lee, Philipp Ketz, Lars Nesheim, Aviv Nevo, Claudia Noack, Vladislav Morozov, Taisuke Otsu, Marcia Schafgans, Pietro Emilio Spini, and Yike Wang for insightful conversations.

†Department of Economics and Public Policy, Imperial College London. j.lieber@ic.ac.uk.

‡Department of Economics, Columbia University and NBER. p.tebaldi@columbia.edu

§Kenneth C. Griffin Department of Economics, University of Chicago. torgovitsky@uchicago.edu

1 Introduction

Discrete choice models are widely used in many areas of economics. Common empirical implementations of these models rely on parametric distributional assumptions that may be controversial. Relaxing these parametric assumptions raises the question of identification. One strand of literature has answered this question by providing conditions for point identification that require substantial variation in observed characteristics or instruments; see, for example, Thompson (1989), Matzkin (1993), Fox and Gandhi (2016), Berry and Haile (2014, 2024), and Borusyak, Chen, Hull, and Lei (2026). Another strand has focused on characterizing identified sets without placing restrictions on the variation of observed characteristics; see, for example, Manski (2007), Tebaldi, Torgovitsky, and Yang (2023), Pakes and Porter (2024), Gu, Russell, and Stringham (2025), and Kamat and Norris (2025). Both strands have had to contend with dimensionality challenges in applications; see Compiani (2022) for an application with point identification and Tebaldi, Torgovitsky, and Yang (2023) for an application with partial identification.

In this paper, we leverage connections between mechanism design and graph theory to provide new algorithms for implementing the partial identification approach at larger scales for a class of quasilinear discrete choice models with exogenous characteristics. We also propose a way to bound the density of the unobserved component of utility—the latent valuations—relative to a researcher-specified reference density. This facilitates a sensitivity analysis on the extent to which common parametric assumptions (e.g. logit) drive empirical conclusions about counterfactuals. In simulations based on administrative data on hospital choice in Austin, Texas—with seven products and several thousands of observations—we obtain tight and informative identified sets on diversion ratios between hospitals, and study sensitivity to logit-like assumptions.

Section 2 defines the class of nonparametric discrete choice models that we analyze in this paper. We consider the assumption that the unknown density of valuations is within a specified distance from a *reference density* that has particular salience to the researcher, such as an i.i.d. type I extreme value distribution. The same idea has been used for partial identification by Christensen and Connault (2023) and Gu and Russell (2024), who employed the Kullback-Leibler divergence and Wasserstein distance, respectively. We instead use the idea of a *density bounded class* from the robust Bayes literature (for example, Lavine, 1991; Wasserman and Kadane, 1992), which preserves the linear-programming characterization of the identified set developed in Tebaldi, Torgovitsky, and Yang (2023). When the resulting identified set is empty, we report the *pseudo-true identified set* (Kaido and Molinari, 2024), which generalizes the idea of a pseudo-true parameter from point-identified to partially identified models.

Because the density bounded class is indexed by a single scalar κ that interpolates between the reference density and the fully nonparametric model, it naturally produces a scalar measure of robustness to parametric assumptions, which we call the *nonparametric robustness criterion* (NRC). This is a number between 0 and 1 that measures how much an empirical claim made under the reference density relies on its specific form. If the NRC is one, then the claim is nonparametric; the same claim would be obtained without restricting the density of valuations. If the NRC is zero, then even small departures from the reference density overturn the claim. The NRC can therefore be used to gauge the extent to which empirical claim from a logit or probit model (for example) rely on the specific functional forms imposed by these models.

As in Tebaldi, Torgovitsky, and Yang (2023), our characterization of the identified set relies on a finite partition of the space of latent valuations called the *minimal relevant partition* (MRP). Individuals with valuations in the same set of the MRP make the same choices under all choice characteristics observed in the data and all choice characteristics relevant for the researcher's counterfactual of interest. The size of the MRP grows quickly as the number of choices and/or observations increases; we prove that, generically, with $J + 1$ choices and T markets the MRP contains $\binom{J+T}{T}$ sets. This makes it increasingly difficult to compute the MRP using existing algorithms.

In Section 3, we provide a new algorithm for computing the MRP that substantially reduces this difficulty. The algorithm leverages a duality for quasilinear models between rationalizability and implementable mechanisms, which converts constructing the MRP to the problem of checking for cyclical monotonicity among all possible price cycles (Rochet, 1987). However, the number of price cycles that need to be checked still grows quickly with the number of choices and observations. To simplify the problem further, we recast it as a node-labeling problem in graph theory as in Vohra (2011). This allows us to use the Bellman-Ford algorithm (Ford Jr, 1956; Bellman, 1958) to check for cyclical monotonicity among price cycles. The resulting algorithm is several orders of magnitude faster than the ones developed in Tebaldi, Torgovitsky, and Yang (2023) and Gu, Russell, and Stringham (2025).

The new algorithm moves the bottleneck from computing the MRP to solving the linear programs that characterize sharp identified sets for counterfactuals. Some of the MRPs we compute have hundreds of millions of sets, which correspond to consumer types making choices consistent with the discrete choice model. Our new algorithm can enumerate these MRPs relatively easily, but the resulting linear programs have an equal number of variables, so are often prohibitively difficult to solve.

Section 4 addresses this problem by developing strategies for constructing informative outer (non-sharp) identified sets based on coarser versions of the MRP. The first strategy builds the MRP with only prices that preserve the "shape" of the target parameter, while still harnessing information from other prices by introducing a set of inequalities between the observed and model-implied choice shares. The resulting linear programs are smaller, but often yield outer identified sets that are nearly identical to the sharp identified set. A second strategy scales further using smaller subsets of prices to build the MRP, while still including a set of inequalities that includes all observed choices. We show that, in settings where the sharp set is feasible to compute, these subset outer sets are typically close to it. The subset outer sets remain feasible to compute even in large applications and seem to often lead to tight identified sets as the number of observations increases by orders of magnitude.

In Section 5, we illustrate our results using a DGP calibrated to administrative hospital records from Austin, Texas. This setting has seven products, thousands of markets, and realistic variation in prices and mean utilities. We find that our subset outer sets are informative and tight enough to imply a narrow range for diversion ratios between hospitals, and to rule out the substitution patterns implied by a misspecified logit model.

Section 6 provides some brief concluding remarks and directions for further research. In particular, we highlight how the computational improvements made in the current paper pave the way for extensions to handle endogenous prices.

2 Model and Identification

In this section, we define the quasilinear discrete choice model that we analyze throughout the paper. We introduce our reference density assumption, then show how to construct sharp identified sets under this assumption.

2.1 Model

An individual i chooses option Y_i from a set $\mathcal{J} \equiv \{0, 1, \ldots, J\}$ of $J + 1$ choices. Each choice is characterized by an observable price $P_{i,j}$ that may or may not vary for each individual i. This "price" could be a linear index constructed from several observable characteristics including price; see Tebaldi, Torgovitsky, and Yang (2023, Section S.2).

We use bold font $\boldsymbol{P}_i = (P_{i,0}, P_{i,1}, \ldots, P_{i,J})$ to denote the vector of prices and assume that the support of $\boldsymbol{P}_i$ is a finite set $\{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_T\}$.

Each individual i has quasilinear preferences characterized by a latent vector of valuations $\boldsymbol{V}_i \equiv (V_{i,0}, V_{i,1}, \ldots, V_{i,J})$. Their choice Y_i is given by

$$Y_i = Y(\boldsymbol{V}_i, \boldsymbol{P}_i) \equiv \underset{j \in \mathcal{J}}{\arg\max}\, V_{i,j} - P_{i,j}\ . \tag{QL}$$

We adopt the standard normalizations $P_{i,0} = V_{i,0} = 0$ for choice $j = 0$, which we view as the outside option of not choosing any of $j = 1, \ldots, J$.

The primitive object of interest in (QL) is the distribution of valuations $\boldsymbol{V}_i$ conditional on prices $\boldsymbol{P}_i$. This conditional distribution is assumed to admit a conditional density $f(\boldsymbol{v}|\boldsymbol{p})$, which ensures that consumers are indifferent between choices with probability zero. For simplicity, we assume throughout that prices are exogenous, so that $\boldsymbol{P}_i$ and $\boldsymbol{V}_i$ are independent.

Assumption EP. (Exogenous prices) $f(\boldsymbol{v}|\boldsymbol{p}) = f(\boldsymbol{v})$ for all $\boldsymbol{p}$ and $\boldsymbol{v}$.

Relaxing Assumption EP and incorporating an instrument doesn't fundamentally change the following analysis (see Tebaldi, Torgovitsky, and Yang, 2023, Assumption IV and Section S.2).

2.2 The Reference Density

Let $\mathcal{F}$ denote a set of densities that the researcher wants to restrict f to be a member of. Common parametric discrete choice models assume that $\mathcal{F}$ is a particular parametric family, typically some sort of generalized extreme value distribution, a multivariate normal distribution, or some mixture of the two. At the other end, a fully nonparametric model allows $\mathcal{F}$ to include densities that are highly irregular and concentrate mass in extreme ways. In Section 2.4 ahead, we show that the sharp bounds in these nonparametric models can be attained under densities that may seem improbable a priori.

One goal of our paper is to systematically explore the middle ground between the two extremes of a tightly parameterized parametric model and a fully unrestricted nonparametric model. We do this by using the following assumption.

Assumption RD. (Reference density) Let g be a known *reference density* of valuations and let $\kappa \geq 0$ be chosen by the researcher. Let $\mathcal{F}(\kappa)$ denote the set of densities for which the following is true:

$$(1-\kappa)g(\boldsymbol{v}) \leq f(\boldsymbol{v}) \leq (1+\kappa)g(\boldsymbol{v}), \text{ for almost all } \boldsymbol{v} \in \mathbb{R}^J. \quad \text{(RD)}$$

Then $\mathcal{F} = \mathcal{F}(\kappa)$.

Assumption RD defines a parameter κ that controls the maximum pointwise difference that a density of valuations f can have from a fixed reference density g. We specify this difference in terms of the difference between the ratio $f(v)/g(v)$ and one. The same construction is used in the robust Bayes literature, where g is set to be some baseline prior and the resulting set of densities f defined through Assumption RD is called a *density bounded class* (for example, Lavine, 1991; Wasserman and Kadane, 1992).[1]

In practice, we take g to be a density estimated from a baseline parametric model, which we keep fixed while varying κ. Setting $\kappa = 0$ then collapses our model to that baseline parametric model: $\mathcal{F}(0) = \{g\}$. If g has full support, then taking $\kappa = \infty$ recovers the nonparametric model $\mathcal{F}(\infty)$ in which any density f is allowed. Christensen and Connault (2023) and Gu and Russell (2024) used similar assumptions for partial

[1]Alternatively and equivalently, Assumption RD restricts f to lie within a g-relative (essential) sup-norm ball of radius κ: $\sup_v |f(v)/g(v)-1| \leq \kappa$. A third interpretation is in terms of the Rényi ∞-divergence between f and g, defined as $D_\infty(f\|g) \equiv \log \sup_v f(v)/g(v)$ (see, for example, van Erven and Harremoes, 2014). Assumption RD restricts attention to f for which $D_\infty(f\|g) \leq \log(1+\kappa)$ and $D_\infty(g\|f) \leq -\log(1-\kappa)$.

identification, except they defined discrepancies with the baseline measure using the Kullback-Leibler divergence and Wasserstein distance, respectively. Our primary motivation for using the density bounded class in (RD) is that it is straightforward to interpret and convenient for computation at scale. As we show below, it is also a natural way to prevent the type of extreme "spiked" densities that characterize nonparametric bounds. A similar motivation is used in the robust Bayes literature to avoid considering priors (f in that context) which have densities equal to zero or infinity (Lavine, 1991).

Assumption RD serves two purposes for our analysis. The first is to tighten bounds by disciplining extreme behavior that is allowed in a fully nonparametric analysis ($\kappa = \infty$). The second is to give researchers a way to assess the sensitivity of parametric estimates to small deviations from the parametric assumptions ($\kappa \approx 0$). Changing κ allows one to smoothly trace out this range.

2.3 The Identified Set

Suppose that the researcher observes choice shares $s_j(\boldsymbol{p}) \equiv \mathbb{P}[Y_i = j | \boldsymbol{P}_i = \boldsymbol{p}]$ for every $j \in \mathcal{J}$ and $\boldsymbol{p} \in \{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_T\}$. If the density of valuations were f, then the choice shares implied by model (QL) would be

$$\sigma_j(\boldsymbol{p}; f) \equiv \int_{\mathcal{V}_j(\boldsymbol{p})} f(\boldsymbol{v})\, d\boldsymbol{v}, \quad \text{where } \mathcal{V}_j(\boldsymbol{p}) \equiv \left\{\boldsymbol{v} \in \mathbb{R}^J : v_j - p_j \geq v_k - p_k \ \text{ for all } k\right\}. \tag{1}$$

We say that a density f is consistent with the observed choice shares ("matches the data") if it satisfies

$$s_j(\boldsymbol{p}_t) = \sigma_j(\boldsymbol{p}_t; f) \quad \text{for all } j \in \mathcal{J} \text{ and } t = 1, \ldots, T. \tag{MD}$$

The identified set of f is then defined as the subset of $\mathcal{F}(\kappa)$ that matches the data:

$$\mathcal{F}^\star(\kappa) \equiv \{f \in \mathcal{F}(\kappa) : f \text{ satisfies (MD)}\}.$$

The researcher is interested in a lower-dimensional target parameter $\theta(f)$ taking values in $\mathbb{R}^{d_\theta}$, usually with $d_\theta = 1$. This target parameter often describes an aspect of a counterfactual, such as a choice probability, semielasticity, or diversion ratio. The identified set for the target parameter is the image of $\mathcal{F}^\star(\kappa)$ under θ, which we denote as

$$\Theta^\star(\kappa) \equiv \{\theta(f) : f \in \mathcal{F}^\star(\kappa)\}.$$

Our goal is to characterize and compute $\Theta^\star(\kappa)$. To make this problem feasible, we need to define a set of *relevant prices* $\mathcal{P}$ that is large enough to evaluate both the target parameter and (MD). In particular, $\mathcal{P}$ contains $\{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_T\}$.

Tebaldi, Torgovitsky, and Yang (2023, TTY in the following) showed how to compute $\Theta^\star(\infty)$ when g has full support. We refer to this nonparametric identified set as $\Theta^\star$ in the following. Their argument used the concept of the *minimal relevant partition* of valuations induced by the relevant prices $\mathcal{P}$, which is a collection of subsets of $\mathbb{R}^J$ that we denote by $\text{MRP}(\mathcal{P})$. A subset $\mathcal{V} \subseteq \mathbb{R}^J$ is an element of $\text{MRP}(\mathcal{P})$ if the following property holds for (Lebesgue) almost every $\boldsymbol{v}, \boldsymbol{v}' \in \mathbb{R}^J$:

$$\boldsymbol{v}, \boldsymbol{v}' \in \mathcal{V} \iff Y(\boldsymbol{v}, \boldsymbol{p}) = Y(\boldsymbol{v}', \boldsymbol{p}) \quad \text{for all } \boldsymbol{p} \in \mathcal{P}. \tag{MRP}$$

In words, $\text{MRP}(\mathcal{P})$ is the coarsest partition of $\mathbb{R}^J$ into sets such that all valuations in the same set imply identical choices for all prices in $\mathcal{P}$. The definition of $\text{MRP}(\mathcal{P})$ is unique up to some technical caveats about measure zero boundaries (see Section S4 of TTY).[2]

[2]Throughout the paper, set operations on the space of valuations are interpreted modulo Lebesgue

TTY observed that because $\text{MRP}(\mathcal{P})$ partitions valuations into sets that produce identical choice behavior under all relevant prices, one can replace the density f by a probability mass function ϕ defined on $\text{MRP}(\mathcal{P})$. Let Φ denote the set of all functions $\phi : \text{MRP}(\mathcal{P}) \to [0,1]$ that could represent a probability mass function over the elements of $\text{MRP}(\mathcal{P})$, that is, such that $\sum_{\mathcal{V} \in \text{MRP}(\mathcal{P})} \phi(\mathcal{V}) = 1$. (Alternatively, Φ is the simplex with dimension equal to the number of elements in $\text{MRP}(\mathcal{P})$.) Then each density $f \in \mathcal{F}(\infty)$ induces a mass function $\phi^f \in \Phi$ via $\phi^f(\mathcal{V}) \equiv \int_{\mathcal{V}} f(\boldsymbol{v})\, d\boldsymbol{v}$. We assume that the target parameter can be evaluated with only knowledge of this mass function. We let $d_\phi \equiv |\text{MRP}(\mathcal{P})|$ and write $\Phi^{\text{e}} \equiv \mathbb{R}^{d_\phi}$ for the space of all real-valued functions on $\text{MRP}(\mathcal{P})$, dropping the requirement that they sum to one, so that $\Phi \subset \Phi^{\text{e}}$.

Assumption TP. (Target parameter) There exists a known function $t : \Phi^{\text{e}} \to \mathbb{R}^{d_\theta}$ such that $\theta(f) = t(\phi^f)$ for every $f \in \mathcal{F}(\infty)$.

Assumption TP is fairly innocuous because $\mathcal{P}$ can be chosen flexibly as long as it contains the observed prices $\{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_T\}$.

The definitions of the relevant prices $\mathcal{P}$ and minimal relevant partition $\text{MRP}(\mathcal{P})$ ensure that any mass function $\phi \in \Phi$ produces a set of choice shares for each $\boldsymbol{p} \in \text{supp}(\boldsymbol{P})$. Then the choice shares induced by a $\phi \in \Phi$ are

$$\mathscr{s}_j(\boldsymbol{p}; \phi) \equiv \sum_{\mathcal{V} \in \text{MRP}(\mathcal{P})} \mathbf{1}\left[\mathcal{V} \subseteq \mathcal{V}_j(\boldsymbol{p})\right] \phi(\mathcal{V}).$$

TTY prove that $\Theta^\star$ can be characterized as

$$\Theta^\star = \left\{ \vartheta \in \mathbb{R}^{d_\theta} : \vartheta = t(\phi) \quad \text{for some } \phi \in \Phi \text{ s.t. } \mathscr{s}_j(\boldsymbol{p}_t; \phi) = s_j(\boldsymbol{p}_t) \text{ for all } j \text{ and } t \right\},$$

a set which can typically be computed by solving two linear programs, one which min-

null sets. That is, if Leb denotes Lebesgue measure on $\mathbb{R}^J$, then $\mathcal{V} \subseteq \mathcal{V}'$ means $\text{Leb}(\mathcal{V} \setminus \mathcal{V}') = 0$, $\mathcal{V} \cap \mathcal{V}' \neq \emptyset$ means $\text{Leb}(\mathcal{V} \cap \mathcal{V}') > 0$, and set equality means symmetric a.e. inclusion.

imizes $t(\phi)$ and another which maximizes $t(\phi)$, both subject to the constraints that the implied shares $\mathscr{s}_j(\boldsymbol{p}_t; \phi)$ are equal to the observed shares $s_j(\boldsymbol{p}_t)$.

Extending this characterization to finite κ requires first observing that if $f \in \mathcal{F}(\kappa)$, then the implied mass function ϕ^f must lie in the set

$$\Phi(\kappa) \equiv \left\{ \phi \in \Phi : (1-\kappa) \int_{\mathcal{V}} g(\boldsymbol{v})\, d\boldsymbol{v} \leq \phi(\mathcal{V}) \leq (1+\kappa) \int_{\mathcal{V}} g(\boldsymbol{v})\, d\boldsymbol{v} \quad \text{for all } \mathcal{V} \in \text{MRP}(\mathcal{P}) \right\}.$$

This suggests that the sharp identified set for finite κ is simply

$$\Theta^\star(\kappa) = \left\{ \vartheta \in \mathbb{R}^{d_\theta} : \vartheta = t(\phi) \quad \text{for some } \phi \in \Phi(\kappa) \text{ s.t. } \mathscr{s}_j(\boldsymbol{p}_t; \phi) = s_j(\boldsymbol{p}_t) \text{ for all } j \text{ and } t \right\},$$

a statement which we establish below. This characterization of $\Theta^\star(\kappa)$ nests the case when $\kappa = \infty$ and g has full support, so that $\Phi(\kappa) = \Phi$. It shows that $\Theta^\star(\kappa)$ can be computed by solving two linear programs as long as t is a linear function of ϕ and $d_\theta = 1$.

It is possible for $\Theta^\star(\kappa)$ to be empty when the reference density differs from the true density that generated the shares, even maintaining the assumption that these shares were generated by (QL). Instead of reporting empty sets, we consider the notion of a *pseudo-true identified set*. The pseudo-true identified set for the target parameter is defined as

$$\Theta^+(\kappa) \equiv \left\{ \vartheta \in \mathbb{R}^{d_\theta} : \vartheta = t(\phi) \quad \text{for some } \phi \in \operatorname*{arg\,min}_{\phi' \in \Phi(\kappa)} Q(\phi') \right\}, \tag{2}$$

where Q is a criterion function that measures the extent to which a given mass function ϕ fails to match the observed choice shares. Kaido and Molinari (2024) considered pseudo-true identified sets in the context of partially identified likelihood-based models. Our definition (2) is the natural generalization to identified sets described by more general criterion functions.

