Robustly Optimal Mechanisms for Selling Multiple Goods

Yeon-Koo Che and Weijie Zhong*

August 30, 2024

Abstract

We study robustly optimal mechanisms for selling multiple items. The seller maximizes revenue against a worst-case distribution of a buyer's valuations within a set of distributions, called an "ambiguity" set. We identify the exact forms of robustly optimal selling mechanisms and the worst-case distributions when the ambiguity set satisfies various moment conditions on the values of subsets of goods. The analysis reveals general properties of the ambiguity set that justifies categorical bundling, which includes separate sales and pure bundling as special cases.

1 Introduction

How should a seller sell multiple goods to a buyer? The answer to this seemingly simple question remains elusive. Unlike the single-good case, the optimal selling mechanism for multiple goods is difficult to identify. Traditional Bayesian approaches often lead to intricate mechanisms that are highly sensitive to the buyer's value distribution, as highlighted by Daskalakis et al. (2017) and Manelli and Vincent (2007). Moreover, simple mechanisms like item pricing or bundled pricing can significantly underperform compared to these theoretical optima (Briest et al. (2010); Hart and

*Che: Department of Economics, Columbia University (email: yeonkooche@gmail.com); Zhong: Graduate School of Business, Stanford University (email: weijie.zhong@stanford.edu)

Nisan (2013)), and it is challenging to identify when such simple approaches are effective (see Daskalakis et al. (2017), Manelli and Vincent (2006)).

Relaxing the Bayesian assumptions yields different optimal simple mechanisms in specific scenarios. Carroll (2017) demonstrates that, when the seller knows only the marginal distributions of item values, separate sales at monopoly prices maximize worst-case revenue. In a concurrent study, Deb and Roesler (2023) show that selling a grand bundle is optimal for worst-case revenue when the seller knows the value distribution but lacks information about the buyer's knowledge.

In practice, product bundling exhibits greater complexity. Sellers rarely offer all products individually or as a single grand bundle. Instead, they commonly group products into distinct categories and sell them as (separate) bundles.[1] Video streaming and cable TV services exemplify this, offering bundles of channels categorized by genre, such as news, entertainment, or sports. Similarly, investment banks and financial companies bundle assets into securities based on sectors like technology, energy, or healthcare.

In line with recent literature, this paper adopts a robustness approach by departing from Bayesian assumptions. Our distinctive contribution lies in characterizing the seller's knowledge structure that directly results in a specific pattern and scope of product bundling. Through this analysis, we provide unified insights into two extreme scenarios: full separation (selling all goods individually) and pure bundling (selling only a grand bundle). We achieve this by identifying the precise knowledge structures that underpin these contrasting bundling mechanisms.

The seller in our model has $n \geqslant 2$ heterogeneous goods to sell to a buyer. The buyer has a quasilinear utility function additive in his valuations $(v_1, ..., v_n) \in \mathbb{R}_+^n$ of the goods. The valuations are the buyer's private information, unobserved by the seller. In modeling a seller's information, we focus on the realistic scenario in which the seller accesses basic summary statistics such as means and variances of valuations along several categories of goods. The goods are partitioned into *categories* $K \in \mathcal{K}$, where $\mathcal{K}$ is an arbitrary partition, such that the seller knows only the means and some convex dispersion moments (to be defined) of each category value—the total valuation

[1] Also common is the bundling of complementary products such as computer hardware and software, hotels and flights, razors, and razor blades. While our main model focuses on the additive value setting, we show in Corollary 3 that such complementarity can be easily incorporated into our model and is inconsequential for the prediction.

of the goods in each category. The partition structure describes the *granularity* of the seller's data, representing the smallest groups of goods on which the seller may infer about the group-level valuation. Meanwhile, the dispersion moments describe the *precision* of the seller's inference of the summary statistics, generating "confidence intervals" for valuation estimates.

We study the robustly optimal selling mechanism that maximizes the seller's expected revenue under the worst-case joint distribution consistent with the moment conditions. For analysis, we consider a zero-sum game played by the seller seeking to maximize her revenue and the adversarial nature seeking to minimize it. The equilibrium of this game, or a saddle point, identifies the optimal revenue guarantee for the seller.

Our first main result, Theorem 1, shows that the robustly-optimal mechanism consists of $\mathcal{K}$**-bundled sales**: *each bundle $K \in \mathcal{K}$ of items is sold separately at an independently distributed random price, or equivalently via a menu of lotteries with distinct prices.* As a direct corollary, a separate selling mechanism and a pure bundling mechanism are robustly optimal when $\mathcal{K}$ are the finest partition and the coarsest partition, respectively. The intuition for Theorem 1 is that our seller faces three layers of uncertainties, and the $\mathcal{K}$-bundled sales mechanism hedges against each. First, the independent pricing of alternative categories guards the seller against the first layer of ambiguity concerning the correlation of their valuations across categories. Then, the bundling of each product category hedges against the second layer of ambiguity concerning how the value of each category is divided across its constituent items. Finally, the randomization of bundle prices, or a menu of prices for each bundle, hedges against the distributional ambiguity of each category value, consistent with moment conditions.[2]

Our second main result, Theorem 2, shows that the $\mathcal{K}$-bundling structure is not only sufficient but also necessary for the robust optimality of the mechanism: it is *not* robustly optimal either to separate items within each product category $K \in \mathcal{K}$

[2]The intuition behind the first layer of uncertainty aligns with Carroll (2017), where each item represents a distinct category and only marginal distributions are known. Without the other two layers of ambiguity, this naturally leads to full separation as the optimal mechanism. Meanwhile, the insights related to the second and third layers resonate with those provided by Deb and Deb and Roesler (2023). They demonstrate that similar uncertainties, arising from the buyer's information instead, justify pure bundling when the first layer of uncertainty is absent.

or to bundle multiple product categories in $\mathcal{K}$. These other mechanisms may also be optimal against the distribution that justifies the use of $\mathcal{K}$-bundled sales. What fails them is robustness: they perform poorly against a different "counterfactual" distribution, the analysis of which reveals the seller's true motive for the use of $\mathcal{K}$-bundled sales. Specifically, a bundled sale of items within each category is motivated by the fear that a certain negative correlation across values would lead to revenue loss if items were sold separately. This finding harks back to the classic insight by Adams and Yellen (1976). By contrast, separation across distinct categories is motivated by the fear of asymmetric distributions of the buyer's valuations across categories. If different categories are bundled, the bundle screens the buyer symmetrically across distinct categories, which results in screening inefficiency and revenue loss under the counterfactual distribution.