We specify Q using an ℓ_1 distance:

$$Q(\phi) \equiv \sum_{t=1}^{T} \sum_{j \in \mathcal{J}} \left| \mathfrak{s}_j(\boldsymbol{p}_t; \phi) - s_j(\boldsymbol{p}_t) \right| . \tag{3}$$

If t is a continuous, scalar-valued function, then $\Theta^+(\kappa)$ is an interval with endpoints given by the optimal values of the programs

$$\min_{\phi \in \Phi(\kappa)} / \max \ t(\phi) \quad \text{s.t.} \quad Q(\phi) = \min_{\phi' \in \Phi(\kappa)} Q(\phi'), \tag{4}$$

which can be reformulated as linear programs if t is a linear function of ϕ. If the minimum value of $Q(\phi)$ over $\Phi(\kappa)$ is zero, then $\Theta^+(\kappa)$ is equal to $\Theta^\star(\kappa)$. The implication is that (4) provides a computationally tractable characterization of the (sharp) identified set $\Theta^\star(\kappa)$ when this set is non-empty and of the pseudo-true identified set $\Theta^+(\kappa)$ when $\Theta^\star(\kappa)$ is empty. The following proposition proves these claims.

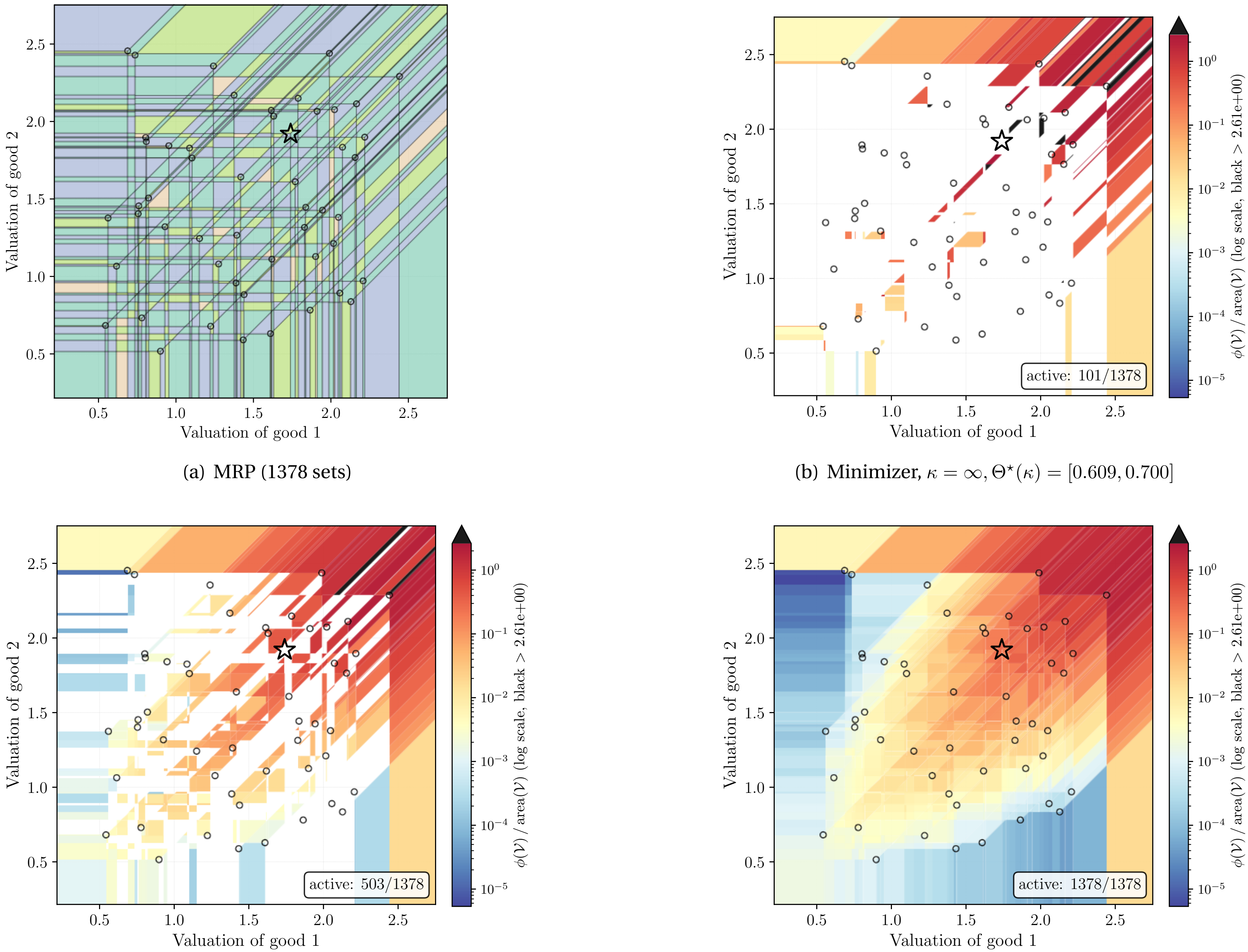
Proposition 1. Suppose that Assumptions EP, RD, and TP are satisfied. If $\min_{\phi' \in \Phi(\kappa)} Q(\phi') = 0$, then $\Theta^+(\kappa) = \Theta^\star(\kappa)$. If t is continuous and scalar ($d_\theta = 1$), then $\Theta^+(\kappa)$ is an interval with endpoints given by the optimal values of (4).

Proof. See Appendix D. □

2.4 Example

We illustrate the impact of the reference density (Assumption RD) through a small-scale example. Suppose that $J = 2$, so that there are two inside goods. Suppose that the data is generated by consumers with logit valuations $V_{i,j} = .25(\delta_j + \varepsilon_{i,j} - \varepsilon_{i,0})$, where $\delta_j = 40$ for both $j = 1, 2$, and $\varepsilon_{i,j}$ are i.i.d. type I extreme value. We generate $T = 50$ markets, drawing the prices $\boldsymbol{p}_t$ from a uniform distribution over the square $[0.5, 2.5] \times [0.5, 2.5]$. These prices are shown in Figure 1 as hollow circles.

Figure 1: The Impact of Assumption RD



(a) MRP (1378 sets)

(b) Minimizer, $\kappa = \infty, \Theta^\star(\kappa) = [0.609, 0.700]$

(c) Minimizer, $\kappa = 1.5, \Theta^\star(\kappa) = [0.619, 0.693]$

(d) Minimizer, $\kappa = 0.25, \Theta^\star(\kappa) = [0.644, 0.660]$

Suppose that the researcher's target parameter of interest is the share of good one at the counterfactual price $\boldsymbol{p}^\star = [1.74, 1.92]$, represented with a star in Figure 1. The relevant price set $\mathcal{P}$ consists of the 50 market prices together with $\boldsymbol{p}^\star$, for a total of 51 prices. The minimal relevant partition $\text{MRP}(\mathcal{P})$ is shown in panel (a). It consists of 1378 sets, many of which are evidently quite "thin" as subsets of $\mathbb{R}^2$.

Panel (b) of Figure 1 is a heat map that depicts a solution to (4) in the nonparametric case when $\kappa = \infty$. The solution places no mass on the majority of the 1378 sets in the MRP; only 101 sets receive any mass at all. These 101 sets are scattered across $\mathbb{R}^2$ in a disconnected way, implying that it can only be produced by a density with a disconnected support. Of the few MRP sets that do have positive mass, some have a large amount of mass concentrated on a small area, while others have a small amount of mass spread out over a large area.

Panel (c) of Figure 1 shows the same heat map when $\kappa = 1.5$ with the reference density g set equal to the true logit distribution used to generate the choice shares. Lowering κ rules out the solution in panel (b), shrinking the identified set from $\Theta^\star(\infty) = [0.609, 0.700]$ to $\Theta^\star(1.5) = [0.619, 0.693]$. In this example, it does not lead to a pseudo-true identified set, because g was used to generate the choice shares. Compared to the minimizer in panel (b), the minimizer in panel (c) distributes mass over a larger number of sets (637), although it still implies a density with many gaps in its support and extreme peaks over some small sets.

In Panel (d) we decrease κ further to 0.25, which smoothes out the gaps and peaks, leading to a bound of $[0.644, 0.660]$ that is tight around the true value of 0.652. Because $\kappa < 1$ and g has full support, all sets in the MRP must have non-zero mass for any feasible $\phi \in \Phi(\kappa)$, including the minimizer shown in panel (d). The mass is closer to the baseline logit, but still different, as evidenced by the non-negligible difference between the lower and upper bounds. We expect that valuation densities like those in panel (d) are more plausible to economists than those in panels (b) and (c). They capture a notion of smoothness over preferences which seems intuitively appealing.

2.5 A Nonparametric Robustness Criterion

The example in the previous section shows how Assumption RD can be used to discipline extreme behavior that is allowed for in fully nonparametric analyses. Another use of Assumption RD is as a way to measure the sensitivity of empirical conclusions to deviations from parametric assumptions. In this section, we develop and illustrate a scalar measure of this sensitivity that we call the *nonparametric robustness criterion* (NRC).

The first ingredient to the NRC is an empirical claim about the target parameter, which we describe using the set Θ^0. We think of the claim as being supported by the model with a given κ if $\Theta^+(\kappa) \subseteq \Theta^0$ and being unsupported otherwise. For example, if $\theta(f)$ is a diversion ratio for a given discrete price change, then setting $\Theta^0 = \{\vartheta : \vartheta \leq .20\}$ would reflect the claim that this diversion ratio is smaller than $.20$. We can then define the largest κ under which the claim Θ^0 is supported as

$$\underline{\kappa}(\Theta^0) \equiv \sup \left\{ \kappa : \Theta^+(\kappa) \subseteq \Theta^0 \right\}. \tag{5}$$

This quantity can be thought of as the (identification) *breakdown point* of the empirical claim Θ^0 with respect to the parametric assumption represented by the reference density g (Horowitz and Manski, 1995; Kline and Santos, 2013; Masten and Poirier, 2020).

There are two reasons that $\underline{\kappa}(\Theta^0)$ could be large. One is that the pseudo-true identified set $\Theta^+(\kappa)$ doesn't change much as κ increases. Another is that the claim Θ^0 is so conservative that it remains satisfied even while $\Theta^+(\kappa)$ changes rapidly; for example if we were to change the diversion ratio claim to instead be $\Theta^0 = \{\vartheta : \vartheta \leq .80\}$. To separate these two explanations, we measure $\underline{\kappa}(\Theta^0)$ relative to the smallest value of κ at which

$\Theta^+(\kappa) = \Theta^\star$, that is, the smallest value of κ that leads to the nonparametric identified set. We define this quantity as

$$\overline{\kappa} \equiv \inf\left\{\kappa : \Theta^+(\kappa) = \Theta^\star\right\}. \tag{6}$$

Our NRC measure combines (5) and (6) as follows.

Definition NRC. The *nonparametric robustness criterion* of a claim Θ^0 is defined as

$$\text{NRC}(\Theta^0) \equiv \min\left\{\underline{\kappa}(\Theta^0)/\overline{\kappa}, 1\right\},$$

with the conventions that $0/0 = 0$ and $\infty/\infty = 1$.

If $\text{NRC}(\Theta^0) \approx 0$, then the claim Θ^0 is fully driven by the choice of reference density g in the sense that valuation densities close to g overturn the claim. One case in which this would typically happen is if Θ^0 is a point statement, such as $\Theta^0 = \{.2\}$, because changing κ even a small amount would lead to a pseudo-true identified set that consists of points other than $.2$. Alternatively, if $\text{NRC}(\Theta^0) = 1$, then Θ^0 is nonparametrically robust, because the nonparametric identified set $\Theta^\star$ is contained in Θ^0. Note that this could happen either when $\overline{\kappa}$ is large—indicating that the nonparametric bounds are attained at densities quite different from g—or when $\overline{\kappa}$ is small; as long as the conclusion Θ^0 remains satisfied at $\overline{\kappa}$. Larger values of $\text{NRC}(\Theta^0)$ indicate conclusions Θ^0 that are more nonparametrically robust to g because they require moving κ closer to its nonparametric boundary $\overline{\kappa}$ before they are overturned.

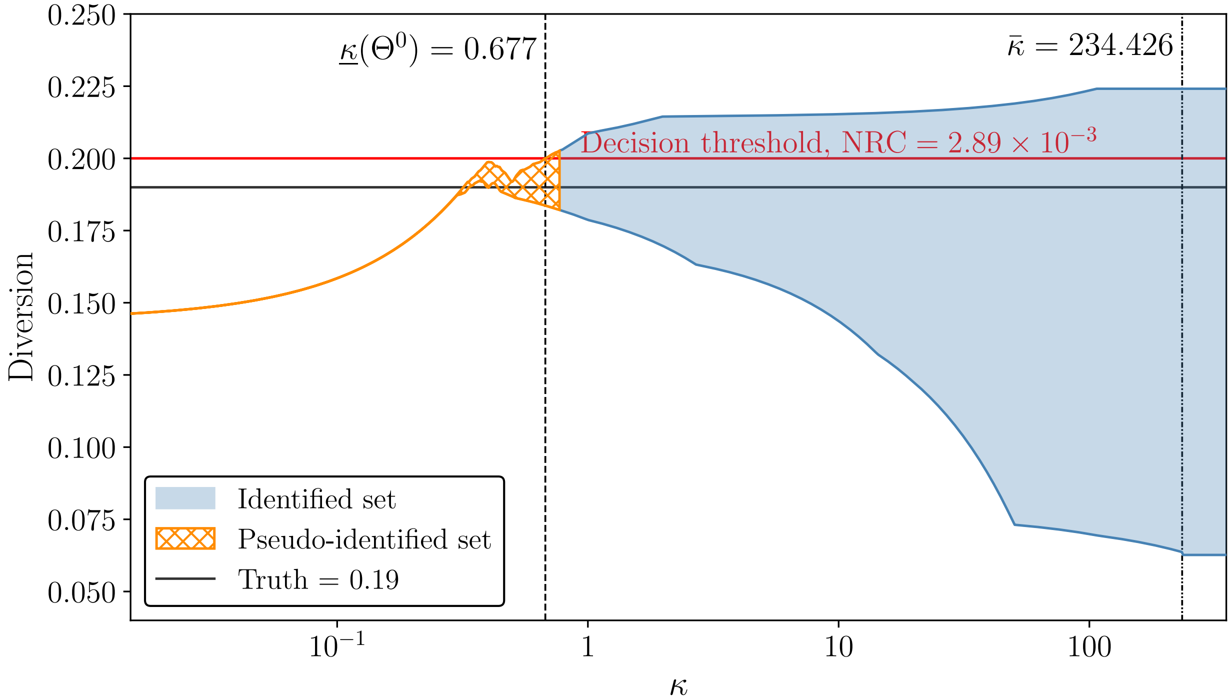
We illustrate the NRC in a case with $J = 2$ inside goods. We draw 100 price vectors i.i.d. from a uniform distribution over the square $[2.0, 2.5]^2$. We set the target parameter to be the proportion of buyers who choose $j = 2$ when $\boldsymbol{p}^0 = [2.4, 2.4]$, but who would switch to $j = 1$ when $\boldsymbol{p}^\star = [2.4, 2.5]$—that is, $\mathbb{P}[Y_i(\boldsymbol{V}_i, \boldsymbol{p}^\star) = 1 | Y_i(\boldsymbol{V}_i, \boldsymbol{p}^0) = 2]$, which can be seen as a measure of diversion. We take the claim to be $\Theta^0 = \{\vartheta : \vartheta \leq 0.2\}$—fewer than 20% of consumers of good one switch to good two—representing the statement that competition is somewhat weak.

In panel (a) of Figure 2, we generate choice shares from a distribution of valuations that follows a two-type mixed logit with $V_{i,j} = \sigma_i(\delta_{i,j} + \varepsilon_{i,j} - \varepsilon_{i,0})$, where 70% of consumers i have $\sigma_i = 1/6$ and $\delta_{i,1} = \delta_{i,2} = 96$, while the remaining 30% of consumers have $\sigma_i = 50$ and $\delta_{i,1} = \delta_{i,2} = 0.01$. This produces a true diversion measure of 0.19. The reference density g is set to be the standard (not mixed) logit estimated from the resulting shares, which is a misspecified model. The point estimate of the diversion measure using g is $\Theta^+(0) = \{0.15\}$, under which Θ^0 is true.

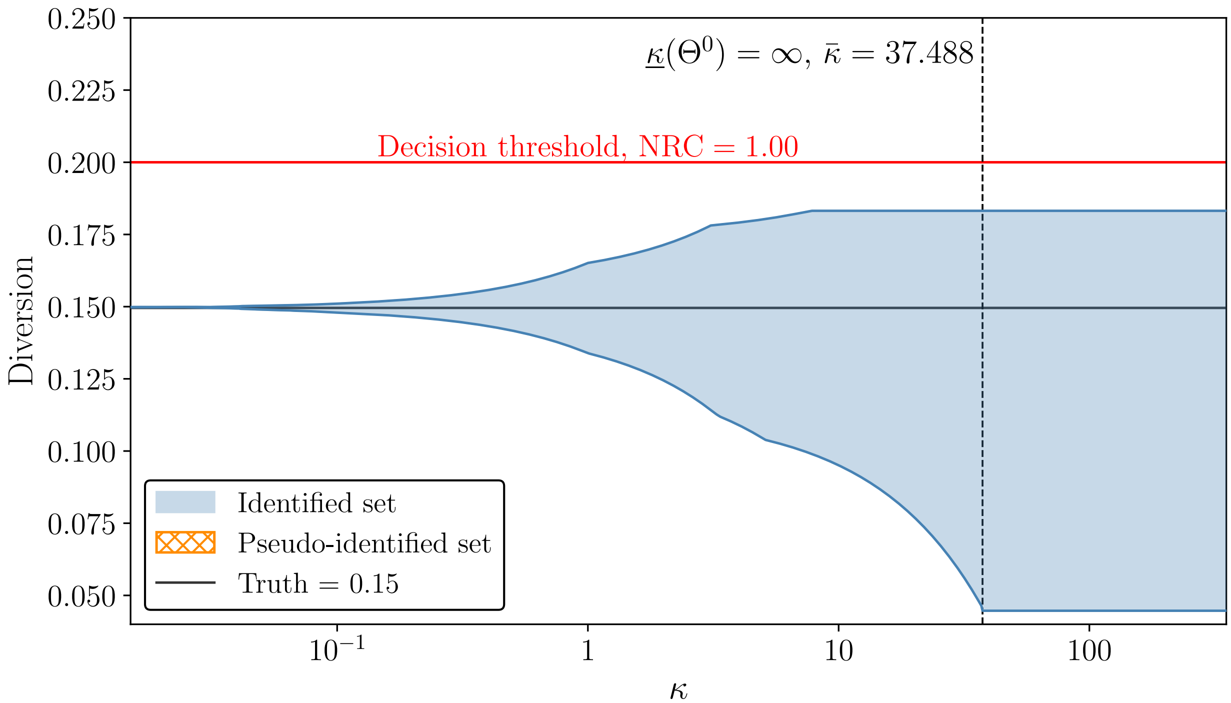
As we increase κ, the pseudo-true identified set remains a singleton but its singleton value increases steadily towards the true value. Eventually, $\Theta^+(\kappa)$ becomes a bona fide set, while still being contained in Θ^0. Once $\kappa \geq \underline{\kappa}(\Theta^0) \approx 0.677$, the pseudo-true identified set contains values larger than 0.20, so that Θ^0 is no longer true. Shortly after that, the pseudo-true identified set becomes a proper identified set (shown in blue), because it is possible to find densities that match all of the choice shares exactly. The identified set continues to expand as κ increases until $\overline{\kappa} \approx 234.426$, at which point it becomes equal to the nonparametric identified set $\Theta^\star$. The NRC is 0.003, indicating that the claim is not very robust to the (incorrect) assumption that the distribution of valuations follows a standard logit. Changing Θ^0 to be $\{\vartheta : \vartheta \leq 0.25\}$ would lead to a NRC of one.

We repeat this exercise in panel (b), keeping the same price vectors, but rebuilding the observed choice shares from the logit estimates constructed in panel (a): $\sigma_i = 0.267$ and $\delta_{i,j} = 37$, for all i, and $j = 1, 2$. The true value of the diversion measure is now 0.15. The reference density g now matches the data exactly because it generated the data, so the pseudo-true identified set reduces to the sharp identified set, $\Theta^+(\kappa) = \Theta^\star(\kappa)$, for all values of κ. Increasing κ still widens $\Theta^\star(\kappa)$, but even at the widest point of $\overline{\kappa} = 37.488$, its

Figure 2: The Pseudo-True Identified Set and the NRC for a Diversion Measure



(a) DGP with $f \neq g$



(b) DGP with $f = g$

upper bound is smaller than 0.20, so that Θ^0 remains supported. This implies $\underline{\kappa}(\Theta^0) = \infty$, so that $\text{NRC}(\Theta^0) = 1$. The conclusion Θ^0 is fully nonparametrically robust to relaxing the logit assumption. A more aggressive conclusion like $\Theta^0 = \{\vartheta : \vartheta \leq 0.175\}$ would be less robust, with an NRC somewhere strictly between zero and one.

3 An Efficient Algorithm for Constructing the MRP

Computing the identified sets or pseudo-true identified sets requires an algorithm for constructing $\text{MRP}(\mathcal{P})$. Tebaldi, Torgovitsky, and Yang (2023) proposed a divide-and-conquer algorithm. Gu, Russell, and Stringham (2025) developed a more general cell-enumeration algorithm, which they showed how to apply to the type of discrete choice models that we consider. In this section, we develop a new graph-theoretic algorithm for constructing $\text{MRP}(\mathcal{P})$, which we show is several orders of magnitude more efficient than either of the existing algorithms.

3.1 Price Cycles

The goal is to find an efficient procedure that is guaranteed to return $\text{MRP}(\mathcal{P})$ for any $\mathcal{P}$. Each set in $\text{MRP}(\mathcal{P})$ can be viewed as a choice function that is consistent with choices made under the quasilinear model (QL). Our algorithm leverages a duality between rationalizability and mechanism design for quasilinear models that was developed by Rochet (1987) and linked to graph theory in Vohra (2011).

Let $\mathcal{P}$ be represented as a matrix with T rows, each corresponding to a price vector $\boldsymbol{p}_t = (p_{t,1}, \ldots, p_{t,J}) \in \mathbb{R}^J$, $t = 1, \ldots, T$, where we omit the price of the outside option (zero). For a set $\mathcal{V} \subseteq \mathbb{R}^J$ to be an MRP set it must satisfy (MRP), which means that for each $t = 1, \ldots, T$ there is a common value $y_t(\mathcal{V}) \in \mathcal{J}$ such that $Y(\boldsymbol{v}, \boldsymbol{p}_t) = y_t(\mathcal{V})$ for almost every $\boldsymbol{v} \in \mathcal{V}$. The set of candidate MRP sets can therefore be identified with elements in $\mathcal{J}^T$, that is, vectors $\boldsymbol{y}(\mathcal{V}) \in \mathcal{J}^T$ that describe the choice that an individual with valuations $\boldsymbol{v} \in \mathcal{V}$ would make at each price vector $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_T$. Then $\text{MRP}(\mathcal{P})$ contains all such candidate $\mathcal{V}$ for which these choices $\boldsymbol{y}(\mathcal{V})$ are rationalizable under (QL). More formally,

$$\mathcal{V} \in \text{MRP}(\mathcal{P}) \iff \exists\, \boldsymbol{v} \in \mathbb{R}^J : Y(\boldsymbol{v}, \boldsymbol{p}_t) = j \text{ for all } t \in \mathcal{T}_j(\mathcal{V}), \text{ all } j \in \mathcal{J}, \tag{7}$$
$$\text{where } \mathcal{T}_j(\mathcal{V}) \equiv \{t = 1, \ldots, T : y_t(\mathcal{V}) = j\}.$$

That is, $\mathcal{T}_j(\mathcal{V})$ is the set of observed price vectors for which $\mathcal{V}$ would lead to choice j.

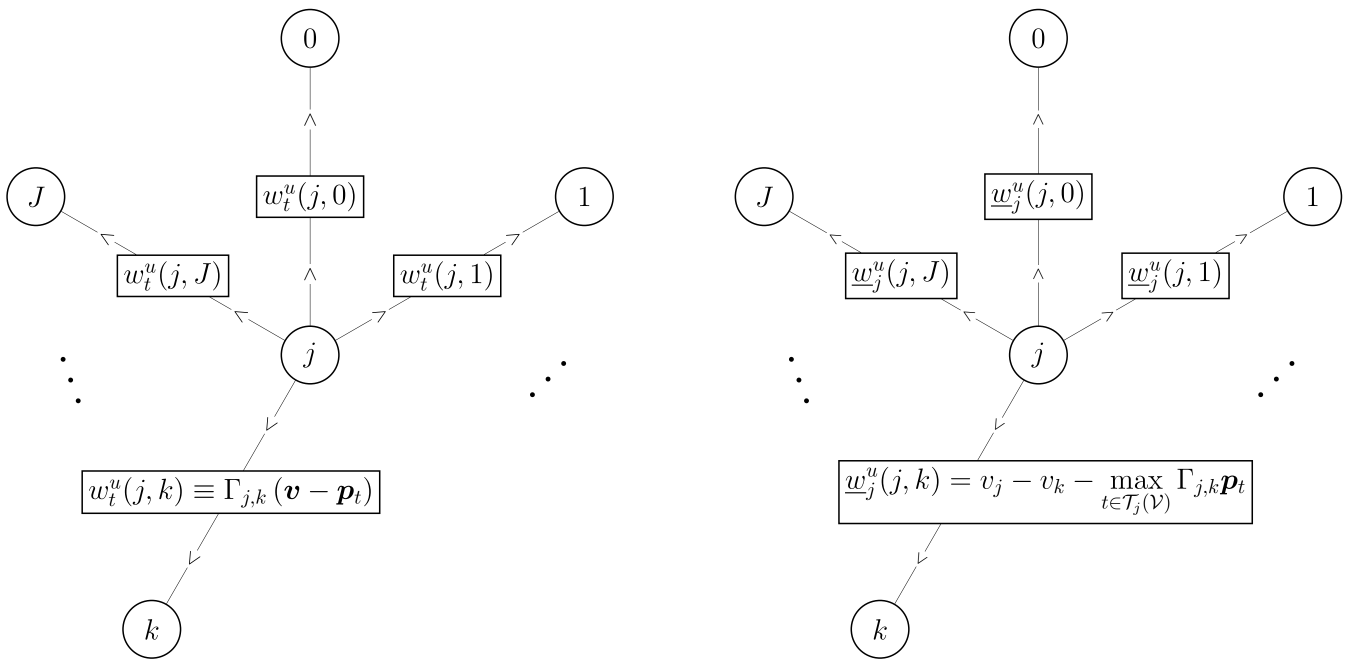
Rochet (1987) highlighted the mathematical equivalence between the rationalizability problem in (7) and a mechanism design problem. The intuition is that $\boldsymbol{v}$ and $-\boldsymbol{p}$ can be thought of as perfect substitutes in (QL). An algorithm that verifies whether $\mathcal{V} \in \text{MRP}(\mathcal{P})$ can be thought of as a mechanism designer who must find a vector of transfers $\boldsymbol{v} \in \mathbb{R}^J$ such that $Y(\boldsymbol{v}, \boldsymbol{p}_t) = y_t(\mathcal{V})$ for every type t. The goal of the mechanism is to implement the outcome $y_t(\mathcal{V})$ for every type t. Transfers can only depend on the realization of the outcome, and they cannot vary across (private) types. If such transfers exist, then $\boldsymbol{y}(\mathcal{V})$ is implementable given $\mathcal{P}$, $\boldsymbol{y}(\mathcal{V})$ is rationalizable given $\mathcal{P}$, and $\mathcal{V} \in \text{MRP}(\mathcal{P})$.

Rochet (1987) showed that a necessary and sufficient condition for implementation can be stated in terms of cyclical monotonicity. In particular, his result implies that $\mathcal{V} \in \text{MRP}(\mathcal{P})$ if and only if

$$\sum_{\ell=1}^{n-1} -p_{\ell+1, y_\ell(\mathcal{V})} + p_{\ell, y_\ell(\mathcal{V})} \leq 0$$

for all finite cycles $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n$ with $\boldsymbol{p}_1 = \boldsymbol{p}_n$ that can be constructed from prices in $\mathcal{P}$. This makes the problem of checking if $\mathcal{V} \in \text{MRP}(\mathcal{P})$ a finite one of evaluating price cycles. However, the number of possible cycles is $\frac{T(T-1)}{2} + \sum_{k=3}^{T} \frac{T!}{2k(T-k)!}$, which grows quickly in T.

Figure 3: Star Graph Representations of "$\mathcal{V}$ Chooses j"



(a) Choose j at $\boldsymbol{p}_t \Leftrightarrow w_t^u(j,k) \geq 0$ for all k

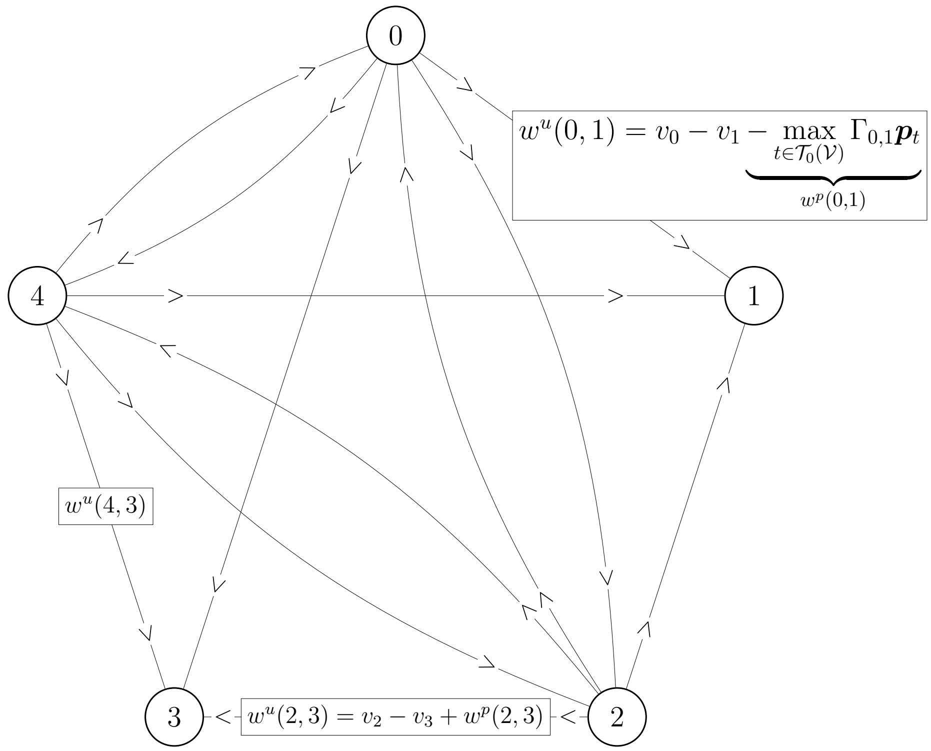
(b) Choose j for all $t \in \mathcal{T}_j(\mathcal{V}) \Leftrightarrow \underline{w}_j^u(j,k) \geq 0$ for all k

3.2 Graph Theory

To simplify the problem further, we turn to a graph-theoretic representation, following Vohra (2011, Section 4.2). The idea is illustrated in Figure 3. Let $\Gamma_{j,k}$ denote a "first-difference" operator such that, for any $\boldsymbol{x} \in \mathbb{R}^J$, $\Gamma_{j,k}\boldsymbol{x} = x_j - x_k$ (cf. Thompson, 1989; Berry and Haile, 2014). Figure 3(a) illustrates how the statement "$\mathcal{V}$ chooses j at $\boldsymbol{p}_t$", or $y_t(\mathcal{V}) = j$, or $t \in \mathcal{T}_j(\mathcal{V})$, can be represented by a weighted directed (star) graph. The nodes of the graph are elements of $\mathcal{J}$. For all $k \neq j$, an edge points at k from j with weight $w_t^u(j,k) \equiv v_j - v_k - (p_{t,j} - p_{t,k}) = \Gamma_{j,k}(\boldsymbol{v} - \boldsymbol{p}_t)$. All such weights must be non-negative.

In order for $\mathcal{V} \in \text{MRP}(\mathcal{P})$ it must be that "$\mathcal{V}$ chooses j at $\boldsymbol{p}_t$ for all $t \in \mathcal{T}_j(\mathcal{V})$." This condition implies that all edges of the star graph in Figure 3(b) are non-negative. Each of these edges, $\underline{w}_j^u(j,k)$, is constructed by taking the minimum over $t \in \mathcal{T}_j(\mathcal{V})$ of $w_t^u(j,k)$:

Figure 4: Example of Directed Weighted Graph to Check "$\mathcal{V} \in \text{MRP}(\mathcal{P})$", $\mathcal{T}_1(\mathcal{V}) = \mathcal{T}_3(\mathcal{V}) = \emptyset$



$\underline{w}^u_j(j,k) \equiv \min_{t \in \mathcal{T}_j(\mathcal{V})} w^u_t(j,k) = v_j - v_k - \max_{t \in \mathcal{T}_j(\mathcal{V})} \Gamma_{j,k} \boldsymbol{p}_t$. Imposing $\underline{w}^u_j(j,k) \geq 0$ requires the difference in valuations $v_j - v_k$ to be at least as large as the largest price difference $p_{t,j} - p_{t,k}$ across the prices at which $\mathcal{V}$ would choose j.

Extending the argument to all $j \in \mathcal{J}$ leads to the conclusion that $\mathcal{V} \in \text{MRP}(\mathcal{P})$ if and only if all edges have non-negative weights in the (complete) graph in which the edge pointing from j to k has weight $w^u(j,k)$ equal to $\underline{w}^u_j(j,k)$. We adopt two conventions for missing edges (which are not represented graphically): first, if j is never chosen by $\mathcal{V}$, so that $\mathcal{T}_j(\mathcal{V}) = \emptyset$, $w^u(j,k) = \infty$ for all $k \in \mathcal{J}$; second, $w^u(j,j) = \infty$ for all $j \in \mathcal{J}$. Figure 4 shows an example of this graph for $J = 4$, considering a situation in which $\mathcal{T}_1(\mathcal{V}) = \mathcal{T}_3(\mathcal{V}) = \emptyset$.

Asking whether $\mathcal{V} \in \text{MRP}(\mathcal{P})$ then amounts to asking whether there exists a "labeling" function $d : \mathcal{J} \to \overline{\mathbb{R}}$ such that $d(j) - d(k) - \max_{t \in \mathcal{T}_j(\mathcal{V})} \Gamma_{j,k} \boldsymbol{p}_t \geq 0$ for all j and all k. If such a function exists, the vector $\boldsymbol{v}$ with $v_j = d(j) - d(0)$ satisfies (7). This is precisely the node labeling problem described in Appendix A, applied to the weighted directed graph with edges' weights $w^p(j,k) \equiv -\max_{t\in\mathcal{T}_j(\mathcal{V})}\Gamma_{j,k}\boldsymbol{p}_t$, with $w^p(j,k)=\infty$ whenever $\mathcal{T}_j(\mathcal{V})=\emptyset$.

In Appendix A, we adapt the Bellman-Ford algorithm (Bellman, 1958; Ford Jr, 1956) to produce Algorithm BF, which quickly checks whether $\mathcal{V}\in \textsc{mrp}(\mathcal{P})$ by ruling out the presence of negative cycles in this graph. Since the graph has $J+1$ nodes regardless of the number of prices T, this problem is significantly simpler and less demanding in terms of time and memory resources than verifying cyclical monotonicity in the space of price vectors.[3]

We can now use these findings to introduce our new algorithm to compute $\textsc{mrp}(\mathcal{P})$. A convenient way to collect (and represent) sets $\mathcal{V}\in \textsc{mrp}(\mathcal{P})$ is through the choices dictated by $\boldsymbol{v}\in\mathcal{V}$. Given a matrix of price vectors $\mathcal{P}\in\mathbb{R}^{T\times J}$, the collection

$$\mathbb{V}(\mathcal{P}) \equiv \left\{\tilde{\boldsymbol{y}}\in\mathcal{J}^T : \exists\mathcal{V}\in\textsc{mrp}(\mathcal{P}) \text{ such that } \tilde{y}_t = y_t(\mathcal{V}) \text{ for all } t=1,2,\dots,T\right\} \tag{8}$$

can be mapped one-to-one with $\textsc{mrp}(\mathcal{P})$, and it can be stored as a list of vectors in $\mathcal{J}^T$.

To build $\mathbb{V}(\mathcal{P})$, Algorithm 1 proceeds by looping over price vectors in $\mathcal{P}$ and progressively refining the corresponding partition. For the first price vector, say $\boldsymbol{p}_1$, all choices are rationalizable, so that $\mathbb{V}(\boldsymbol{p}_1)=\mathcal{J}$. Next, when considering $\boldsymbol{p}_2$, (j,k) must be added to $\mathbb{V}(\{\boldsymbol{p}_1,\boldsymbol{p}_2\})$ if choosing j at $\boldsymbol{p}_1$ and k at $\boldsymbol{p}_2$ is rationalizable. This can be checked by ruling out negative cycles in the graph with edges' weights $w^p(j,k)=p_{1,k}-p_{1,j}$ and $w^p(k,j)=p_{2,j}-p_{2,k}$, or simply by checking that $p_{2,j}-p_{2,k}\geq p_{1,k}-p_{1,j}$. The algorithm then proceeds recursively.

[3] In particular, if $N(\mathcal{V})$ is the number of distinct choices ever made by $\mathcal{V}$, the time complexity of verifying whether $\mathcal{V}\in\textsc{mrp}(\mathcal{P})$ is bounded above by $O\left((J+2)\left(J\left(N(\mathcal{V})+1\right)+1\right)\right)$, and memory requirements are limited to storing $(J+2)^2+(J+2)$ values. See Appendix A for more details.

Algorithm 1 Construction of $\mathbb{V}(\mathcal{P})$

Require: Matrix $\mathcal{P} \in \mathbb{R}^{T \times J}$

Initialize $\mathbb{V}^1 = \{0, 1, \ldots, J-1, J\}$
for $t = 2, \ldots, T$ **do**
  Initialize $\mathbb{V}^t = \emptyset$, set $\mathcal{P}_{1:t}$ = rows 1 through t of $\mathcal{P}$
  parfor $\tilde{\boldsymbol{y}} \in \mathbb{V}^{t-1}$ **do**
    for $j \in \mathcal{J}$ **do**
      Set $\boldsymbol{y}^{\text{test}} = (\tilde{\boldsymbol{y}}, j)$
      for $k \in \mathcal{J}$ **do**
        Set $T_k^{\text{test}} = \{\ell = 1, 2, \ldots, t : y_\ell^{\text{test}} = k\}$
        for $k' \in \mathcal{J} \setminus \{k\}$ **do**
          if $T_k^{\text{test}} \neq \emptyset$ **then**
            Set $w(k, k') = -\max_{\ell \in T_k^{\text{test}}}(p_{\ell,k} - p_{\ell,k'})$
          else
            Set $w(k, k') = \infty$
          end if
        end for
        Set $w(k, k) = \infty$
      end for
      Run Algorithm BF (Appendix A) using $\mathcal{J}$ and w as inputs
      if `Negative Cycles` = `No` **then**
        Add $\boldsymbol{y}^{\text{test}}$ to $\mathbb{V}^t$
      end if
    end for
  end parfor
end for
Return $\mathbb{V}(\mathcal{P}) = \mathbb{V}^T$

3.3 Comparing Algorithms

In this section, we compare the performance of three algorithms. The first algorithm is Algorithm 1. The second is the divide and conquer algorithm used by TTY, which uses linear programs to check whether sequences of choices are rationalizable.[4] The third is

[4]This algorithm begins by dividing the full set of price vectors $\mathcal{P}$ into smaller blocks $\mathcal{P}^{(1)}, \mathcal{P}^{(2)} \ldots, \mathcal{P}^{(M)}$. Then, within each block, it builds the MRP using only the prices in that block by sequentially verifying the existence of valid solutions to the linear programs implied by stacking all revealed preference inequalities. Beginning with sequences of two choices, and keeping only the surviving subset prior to moving to sequences of three choices, and so on and so forth, recursively. After obtaining each of the coarser partitions, one for every block $\mathcal{P}^{(m)}$, $m = 1, 2, \ldots, M$, the algorithm considers, in sequence, the much bigger linear programs implied by stacking every pair of choice sequences for every two blocks, keeping only the surviving ones. Iteratively, the procedure moves to a smaller number of increasingly large linear programs, until all choice sequences that cannot be rationalized have been ruled out.

the cell-enumeration algorithm developed by (Gu, Russell, and Stringham, 2025, "GRS" in the following). The GRS algorithm can be used to partition spaces of latent distributions for a wider range of economic applications that leverage revealed preferences in additively separable models (see also Agarwal, Li, and Somaini, 2023, for a discussion of identification of these models). In addition to discrete choice with full information, these include discrete choice with limited information, panel data with discrete responses, and static games of incomplete information.

We briefly review the algorithm proposed by GRS as applied to (QL). The algorithm proceeds in two steps. The first step builds a partition that collects all half-spaces defined by all possible revealed-preference inequalities (binary comparisons) implied by the model. This partition—which GRS call the *hyperplane arrangement cells*—can be expressed as

$$\mathcal{G}(\mathcal{P}) \equiv \bigcap_{t=1}^{T} \bigcap_{j \in \mathcal{J}} \bigcap_{k \in \mathcal{J} \setminus \{j\}} \left\{ \left\{ \boldsymbol{v} \in \mathbb{R}^J : v_j - v_k > p_{t,j} - p_{t,k} \right\}, \left\{ \boldsymbol{v} \in \mathbb{R}^J : v_j - v_k < p_{t,j} - p_{t,k} \right\} \right\},$$

where intersections of partitions are element-wise intersections of their elements.[5] An important property of the algorithm proposed by GRS to build $\mathcal{G}$ is that, unlike TTY, it does not require solving linear programs. Instead, it only keeps track of one "witness point" $\boldsymbol{v} \in G$ for every $G \in \mathcal{G}$.

The number of witness points increases as the partition is progressively refined by including more prices. The resulting partition $\mathcal{G}$ is *not minimal*, and it is typically much larger than $\text{MRP}(\mathcal{P})$. For instance, if $J = T = 2$ then $|\text{MRP}(\mathcal{P})| = 6$ for typical (generic) configurations of prices, while $|\mathcal{G}(\mathcal{P})| = 16$. The second step of the algorithm amounts to "stitching together" elements of $\mathcal{G}$ that lead to the same choices for all prices in $\mathcal{P}$.[6]

Table 1 compares the performance of the three algorithms for different numbers of

[5] If $\mathcal{A}$ and $\mathcal{B}$ are two partitions of $\mathbb{R}^J$, $\mathcal{A} \cap \mathcal{B} \equiv \left\{ X \subset \mathbb{R}^J : X = A \cap B \text{ for some } A \in \mathcal{A}, B \in \mathcal{B} \right\}$.

[6] Stitching can be done recursively by taking unions of sets that prescribe the same choice for all prices.