We explore two extensions of our baseline model in Sections 5 and 6. The first extension adapts our framework to study "informational ambiguity," where the seller has an unambiguous prior about the buyer's valuations but faces ambiguity with regard to the information the buyer has on his own valuations. Such informational ambiguity is characterized by a collection of convex moment conditions, rendering our framework readily applicable. We show that if the valuation distribution is *stochastically comonotonic*, i.e., they are obtainable by a suitable garbling of comonotonic distribution, then the worst-case signal that the buyer may have coincides with the worst-case distribution that justifies the use of pure bundling. Consequently, pure bundling is informationally robust. The second extension goes beyond the moment restrictions assumed in the baseline model and characterizes (more general) distributional restrictions, called **$\mathcal{K}$-Knightian ambiguity**, that justify the use of $\mathcal{K}$-bundled sales. These two extensions demonstrate that the insights obtained from our analysis apply more broadly and resiliently beyond the specific settings considered in Section 3.

The current paper intersects with two broad strands of literature. First, it contributes to the multiproduct monopoly literature and, more broadly, the multidimensional screening and mechanism design literature. Representative works include McAfee and McMillan (1988), Armstrong (1996, 1999), Manelli and Vincent (2006, 2007), Rochet and Chone (1988), Daskalakis et al. (2013, 2017), Hart and Reny (2015, 2019), Menicucci et al. (2015), and Haghpanah and Hartline (2021). The current paper departs from this literature by taking a robustness approach.

Second, the current paper contributes to the literature on robust mechanism design. Many authors study optimal mechanisms under the worst-case distribution of states. To the best of our knowledge, Scarf (1957) was the first to adopt this approach in inventory management. Carroll (2015, 2019) apply the approach to contracting settings. Bergemann and Schlag (2011) and Carrasco et al. (2018) solve the single-item monopoly problem with neighborhood restrictions and moment conditions, respectively. Koçyiğit et al. (2019), Brooks and Du (2021a), He and Li (2022) and Che (2022) extend the framework to the multi-buyer auction setting, but still with one item. As already discussed, Carroll (2017) applies the robust mechanism design approach to a multi-item sale problem with known marginals, making it the closest antecedent of the current paper. We develop his framework further and provide a robustness-based rationale for general forms of categorical bundling, which include separate sales and pure bundling as special cases.[3] In addition, to our knowledge, our necessity result of a robustly optimal mechanism is new to the literature.

Recent authors have also studied the optimal mechanism in the worst-case scenario in terms of the information possessed by agents; see Du (2018), Bergemann et al. (2016), Brooks and Du (2021a,b, 2023), and Deb and Roesler (2023). These papers assume that a seller is ambiguous about the buyer's information regarding the values of items and chooses an optimal mechanism robust with respect to the buyer's information.[4] Such a model can be seen as a robust mechanism design problem in which the ambiguity set is determined by the seller's prior belief in a particular way. Among them, Deb and Roesler (2023) deals with the multi-item selling problem. They show that pure bundling is informationally robust when the prior belief is exchangeable across alternative items. We identify a more general stochastic comonotonicity

[3] Although worst-case revenue maximization is a natural way to extend the standard Bayesian framework, several authors have also considered other notions of robustness in mechanism design. Bergemann and Schlag (2008), Guo and Shmaya (2023) and Koçyiğit et al. (2021) study the minimization of regret—namely, a revenue shortfall of the chosen mechanism relative to the complete-information optimal mechanism. In particular, Koçyiğit et al. (2021) finds a regret-minimizing mechanism for selling multiple items with known means and rectangular domain, which parallels the case treated in Appendix B.6. Another objective popular in algorithmic mechanism design is the revenue ratio of simple mechanisms (often separate sales and pure bundling) to all mechanisms across all or a restricted set of valuation distributions. As the number of items grows large, the ratio tends to zero when the distributed is unrestricted (Briest et al. (2010); Hart and Nisan (2013)) and is bounded away from zero when item values are independently distributed (Babaioff et al. (2014); Hart and Nisan (2012); Li and Yao (2013)).

[4] Brooks and Du (2021a) employs both distributional and informational uncertainties.

condition for the informational robust optimality of pure bundling, which nests the exchangeable prior assumption as a special case but also permits highly asymmetric priors.[5] Brooks and Du (2023) proves that in multi-item auctions, informational ambiguity implies the robust optimality of indirect mechanisms with one-dimensional message space.

The rest of the paper is organized as follows. Section 2 introduces a model of multi-item sale and defines a notion of robust optimality. Section 3 considers an ambiguity set defined by a combination of moment conditions and establishes the robust optimality of $\mathcal{K}-$bundled sales, which specialize to separate sales and pure bundling when $\mathcal{K}$ are the finest and coarsest partitions, respectively. Section 4 establishes that the main qualitative features of $\mathcal{K}$-bundled sales are necessary. Sections 5 and 6 extend the optimality of $\mathcal{K}$-bundled sales to a setting with informational ambiguity and a setting with more general distributional ambiguity, respectively. Section 7 concludes.

2 Model

A seller wishes to sell n items to a single buyer. The buyer has values $\boldsymbol{v} := (v_1, ..., v_n)$ for the items whose distribution is unknown to the seller.[6] The seller simply knows that the distribution lies within some *ambiguity set* $\mathcal{F} \subset \Delta(\mathbb{R}^n_+)$ defined by a set of moment conditions.

Moment conditions: The moment conditions are defined in terms of means and dispersions on a joint distribution. To define them, fix any joint distribution $F \in \Delta(\mathbb{R}^n_+)$. First, we assume the seller has some knowledge about the means of item values. Given F, let $\mu_i(F) := \mathbb{E}_F[v_i]$ denote the mean value of item i. Next, the seller has some knowledge about the dispersion of values of arbitrary subsets of items. Specifically, let $\mathcal{K}$ be an arbitrary partition of the goods, with its element $K \in \mathcal{K}$

[5] Deb and Roesler (2023) also explores the buyer-optimal information structure facing a Bayesian seller, which is not the main focus of our paper.