Table 1: A Comparison Between Different Algorithms to Build the MRP

		Running Times (seconds, capped at 4 hours)				**Partition Sizes**	
				GRS			
T	J	Algorithm 1	TTY	Step 1	Stitch	$\mathbb{G}(\mathcal{P})$	$\text{MRP}(\mathcal{P})$
	3	< 0.001	4.62	6.05	0.43	32,682	286
10	5	0.002	59.3	>4 hours	—	—	3,003
	10	1.66	5001	OOM	—	—	184,756
	3	0.010	89.6	88.0	5.83	258,562	1,771
20	5	0.680	$6,350$	OOM	—	—	53,130
	10	740	>4 hours	OOM	—	—	30,045,015
	3	0.870	>4 hours	$3,810$	194	4,015,402	23,426
50	5	176	>4 hours	OOM	—	—	3,478,761
	10	OOM	OOM	OOM	—	—	75,394,027,566
	3	19.4	>4 hours	>4 hours	—	—	176,851
100	5	$14,200$	>4 hours	OOM	—	—	96,560,646
	10	OOM	OOM	OOM	—	—	46,897,636,623,981

Notes: OOM indicates out of memory. We randomly draw prices iid from a standard normal distribution, and use a machine with 512 Gb of RAM and an Intel(R) Xeon(R) W-3345 CPU @ 3.00GHz, 3000 Mhz, 24 Core(s), 48 Logical Processor(s). The size of the $\text{MRP}(\mathcal{P})$ is equal to $\binom{T+J}{J}$, as shown in Appendix B; lack of ties is confirmed whenever we can build the MRP using one of the algorithms.

choices J and price vectors T. Algorithm 1 is the clear winner in terms of runtime. It is also able to build MRPs without running out of memory for larger values of J and T relative to the alternatives. We find that GRS and TTY are comparable in terms of runtime for $J = 3$ and $T = 20$, but for $T = 50$, GRS becomes more than 4 times faster. For $J = 5$, however, GRS runs out of memory even with $T = 20$. TTY is still able to build the MRP, albeit with a runtime of more than 2 hours. By contrast, Algorithm 1 builds the MRP with $J = 5$ and $T = 20$ in a fraction of a second.

This achieves the goal since any $\mathcal{V}$ in $\text{MRP}(\mathcal{P})$ can be obtained as

$$\mathcal{V} = \bigcup \{G \in \mathcal{G} : \boldsymbol{v} \in G \Rightarrow Y(\boldsymbol{v}, \boldsymbol{p}_t) = y_t(\mathcal{V}) \text{ for all } t = 1, 2, \dots, T\} .$$

4 Outer Sets

Table 1 shows that Algorithm 1 can be used to build $\textsc{mrp}(\mathcal{P})$ for values of J and T that are large enough to encompass many applications. However, the number of elements in $\textsc{mrp}(\mathcal{P})$ can be quite large; in Appendix B we prove that if prices are drawn from a continuous distribution then the generic number of sets in $\textsc{mrp}(\mathcal{P})$ is $\binom{T+J}{J}$. This makes the linear programs in Proposition 1 prohibitively difficult to solve. In this section, we develop strategies for constructing informative outer sets. These sets are not necessarily sharp, but they are feasible to implement for larger values of J and T. As we will see, they often end up being close to the sharp set.

4.1 Tailoring the MRP to the Target Parameter

The MRP characterizes all possible choice types at all prices included in $\mathcal{P}$. Many of these prices are support points of the observed prices P_i, which are not directly relevant for describing the target parameter, θ. Including these prices helps tighten identified sets through the data-matching condition (MD), but also contributes to the large number of sets in the MRP. To reduce dimensionality it seems natural to relax some of these data-matching conditions while preserving the MRP sets that are essential for characterizing θ.

We introduce some definitions that we use to characterize the MRP sets that are relevant for θ. Recall that t is the function defined on Φ^{e} that is induced by θ and $\textsc{mrp}(\mathcal{P})$ using all of the prices $\mathcal{P}$. We define the *θ-relevant set of valuations* $\mathcal{V}^{\theta}(\mathcal{P}) \subseteq \mathbb{R}^{J}$ as the smallest union of sets in $\textsc{mrp}(\mathcal{P})$ such that the following condition holds:

$$\text{if } \phi, \phi' \in \Phi^{\text{e}} \text{ satisfy } \phi(\mathcal{V}) = \phi'(\mathcal{V}) \text{ for all } \mathcal{V} \subseteq \mathcal{V}^{\theta}(\mathcal{P}), \text{ then } t(\phi) = t(\phi'). \tag{9}$$

When t is linear, $t(\phi) = \sum_{\mathcal{V} \in \textsc{mrp}(\mathcal{P})} c(\mathcal{V})\phi(\mathcal{V})$ with coefficients $c(\mathcal{V}) \in \mathbb{R}^{d_\theta}$, and (9) reduces to $\mathcal{V}^\theta(\mathcal{P}) = \bigcup\{\mathcal{V} \in \text{MRP}(\mathcal{P}) : c(\mathcal{V}) \neq 0\}$.

We then define a price to be θ*-relevant* if either (i) it is a price that was added for the purpose of evaluating the target parameter, that is, $\boldsymbol{p} \in \mathcal{P}^\circ \equiv \mathcal{P} \setminus \{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_T\}$, or (ii) it is an observed price $\boldsymbol{p}_t$ for which at least one choice region $\mathcal{V}_j(\boldsymbol{p}_t)$ non-trivially intersects $\mathcal{V}^\theta(\mathcal{P})$—meaning the two sets overlap but $\mathcal{V}^\theta(\mathcal{P})$ is not contained in $\mathcal{V}_j(\boldsymbol{p}_t)$. Formally, the set of θ*-relevant prices* $\mathcal{P}^\theta$ is defined as

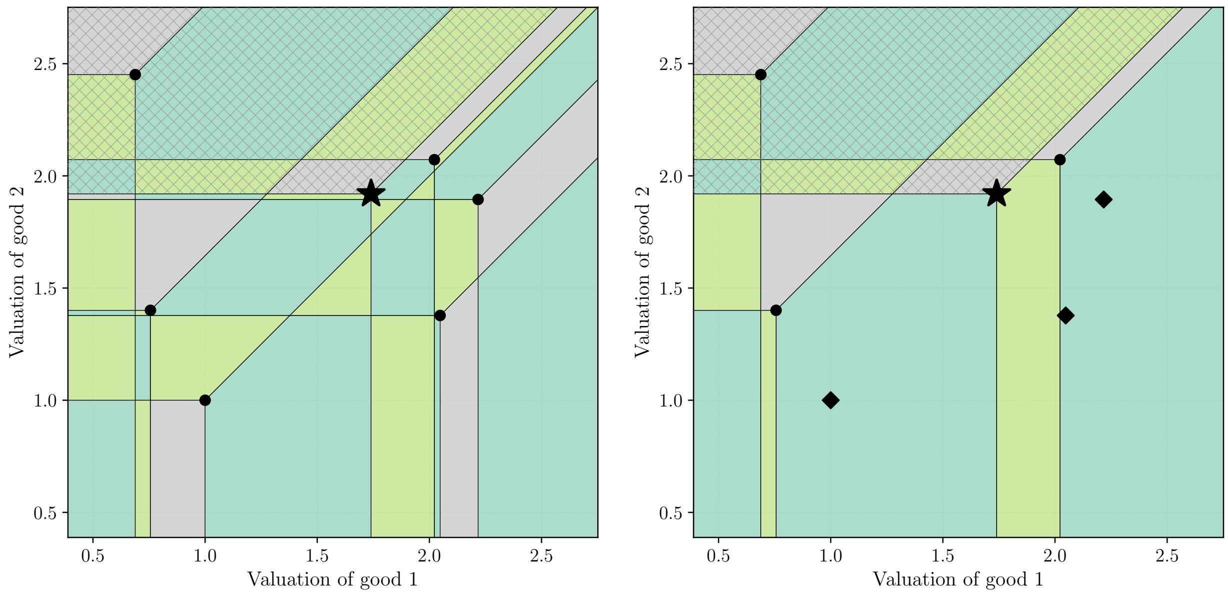
$$\mathcal{P}^\theta \equiv \mathcal{P}^\circ \cup \left\{\boldsymbol{p}_t : \mathcal{V}^\theta(\mathcal{P}) \cap \mathcal{V}_j(\boldsymbol{p}_t) \neq \emptyset \text{ and } \mathcal{V}^\theta(\mathcal{P}) \not\subseteq \mathcal{V}_j(\boldsymbol{p}_t) \text{ for some } j \in \mathcal{J}\right\}.$$

We call $\text{MRP}(\mathcal{P}^\theta)$—the MRP generated from the θ-relevant prices—the θ-relevant partition, or $\theta\text{RP}(\mathcal{P})$.

Figure 5 illustrates these definitions in an example with $J = 2$ and $T = 6$. The six observed prices are shown in panel (a) as dots. The target parameter is the share of individuals who choose $j = 2$ at the counterfactual price vector $\boldsymbol{p}^\star$ denoted by the star. Panel (a) shows the full MRP constructed from $\mathcal{P} = \mathcal{P}^\circ \cup \{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_T\}$, where $\mathcal{P}^\circ = \{\boldsymbol{p}^\star\}$. The MRP consists of 36 sets. The θ-relevant set $\mathcal{V}^\theta(\mathcal{P})$ is shown as the crosshatched region. Panel (b) shows the three prices that are *not* θ-relevant as diamonds. Each of these prices produces a $\mathcal{V}_0(\boldsymbol{p}), \mathcal{V}_1(\boldsymbol{p})$ that does not intersect $\mathcal{V}^\theta(\mathcal{P})$ and produces a $\mathcal{V}_2(\boldsymbol{p})$ that fully contains $\mathcal{V}^\theta(\mathcal{P})$; none of these non-trivially intersect $\mathcal{V}^\theta(\mathcal{P})$, so these prices are not θ-relevant. Using the remaining four θ-relevant prices leads to $\text{MRP}(\mathcal{P}^\theta) \equiv \theta\text{RP}(\mathcal{P})$ that consists of the 15 sets shown in panel (b).

Constructing $\theta\text{RP}(\mathcal{P})$ does not require constructing the larger $\text{MRP}(\mathcal{P})$. Appendix C shows how to use Algorithm 1 to construct $\theta\text{RP}(\mathcal{P})$ iteratively, starting from $\text{MRP}(\mathcal{P}^\circ)$.

Figure 5: Example of θ-Reduced Partition with $J = 2$, $T = 6$, and $\theta(f) = \sigma_2(\boldsymbol{p}^{\star}; f)$



(a) $\text{MRP}(\mathcal{P})$, 36 sets, and $\mathcal{V}^{\theta}(\mathcal{P})$, crosshatched

(b) $\mathcal{P}^{\theta}$, 4 price vectors, and $\theta\text{RP}(\mathcal{P})$, 15 sets

4.2 Computing Outer Sets

Next, we use $\theta\text{RP}(\mathcal{P})$ to compute valid outer sets by solving linear programs that have more constraints but many fewer variables. Let Φ^{θ} denote the set of all functions $\phi : \theta\text{RP}(\mathcal{P}) \to [0, 1]$ that could represent a probability mass function over the elements of $\theta\text{RP}(\mathcal{P})$, that is, such that $\sum_{\mathcal{V} \in \theta\text{RP}(\mathcal{P})} \phi(\mathcal{V}) = 1$. As before, we define $\Phi^{\theta}(\kappa)$ as

$$\Phi^{\theta}(\kappa) \equiv \left\{ \phi \in \Phi^{\theta} : (1 - \kappa) \int_{\mathcal{V}} g(\boldsymbol{v})\, d\boldsymbol{v} \leq \phi(\mathcal{V}) \leq (1 + \kappa) \int_{\mathcal{V}} g(\boldsymbol{v})\, d\boldsymbol{v} \quad \text{for all } \mathcal{V} \in \theta\text{RP}(\mathcal{P}) \right\},$$

with the difference being that we use $\theta\text{RP}(\mathcal{P})$ instead of the finer $\text{MRP}(\mathcal{P})$.

First, we define the induced target parameter function on Φ^{θ}. Every θ-relevant set $\mathcal{V} \subseteq \mathcal{V}^{\theta}(\mathcal{P})$ is an element of both $\text{MRP}(\mathcal{P})$ and $\theta\text{RP}(\mathcal{P})$: within $\mathcal{V}^{\theta}(\mathcal{P})$ the two partitions coincide. Given any $\phi \in \Phi^{\theta}$, let $e(\phi)$ be the function in Φ^{e} that extends ϕ to $\text{MRP}(\mathcal{P})$ by assigning zero mass to all sets in $\text{MRP}(\mathcal{P})$ that are not in $\theta\text{RP}(\mathcal{P})$, and define

$$t^{\theta} : \Phi^{\theta} \to \mathbb{R}^{d_{\theta}}, \qquad t^{\theta}(\phi) \equiv t(e(\phi)). \tag{10}$$

By (9), $t^{\theta}(\phi) = t(\tilde{\phi})$ for *every* $\tilde{\phi} \in \Phi^{\mathrm{e}}$ that agrees with ϕ on the θ-relevant sets, that is, with $\tilde{\phi}(\mathcal{V}) = \phi(\mathcal{V})$ for all $\mathcal{V} \subseteq \mathcal{V}^{\theta}(\mathcal{P})$, since any such $\tilde{\phi}$ coincides with $e(\phi)$ on those sets.

Second, we consider the implications that a given $\phi \in \Phi^{\theta}$ would have for the observed choice shares. Define

$$\begin{aligned} \underline{s}_j(\boldsymbol{p}; \phi) &\equiv \sum_{\mathcal{V} \in \theta\text{RP}(\mathcal{P})} \mathbf{1}\left[\mathcal{V} \subseteq \mathcal{V}_j(\boldsymbol{p})\right] \phi(\mathcal{V}), \text{ and} \\ \overline{s}_j(\boldsymbol{p}; \phi) &\equiv \sum_{\mathcal{V} \in \theta\text{RP}(\mathcal{P})} \mathbf{1}\left[\mathcal{V} \cap \mathcal{V}_j(\boldsymbol{p}) \neq \emptyset\right] \phi(\mathcal{V}), \end{aligned}$$

so that $\underline{s}_j(\boldsymbol{p}; \phi)$ collects the mass that ϕ places on sets contained in $\mathcal{V}_j(\boldsymbol{p})$ and $\overline{s}_j(\boldsymbol{p}; \phi)$ collects the mass that ϕ places on sets that intersect $\mathcal{V}_j(\boldsymbol{p})$. Notice that $\underline{s}_j(\boldsymbol{p}; \phi)$ and $\overline{s}_j(\boldsymbol{p}; \phi)$ can be evaluated for *every* $\boldsymbol{p} \in \mathcal{P}$—not just $\boldsymbol{p} \in \mathcal{P}^{\theta}$—using a mass function $\phi \in \Phi^{\theta}$ that is defined only over $\theta\text{RP}(\mathcal{P})$. The definition of an MRP—in particular, of $\theta\text{RP}(\mathcal{P})$—ensures that for any $\boldsymbol{p} \in \mathcal{P}^{\theta}$ we have $\underline{s}_j(\boldsymbol{p}; \phi) = \overline{s}_j(\boldsymbol{p}; \phi)$ for every $j \in \mathcal{J}$. For other prices—in particular, the observed prices that are not θ-relevant—we know that $\underline{s}_j(\boldsymbol{p}; \phi) \leq \overline{s}_j(\boldsymbol{p}; \phi)$ by definition, but the inequalities might be strict.

This reasoning suggests that the identified set for the target parameter $\Theta^{\star}(\kappa)$ must be contained in the set

$$\begin{aligned} \Theta_{\text{OUT}}(\kappa) \equiv \Big\{ \vartheta \in \mathbb{R}^{d_{\theta}} : \vartheta = t^{\theta}(\phi) \quad & \text{for some } \phi \in \Phi^{\theta}(\kappa) \text{ s.t.} \\ & \underline{s}_j(\boldsymbol{p}_t; \phi) \leq s_j(\boldsymbol{p}_t) \leq \overline{s}_j(\boldsymbol{p}_t; \phi) \text{ for all } j \text{ and } t \Big\}, \end{aligned}$$

which would make $\Theta_{\text{OUT}}(\kappa)$ a valid outer set (also called an outer region; see, for example

Molinari, 2020). We measure satisfaction of the inequality constraints in $\Theta_{\text{OUT}}(\kappa)$ using the criterion function

$$Q_{\text{OUT}}(\phi) \equiv \sum_{t=1}^{T} \sum_{j \in \mathcal{J}} \max\{0, \underline{s}_j(\boldsymbol{p}_t; \phi) - s_j(\boldsymbol{p}_t)\} + \max\{0, s_j(\boldsymbol{p}_t) - \overline{s}_j(\boldsymbol{p}_t; \phi)\}.$$

If t^θ is scalar-valued, then the endpoints of $\Theta_{\text{OUT}}(\kappa)$ (if non-empty) are given by the optimal values of the programs

$$\min_{\phi \in \Phi^\theta(\kappa)}/\max t^\theta(\phi) \quad \text{s.t.} \quad Q_{\text{OUT}}(\phi) = 0, \tag{11}$$

which reduce to linear programs—via slack variables for the constraints—whenever t^θ is linear. The following proposition verifies the suggestion that $\Theta_{\text{OUT}}(\kappa)$ is an outer set.

Proposition 2. Suppose that Assumptions EP, RD, and TP are satisfied. Then $\Theta^\star(\kappa) \subseteq \Theta_{\text{OUT}}(\kappa)$.

Proof. See Appendix D. □

4.3 Outer Sets with Subset Prices

The number of variables in the linear program (11) is determined by $\theta\text{RP}(\mathcal{P})$, which is often much smaller than $\text{MRP}(\mathcal{P})$. The cost is in replacing the equality constraints for observed prices not in $\mathcal{P}^\theta$ by two inequalities, which is what leads these sets to be potentially outer. Nevertheless, as T increases, the number of prices in $\mathcal{P}^\theta$—and therefore the number of sets in $\theta\text{RP}(\mathcal{P})$—may still become prohibitively large.

To address this issue, one can use subsets of the θ-relevant prices. Let $\mathcal{P}^\theta_{\text{SUB}} \subseteq \mathcal{P}^\theta$ denote a subset of $\mathcal{P}^\theta$. We always choose $\mathcal{P}^\theta_{\text{SUB}}$ such that $\mathcal{P}^\circ \subseteq \mathcal{P}^\theta_{\text{SUB}}$, so that the subset of θ-relevant prices contains the set of prices added for evaluating the target parameter. Then we replace $\theta\text{RP}(\mathcal{P})$ with the coarser partition $\text{MRP}(\mathcal{P}^\theta_{\text{SUB}})$. The rest of the construction and results in the previous section proceeds the same way with $\text{MRP}(\mathcal{P}^{\theta}_{\text{SUB}})$ in place of $\theta\text{RP}(\mathcal{P})$. The outer bounds can be made easier to compute by making $\mathcal{P}^{\theta}_{\text{SUB}}$ smaller, but are inherently weakly wider as a consequence.

Depending on how $\mathcal{P}^{\theta}_{\text{SUB}}$ is chosen, the bounds can be considerably wider than those constructed with $\theta\text{RP}(\mathcal{P})$ or nearly the same. In lieu of a more sophisticated rule, we draw several random samples of $\mathcal{P}^{\theta}_{\text{SUB}}$ and compute outer bounds under each. Then, we take the intersection of these bounds.

4.4 Simulation

We expand the example introduced in Section 2 to a setting with $J = 2, 4, 6$, or 10 inside products, and $T = 25, 50, 100$, or 500 observed markets. For each (J, T) pair, we run $M = 10$ simulations with price vectors drawn from $[2, 2.5]^J$ as a way to measure timing across different configurations. Choice shares are generated from a distribution of valuations that follows a two-type mixed logit with $V_{i,j} = \sigma_i(\delta_{i,j} + \varepsilon_{i,j} - \varepsilon_{i,0})$, where $\varepsilon_{i,j}$ are i.i.d. draws from the type I extreme value distribution; 70% of consumers i have $\sigma_i = 1/6$, $\delta_{i,1} = \delta_{i,2} = 8/3$, and $\delta_{i,j} = 2/3$ for $j > 2$, while the rest have $\sigma_i = 50$, $\delta_{i,1} = \delta_{i,2} = 25$, and $\delta_{i,j} = -50$ for $j > 2$.