[6] To be precise, the "values" $\boldsymbol{v} := (v_1, ..., v_n)$ need not be true values but rather the estimates the buyer assigns to items. In this sense, the ambiguity the seller faces is ultimately an informational one, arising from her ignorance of what the buyer "knows."

interpreted as a *product category*. For each product category $K \in \mathcal{K}$, we let

$$\sigma_K(F) := \mathbb{E}_F\left[\phi_K\left(\textstyle\sum_{i\in K} v_i\right)\right]$$

be the *dispersion* of category K's value under F, where $\phi_K : \mathbb{R}_+ \to \mathbb{R}_+$ is a twice-differentiable convex function satisfying $\phi_K'' \geqslant \varepsilon$ for some $\varepsilon > 0$. We will refer to such a function as *convex moment function*.

We assume that the seller knows item value means and dispersion lie in some arbitrary nonempty convex and compact set $\Omega \subset \mathbb{R}_+^{n+|\mathcal{K}|}$.[7] Formally, the seller faces an ambiguity set

$$\mathcal{F} := \left\{F \in \Delta(\mathbb{R}_+^n) : (\mu_i(F), \sigma_K(F))_{i\in N, K\in\mathcal{K}} \in \Omega\right\}. \tag{1}$$

An example, which will be used throughout, illustrates the nature of ambiguity facing the seller.

Example 1 *Suppose the seller has* 3 *goods with* $N = \{1, 2, 3\}$*. The seller does not know the distribution of the items' valuations to the buyer except that she knows the means* $\mathbb{E}[v_1] = 0.5,\ \mathbb{E}[v_2] = \mathbb{E}[v_3] = 0.3$*, and variances* $\mathbb{V}[v_1] = \mathbb{V}[v_2 + v_3] = 0.1$ *for two categories, given by the partition* $\mathcal{K} = \{\{1\}, \{2, 3\}\}$*. In this case, the convex moment functions are* $\phi_{\{1\}}(v) = \phi_{\{2,3\}}(v) = v^2$*, and the moment constraint set is* $\Omega = \{(0.5, 0.3, 0.3, 0.35, 0.46)\}$.

Feasible mechanisms: The seller is free to choose any selling mechanism. By the revelation principle, it is without loss to focus on direct revelation mechanisms, denoted by $M = (q(\boldsymbol{v}), t(\boldsymbol{v}))$, where the *allocation rule* $q : \boldsymbol{v} \mapsto [0, 1]^n$ specifies the probability of allocating each item to the buyer, and the *payment rule* $t : \boldsymbol{v} \mapsto \mathbb{R}^+$ specifies the expected payment received from the buyer, both as Borel measurable functions of the vector $\boldsymbol{v}$ of values reported by the buyer. The mechanism satisfies *incentive compatibility* and *individual rationality*:

$$\boldsymbol{v} \cdot q(\boldsymbol{v}) - t(\boldsymbol{v}) \geqslant \sup_{\boldsymbol{v}' \in \mathbb{R}_+^n} \boldsymbol{v} \cdot q(\boldsymbol{v}') - t(\boldsymbol{v}') \tag{IC}$$

$$\boldsymbol{v} \cdot q(\boldsymbol{v}) - t(\boldsymbol{v}) \geqslant 0 \tag{IR}$$

[7]Given the linearity of a mean, our ambiguity set allows a mean condition to apply to the value of each category, instead of the value of each item. As is clear from the proof of Theorem 1, this has no effect for the qualitative features of the robustly optimal mechanism.

for each $\boldsymbol{v} \in \mathbb{R}_+^n$. Let $\mathcal{M}$ denote the set of all direct mechanisms satisfying the (IC) and (IR) constraints—called *feasible* mechanisms.[8]

$\mathcal{K}$**-bundled sales**: Among the feasible mechanisms, certain types of mechanisms will be of special interest to us. Consider the partition $\mathcal{K}$. The seller may bundle each category $K \in \mathcal{K}$ of items and sell that bundle separately from the other categories of items. For each item i, let $K(i)$ be the category $K \in \mathcal{K}$ containing it. Formally, we say a feasible mechanism $M := (q, t) \in \mathcal{M}$ is a $\mathcal{K}$-*bundled sales* mechanism if, for each $K \in \mathcal{K}$, there exists a feasible (one-dimensional) mechanism $q_K : \mathbb{R}_+ \to [0, 1]$ and $t_K : \mathbb{R}_+ \to \mathbb{R}$ such that $t(\boldsymbol{v}) = \sum_{K \in \mathcal{K}} t_K(\sum_{j \in K} v_j)$ and $q_i(\boldsymbol{v}) = q_{K(i)}(\sum_{j \in K} v_j)$.That is, the mechanism sells each bundle K with probability q_K and collects expected payment t_K. Let $\mathcal{M}_{\mathcal{K}}$ denote the set of all feasible $\mathcal{K}$-bundled sales mechanisms.

In our leading example, the $\mathcal{K}$-bundled sales mechanism involves two bundles: the first bundle is good 1 only and the second bundle comprises goods 2 and 3, priced independently according to distributions g_1 and g_{23}.

$\mathcal{K}$-bundled sales include two canonical mechanisms as special cases. When $\mathcal{K}$ is the finest partition, namely when $\mathcal{K} = \{\{1\}, ..., \{n\}\}$, $\mathcal{K}$-bundled sales reduce to selling each item separately; we will refer to this as a *separate sales* mechanism. When $\mathcal{K}$ is the coarsest partition, namely when $\mathcal{K} = \{\{1, ..., n\}\}$, $\mathcal{K}$-bundled sales reduces to selling all items as a single grand bundle; we will call such a mechanism *pure bundling*.

Robustness solution concept: The seller's revenue from a mechanism $M \in \mathcal{M}$ given value distribution F is $R(M, F) := \int t(\boldsymbol{v}) F(\mathrm{d}\boldsymbol{v})$.[9] Let $R \in \mathbb{R}$ be a *revenue guarantee* if there exists a mechanism $M \in \mathcal{M}$ such that $R(M, F) \geqslant R$ for all $F \in \mathcal{F}$. The seller's objective is to maximize the revenue guarantee. Let

$$R^* := \sup_{M \in \mathcal{M}} \inf_{F \in \mathcal{F}} R(M, F)$$

[8] It is without loss to require (IC) and (IR) for all types in $\mathbb{R}_+^n$, rather than only for $\boldsymbol{v} \in \bigcup_{F \in \mathcal{F}} \operatorname{supp}(F)$. Proposition B.1 shows that, for any feasible mechanism defined on $\bigcup_{F \in \mathcal{F}} \operatorname{supp}(F)$, one can find a Borel measurable extension that satisfies (IC) and (IR) for all types in $\mathbb{R}_+^n$ and implements the same outcome for the types in the original domain.