For each simulation, we consider the target parameter that measures the substitution between products 1 and 2,

$$\theta(f) = \frac{\sigma_1(p^\star; f) - \sigma_1(p^0; f)}{\sigma_2(p^0; f)}$$

with $p^0 = (2.4, 2.4, 2.4, \dots)$ and $p^\star = (2.4, 2.5, 2.4, \dots)$. We estimate a simple logit model that we use as a reference density and then compute the pseudo-true identified set and the identified set using—depending on computational feasibility—sharp, outer, and subset outer bounds using 20 draws of 10 prices each. In each instance we also com-

Table 2: Comparison of Sharp, Outer, and Subset Outer Bounds

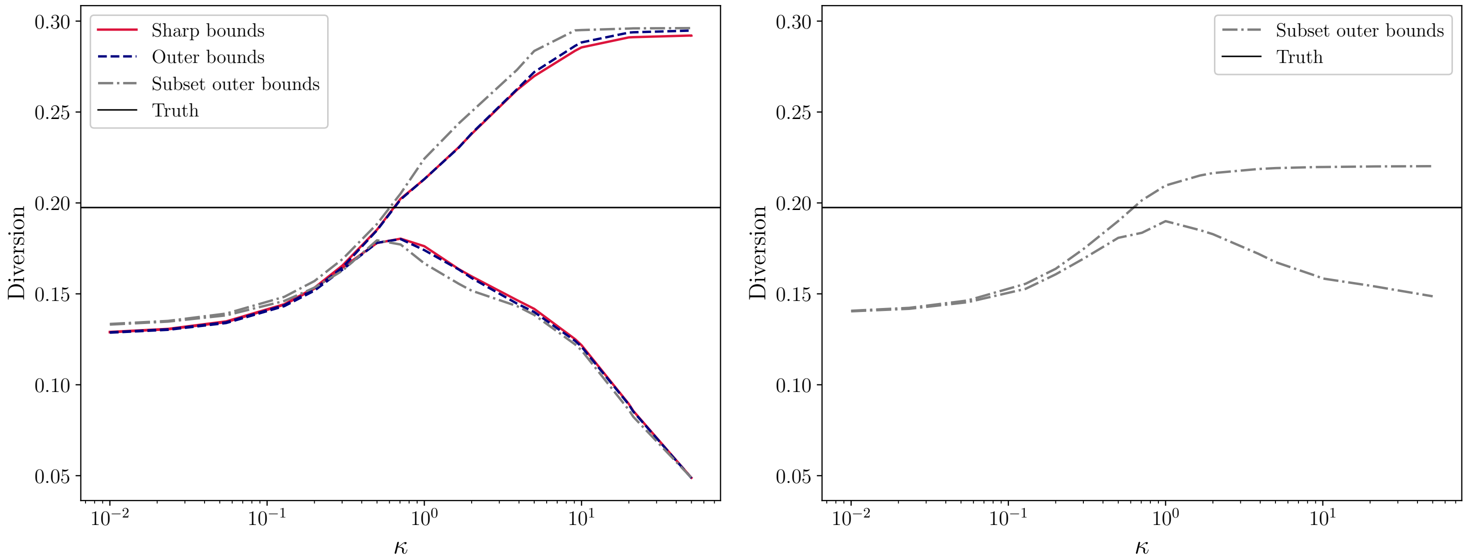
		Lower Bound					Upper Bound			NRC for $\theta \leq 0.2$		
J	T	Subs Out	Outer	Sharp	Logit	Truth	Sharp	Outer	Subs Out	Sharp	Outer	Subs Out
2	25	0.033	0.034	0.034	0.144	0.190	0.302	0.308	0.308	0.006	0.013	0.013
	50	0.061	0.062	0.064	0.145	0.190	0.253	0.254	0.255	0.039	0.058	0.031
	100	0.110	0.116	0.119	0.145	0.190	0.226	0.227	0.227	0.008	0.022	0.037
	500	0.152	–	–	0.144	0.190	–	–	0.207	–	–	0.315
4	25	0.041	0.041	0.041	0.132	0.197	0.293	0.295	0.296	0.040	0.040	0.036
	50	0.087	0.090	–	0.132	0.197	–	0.245	0.247	–	0.048	0.071
	100	0.109	–	–	0.132	0.197	–	–	0.236	–	–	0.052
	500	0.146	–	–	0.132	0.197	–	–	0.222	–	–	0.071
6	25	0.029	0.029	–	0.124	0.203	–	0.296	0.296	–	0.012	0.021
	50	0.079	–	–	0.122	0.203	–	–	0.272	–	–	0.032
	100	0.094	–	–	0.123	0.203	–	–	0.257	–	–	0.015
	500	0.152	–	–	0.123	0.203	–	–	0.226	–	–	0.061
10	25	0.034	–	–	0.113	0.212	–	–	0.355	–	–	0.006
	50	0.062	–	–	0.112	0.212	–	–	0.280	–	–	0.011
	100	0.098	–	–	0.111	0.212	–	–	0.253	–	–	0.016
	500	0.154	–	–	0.113	0.212	–	–	0.237	–	–	0.031

pute the NRC for the claim $\Theta^0 = \{\vartheta : \vartheta \leq 0.2\}$.

Table 2 shows the results averaged over the M simulations for each (J, T) pair. Figure 6 illustrates the three sets of bounds for different values of κ for the case in which $J = 4$, varying T between 25, when all bounds are computationally feasible, and 500, when only subset outer bounds can be computed quickly with reasonable resources. The overall pattern is clear. While the outer sets are somewhat wider, they still return informative sets that are quite close to the sharp sets. Moreover, they can be computed in cases with larger T in which the sharp identified sets cannot.

The true value of the diversion ratio (θ) in our DGP is around 0.2 for all J and T. The logit estimates are always biased downward, with the bias increasing in J from 0.045 at $J = 2$ to 0.1 at $J = 10$. All three sets of bounds are always valid and cover the true value. The values of the NRC indicate that our approach would flag that the claim $\Theta^0 =$

Figure 6: Comparison of Sharp, Outer, and Subset Outer Sets for $J = 4$



(a) $T = 25$ (b) $T = 500$

$\{\theta \leq 0.2\}$ has little support when relaxing the logit assumption. Most importantly, the gaps between sharp and outer bounds, and between outer and subset outer bounds, are always small. The extent of this slack is dwarfed by the gains in considering larger values of T.

For example, averaging across simulations at $J = 4$ and $T = 25$ (see also panel (a) of Figure 6), the sharp set, based on 24,813 $\text{MRP}(\mathcal{P})$ elements in 1.61 seconds, is $[0.041, 0.293]$. The outer set is nearly identical, at $[0.041, 0.295]$, while its characterization only requires, on average, 4,623 elements of $\text{MRP}(\mathcal{P}^{\theta})$ and 0.5 seconds. Lastly, the subset outer set is $[0.041, 0.296]$, and it can be computed in 0.04 seconds based on only 151.9 elements of $\text{MRP}(\mathcal{P}^{\text{sub}})$.[7] When increasing T to 500 as in panel (b) of Figure 6, the subset outer set runs in 2.5 seconds, nearly as fast as the sharp set at $T = 25$, but the bounds become $[0.146, 0.222]$, an interval that is tighter around the true value and that rules out the logit estimate.

[7]The discrepancy evident in Figure 6 between the subset and outer bounds for low values of κ is due to numerical error in integration of the reference density.

Table 3: Computational Complexity of Sharp, Outer, and Subset Outer Sets

		Sharp			Outer			Subset Outer		
		Size	Runtime		Size	Runtime		Size	Runtime	
J	T	$\lvert\text{MRP}(\mathcal{P})\rvert$	Partition	Program	$\lvert\text{MRP}(\mathcal{P}^{\theta})\rvert$	Partition	Program	$\lvert\text{MRP}(\mathcal{P}^{\text{sub}})\rvert$	Partition	Program
2	25	365.3	0.03	0.03	34.1	0.01	0.01	10.0	0.07	0.04
	50	1353.5	0.08	0.15	86.8	0.02	0.01	10.2	0.08	0.05
	100	5203.5	0.45	1.38	250.1	0.13	0.04	9.7	0.07	0.11
	500	*1.3×10^5*	–	–	–	–	–	9.7	0.28	0.28
4	25	24813	0.41	1.20	4622.8	0.30	0.19	151.9	0.02	0.02
	50	*3.4×10^5*	–	–	49333.4	7.94	10.71	145.4	0.03	0.03
	100	*4.8×10^6*	–	–	–	–	–	133.7	0.45	0.37
	500	*2.7×10^9*	–	–	–	–	–	141.3	1.19	1.32
6	25	*9.1×10^5*	–	–	209773.9	13.60	22.01	1040.5	2.12	0.50
	50	*3.6×10^7*	–	–	–	–	–	1038.4	2.61	0.96
	100	*1.8×10^9*	–	–	–	–	–	964.2	3.29	1.89
	500	*2.3×10^{13}*	–	–	–	–	–	941.2	10.03	12.62
10	25	*2.5×10^8*	–	–	–	–	–	11404.7	37.00	6.70
	50	*9.0×10^{10}*	–	–	–	–	–	11414.7	47.59	15.46
	100	*5.2×10^{13}*	–	–	–	–	–	11861.4	71.87	47.71
	500	*3.1×10^{20}*	–	–	–	–	–	12904.6	70.90	684.31

Note: Italicized numbers in column 3 show MRP sizes computed using the formula from Proposition A1.

5 Hospital Choice in Austin, Texas

In this section, we use administrative data on inpatient hospital visits in Austin, Texas for 2017–2018 to create a data generating process closely tied to a real empirical setting. Our focus is on identification, so we continue to simulate choice data that satisfy (QL) and assume that we know the population-level choice shares. See TTY for a discussion of estimation and inference.

5.1 Setting and Data

The data, which is summarized in Table 4, contains anonymized individual-level visit records that include the patient's age (in bins), gender, zipcode of residence, and sever-

Table 4: Summary Statistics for the Austin Hospital Choice Dataset

	Mean	Std. Dev.	Min	Max
Charges	20701	21375	0	99980
Distance	6.308	4.724	1	27.91
Severity Index (0-4)	1.740	0.773	0	4
Demographic indicators:				
Age 0-17	0.306			
18-30	0.148			
31-44	0.199			
45-64	0.138			
65+	0.209			
Female	0.643			

	Market Share		
Hospital Name	Male	Female	All
St Davids Hospital	0.104	0.136	0.125
Seton Medical Center	0.151	0.195	0.179
St Davids South Austin Hospital	0.137	0.136	0.136
St David North Austin	0.161	0.206	0.190
Dell Childrens Medical Center	0.078	0.038	0.052
Seton Medical Center at The U of TX	0.058	0.032	0.041
Other	0.312	0.256	0.276
Number of hospitalizations	103,666		

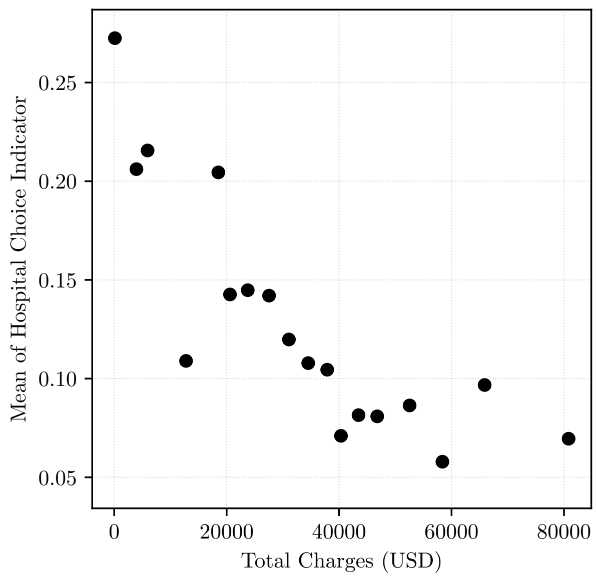
ity index at the time of admission.[8] Each record also includes identifiers for the hospital and total charges for the visit. We keep visit records with charges between $0 and $100,000 (88% of all visits), which leads to a sample with mean charge of $20,701. (Many of these charges are paid by insurers.) We focus on the largest six hospitals in the area (see bottom panel of Table 4), which account for 70% of all visits, and bundle the remaining hospitals in the outside option. This leaves us with $J = 6$ hospitals.

Distance is well-known to be an important determinant of hospital choice (Ho, 2006, 2009; Ho and Pakes, 2014). We augment the visit record data with the distance between the centroid of each hospital's zip code and the patient's own zip code.[9] After keeping

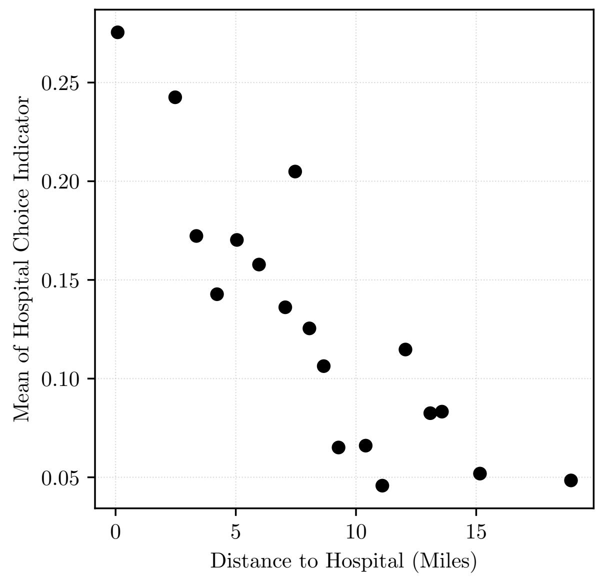
[8] The data is publicly available and freely-provided by the Texas Department of State Health Services. See https://www.dshs.texas.gov/center-health-statistics/; last accessed in May 2026.

[9] Source: https://data.nber.org/distance/2017/50miles/; last accessed in May 2026.

Figure 7: Binscatter of Hospital Choice Against Charges and Distance in the Raw Data



(a) Hospital Choice vs. Charges



(b) Hospital Choice vs. Distance

observations for which we observe all variables, we are left with 103,666 visit records.

Figure 7 shows a strong negative relationship between hospital choice and both distance and charges. Moving from the bottom 5% to the top 5% of charges or distance is associated with a drop in the probability of choosing a hospital of around 0.20 (or 80%).

5.2 Estimates and Simulated DGP

We estimate a standard hospital choice model following Ho (2006), in which the mean utility of hospital j for patient i is given by

$$u_{i,j} = \alpha(\boldsymbol{X}_i)\log(\text{Miles}_{i,j}) + \beta(\boldsymbol{X}_i)\log(\text{Charges}_{i,j}) + \mu_j(\boldsymbol{X}_i) + \varepsilon_{i,j}, \text{ for } j > 0, \tag{12}$$

while $u_{i,0} = \varepsilon_{i,0}$. In this expression, $\boldsymbol{X}_i$ includes the patient's age indicators, gender, and severity, $\mu_j(\boldsymbol{X}_i)$ includes hospital fixed effects interacted with patient characteristics,

Table 5: MLE Estimates of Hospital Choice Model in Equation (12)

	(1)	(2)	(3)
log-Miles	-1.375 (0.010)		
log-Charges	-0.596 (0.008)		
log-Miles ×			
Female		0.002 (0.021)	-0.067 (0.022)
Severity Index		-0.035 (0.027)	0.006 (0.027)
Aged 18–30		-2.126 (0.028)	-1.938 (0.029)
Aged 31–44		-1.650 (0.030)	-1.593 (0.030)
Aged 45–64		-1.157 (0.024)	-1.126 (0.025)
Aged 65+		-1.044 (0.020)	-1.144 (0.022)
log-Charges ×			
Female		0.026 (0.005)	-0.123 (0.018)
Severity Index		0.037 (0.006)	0.030 (0.006)
Aged 18–30		-0.450 (0.010)	0.151 (0.028)
Aged 31–44		-0.542 (0.010)	-0.497 (0.021)
Aged 45–64		-0.604 (0.010)	-0.555 (0.021)
Aged 65+		-0.639 (0.009)	-1.025 (0.016)
Fixed effects	Hospital	Hospital	Hospital-Age, Hospital-Sex
Hospitalizations	103,666	103,666	103,666

Standard errors in parentheses.

and $\varepsilon_{i,j}$ is a type I extreme value idiosyncratic error term. Maximum likelihood estimates are reported in Table 5. The estimated coefficients on distance and charges are negative, statistically significant, and heterogeneous across different groups of patients.

For our simulations, we set the market index t to be a unique combination of quarter, zipcode, age group, gender, admission severity, and diagnosis. We then compute

$$P_{j,t} = \mathbb{E}\left[\alpha(\boldsymbol{X}_i)\log(\text{Miles}_{i,j}) + \beta(\boldsymbol{X}_i)\log(\text{Charges}_{i,j})\middle|\boldsymbol{X}_i \in t\right],$$

and use this to construct the price vector $\boldsymbol{P}_t$ for each market t. Similarly, we set $\overline{V}_{j,t} = \mathbb{E}\left[\mu_j(\boldsymbol{X}_i)|\boldsymbol{X}_i \in t\right]$, and complete our model by specifying $\boldsymbol{V}_{i,t} = \overline{\boldsymbol{V}}_t + \boldsymbol{\omega}_{i,t} - \omega_{i,0,t}$ for a chosen distribution of the vector $\boldsymbol{\omega}_{i,t} \in \mathbb{R}^{J+1}$. This implies that $\boldsymbol{P}_t$ and $\boldsymbol{V}_t$ are not unconditionally independent, but that Assumption EP holds after conditioning on age, gender, and severity, which we collect in the vector $\boldsymbol{W}_t$, for each market t. Formally, $f(\boldsymbol{v}|\boldsymbol{p},\boldsymbol{w}) = f(\boldsymbol{v}|\boldsymbol{w})$, and, in what follows, we condition on $\boldsymbol{W}_t = \boldsymbol{w}$ throughout.

We simulate choice shares under a three-nests, two-level generalized extreme value distribution, i.e., setting the CDF of $\boldsymbol{\omega}_{i,t}$ as

$$F(\boldsymbol{\omega}_{i,t}) = \exp\left\{-\sum_{b=0}^{B}\left(\sum_{j\in\mathcal{J}_b} e^{-\omega_{i,j,t}/\lambda}\right)^{\lambda}\right\}, \text{ where } \lambda = 1-\rho \in (0,1].$$

We specify $B = 3$, and $\mathcal{J}_0 = \{0\}$, $\mathcal{J}_1 = \{1,2\}$, $\mathcal{J}_2 = \{3,4,5\}$, $\mathcal{J}_3 = \{6\}$, and set the within-nest correlation parameter ρ equal to 0.6.

When imposing Assumption RD, we do so using the misspecified simple logit conditional on $\boldsymbol{W}_t$, estimating

$$s_{j,t} = \frac{\exp\{-\gamma(\boldsymbol{W}_t)P_{j,t} + \eta_j(\boldsymbol{W}_t)\}}{1+\sum_{k=1}^{6}\exp\{-\gamma(\boldsymbol{W}_t)P_{k,t} + \eta_k(\boldsymbol{W}_t)\}}$$

via maximum likelihood, and setting g to be the density of $\boldsymbol{\varepsilon}_{i,t} - \varepsilon_{i,0,t}$, where $\boldsymbol{\varepsilon}_{i,t}$ collects 7 independent extreme value type I draws with the vector of locations $(0, \eta_1(\boldsymbol{W}_t)/\gamma(\boldsymbol{W}_t), \ldots, \eta_6(\boldsymbol{W}_t)/\gamma(\boldsymbol{W}_t))$ and common scale $1/\gamma(\boldsymbol{W}_t)$.

5.3 Outer Sets and NRC on Diversion Between Hospitals

Figure 8 illustrates outer bounds and the NRC for the diversion between hospitals 2 (Seton Medical Center) and 1 (St Davids Hospital) in response to a 10% increase in $P_{2,t}$ from a baseline price vector $\overline{\boldsymbol{p}}_{\boldsymbol{w}}$ for a few values of $\boldsymbol{W}_t = \boldsymbol{w}$. For each value of $\boldsymbol{w}$, we take $\overline{\boldsymbol{p}}_{\boldsymbol{w}}$ to be the observed price vector closest to the mean price among patients with that age, gender, and severity.[10] Formally, and omitting the conditioning on $\boldsymbol{W}_t = \boldsymbol{w}$ for nota-

[10] For each $\boldsymbol{w}$, we determine $\overline{\boldsymbol{p}}_{\boldsymbol{w}}$ by letting $\bar{p}_{j,w}$ denote the mean price of product j conditional on $\boldsymbol{W}_t = \boldsymbol{w}$, then picking the t that minimizes $\frac{1}{J}\sum_j |p_{j,t} - \bar{p}_{j,w}|$.

tional ease in what follows, the target parameter is

$$\theta(f) = \frac{\sigma_1(\boldsymbol{p}^\star; f) - \sigma_1(\overline{\boldsymbol{p}}; f)}{\sigma_2(\overline{\boldsymbol{p}}; f)}. \tag{13}$$

We compute outer sets based on subsets of 10 vectors intersected over 10 random draws.

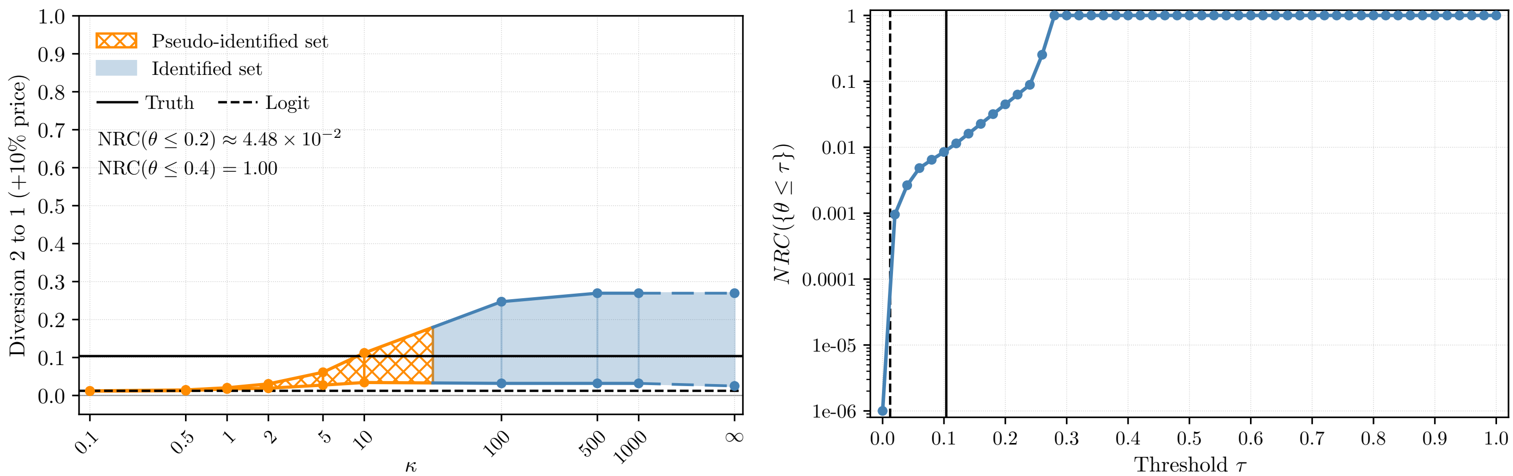
On the left of panel (a) of Figure 8, we show outer sets for the group of over-65 males with a severity index of 3 out of 4. Under the data-generating process, approximately 10% of these patients would divert from hospital 2 toward hospital 1 if the former increased its price (or, equivalently, was farther away) to induce a 10% increase in P_{2t}. Nevertheless, the logit model that we use for g would predict a lower, near-zero diversion for this group. With $T = 1860$ markets for this group of patients, the logit model's prediction lies outside the nonparametric outer sets, and the truth is contained in the pseudo-true outer sets for any value of κ greater than 10.

The NRC for the claim $\Theta^0 = \{\vartheta : \vartheta \leq \tau\}$ that the diversion is less than $\tau = 0.2$ is around 0.05, while for the weaker claim that the diversion is less than 0.4 the criterion increases to 1. In the top-right panel we show how this varies with τ using a logarithmic scale on the vertical axis. This shows that the criterion quickly drops to zero for low values of τ and becomes one for any $\tau > 0.3$.

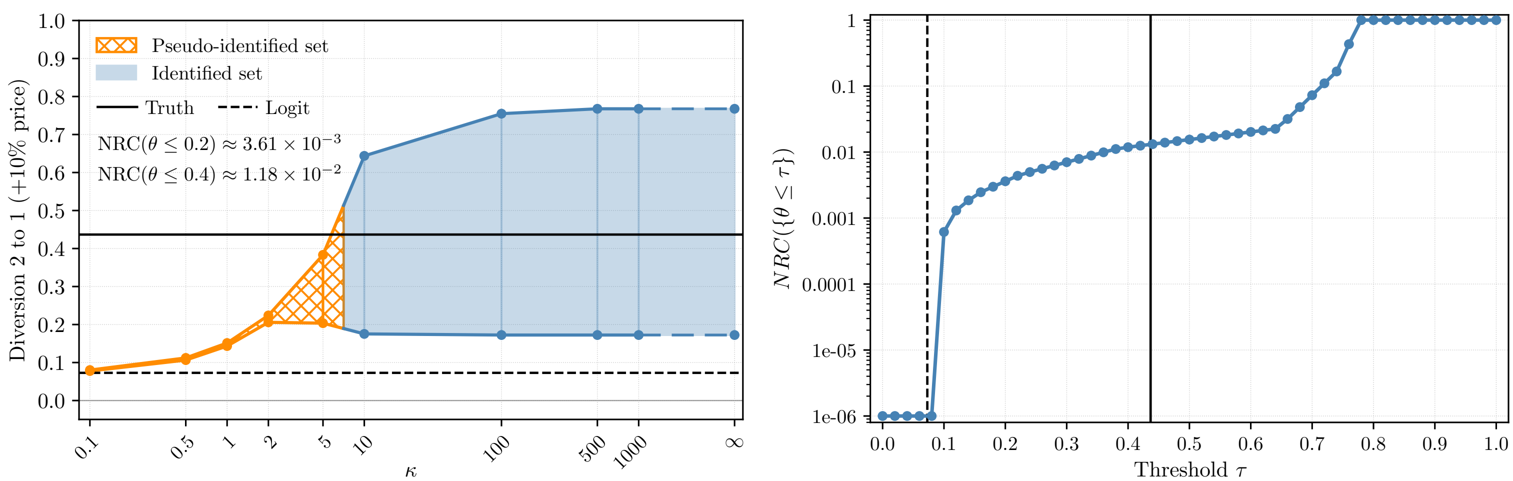
Panels (b) and (c) of Figure 8 repeat this procedure for other groups, the main difference being that the true diversion is much higher in both cases (around 0.5), while the logit prediction is around 0.1. A key difference between panels (b) and (c) is that the former has more than twice as many markets ($T = 2935$) as the latter ($T = 1336$). This results in the outer sets in panel (b) being more informative across the entire range of κ. Moreover, in the group with a larger number of observed markets, the NRC is 1 for any value of $\tau > 0.8$, while in panel (c) the NRC remains lower than 1 as long as $\tau \leq 0.94$.

Our approach can also be used to characterize the identified set for higher-

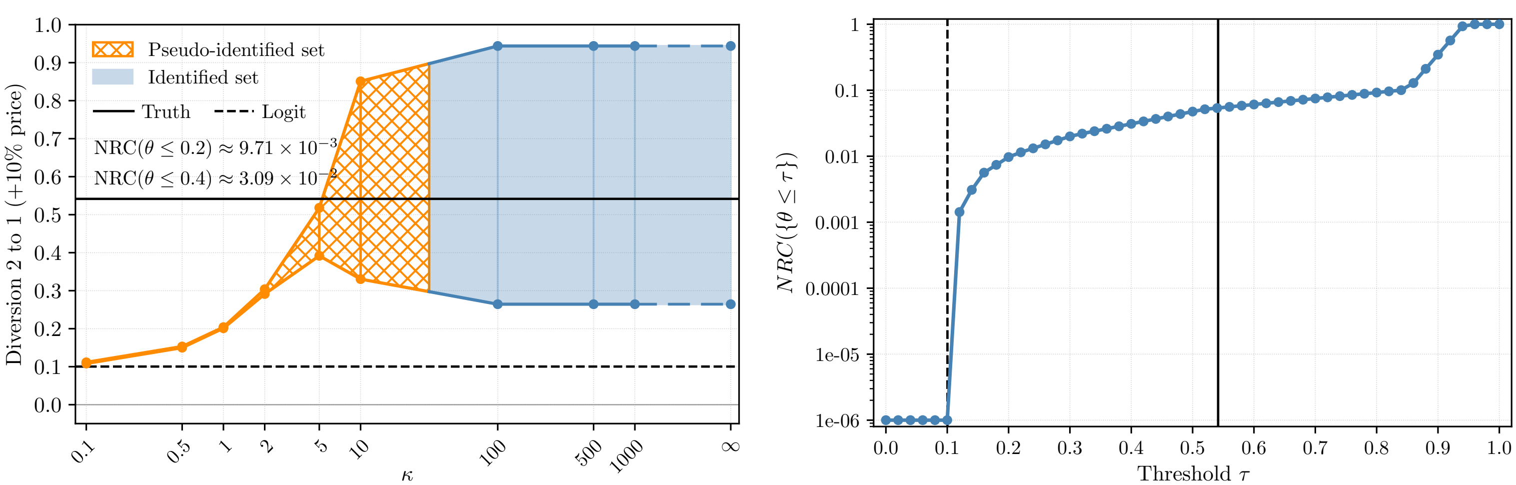
Figure 8: Outer Sets and NRC in Austin Hospital Choice Simulations



(a) Over-65, Severity Index 3, Male; $T_{\boldsymbol{w}} = 1860$

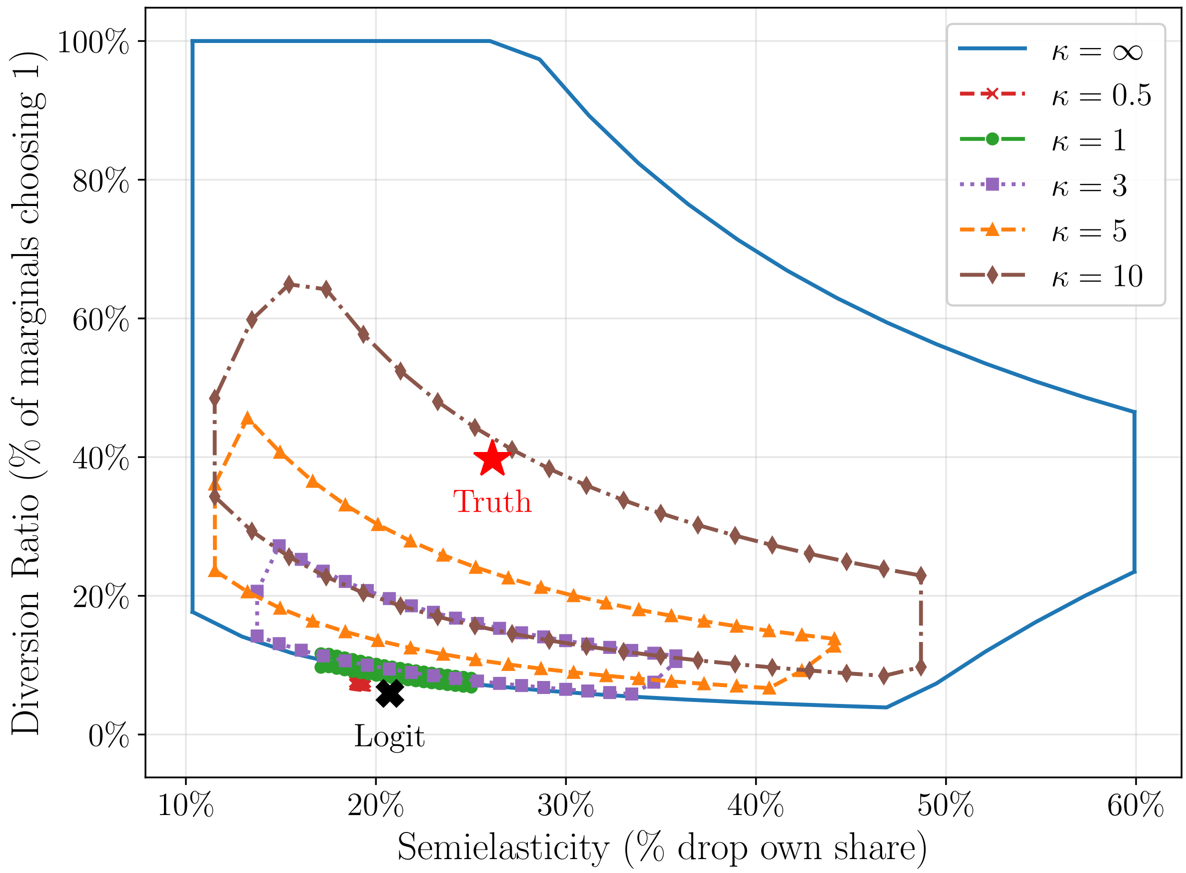


(b) 18-64, Severity Index 2, Female; $T_{\boldsymbol{w}} = 2935$



(c) Under 18, Severity Index 2, Female; $T_{\boldsymbol{w}} = 1336$

Figure 9: Outer Sets for Own-Price Semielasticity and Diversion Ratio



dimensional parameters. To this end, in Figure 9 we focus on the group in the top panel of Figure 8 and show the joint outer sets for the own-price semielasticity of hospital 2 (horizontal axis) and the diversion ratio between hospitals 2 and 1 (vertical axis), which measures the share of marginal buyers who, after leaving hospital 2, would choose hospital 1. This is relevant for assessing the competitive effects of a merger.

The true parameter $\boldsymbol{\theta}(f) = (0.27, 0.4)$ is far from the logit prediction $\boldsymbol{\theta}(g) = (0.2, 0.08)$. Over the grid of κ values considered in this figure, the outer sets contain neither the truth nor the logit for κ between 1 and 5. For $\kappa \geq 10$, they contain $\boldsymbol{\theta}(f)$ while continuing to exclude $\boldsymbol{\theta}(g)$.

6 Conclusion

We developed a new algorithm for characterizing identified sets in a class of nonparametric discrete choice models. We used the computational power afforded by the algorithm to analyze sensitivity to parametric assumptions on latent valuations by considering increasingly different families of distributions around a reference density. In simulations using both artificial and calibrated DGPs, we found that it was tractable to compute informative outer identified sets.

References

AGARWAL, N., P. Z. LI, AND P. J. SOMAINI (2023): "Identification using Revealed Preferences in Linearly Separable Models," Discussion paper, National Bureau of Economic Research.

AHUJA, R. K., T. L. MAGNANTI, AND J. B. ORLIN (1993): *Network Flows: Theory, Algorithms, and Applications*. Prentice hall.

ARDILA, F., AND M. DEVELIN (2009): "Tropical hyperplane arrangements and oriented matroids," *Mathematische Zeitschrift*, 262(4), 795–816.

BELLMAN, R. (1958): "On a routing problem," *Quarterly of applied mathematics*, 16(1), 87–90.

BERRY, S. T., AND P. A. HAILE (2014): "Identification in differentiated products markets using market level data," *Econometrica*, 82(5), 1749–1797.

——— (2024): "Nonparametric identification of differentiated products demand using micro data," *Econometrica*, 92(4), 1135–1162.

BORUSYAK, K., J. CHEN, P. HULL, AND L. LEI (2026): "Nonparametric Identification of Demand Without Exogenous Product Characteristics," Discussion paper, National Bureau of Economic Research.

CHRISTENSEN, T., AND B. CONNAULT (2023): "Counterfactual sensitivity and robustness," *Econometrica*, 91(1), 263–298.

COMPIANI, G. (2022): "Market Counterfactuals and the Specification of Multiproduct Demand: A Nonparametric Approach," *Quantitative Economics*, 13(2), 545–591.

DEVELIN, M., AND B. STURMFELS (2004): "Tropical Convexity," *Documenta Mathematica*, 9, 1–27.

FARKAS, J. (1902): “Theorie der einfachen Ungleichungen.,” *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 1902(124), 1–27.

FORD JR, L. R. (1956): “Network Flow Theory,” *Rand Corporation Paper, Santa Monica.*

FOX, J. T., AND A. GANDHI (2016): “Nonparametric identification and estimation of random coefficients in multinomial choice models,” *The RAND Journal of Economics*, 47(1), 118–139.

GALLAI, T. (1958): “Maximum-minimum sätze über graphen,” *Acta Mathematica Hungarica*, 9(3-4), 395–434.

GU, J., AND T. RUSSELL (2024): “Wasserstein-Robust Counterfactuals,” *Available at SSRN 4517842.*

GU, J., T. M. RUSSELL, AND T. STRINGHAM (2025): “Counterfactual identification and latent space enumeration in discrete outcome models,” *Review of Economic Studies*, p. rdaf058.

HO, K. (2006): “The welfare effects of restricted hospital choice in the US medical care market,” *Journal of Applied Econometrics*, 21(7), 1039–1079.

——— (2009): “Insurer-provider networks in the medical care market,” *American Economic Review*, 99(1), 393–430.

HO, K., AND A. PAKES (2014): “Hospital choices, hospital prices, and financial incentives to physicians,” *American Economic Review*, 104(12), 3841–3884.

HOROWITZ, J. L., AND C. F. MANSKI (1995): “Identification and Robustness with Contaminated and Corrupted Data,” *Econometrica*, 63(2), 281–302.

KAIDO, H., AND F. MOLINARI (2024): “Information Based Inference in Models with Set-Valued Predictions and Misspecification,” .

KAMAT, V., AND S. NORRIS (2025): “Estimating Welfare Effects in a Non-Parametric Choice Model: The Case of School Vouchers,” *Review of Economic Studies*, p. rdaf064.

KLINE, P., AND A. SANTOS (2013): “Sensitivity to missing data assumptions: Theory and an evaluation of the U.S. wage structure: Sensitivity to missing data assumptions,” *Quantitative Economics*, 4(2), 231–267.

LAVINE, M. (1991): “Sensitivity in Bayesian Statistics: The Prior and the Likelihood,” *Journal of the American Statistical Association*, 86, 396–399.

MACLAGAN, D., AND B. STURMFELS (2015): *Introduction to Tropical Geometry*, vol. 161. American Mathematical Society.

MANSKI, C. F. (2007): “Partial identification of counterfactual choice probabilities,” *International Economic Review*, 48(4), 1393–1410.

MASTEN, M. A., AND A. POIRIER (2020): "Inference on breakdown frontiers," *Quantitative Economics*, 11(1), 41–111.

MATZKIN, R. L. (1993): "Nonparametric Identification and Estimation of Polychotomous Choice Models," *Journal of Econometrics*, 58(1-2), 137–168.

MOLINARI, F. (2020): "Microeconometrics with Partial Identification," *Handbook of Econometrics*, 7, 355–486.

PAKES, A., AND J. PORTER (2024): "Moment inequalities for multinomial choice with fixed effects," *Quantitative Economics*, 15(1), 1–25.

ROCHET, J.-C. (1987): "A necessary and sufficient condition for rationalizability in a quasi-linear context," *Journal of mathematical Economics*, 16(2), 191–200.

TEBALDI, P., A. TORGOVITSKY, AND H. YANG (2023): "Nonparametric estimates of demand in the california health insurance exchange," *Econometrica*, 91(1), 107–146.

THOMPSON, T. S. (1989): "Identification of semiparametric discrete choice models," .

VAN ERVEN, T., AND P. HARREMOES (2014): "Rényi Divergence and Kullback-Leibler Divergence," *IEEE Transactions on Information Theory*, 60, 3797–3820.

VOHRA, R. V. (2011): *Mechanism design: a linear programming approach*, vol. 47. Cambridge University Press.

WASSERMAN, L., AND J. B. KADANE (1992): "Computing bounds on expectations," *Journal of the American Statistical Association*, 87(418), 516–522.

Appendix

A Labeling and Negative Cycles on Directed Weighted Graphs

As discussed in Ahuja, Magnanti, and Orlin (1993), an important literature within operations research has actively studied procedures to detect shortest paths in weighted directed graphs, and the dual problem of finding "labels" $d : \mathcal{J} \to \overline{\mathbb{R}}$ (where $\overline{\mathbb{R}} \equiv \mathbb{R} \cup \{+\infty\}$ denotes the extended reals) for the nodes $\mathcal{J}$ of such graphs to obtain $w(j,k) + d(j) - d(k) \geq 0$ for all $j \in \mathcal{J}$, all $k \in \mathcal{J}$, where $w(j,k)$ is the real-valued weight assigned to the edge pointing at k from j.[11] We adopt the usual convention by which $w(j,k) = \infty$ if there is no edge from j to k, and $w(j,j) = \infty$. Leveraging Farkas (1902), Gallai (1958) proved that a necessary and sufficient condition for the labeling problem to admit a solution is the absence of negative cycles in the graph.

To see this constructively, in the absence of negative cycles, a valid solution $d(\cdot)$ for the labeling problem is obtained by setting $d(j)$ equal to the minimum distance (or length of the shortest path) to node j from a source node ι that is artificially added to the graph, and linked in one direction to all nodes with $w(\iota, k) = 0$, and $w(k, \iota) = \infty$ for all $k \in \mathcal{J}$. If there are no negative cycles, the lengths of all shortest paths from ι are finite and well-defined. For any two nodes j and k it must then be that $d(k) \leq d(j) + w(j,k)$, otherwise following the shortest path from ι to j and then traveling the edge from j to k would be shorter than the shortest path from ι to k. If there is a negative cycle $j_1, j_2, \ldots, j_{n-1}, j_n = j_1$, then for any solution $d(\cdot)$ of the labeling problem one has the

[11] Particularly relevant to our purposes is that this labeling problem is equivalent to finding a vector $\boldsymbol{d} \in \overline{\mathbb{R}}^{|\mathcal{J}|}$ satisfying a system of linear inequalities $A\boldsymbol{d} \leq \boldsymbol{w}$, where each row of the incidence-like matrix A contains exactly one $+1$ and one -1 (a Thompson-like operator; cf. Thompson, 1989), and the right-hand side $\boldsymbol{w}$ is the vector of edge weights $w(j,k)$ indexed by ordered pairs. This equivalence is discussed in Application 4.5 in Ahuja, Magnanti, and Orlin (1993).

following contradiction:

$$0 > \sum_{\ell=1}^{n-1} w(j_\ell, j_{\ell+1}) = \sum_{\ell=1}^{n-1} w(j_\ell, j_{\ell+1}) + d(j_\ell) - d(j_{\ell+1}) \geq 0.$$

Therefore, to rule out negative cycles and ensure that the labeling problem admits a solution, one can adapt the efficient Bellman-Ford algorithm (Bellman, 1958; Ford Jr, 1956) which solves the single-source (ι) shortest path problem:

Algorithm BF Detect negative cycles in a directed weighted graph

Require: Set of nodes $\mathcal{J}$ and directed edges' weights $w : \mathcal{J} \times \mathcal{J} \to \overline{\mathbb{R}}$

Add single source node ι to graph
Set $w(\iota, j) = 0$ and $w(j, \iota) = \infty$ for all $j \in \mathcal{J}$
Initialize $d : \mathcal{J} \cup \{\iota\} \to \overline{\mathbb{R}}$ with $d(j) = \infty$ for all $j \in \mathcal{J}$, $d(\iota) = 0$
for $m = 1, \ldots, |\mathcal{J}|$ **do**
    for all edges $(j, k) \in (\mathcal{J} \cup \{\iota\}) \times (\mathcal{J} \cup \{\iota\})$ with $w(j, k) < \infty$ **do**
        if $d(j) + w(j, k) < d(k)$ **then**
            $d(k) \leftarrow d(j) + w(j, k)$
        end if
    end for
end for
Initialize "`Negative Cycles`" $\leftarrow$ `NO`
for all edges $(j, k) \in (\mathcal{J} \cup \{\iota\}) \times (\mathcal{J} \cup \{\iota\})$ with $w(j, k) < \infty$ **do**
    if $d(j) + w(j, k) < d(k)$ **then**
        "`Negative Cycles`" $\leftarrow$ `YES`
        break
    end if
end for
return `Negative Cycles`

The key intuition behind the algorithm is that since $w(\iota, j) = 0$ for all j, iterating over all edges $|\mathcal{J}|$ times guarantees that all simple paths (paths with no repeated vertices) are considered. The final step is then able to verify the presence of negative cycles, which are the only way a path from ι could still be shortened at iteration $|\mathcal{J}| + 1$.