[9] Here, we implicitly assume that the seller values each item at zero. This is without loss. If there are unit costs $\boldsymbol{c} = (c_i) \geqslant 0$ for the items, then the problem facing the seller is exactly the same as in our model in which she faces $\boldsymbol{w} = \boldsymbol{v} - \boldsymbol{c}$ as the buyer's valuations and zero costs. Robustly optimal mechanisms are then obtained upon an appropriate change of variables. Specifically, a saddle point (M^*, F^*) in our original model without costs remains a saddle point in terms of $\boldsymbol{w}$ in the new model.

be the *optimal revenue guarantee*, and we say the mechanism attaining R^* is *robustly optimal*.

2.1 Discussions of Model

The main elements of the model are motivated as follows.

Partitional knowledge structure: As illustrated in the introduction, the partition of goods $\mathcal{K}$ describes the granularity of the seller's knowledge about the value distribution. In practice, the partitional structure of knowledge often reflects the intrinsic characteristics of goods. For instance, a financial broker may categorize stocks according to the sector possibly because stocks within a sector are influenced by common factors, making them more comparable to each other than to stocks from different sectors. Similarly, in media and entertainment, products such as movies, music, and books are grouped by genres, acknowledging the shared preferences of consumers for content within the same genre. Wholesale distributors who supply many goods to retail grocers categorize goods according to their industry, brand, and grade. Market research then employs this categorization to derive key summary statistics, such as the mean and variance of consumer valuations for these product clusters, resulting in a partitional knowledge structure.

$\mathcal{K}-$**bundled sales in practice:** Besides the two extreme cases (fully separation and pure bundling), bundled sales with a non-degenerate partition structure are widely adopted in practice. Intriguingly, they are commonly adopted in the examples we introduced where the seller's knowledge exhibits the partitional structure. Content providers often sell bundled subscriptions based on the genres of the content. For example, the Wall Street Journal offers their professional clients six bundles of news partitioned based on the industry.[10] Similarly, cable TVs and video streaming services typically categorize their channels into three bundles news, sports, and movies. In the supply chain of groceries, a wholesale distributor often offers a menu consisting of *product lines*, each of which is a bundle of commodities of a specific brand.

[10]See https://wsjpro.com/.

Moment conditions: We consider the ambiguity set defined by moment conditions because moments (such as mean and variance) are the most natural information the seller may receive in practice. However, our assumption on $\mathcal{F}$ covers more than conventional moment conditions.

First, we consider a general convex moment function ϕ, instead of conventional power moment functions (e.g. Carrasco et al. (2018)). This generality enables us to extend our results to other types of ambiguities. In Section 5, we demonstrate that the so-called *informational ambiguity*—the seller being ambiguous about what the buyer knows about—can be represented as a dispersion condition corresponding to a particular convex moment function. Additionally, in Appendix B.6, we show that the case of *domain restriction*, where the value of each bundle $K \in \mathcal{K}$, $\sum_{i \in K} v_i$, lies in some interval $[0, \bar{v}_K]$, can also be analyzed as a limiting case of convex moments.

Second, the generality of the "confidence set" Ω allows us to capture a wide range of scenarios regarding the seller's ambiguity. For instance, Ω could be arbitrarily close to $\mathbb{R}^{|\mathcal{K}|}$ when projected to the last $|\mathcal{K}|$ dimensions, in which case the convex moments would be unrestricted.[11] At the other extreme, Ω could be a singleton; the seller then knows the exact means and dispersion of individual item values. As another example, Ω could be characterized by a system of inequalities: $\psi_j(\mu_1(F), ..., \mu_n(F)) \geqslant 0$, for some concave functions $\psi_j, j = 1, ..., n$. This allows for cases in which the seller knows the average values of subsets of items.[12] Generally, one can view the size of Ω as capturing the magnitude of ambiguity the seller faces concerning the relevant bundle values.

Robustness solution concept: Throughout the paper (except Section 6) we find the robustly optimal mechanism by solving a simultaneous-move zero-sum game. The equilibrium $(M^*, F^*) \in (\mathcal{M}, \mathcal{F})$ of the zero-sum game is a *saddle point*; i.e., $\forall M \in \mathcal{M}$, $\forall F \in \mathcal{F}$,

$$R(M, F^*) \leqslant R(M^*, F^*) \leqslant R(M^*, F). \tag{2}$$

It is well-known that a saddle point gives rise to an optimal revenue guarantee (see Osborne and Rubinstein (1994), Proposition 22.2-b) and the guarantee does not change

[11] In this special case, the ambiguity set is compatible with any arbitrary partition $\mathcal{K}'$. Our Theorem 1 would then imply that all bundling structures (including pure bundling and full separation) are robustly optimal.

[12] For instance, we could have $\sum_{i \in K} \mathbb{E}_F[v_i] = m_K$ for each $K \in \mathcal{K}$.

with the order of moves:

$$R(M^*, F^*) = \max_{M \in \mathcal{M}} \min_{F \in \mathcal{F}} R(M, F) = \min_{F \in \mathcal{F}} \max_{M \in \mathcal{M}} R(M, F) = R^*. \tag{3}$$

3 Robust Optimality of $\mathcal{K}$-bundled sales

In this section, we solve for the robustly optimal selling mechanism for the seller who faces the ambiguity set indexed by an arbitrary partition $\mathcal{K}$ and the set Ω of possible moments. We begin with the main theorem:

Theorem 1 *It is robustly optimal for the seller to use a $\mathcal{K}$-bundled sales mechanism.*

PROOF: See Appendix A.1. ■

To prove Theorem 1, we construct a $\mathcal{K}$-bundled sales mechanism $M^* \in \mathcal{M}$ together with the distribution $F^* \in \mathcal{F}$ such that they form mutual best responses. Here, we illustrate the main economic intuition behind the construction using Example 1; the general construction of (F^*, M^*) and proof appear in Appendix A.1.