In terms of runtime, if $E \leq |\mathcal{J}|^2$ is the number of finite edges in the original graph, an upper bound on the time complexity of the algorithm is $O\left(\left(|\mathcal{J}| + 1\right)\left(E + |\mathcal{J}|\right)\right)$.[12] Ef-

[12] Runtime can be lower since the routine can be stopped and return "`Negative Cycles` = `No`" as soon

ficiency in terms of memory requirements is excellent, since one only needs to store the edges' weights $w(\cdot,\cdot)$ of size $(|\mathcal{J}|+1)^2$, and the labels $d(\cdot)$ of size $|\mathcal{J}|+1$.[13]

B A Formula for $|\mathbf{MRP}(\mathcal{P})|$

Proposition A1. Fix $t \in \mathcal{T}$ and $j \in \mathcal{J} \setminus \{0\}$. Suppose that $P_{j,t}$ is drawn from a continuous distribution $F_{j,t}$, independently over t and j. Then the number of elements of the MRP induced by P is almost surely constant and given by

$$\binom{J+T}{T}.$$

Proof. For this proof only, we denote the set of markets by $\mathcal{T} = \{1, \ldots, T\}$. The proof relies on tropical geometry, see Maclagan and Sturmfels (2015) for an excellent textbook treatment. Since tropical geometry is not part of the standard toolkit of economics, we here give a brief introduction. The key result is established Develin and Sturmfels (2004), but we use the notation of Ardila and Develin (2009) since it more transparently maps to our setting.

We define for $\boldsymbol{v}, \boldsymbol{p} \in \mathbb{R}^J$

$$U_{\boldsymbol{p}}(\boldsymbol{v}) = \max\{0, v_1 - p_1, \ldots, v_J - p_J\}. \tag{14}$$

We define the tropical hypersurface of $U_{\boldsymbol{p}}$ as

$$H_{\boldsymbol{p}} \equiv \{\boldsymbol{v} \in \mathbb{R}^J \quad | \quad \text{the maximum in (14) is attained at least twice.}\}$$

For $J = 2$, the tropical hypersurface is a collection of three lines starting at $\boldsymbol{p}$ as illus-

as iterating over all edges does not update any label.

[13]These runtime and memory efficiency properties deliver the scalability of our use of Algorithm BF to build the MRPs as the number of possible choices J increases (Section 3).

trated in figure 1 of Ardila and Develin (2009). In words, $H_{\boldsymbol{p}}$ is the union of those pieces of the ordinary indifference hyperplanes $\{\boldsymbol{v} \mid v_j - p_j = v_k - p_k\}$ on which the corresponding indifference $v_j - a_j = v_k - a_k$ is in fact the running maximum over all of $\mathcal{J}$. The complement $\mathbb{R}^J \setminus H_a$ splits into $J+1$ open connected components, one for each choice of a consumer with valuation in that component.

Given a price matrix $P \in \mathbb{R}^{T \times J}$, $\mathcal{A}(P) \equiv \{H_{\boldsymbol{P}_{t,\cdot}}\}_{t=1}^{T}$ is the *tropical hyperplane arrangement* associated with P. Its *chambers* are the connected components of $\mathbb{R}^J \setminus \bigcup_{t=1}^{T} H_{\boldsymbol{P}_{t,\cdot}}$. These chambers have a one-to-one correspondence to MRP choice keys. We henceforth identify the MRP cells with the chambers of $\mathcal{A}(P)$. We thus have to study the number of chambers in $\mathcal{A}(P)$. For this, we introduce some notation and concepts.

- Δ_{T-1} is the standard $(T-1)$-simplex, with T vertices labelled $1, \ldots, T$.
- Δ_J denotes the standard J-simplex, with $J+1$ vertices labelled $0, 1, \ldots, J$.
- $\Delta_{T-1} \times \Delta_J$ is their Cartesian product, a polytope of dimension $T+J-1$ whose $T(J+1)$ vertices are the pairs $(t, j) \in \mathcal{T} \times \mathcal{J}$.
- *P-Vertex-lifting* is the map $\ell : \mathcal{T} \times \mathcal{J} \to \mathcal{T} \times \mathcal{J} \times \mathbb{R}$ defined by $\ell(t, j) = (t, j, P_{t,j})$.
- Let $\mathcal{Q}$ be the convex hull of all P-lifted vertices. Then $\mathcal{Q}$ is a polytope, i.e., a bounded polyhedron. Denote $\mathcal{W} := \Delta_{T-1} \times \Delta_J \times \mathbb{R}$.
- A subset $F \subseteq \mathcal{Q}$ is called a *face* of $\mathcal{Q}$ if there exists a hyperplane

  $$H = \{w \in \mathcal{W} : a^\top w = c\}$$

  such that $\mathcal{Q}$ is contained in one of the two closed halfspaces determined by H, and

  $$F = \mathcal{Q} \cap H.$$

Equivalently, F is the subset of $\mathcal{Q}$ where some valid linear inequality is binding.

- We call a face F "upward facing" if for all $(t,j,h) \in F$ and all $\varepsilon > 0$ it holds that $(t,j,h) + (0,0,\varepsilon) \notin \mathcal{Q}$. In the example of a three-dimensional cube with vertices $\{(0,0,0),(1,0,0),(0,1,0),(0,0,1),(0,1,1),(1,0,1),(1,1,0),(1,1,1)\}$, the upward facing face is the face spanned by $\{(0,0,1),(0,1,1),(1,0,1),(1,1,1)\}$.

- Let $\pi : \mathcal{W} \to \Delta_{T-1} \times \Delta_J$ be the projection, i.e., $\pi(t,j,h) = (t,j)$.

- The π-projections of the upward-facing faces form the cells of the regular P-subdivision of $\Delta_{T-1} \times \Delta_J$.

- For each $t \in \{1,\dots,T\}$, the *t-th copy of* Δ_J inside $\Delta_{T-1} \times \Delta_J$ is the face spanned by the $J+1$ vertices $\{(t,0),(t,1),\dots,(t,J)\}$. There are T such copies in total, one per market.

Then we can state the key result on the tropical hyperplane arrangement by relating it to regular P-subdivisions.

Theorem (Theorems 2.3 and 2.4 of Ardila and Develin (2009))**.**
The tropical hyperplane arrangement $\mathcal{A}(P)$ is combinatorially isomorphic to the subcomplex of the regular P-subdivision of $\Delta_{T-1} \times \Delta_J$ consisting of faces that contain at least one vertex from each of the T copies of Δ_J.

All that remains is to study the number of regular P-subdivisions.

Theorem (Corollary 25 of Develin and Sturmfels (2004))**.**
Fix $t \in \mathcal{T}$ and $j \in \mathcal{J} \setminus \{0\}$. Suppose that $P_{j,t}$ is drawn from a continuous distribution $F_{j,t}$, independently over t and j. The with probability one, every regular P-subdivision

of $\Delta_{T-1} \times \Delta_J$ has the same number of chambers. This number is given by

$$\sum_{k=0}^{J} \binom{T+J+1-k-2}{T-k-1, J+1-k-1, k} = \binom{J+T}{T},$$

where the formula uses Vandermonde's identity.

This completes the proof. $\square$

C Unique Construction of θRP

We first note a simple permutation-invariance property of the MRP. Recall the equivalent representation of $\text{MRP}(\mathcal{P})$ as $\mathbb{V}(\mathcal{P})$ introduced in (8). For a collection of prices $\mathcal{P} = (P_t)_{t\in\mathcal{T}}$, and any permutation $\pi : \mathcal{T} \to \mathcal{T}$, we have that $\mathbb{V}(\mathcal{P}) = \mathbb{V}(\mathcal{P}_\pi)$, where $\mathcal{P}_\pi = (P_{\pi(t)})_{t\in\mathcal{T}}$. In other words, letting $\boldsymbol{y}^\pi$ be defined by setting $y^\pi_t = y_t$ for all $t \in \mathcal{T}$, we have that $\boldsymbol{y} \in \mathbb{V}(\mathcal{P})$ if and only if $\boldsymbol{y}^\pi \in \mathbb{V}(\mathcal{P}_\pi)$.

Now consider the setting of section 4. This permutation-invariance property shows that the first ingredient to the algorithm for building $\theta\text{RP}(\mathcal{P})$, namely the initialization $\theta\text{RP}^0 = \text{MRP}(\boldsymbol{p}^\star)$ using Algorithm 1, is permutation-invariant. If the set of included prices is invariant to the order of prices, then the permutation-invariance property above, applied to the selected prices, implies that the MRP is permutation-invariant.

Next, we show that for any two permutations of prices, the set of included prices is constant. To see this, note that a price $\check{\boldsymbol{p}}$ splits a set in $\mathcal{V}^\theta(\boldsymbol{p}^\star)$ if and only if it splits a set in $\mathcal{V}^\theta(\boldsymbol{p}^\star \cup \tilde{P})$ for any finite set of prices $\tilde{P}$ which does not contain $\check{\boldsymbol{p}}$ and such that the full set of prices is in general position. This follows from the fact that, under general position, any MRP element induced by a finite set of prices has nonempty interior. If an MRP element in $\mathcal{V}^\theta(\boldsymbol{p}^\star)$ is split by $\check{\boldsymbol{p}}$, then adding the set of prices $\tilde{P}$ to the partition either leaves the set unchanged, in which case there is nothing to prove, or refines the set into multiple MRP elements, all of which have nonempty interior. Since $\check{\boldsymbol{p}}$ split the original set, there exist distinct $k_1, k_2 \in \mathcal{J}$ such that the original set intersects both $\mathcal{V}_{k_1}(\check{\boldsymbol{p}})$ and $\mathcal{V}_{k_2}(\check{\boldsymbol{p}})$. As long as prices are in general position, at least one refined MRP element must intersect both regions and is therefore split by $\check{\boldsymbol{p}}$.

Finally, we note that $\check{\boldsymbol{p}}$ splitting a set $\mathcal{V} \in \mathcal{V}^\theta(\boldsymbol{p}^\star)$ is equivalent to non-existence of a $j \in \mathcal{J}$ such that

$$\mathcal{V} \subset \mathcal{V}_j(\check{\boldsymbol{p}}).$$

Suppose $\check{\boldsymbol{p}}$ splits an MRP set $\mathcal{V} \in \mathcal{V}^\theta(\boldsymbol{p}^\star)$ into at least two MRP elements, appending choice keys $k_1 \in \mathcal{J}$ to one and $k_2 \in \mathcal{J}$, with $k_2 \neq k_1$, to the other. Because $(\mathcal{V}_j(\check{\boldsymbol{p}}))_{j\in\mathcal{J}}$ is a partition, there cannot exist any $j \in \mathcal{J}$ such that $\mathcal{V}_j(\check{\boldsymbol{p}})$ contains $\mathcal{V}$ and therefore $\mathcal{V}^\theta(\boldsymbol{p}^\star)$: the MRP element appended by k_1 is contained in $\mathcal{V}_{k_1}(\check{\boldsymbol{p}})$, while the MRP element appended by k_2 is contained in $\mathcal{V}_{k_2}(\check{\boldsymbol{p}})$. Since $k_1 \neq k_2$, no single partition element can contain $\mathcal{V}$. Thus, the inclusion property fails.

D Proofs of Proposition 1 and Proposition 2

D.1 Proof of Proposition 1

Note first that $\Phi(\kappa)$ is a non-empty, compact, connected set, so the feasible set of (4) is also non-empty, compact, and connected. If t is continuous then its image over this feasible set is non-empty, compact, and connected as well. If $d_\theta = 1$, then this image—which is $\Theta^+(\kappa)$—is a closed interval with endpoints given by the optimal values of (4).

Now suppose that $\min_{\phi' \in \Phi(\kappa)} Q(\phi') = 0$. Given the compactness of the feasible set, this implies that there exists a $\phi \in \Phi(\kappa)$ such that $Q(\phi) = 0$. The definition of Q means that this ϕ must satisfy $\mathcal{s}_j(\boldsymbol{p}_t; \phi) = s_j(\boldsymbol{p}_t)$ for all $j \in \mathcal{J}$ and $t = 1, \ldots, T$. We conclude that

$$\Theta^+(\kappa) = \left\{\vartheta \in \mathbb{R}^{d_\theta} : \vartheta = t(\phi) \quad \text{for some } \phi \in \Phi(\kappa) \text{ s.t. } \mathcal{s}_j(\boldsymbol{p}_t; \phi) = s_j(\boldsymbol{p}_t) \text{ for all } j \text{ and } t\right\}.$$

The remainder of the proof shows that $\Theta^+(\kappa) = \Theta^\star(\kappa) \equiv \{\theta(f) : f \in \mathcal{F}^\star(\kappa)\}$.

Suppose first $\vartheta \in \Theta^\star(\kappa)$, so that there exists an $f \in \mathcal{F}^\star(\kappa)$ such that $\theta(f) = \vartheta$. Because f is a density and $\text{MRP}(\mathcal{P})$ is a partition of $\mathbb{R}^J$, $\phi^f(\mathcal{V}) \equiv \int_{\mathcal{V}} f(\boldsymbol{v})\,d\boldsymbol{v} \in [0,1]$ for all $\mathcal{V}$ and

$$\sum_{\mathcal{V}\in\text{MRP}(\mathcal{P})} \phi^f(\mathcal{V}) = \sum_{\mathcal{V}\in\text{MRP}(\mathcal{P})} \int_{\mathcal{V}} f(\boldsymbol{v})\,d\boldsymbol{v} = \int_{\bigcup_{\mathcal{V}\in\text{MRP}(\mathcal{P})}} f(\boldsymbol{v})\,d\boldsymbol{v} = \int f(\boldsymbol{v})\,d\boldsymbol{v} = 1.$$

In addition,

$$\phi^f(\mathcal{V}) \equiv \int_{\mathcal{V}} f(\boldsymbol{v})\,d\boldsymbol{v} \in \left[(1-\kappa)\int_{\mathcal{V}} g(\boldsymbol{v})\,d\boldsymbol{v}, (1+\kappa)\int_{\mathcal{V}} g(\boldsymbol{v})\,d\boldsymbol{v}\right],$$

so that $\phi^f \in \Phi(\kappa)$. Moreover, for any j and t,

$$\begin{aligned}
\mathscr{s}_j(\boldsymbol{p}_t;\phi^f) &\equiv \sum_{\mathcal{V}\in\text{MRP}(\mathcal{P})} \mathbf{1}\left[\mathcal{V} \subseteq \mathcal{V}_j(\boldsymbol{p}_t)\right]\phi^f(\mathcal{V}) \\
&= \sum_{\mathcal{V}\in\text{MRP}(\mathcal{P})} \int_{\mathcal{V}} \mathbf{1}\left[\mathcal{V} \subseteq \mathcal{V}_j(\boldsymbol{p}_t)\right] f(\boldsymbol{v})\,d\boldsymbol{v} \\
&= \int_{\mathcal{V}_j(\boldsymbol{p}_t)} f(\boldsymbol{v})\,d\boldsymbol{v} = s_j(\boldsymbol{p}_t),
\end{aligned}$$

where the third equality follows from the MRP property, which implies that the elements of $\text{MRP}(\mathcal{P})$ contained in $\mathcal{V}_j(\boldsymbol{p}_t)$ form a partition of $\mathcal{V}_j(\boldsymbol{p}_t)$, and the final equality follows because $f \in \mathcal{F}^\star(\kappa)$. Finally, Assumption TP directly ensures that $t(\phi^f) = \theta(f) = \vartheta$. We conclude that $\vartheta \in \Theta^+(\kappa)$.

Conversely, suppose that $\vartheta \in \Theta^+(\kappa)$. Then there exists a $\phi^\star \in \Phi(\kappa)$ that satisfies $t(\phi^\star) = \vartheta$ and $\mathscr{s}_j(\boldsymbol{p}_t;\phi^\star) = s_j(\boldsymbol{p}_t)$ for all $j \in \mathcal{J}$ and $t = 1,\ldots,T$. We use $\phi^\star$ to construct the function

$$f^{\star}(\boldsymbol{v}) \equiv \sum_{\mathcal{V} \in \mathrm{MRP}(\mathcal{P})} \mathbf{1}\left[\boldsymbol{v} \in \mathcal{V}\right] \left(\frac{g(\boldsymbol{v})}{\displaystyle\int_{\mathcal{V}} g(\boldsymbol{v}')\, d\boldsymbol{v}'} \right) \phi^{\star}(\mathcal{V}),$$

which we now show is an element of $\mathcal{F}^{\star}(\kappa)$. Because g is a density and $\phi^{\star} \in \Phi$, we know that $f^{\star}(\boldsymbol{v}) \geq 0$ for all $\boldsymbol{v}$ and that the integral of $f^{\star}$ satisfies

$$\int f^{\star}(\boldsymbol{v})\, d\boldsymbol{v} = \sum_{\mathcal{V} \in \mathrm{MRP}(\mathcal{P})} \phi^{\star}(\mathcal{V}) \int \frac{\mathbf{1}\left[\boldsymbol{v} \in \mathcal{V}\right] g(\boldsymbol{v})}{\displaystyle\int_{\mathcal{V}} g(\boldsymbol{v}')\, d\boldsymbol{v}'} d\boldsymbol{v} = \sum_{\mathcal{V} \in \mathrm{MRP}(\mathcal{P})} \phi^{\star}(\mathcal{V}) = 1.$$

Moreover, because $\phi^{\star} \in \Phi(\kappa)$,

$$f^{\star}(\boldsymbol{v}) \leq \sum_{\mathcal{V} \in \mathrm{MRP}(\mathcal{P})} \mathbf{1}\left[\boldsymbol{v} \in \mathcal{V}\right] \left(\frac{g(\boldsymbol{v})}{\displaystyle\int_{\mathcal{V}} g(\boldsymbol{v}')\, d\boldsymbol{v}'} \right) (1+\kappa) \int_{\mathcal{V}} g(\boldsymbol{v}')\, d\boldsymbol{v}' = (1+\kappa) g(\boldsymbol{v}),$$

and similarly $f^{\star}(\boldsymbol{v}) \geq (1-\kappa) g(\boldsymbol{v})$. Moreover, for any j and t,

$$\begin{aligned}
\sigma_j(\boldsymbol{p}_t; f^{\star}) &\equiv \int_{\mathcal{V}_j(\boldsymbol{p}_t)} f^{\star}(\boldsymbol{v})\, d\boldsymbol{v} \\
&= \sum_{\mathcal{V} \in \mathrm{MRP}(\mathcal{P})} \mathbf{1}\left[\mathcal{V} \subseteq \mathcal{V}_j(\boldsymbol{p}_t)\right] \phi^{\star}(\mathcal{V}) \frac{\displaystyle\int_{\mathcal{V}} g(\boldsymbol{v})\, d\boldsymbol{v}}{\displaystyle\int_{\mathcal{V}} g(\boldsymbol{v}')\, d\boldsymbol{v}'} \\
&= \sum_{\mathcal{V} \in \mathrm{MRP}(\mathcal{P})} \mathbf{1}\left[\mathcal{V} \subseteq \mathcal{V}_j(\boldsymbol{p}_t)\right] \phi^{\star}(\mathcal{V}) \\
&\equiv \mathscr{s}_j(\boldsymbol{p}_t; \phi^{\star}) = s_j(\boldsymbol{p}_t).
\end{aligned}$$

We conclude that $f^\star \in \mathcal{F}^\star(\kappa)$. Assumption TP then implies that $\vartheta \in \Theta^\star(\kappa)$ because

$$\phi^{f^\star}(\mathcal{V}) \equiv \int_{\mathcal{V}} f^\star(\boldsymbol{v})\, d\boldsymbol{v} = \int_{\mathcal{V}} \sum_{\mathcal{V}' \in \text{MRP}(\mathcal{P})} \mathbf{1}[\boldsymbol{v} \in \mathcal{V}'] \frac{g(\boldsymbol{v})}{\int_{\mathcal{V}} g(\boldsymbol{v}')\, d\boldsymbol{v}'} \phi^\star(\mathcal{V}')\, d\boldsymbol{v} = \phi^\star(\mathcal{V}),$$

so that $\theta(f^\star) = t(\phi^{f^\star}) = t(\phi^\star) = \vartheta$.

□

D.2 Proof of Proposition 2

If $\Theta^\star(\kappa)$ is empty, then the statement is trivial, because the empty set is a subset of all other sets. So, suppose that $\Theta^\star(\kappa)$ is non-empty. Let $\vartheta \in \Theta^\star(\kappa)$. Then there exists a $\phi \in \Phi(\kappa)$ such that $t(\phi) = \vartheta$ and $Q(\phi) = 0$. We use this ϕ to construct a $\phi^\theta \in \Phi^\theta$ defined as

$$\phi^\theta(\mathcal{V}) \equiv \sum_{\mathcal{V}' \in \text{MRP}(\mathcal{P})} \mathbf{1}[\mathcal{V}' \subseteq \mathcal{V}] \phi(\mathcal{V}) \quad \text{for any } \mathcal{V} \in \theta\text{RP}(\mathcal{P}).$$

We claim that (i) ϕ^θ satisfies $t^\theta(\phi^\theta) = \vartheta$, (ii) $\phi^\theta \in \Phi^\theta(\kappa)$, and (iii) $Q_{\text{OUT}}(\phi^\theta) = 0$. These three statements together imply that $\vartheta \in \Theta_{\text{OUT}}(\kappa)$, which verifies the claim.

To see that (i) is true, notice that for every θ-relevant set $\mathcal{V} \subseteq \mathcal{V}^\theta(\mathcal{P})$,

$$\phi^\theta(\mathcal{V}) = \sum_{\mathcal{V}' \in \text{MRP}(\mathcal{P})} \mathbf{1}[\mathcal{V}' \subseteq \mathcal{V}] \phi(\mathcal{V}') = \phi(\mathcal{V}),$$

because such a $\mathcal{V}$ is itself a single element of $\text{MRP}(\mathcal{P})$. Thus ϕ agrees with ϕ^θ on every θ-relevant set, so by the invariance property noted after the definition of t^θ in (10), $t^\theta(\phi^\theta) = t(\phi)$. Since $t(\phi) = \vartheta$, this implies (i).

Statement (ii) has two parts. First $\phi^\theta \in \Phi^\theta \equiv \Phi^\theta(\infty)$ because $\phi(\mathcal{V}) \in [0,1]$ and

$$\sum_{\mathcal{V}\in\theta\text{RP}(\mathcal{P})} \phi^\theta(\mathcal{V}) = \sum_{\mathcal{V}'\in\text{MRP}(\mathcal{P})} \left(\sum_{\mathcal{V}\in\theta\text{RP}(\mathcal{P})} \mathbf{1}[\mathcal{V}' \subseteq \mathcal{V}] \right) \phi(\mathcal{V}') = \sum_{\mathcal{V}'\in\text{MRP}(\mathcal{P})} \phi(\mathcal{V}') = 1,$$

where the second equality follows because $\text{MRP}(\mathcal{P})$ is a finer partition of $\theta\text{RP}(\mathcal{P})$, so any fixed $\mathcal{V}' \in \text{MRP}(\mathcal{P})$ is a subset of exactly one $\mathcal{V} \in \theta\text{RP}(\mathcal{P})$. Second, $\phi^\theta \in \Phi^\theta(\kappa)$ because if $\mathcal{V} \in \theta\text{RP}(\mathcal{P})$ then

$$\begin{aligned}
\phi^\theta(\mathcal{V}) &= \sum_{\mathcal{V}'\in\text{MRP}(\mathcal{P})} \mathbf{1}[\mathcal{V}' \subseteq \mathcal{V}]\phi(\mathcal{V}') \\
&\geq \sum_{\mathcal{V}'\in\text{MRP}(\mathcal{P})} \mathbf{1}[\mathcal{V}' \subseteq \mathcal{V}](1-\kappa) \int_{\mathcal{V}'} g(\boldsymbol{v})\, d\boldsymbol{v} = (1-\kappa) \int_{\mathcal{V}} g(\boldsymbol{v})\, d\boldsymbol{v},
\end{aligned}$$

where the final equality follows because $\text{MRP}(\mathcal{P})$ is a finer partition of $\theta\text{RP}(\mathcal{P})$. The opposite inequality shows that $\phi^\theta(\mathcal{V}) \leq (1+\kappa) \int_{\mathcal{V}} g(\boldsymbol{v})\, d\boldsymbol{v}$, so that $\phi^\theta \in \Phi^\theta(\kappa)$.

To verify (iii), observe that

$$\begin{aligned}
\underline{s}_j(\boldsymbol{p}_t; \phi^\theta) &= \sum_{\mathcal{V}\in\theta\text{RP}(\mathcal{P})} \mathbf{1}[\mathcal{V} \subseteq \mathcal{V}_j(\boldsymbol{p}_t)]\phi^\theta(\mathcal{V}) \\
&= \sum_{\mathcal{V}'\in\text{MRP}(\mathcal{P})} \left(\sum_{\mathcal{V}\in\theta\text{RP}(\mathcal{P})} \mathbf{1}[\mathcal{V}' \subseteq \mathcal{V}]\mathbf{1}[\mathcal{V} \subseteq \mathcal{V}_j(\boldsymbol{p}_t)] \right) \phi(\mathcal{V}) \\
&\leq \sum_{\mathcal{V}'\in\text{MRP}(\mathcal{P})} \mathbf{1}[\mathcal{V}' \subseteq \mathcal{V}_j(\boldsymbol{p}_t)]\phi(\mathcal{V}') = s_j(\boldsymbol{p}_t; \phi),
\end{aligned}$$

where the inequality follows because $\text{MRP}(\mathcal{P})$ is a finer partition than $\theta\text{RP}(\mathcal{P})$—so a $\mathcal{V}' \in \text{MRP}(\mathcal{P})$ can be a subset of at most one $\mathcal{V} \in \theta\text{RP}(\mathcal{P})$—and because if $\mathcal{V}' \subseteq \mathcal{V}$ and $\mathcal{V} \subseteq \mathcal{V}_j(\boldsymbol{p}_t)$ for some $\mathcal{V} \in \theta\text{RP}(\mathcal{P})$, then it is also true that $\mathcal{V}' \subseteq \mathcal{V}_j(\boldsymbol{p}_t)$. Applying similar reasoning to the outer probability, we also get that

$$\begin{aligned}
\overline{\mathscr{s}}_j(\boldsymbol{p}_t;\phi^\theta) &= \sum_{\mathcal{V}\in\theta\mathrm{RP}(\mathcal{P})} \mathbf{1}[\mathcal{V}\cap\mathcal{V}_j(\boldsymbol{p}_t)\neq\emptyset]\phi^\theta(\mathcal{V}) \\
&= \sum_{\mathcal{V}'\in\mathrm{MRP}(\mathcal{P})} \left(\sum_{\mathcal{V}\in\theta\mathrm{RP}(\mathcal{P})} \mathbf{1}[\mathcal{V}'\subseteq\mathcal{V}]\mathbf{1}[\mathcal{V}\cap\mathcal{V}_j(\boldsymbol{p}_t)\neq\emptyset] \right) \phi(\mathcal{V}) \\
&\geq \sum_{\mathcal{V}'\in\mathrm{MRP}(\mathcal{P})} \mathbf{1}[\mathcal{V}'\subseteq\mathcal{V}_j(\boldsymbol{p}_t)]\phi(\mathcal{V}') = \mathscr{s}_j(\boldsymbol{p}_t;\phi),
\end{aligned}$$

because if $\mathcal{V}'\subseteq\mathcal{V}$ but $\mathcal{V}\cap\mathcal{V}_j(\boldsymbol{p}_t)=\emptyset$, then $\mathcal{V}'$ cannot be a subset of $\mathcal{V}_j(\boldsymbol{p}_t)$. From these two sets of inequalities, we conclude that

$$\begin{aligned}
Q_{\mathrm{OUT}}(\phi^\theta) &\equiv \sum_{t=1}^{T}\sum_{j\in\mathcal{J}} \max\{0, \underline{\mathscr{s}}_j(\boldsymbol{p}_t;\phi^\theta) - s_j(\boldsymbol{p}_t)\} + \max\{0, s_j(\boldsymbol{p}_t) - \overline{\mathscr{s}}_j(\boldsymbol{p}_t;\phi^\theta)\} \\
&\leq \sum_{t=1}^{T}\sum_{j\in\mathcal{J}} \max\{0, \mathscr{s}_j(\boldsymbol{p}_t;\phi) - s_j(\boldsymbol{p}_t)\} + \max\{0, s_j(\boldsymbol{p}_t) - \mathscr{s}_j(\boldsymbol{p}_t;\phi)\} \\
&= \sum_{t=1}^{T}\sum_{j\in\mathcal{J}} |\mathscr{s}_j(\boldsymbol{p}_t;\phi) - s_j(\boldsymbol{p}_t)| \equiv Q(\phi).
\end{aligned}$$

By hypothesis, $Q(\phi)=0$, so conclude that $Q_{\mathrm{OUT}}(\phi^\theta)=0$ as well.

□

Supplementary Appendix

S1 Details on Importance Sampling Weights for Reference Density

The purpose of this section is to explain the details of the importance sampling procedure and show how it can be implemented for two of the most commonly used latent utility distributions, T1EV and normal. We begin by introducing the notation which helps us discuss the dependence between the latent utility draws in the classical settings.

Consider a demand estimation problem in which consumers have utility

$$u_j = \delta_j + v_j$$

for $j = 0, 1, \ldots, J$. Here, $j = 0$ denotes the outside option for which $\delta_0 = 0$. We also note that this is observationally equivalent to

$$u_j = \delta_j + \eta_j$$

where $u_0 = 0$ and $\eta_j = v_j - v_0$. To simulate from this for Monte Carlo experiments for the logit, one can draw the v_j from a Type 1 Extreme Value distribution and compute the η's. Denote the Gumbel with CDF $F_G(x) = \exp(-e^{-x})$ and pdf $f_G(x) = \exp\big(-(x + e^{-x})\big)$. Our objective is to approximate the probability mass put on many small sets by the logit or probit probabilities. To reduce the variance of these Monte Carlo estimates, we use importance sampling. For any MRP element $\mathcal{V}$, we compute the smallest rectangle

$$R = \{\, \eta \in \mathbb{R}^J : a_j \leq \eta_j \leq b_j,\ j = 1, \ldots, J \,\}$$

which contains our $\mathcal{V}$. Then we draw from the distribution of v truncated to B, check how many of these draws are in our set of interest, and finally rescale by the probability of B to get an unbiased estimate of the probability of the set of interest.

S1.1 Logit Errors

Conditional on v_0, the variables $\eta_1, \ldots, \eta_J$ are i.i.d.. This allows us to decompose the problem of drawing from η conditional on $\eta \in R$ into two simpler problems. First, draw v_0. Then, draw $v_1, \ldots, v_J$ conditional on v_0 and $\eta \in R$. The unconditional joint density of the v's is

$$f_{v_0, v_1, \ldots, v_J}(t, v_1, \ldots, v_J) = f_G(t) \prod_{j=1}^{J} f_G(v_j).$$

The event $\eta \in R$ is equivalent to

$$v_0 + a_j \leq v_j \leq v_0 + b_j \quad \text{for all } j = 1, \ldots, J.$$

The conditional density of v given that $\eta \in R$ is

$$f_{v_0, v_1, \ldots, v_J | \eta \in R}(v_0, v_1, \ldots, v_J) = \frac{f_G(v_0) \ \prod_{j=1}^{J} f_G(v_j) \, \mathbf{1}_{[v_0 + a_j, \ v_0 + b_j]}(v_j)}{\mathbb{P}[\eta \in R]}.$$

The marginal distribution of v_0 given $\eta \in R$ is

$$\begin{aligned} f(v_0 \mid \eta \in R) &= f_G(v_0) \, \frac{\prod_{j=1}^{J} \int_{v_0 + a_j}^{v_0 + b_j} f_G(v_j) \, dv_j}{\mathbb{P}[\eta \in R]} \\ &= f_G(v_0) \, \frac{\prod_{j=1}^{J} \left[F_G(v_0 + b_j) - F_G(v_0 + a_j) \right]}{\mathbb{P}[\eta \in R]}. \end{aligned}$$

Finally,

$$f_{v_1,\dots,v_J|v_0,\,\eta\in R} = \prod_{j=1}^{J} \frac{f_G(v_j)\,\mathbf{1}_{[a_j+v_0,\,b_j+v_0]}(v_j)}{\big[F_G(v_0+b_j) - F_G(v_0+a_j)\big]}\,.$$

We write $\mathbb{P}[\eta \in \mathcal{V}] = \mathbb{P}[\eta \in \mathcal{V} \mid \eta \in R]\,\mathbb{P}[\eta \in R]$ and define

$$a(v_0) := \int_{\mathbb{R}^J} \mathbf{1}_{\mathcal{V}}(v_0, v_1, \dots, v_J)\; f_{v_0,v_1,\dots,v_J|\eta\in R}(v_0, v_1, \dots, v_J)\, dv_1 \cdots dv_J.$$

Then

$$\begin{aligned}
\mathbb{P}[\eta \in \mathcal{V}] &= \int_{\mathbb{R}} a(v_0)\; f(v_0 \mid \eta \in R)\, dv_0\; \mathbb{P}[\eta \in R] \\
&= \int_{\mathbb{R}} a(v_0)\; f_G(v_0)\; \frac{\prod_{j=1}^{J}\big[F_G(v_0+b_j) - F_G(v_0+a_j)\big]}{\mathbb{P}[\eta \in R]}\, dv_0\; \mathbb{P}[\eta \in R] \\
&= \int_{\mathbb{R}} f_G(v_0)\; a(v_0)\; \prod_{j=1}^{J}\big[F_G(v_0+b_j) - F_G(v_0+a_j)\big]\, dv_0
\end{aligned}$$

Now draw $v_0^s \sim f_G$ and drawing $v_1, \dots, v_J$ from $f_{v_1,\dots,v_J|v_0,\,\eta\in R}$. This yields the importance sampling estimator

$$\widehat{\mathbb{P}}[\eta \in \mathcal{V}] = \frac{1}{S}\sum_{s=1}^{S} \mathbf{1}_{\mathcal{V}}(v_0^s, v_1^s, \dots, v_J^s)\; \prod_{j=1}^{J}\big[F_G(v_0^s+b_j) - F_G(v_0^s+a_j)\big].$$

S1.2 Probit Errors (iid)

Assume $v_j \overset{iid}{\sim} \mathcal{N}(0,1)$ for $j = 0, 1, \dots, J$. Define $\eta_j = v_j - v_0$ and let

$$B = \{\, \eta \in \mathbb{R}^J : a_j \le \eta_j \le b_j,\; j = 1, \dots, J \,\}.$$

As before, conditional on v_0, the variables $\eta_1, \dots, \eta_J$ are independent. Equivalently, conditional on v_0 the v_j are independent and the constraint $\eta \in B$ is the set of interval restrictions

$$v_0 + a_j \le v_j \le v_0 + b_j \quad \text{for all } j.$$

Let ϕ and Φ denote the standard normal pdf and cdf. The unconditional joint density of $(v_0, v_1, \ldots, v_J)$ is

$$f_{v_0,v_1,\ldots,v_J}(t, v_1, \ldots, v_J) = \phi(t) \prod_{j=1}^{J} \phi(v_j).$$

Hence the density of v conditional on $\eta \in B$ is

$$f_{v_0,v_1,\ldots,v_J|\eta\in B}(v_0, v_1, \ldots, v_J) = \frac{\phi(v_0) \prod_{j=1}^{J} \phi(v_j)\, \mathbf{1}_{[v_0+a_j,\ v_0+b_j]}(v_j)}{\mathbb{P}[\eta \in B]}.$$

Integrating out $v_1, \ldots, v_J$ gives the marginal of v_0 given $\eta \in B$:

$$\begin{aligned} f(v_0 \mid \eta \in B) &= \phi(v_0)\, \frac{\prod_{j=1}^{J} \int_{v_0+a_j}^{v_0+b_j} \phi(v_j)\, dv_j}{\mathbb{P}[\eta \in B]} \\ &= \phi(v_0)\, \frac{\prod_{j=1}^{J} \left[\Phi(v_0 + b_j) - \Phi(v_0 + a_j)\right]}{\mathbb{P}[\eta \in B]}. \end{aligned}$$

Conditional on v_0 and $\eta \in B$, the coordinates $v_1, \ldots, v_J$ are independent with densities

$$f_{v_j|v_0,\ \eta\in B}(v_j) = \frac{\phi(v_j)\, \mathbf{1}_{[v_0+a_j,\ v_0+b_j]}(v_j)}{\Phi(v_0 + b_j) - \Phi(v_0 + a_j)} \quad \text{for each } j.$$

Thus,

$$\begin{aligned} \mathbb{P}[\eta \in V] &= \int_{\mathbb{R}} \phi(v_0) \left[\int_{\mathbb{R}^J} \mathbf{1}_V(v_0, v_1, \ldots, v_J)\ \prod_{j=1}^{J} \frac{\phi(v_j)\, \mathbf{1}_{[v_0+a_j,\ v_0+b_j]}(v_j)}{\Phi(v_0 + b_j) - \Phi(v_0 + a_j)}\, dv_1 \cdots dv_J \right] dv_0 \\ &= \int_{\mathbb{R}} a(v_0)\, \phi(v_0)\ \prod_{j=1}^{J} \left[\Phi(v_0 + b_j) - \Phi(v_0 + a_j)\right] dv_0, \end{aligned}$$

Using the same derivation for v_j as i.i.d. standard Gaussians, we obtain the importance sampling Monte Carlo estimator

$$\widehat{\mathbb{P}}[\eta \in V] = \frac{1}{S} \sum_{s=1}^{S} \mathbf{1}_V\big(v_0^s, v_1^s, \ldots, v_J^s\big) \; \prod_{j=1}^{J} \big[\Phi(v_0^s + b_j) - \Phi(v_0^s + a_j)\big].$$

S1.3 Probit Errors with General Covariance

Assume $(v_0, v_1, \ldots, v_J)^\top \sim \mathcal{N}(0, \Sigma)$ for a positive definite $(J+1) \times (J+1)$ matrix Σ. Define $\eta_j = v_j - v_0$ and let

$$B = \{\, \eta \in \mathbb{R}^J : a_j \le \eta_j \le b_j,\ j = 1, \ldots, J \,\}.$$

Partition Σ as

$$\Sigma = \begin{pmatrix} \sigma_{00} & \Sigma_{0,1:J} \\ \Sigma_{1:J,0} & \Sigma_{1:J,1:J} \end{pmatrix}, \quad \Sigma_{1:J,0} = \Sigma_{0,1:J}^\top.$$

Let $\varphi_J(\cdot; \mu, \Omega)$ and $\Phi_J(\cdot; \mu, \Omega)$ denote the J-variate normal pdf and cdf with mean μ and covariance Ω, and let $\phi_{\sigma_{00}}$ denote the univariate normal pdf with mean 0 and variance σ_{00}. The Gaussian conditioning formula gives, for $t \in \mathbb{R}$,

$$\begin{aligned} v_{1:J} \mid v_0 = t \sim \mathcal{N}\big(\mu(t),\ \Omega\big), \\ \mu(t) = \Sigma_{1:J,0}\, \sigma_{00}^{-1}\, t, \\ \Omega = \Sigma_{1:J,1:J} - \Sigma_{1:J,0}\, \sigma_{00}^{-1}\, \Sigma_{0,1:J}. \end{aligned}$$

The constraint $\eta \in B$ becomes $t + a_j \le v_j \le t + b_j$ for all j. The joint density conditional on $\eta \in B$ is

$$f_{v_0, v_{1:J} \mid \eta \in B}(t, v_{1:J}) = \frac{\phi_{\sigma_{00}}(t)\ \varphi_J\big(v_{1:J};\ \mu(t), \Omega\big)\ \mathbf{1}_{[t+a,\ t+b]}(v_{1:J})}{\mathbb{P}[\eta \in B]},$$

where $[t+a,\ t+b] = \prod_{j=1}^{J}[t+a_j,\ t+b_j]$. Integrating out $v_{1:J}$ yields

$$f(v_0 = t \mid \eta \in B) = \phi_{\sigma_{00}}(t)\ \frac{\Phi_J\big([t+a,\ t+b];\ \mu(t), \Omega\big)}{\mathbb{P}[\eta \in B]}.$$

Conditional on $v_0 = t$ and $\eta \in B$, the vector $v_{1:J}$ has the truncated multivariate normal density

$$f_{v_{1:J} \mid v_0 = t,\ \eta \in B}(v_{1:J}) = \frac{\varphi_J\big(v_{1:J};\ \mu(t), \Omega\big)\ \mathbf{1}_{[t+a,\ t+b]}(v_{1:J})}{\Phi_J\big([t+a,\ t+b];\ \mu(t), \Omega\big)}.$$

Thus

$$\mathbb{P}[\eta \in V] = \int_{\mathbb{R}} \phi_{\sigma_{00}}(t) \left[\int_{[t+a,\ t+b]} \mathbf{1}_V\big(t, v_{1:J}\big)\ \varphi_J\big(v_{1:J};\ \mu(t), \Omega\big)\ dv_{1:J} \right] dt.$$

This suggests the same one-dimensional importance sampling scheme with proposal $v_0 \sim \mathcal{N}(0, \sigma_{00})$:

1. Draw $v_0^s \sim \mathcal{N}(0, \sigma_{00})$ independently for $s = 1, \ldots, S$.

2. For each s, draw $v_{1:J}^s$ from the truncated multivariate normal $\mathcal{N}\big(\mu(v_0^s), \Omega\big)$ restricted to $[v_0^s + a,\ v_0^s + b]$.

3. Form the estimator

$$\widehat{\mathbb{P}}[\eta \in V] = \frac{1}{S} \sum_{s=1}^{S} \mathbf{1}_V\big(v_0^s, v_{1:J}^s\big)\ \Phi_J\big([v_0^s + a,\ v_0^s + b];\ \mu(v_0^s), \Omega\big).$$

No separate evaluation of $\mathbb{P}[\eta \in B]$ is needed since it cancels.

Nonzero mean terms

It is straightforward to allow a nonzero mean. Suppose $(v_0, \ldots, v_J)^\top \sim \mathcal{N}(m, \Sigma)$ with $m = (m_0,\ m_{1:J})^\top$. Conditioning gives

$$
\begin{aligned}
v_{1:J} \mid v_0 = t \sim \mathcal{N}\big(\mu_m(t),\ \Omega\big),& \\
\mu_m(t) &= m_{1:J} + \Sigma_{1:J,0}\,\sigma_{00}^{-1}\,(t - m_0), \\
\Omega &= \Sigma_{1:J,1:J} - \Sigma_{1:J,0}\,\sigma_{00}^{-1}\,\Sigma_{0,1:J},
\end{aligned}
$$

and $v_0 \sim \mathcal{N}(m_0, \sigma_{00})$. The truncation region remains $[t+a,\ t+b]$. The estimator becomes

$$
\widehat{\mathbb{P}}[\eta \in V] = \frac{1}{S}\sum_{s=1}^{S} \mathbf{1}_V\big(v_0^s, v_{1:J}^s\big)\ \Phi_J\big([v_0^s + a,\ v_0^s + b];\ \mu_m(v_0^s), \Omega\big),
$$

with $v_0^s \sim \mathcal{N}(m_0, \sigma_{00})$ and $v_{1:J}^s$ drawn from the truncated $\mathcal{N}\big(\mu_m(v_0^s), \Omega\big)$. Identification in discrete choice is still achieved by a location normalization on the deterministic utilities, for example $\delta_0 = 0$, since adding a constant to all v_j is observationally equivalent to subtracting it from all δ_j. In practice one may center by $\tilde{v}_j = v_j - m_j$ if desired.

Special case. If $\Sigma = I_{J+1}$ and $m = 0$ then $\mu(t) = 0$ and $\Omega = I_J$, which reduces to the iid probit formulas with Φ and ϕ.