Construction of F^*: Recall that in Example 1, $\mathcal{K} = \{\{1\}, \{2, 3\}\}$, with the ambiguity set $\mathcal{F}$ constrained in terms of means and variances of valuations along two categories, $\{1\}$ and $\{2, 3\}$, of goods. Nature chooses $F^* \in \mathcal{F}$ that has a one-dimensional support depicted in Figure 1: the support forms a ray emanating from the origin and contains a vertical segment toward the end. Specifically, nature draws a one-dimensional random variable X distributed from $[1, \infty)$ according to cdf

$$H(x) := 1 - 1/x,$$

(i.e. Pareto distribution). The valuation of each good $i = 1, 2, 3$ is then determined as

$$v_i = \min\{\alpha_i X, \beta_i\},$$

where (α_i, β_i)'s are chosen (uniquely) to satisfy the moment conditions (i.e., $\mathbb{E}[v_1] = 0.5$, $\mathbb{E}[v_2] = \mathbb{E}[v_3] = 0.3$, and variances $\mathbb{V}[v_1] = \mathbb{V}[v_2 + v_3] = 0.1$).

There are two properties notable about F^*. The first is that the support is comonotonic within each category; see, for example in Figure 1, its projection onto the (v_2, v_3) space. As will be explained in Section 4, this property incentivizes the seller to

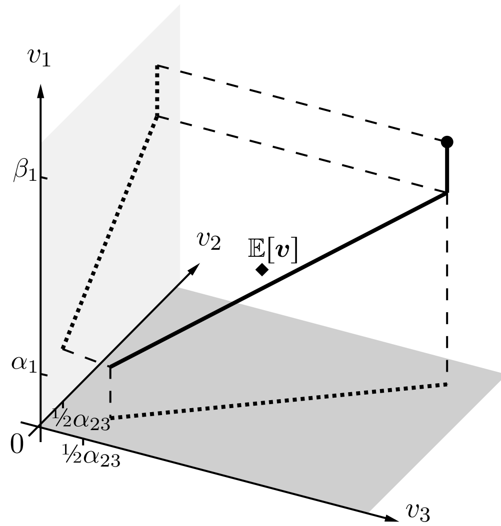

Fig. 1: The solid curve is the support of the joint distribution F^*. The dashed curves are its projections to the subspaces.

bundle each category of goods. Second, F^* is designed to suppress the seller's revenue as much as possible. To see this, note first that, since F^* is one-dimensional, the seller's mechanism design problem reduces to a standard one-dimensional screening problem (with multi-dimensional allocations). The virtual value of each item (as a function of x) is

$$J_i(x) = V_i(x) - V_i'(x)\frac{1 - H(x)}{h(x)} = \begin{cases} 0 & \text{if } x < \beta_i/\alpha_i \\ \beta_i & \text{if } x \geqslant \beta_i/\alpha_i \end{cases}.$$

In words, nature "levels" the seller's virtual valuation. The "flat" virtual value function means that all mechanisms that allocate each item i to buyer type with valuation β_i are optimal. Since any $\mathcal{K}-$bundled sales mechanism that sets bundle $\{1\}$'s price in $[\alpha_1, \beta_1)$ and bundle $\{2,3\}$'s price in $[\alpha_2 + \alpha_3, \beta_2 + \beta_3)$, respectively, satisfy this property, M^* constructed below is optimal given F^*. The resulting revenue is precisely what the seller receives by charging the lowest prices in the support of F^*, i.e., $\alpha_1 + \alpha_2 + \alpha_3$.[13]

Mechanism M^*: The optimal $\mathcal{K}$-bundling mechanism sells the two categories $\{1\}$

[13]This property of the Pareto distribution has been exploited in other papers, Carrasco et al. (2018) and Roesler and Szentes (2017), in the single-good case. Unlike the latter paper, the seller does not choose the lowest price in the support of the worst-case distribution but instead mixes over

and $\{2,3\}$ as two separate bundles, and prices the respective bundles independently and randomly according to density functions:

$$g_1(p) = 2\lambda_1 \cdot \frac{\beta_1 - p}{p}$$

on support $[\alpha_1, \beta_1]$, and

$$g_{23}(p) = 2\lambda_{23} \cdot \frac{\beta_{23} - p}{p}$$

on support $[\alpha_{23}, \beta_{23}]$, where $\alpha_{23} = \alpha_2 + \alpha_3$ and $\beta_{23} = \beta_2 + \beta_3$, and $\boldsymbol{\lambda} = (\lambda_1, \lambda_{23})$ are scale factors chosen so that the densities integrate to ones.

Intuitively, the $\mathcal{K}$-bundling mechanism is designed to “hedge” against possible deviation by nature away from F^*. To see this, let us recall what nature is capable of doing. First, nature can freely shift values across items within the bundle $\{2,3\}$, as the moment condition only constrains its total value. Second, nature can freely correlate the values of categories $\{1\}$ and $\{2,3\}$. Third, nature can redistribute the value of each bundle—bundle 1 and bundle $\{2,3\}$—while maintaining the same dispersion measures.

Our mechanism makes it unprofitable for nature to deviate along these three channels. First, the bundling of items 2 and 3 makes their valuations perfectly substitutable from the revenue standpoint; hence, nature has no incentive to redistribute values within the $\{2,3\}$ bundle. Second, the independent pricing of the two bundles means that nature never gains from manipulating the correlations of values across the bundles. Finally, the prices of the two bundles are randomized according to the particular hyperbolic forms of the densities precisely to eliminate any incentive by nature to redistribute value within each bundle.

The construction and the argument generalize naturally to an arbitrary $\mathcal{K}$, leading to a pair of $\mathcal{K}$-bundled sales mechanisms M^* and a one-dimensional valuation distribution F^* that constitute a saddle point. Two questions remain. First, how do we pin down the parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and $\boldsymbol{\lambda}$, especially under the general set Ω? Second, why are different mechanisms *not* robustly optimal? We relegate the answer to the first question in the formal proof in Appendix A.1. The second question is addressed in Section 4.

the support to guarantee revenue. The role of the Pareto distribution here is therefore to keep the seller's revenue at the level she would enjoy by setting the lowest prices in the support.

3.1 Implications of Theorem 1

Theorem 1 implies a few immediate corollaries. First, we provide a rationale for the use of both separate sales and pure bundling as special cases.

Corollary 1 *If the seller faces the ambiguity set in* (1) *where* $\mathcal{K}$ *is the finest partition, then separate sales of individual items are robustly optimal. If the seller faces ambiguity set in* (1) *where* $\mathcal{K}$ *is the coarsest partition, then a sale of the grand bundle is robustly optimal.*

The next corollary expands the applicability of Theorem 1 beyond the ambiguity set $\mathcal{F}$ in (1).

Corollary 2 *Suppose* (M^*, F^*) *is a saddle point given an ambiguity set* $\mathcal{F}$. *If* $\widetilde{\mathcal{F}} \subset \mathcal{F}$ *such that* $F^* \in \widetilde{\mathcal{F}}$, *then,* (M^*, F^*) *is a saddle point given the ambiguity set* $\widetilde{\mathcal{F}}$.

PROOF: The result follows since $R(M^*, F^*) \leqslant R(M^*, F)$ for any $F \in \widetilde{\mathcal{F}} \subset \mathcal{F}$. ■

This corollary states that F^* remains robustly optimal within any ambiguity set $\widetilde{\mathcal{F}}$ if it is in turn a subset of $\mathcal{F}$ defined in (1). This simple corollary, reminiscent of a revealed preference argument, turns out to be quite useful. For instance, one may find it plausible that item values are positively correlated so that the correlation coefficient between any pair of item values exceeds some number $\theta \in [0, 1)$. Since F^* exhibits high correlation across all the v_i's (they are perfectly correlated in the interior support), it will satisfy this additional restriction for θ small enough, so one may conclude that the mechanism identified in Theorem 1 continues to be robustly optimal given the correlation condition.

Next, we can extend Theorem 1 to allow for the types of complementarities considered by Deb and Roesler (2023).

Corollary 3 *Suppose the buyer's value of each subset of items, J (not necessarily an element of $\mathcal{K}$), is given by* $u_J \cdot \sum_{i \in J} v_i$, *for some* $u_J \in [0, 1]$, *and exhibits* complementarity*:* $u_J = 1$ *if* $J \in \mathcal{K}$. *Then, a* $\mathcal{K}$*-bundled sales mechanism is robustly optimal.*

PROOF: See Appendix A.1.1. ■

In the corollary, the value of all bundles except for those in $\mathcal{K}$ can be discounted by a factor of $u_J \leqslant 1$ and $\mathcal{K}$-bundled sales remains robustly optimal. This generalization is straightforward since the revenue from $\mathcal{K}$-bundled sales is not affected by such complementarity, while all other mechanisms only under-perform.

How does the seller's revenue guarantee vary with the seller's knowledge? Such comparative statistics is interesting in their own right since these parameters may be ex-ante controlled via market research efforts. For instance, one may increase the precision of market forecast by reducing the dispersion; it would be interesting to know how such an investment may improve the revenue guarantee. Thanks to our complete characterization, such comparative statics can be readily performed.

Corollary 4 *Suppose $\Omega = \{(\boldsymbol{m}, \boldsymbol{s})\}$. Let $R^*(\boldsymbol{m}, \boldsymbol{s})$ denote the optimal revenue guarantee as a function of $(\boldsymbol{m}, \boldsymbol{s})$. Let $m_K = \sum_{i \in K} m_i$.*

$$\begin{cases} \dfrac{\mathrm{d}R^*(\boldsymbol{m}, \boldsymbol{s})}{\mathrm{d}m_K} = \lambda_K \phi'_K(\beta_K) > 0; \\ \dfrac{\mathrm{d}R^*(\boldsymbol{m}, \boldsymbol{s})}{\mathrm{d}s_K} = -\lambda_K < 0, \end{cases} \tag{4}$$

where $\lambda_K > 0, \beta_K > m_K$ are parameters pinned down in the proof of Theorem 1.

Proof: See derivation in Appendix B.1. ■

Without loss, we can normalize ϕ_K so that it is centered at m_K (i.e. $\phi'_K(m_K) = 0$), then $\phi''_K > 0$ implies $\phi'_K(\beta_K) > 0$. We obtain a complete comparative statics result when Ω is a singleton: The optimal revenue guarantee decreases strictly in the dispersion of the valuation distribution and increases strictly in the mean of the valuation.

Moreover, for general Ω, (4) quantifies the tradeoff between the revenue guarantee gain from improving the "precision of estimation" for different summary statistics. Consider the special case when Ω is a product set—the seller knows a confidence interval for each mean and moment estimation—then, (4) implies that $\phi'_K(\beta_K)$ is exactly the marginal rate of substitution between shrinking the confidence interval of the mean versus the dispersion moment of a product group.[14]

[14]More precisely, lowering the upper bound of dispersion s_K by a unit is revenue-equivalent to an increase in the lower bound of the mean m_K by $\phi'_K(\beta_K)$ units.

4 Necessity of $\mathcal{K}$-bundled sales

Theorem 1 establishes the robust optimality of a $\mathcal{K}$-bundled sales mechanism. However, it leaves open the possibility that another mechanism may attain that same revenue guarantee. We next show that this is not the case and therefore the main qualitative feature of $\mathcal{K}$-bundled sales is *essential* for achieving that optimal revenue guarantee. This analysis also reveals what motivates the seller to choose the predicted mechanism.

Theorem 2 *Fix an ambiguity set $\mathcal{F}$ defined relative to the partition $\mathcal{K}$.*

1. *Suppose there is $K \in \mathcal{K}$ with $|K| \geqslant 2$. Then, for any nonempty sets $J, J' \subset K$ with $J \cap J' = \emptyset$, no mechanism in $\mathcal{M}_{\mathcal{K}'}$ is robustly optimal if $\mathcal{K}'$ separates J and J'; i.e., $J, J' \in \mathcal{K}'$.*
2. *Let $\boldsymbol{\alpha}, \boldsymbol{\beta}$ be the parameters derived in Theorem 1. Suppose there are $K, K' \in \mathcal{K}$ such that $\beta_K/\alpha_K \neq \beta_{K'}/\alpha_{K'}$ (as defined in (6) and (7)). Then, no mechanism in $\mathcal{M}_{\mathcal{K}'}$ is robustly optimal if $\mathcal{K}'$ bundles K and K' together; i.e, $K \cup K' \in \mathcal{K}'$.*

While we relegate the formal proof of Theorem 2 to Appendix A.2, we illustrate the intuition of the proof via Example 1, where $\mathcal{K} = \{\{1\}, \{2, 3\}\}$. In this case, theorem 2 rules out two alternative types of mechanisms: the fully separated mechanism (corresponding to $\mathcal{K}' = \{\{1\}, \{2\}, \{3\}\}$) and the pure bundling mechanism (corresponding to $\mathcal{K}'' = \{\{1, 2, 3\}\}$). Recall that, given F^*, any mechanism that allocates item i when $v_i = \beta_i$ with probability 1 is optimal. Therefore, the two types of candidate mechanisms are optimal when the prices are chosen not too high. However, they are *not* robustly optimal, as each of them is susceptible to a type of deviation by nature that lowers the revenue guarantee. These deviations reveal the potential distribution F that motivates the seller to choose the "correct" $\mathcal{K}$-bundled sales.

Why is full separation not robustly optimal? Suppose instead of the $\{\{1\}, \{2, 3\}\}$-bundling, the seller sells all three goods separately, in particular, separating items 2 and 3. The separate sale is vulnerable to the following deviation by nature. Consider a distribution $\tilde{F}$, which is the same as F^*, except that a small mass ε is transferred from $(\beta_1, \frac{1}{2}\beta_{23}, \frac{1}{2}\beta_{23})$ (the point mass at the top) to $(\beta_1, \beta_{23}, 0)$ and $(\beta_1, 0, \beta_{23})$, each with respective masses of $\frac{1}{2}\varepsilon$. See Figure 2.

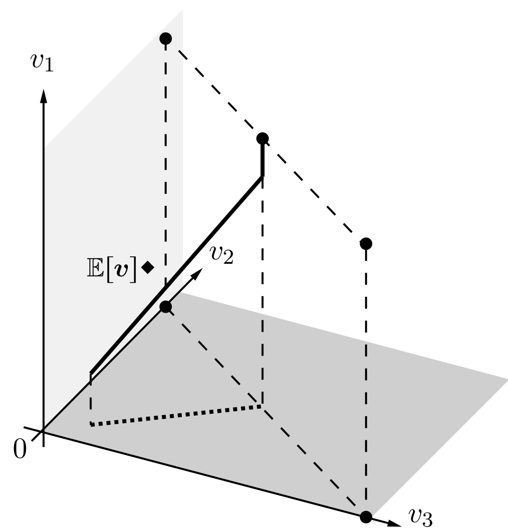

Fig. 2: The solid curve is the support of distribution $\tilde{F}$

This change keeps all constraints satisfied and does not alter the revenue of M^* since the distributions of v_1 and v_2+v_2 remain the same. Yet the change has increased the dispersion of each individual item value in a way that makes separate sales less profitable. For ε sufficiently small, the seller will never wish to charge prices 0 or β_{23} for either item 2 or item 3. For any other price in the support, the seller loses revenue $p \cdot \frac{1}{2}\varepsilon$, when compared with bundling items 2 and 3. Consequently, facing distribution $\tilde{F}$, the seller earns strictly below $\alpha_1 + \alpha_{23}$ by selling the three items separately. In essence, the fear of this "negatively-correlated" counterfactual distribution motivates the seller to bundle goods 2 and 3.

Why is pure bundling not robustly optimal? Suppose now the seller bundles all three items. As observed earlier, given the same F^*, the grand bundle yields the same optimal revenue when it is sold at price p within $\left[\alpha_1 + \alpha_{23}, \beta_{23}\left(1 + \frac{\alpha_1}{\alpha_{23}}\right)\right]$; see Figure 3. However, the same figure hints at why selling the grand bundle is not robustly optimal. Suppose the seller charges an even higher price $p > \beta_{23}\left(1 + \frac{\alpha_1}{\alpha_{23}}\right)$ for the bundle. Then, the revenue would be strictly lower! This is because bundling entails inefficient screening at that price, specifically in the vertical segment of the support depicted in Figure 4: the purchases of all goods are now tied so that the buyer will refuse to buy the bundle *even when* he has the highest value β_{23} for the bundle $\{2, 3\}$, if his value of good 1 is less than $p - \beta_{23}$, resulting in the seller not being able to sell goods 2 and 3 in that case. This is clearly inefficient and this inefficiency never occurs under separate sales of $\{1\}$ and $\{2,3\}$, since the seller would never charge more than β_{23} for the bundle $\{2,3\}$.

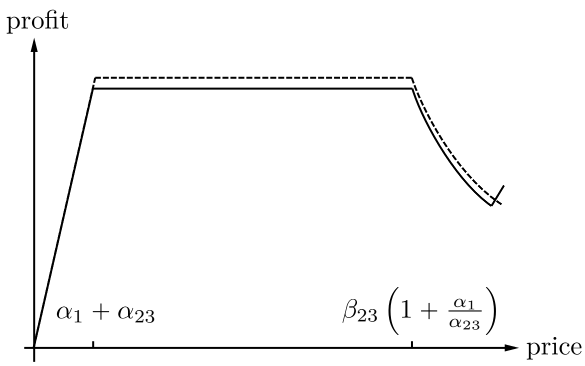

Fig. 3: Profit from pure bundling mechanism under F^* (dashed) and $\tilde{F}$ (solid)

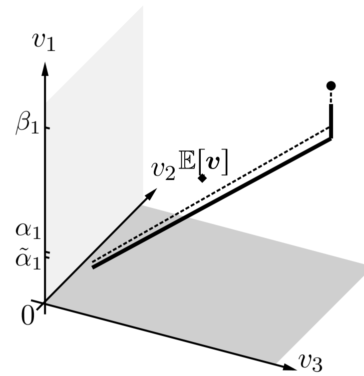

Fig. 4: The support of distribution F^* (dashed) and $\tilde{F}$ (solid)

Nature can exploit this "weakness" of the grand bundling by shifting mass toward that vertical segment. Consider a new distribution $\tilde{F}$ supported on the solid curve in Figure 4. Compared with F^*, this new distribution lowers the infimum of v_1 from α_1 to $\tilde{\alpha}_1$, thus lowering the value $V_1(s)$ of good 1 on the interior segment of the support. This reduces the revenue the seller can collect by charging a low bundle price p. Of course, nature cannot lower the value of good 1 uniformly across the board, because this will violate the mean condition. To satisfy the latter, $\tilde{F}$ must therefore put larger mass at its supremum value β_1 of good 1. The seller cannot take advantage of this increased mass at β_1 under pure bundling since the profit at p in the neighborhood of $\beta_1 + \beta_{23}$ was strictly lower than $\alpha_1 + \alpha_{23}$, as can be seen in Figure 3.[15] Hence, the new distribution keeps the seller's revenue strictly below $\alpha_1 + \alpha_{23}$ no matter the price of the bundle. Intuitively, the distribution $\tilde{F}$ exacerbates the ex ante asymmetry across the two bundles, and the *fear* of such an asymmetric distribution motivates the seller to choose separate sales mechanism.

Analyzing the counterfactual suboptimal mechanisms teaches us two important lessons about the robustly optimal mechanism. First, it is well known that negatively-correlated item values make bundling desirable in the standard Bayesian context (see

[15]The robust optimality of M^* means that the seller would receive at least $\alpha_1 + \alpha_{23}$ from M^* given $\tilde{F}$.

Adams and Yellen (1976)). In light of this, one may find it surprising that the item values under distribution F^* are instead positively correlated. Theorem 2-1 clarifies this issue: it is a *possible* negative correlation "off the path" that motivates the seller to use bundling in the current environment.[16] Second, it is also well-known that perfectly correlated (comonotonic) item values are the worst-case distribution that justifies full separation under correlational ambiguity (see Carroll (2017)). Theorem 2-2 further explains what deters bundling: nature may "deviate" to distributions that are still comonotonic but strongly asymmetric, which requires the seller to screen different dimensions asymmetrically to attain the maximal revenue.

5 Informational Robustness

A rather surprising application of our analysis is informational ambiguity, where the source of ambiguity for the seller is not the prior on the buyer's valuations but rather the information the latter has about the valuations. A growing number of recent papers study mechanisms that are robust with respect to such ambiguity; see, for example, Du (2018); Brooks and Du (2021b,a); Roesler and Szentes (2017); Ravid et al. (2022); Bergemann et al. (2019).

To fix the idea, suppose the seller has prior distribution $G \in \Delta(\mathbb{R}^n_+)$ on the valuations of the goods, but she has ambiguity on the information the buyer himself has about the valuations. By Blackwell (1951), a possible signal the buyer may have is characterized by a mean-preserving contraction of G. Hence, one can describe the seller's ambiguity by a set of convex moment constraints:

$$\mathcal{F}_G := \left\{ F \in \Delta(\mathbb{R}^n_+) \middle| \mathbb{E}_F[\phi(\boldsymbol{v})] \leqslant \mathbb{E}_G[\phi(\boldsymbol{v})], \forall \phi \text{ convex} \right\}. \tag{5}$$

A robustly optimal mechanism given the ambiguity set $\mathcal{F}_G$ in (5) is then called **informationally robust**. Formulated in this manner, our analysis can be readily applied to find an informationally robust mechanism. Consider the following assumption on the prior distribution G.

Assumption 1 (Stochastic Comonotonicity) *There exists* $(\xi_1, ..., \xi_n) \in \mathbb{R}^n_+$ *with*

[16] When we state "off-the-path", we are invoking the definition of R^*: the seller acts first, knowing nature's response.

$\sum_i \xi_i = 1$ *such that for each* $i = 1, ..., n$

$$\mathbb{E}\left[v_i \middle| \sum_{j=1}^{n} v_j\right] = \xi_i \left(\sum_{j=1}^{n} v_j\right).$$

Stochastic comonotonicity means that the expected value of each item conditional on the total value of all items is simply a fixed fraction of the latter. Geometrically, the conditional mean of each item value forms a linear ray, as depicted in Figure 5. Effectively, the condition means that the item values are a garbling—a mean-preserving spread—of such a ray, as illustrated in Figure 5.

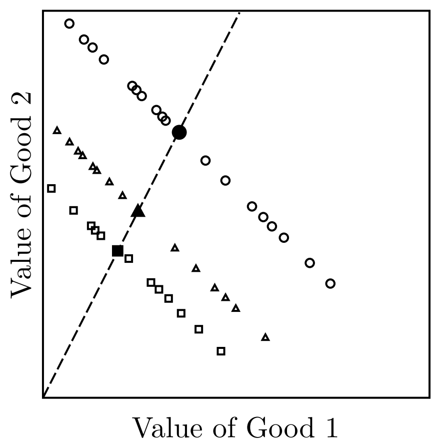

Fig. 5: The solid circle (triangle/square) is the conditional expectation of the values, conditional on the total value being the same as the solid circle, illustrated by the hollow circles (triangles/squares). In this example, such conditional expectations are aligned on the ray $v_2 = 2v_1$.

Theorem 3 *If the seller's prior distribution* G *is stochastically comonotonic, then pure bundling is informationally robust.*

To prove Theorem 3, we invoke Theorem 1 with $\mathcal{K} = \{N\}$. The application is not trivial, however, since the ambiguity set $\mathcal{F}_G$ requires a collection of convex moment conditions instead of a single convex moment condition as required by the ambiguity set in (1). Consequently, in order to apply Theorem 1, we identify a single "binding" convex moment function of the form ϕ_N—namely, one that applies to the total sum of values—out of all convex moment conditions in $\mathcal{F}_G$. Let $\mathcal{F}$ be the ambiguity set that results from imposing only that binding condition. Then, $\mathcal{F}$ conforms to the form assumed in (1) with $\mathcal{K} = \{N\}$, so Theorem 1 can be applied against the ambiguity set $\mathcal{F}$. This means that if the seller were to face $\mathcal{F}$ as the ambiguity set, then pure bundling is robustly optimal against the worst-case distribution F^*, which has comonotonic support.

Note, however, that $\mathcal{F}$ is not the true ambiguity set; instead, the seller's ambiguity set is $\mathcal{F}_G$, a subset of $\mathcal{F}$. Here is where the stochastic comonotonicity of G is required. If G is stochastically comonotonic, then G is a mean preserving spread of F^* (see Figure 5 for an illustration), so $\mathcal{F}_G$ does indeed contain F^*. This means that the pure bundling that forms a saddle point along with F^* given $\mathcal{F}$ also forms a saddle point given $\mathcal{F}_G$ (by Corollary 2), which proves that it is max-min optimal against $\mathcal{F}_G$.

Stochastic comonotonicity is fairly general. In particular, it accommodates *exchangeable prior* as a special case. To see this, suppose G is exchangeable; namely, for all permutations $(i_1, \ldots, i_n)$ of $(1, \ldots, n)$, $G(v_1, \ldots, v_n) = G(v_{i_1}, \ldots, v_{i_n})$. Then, for each i,

$$\mathbb{E}\left[v_i \middle| \textstyle\sum_{j=1}^n v_j\right] = \frac{1}{n}\textstyle\sum_k \mathbb{E}\left[v_k \middle| \textstyle\sum_j v_j\right] = \frac{1}{n}\mathbb{E}\left[\textstyle\sum_k v_k \middle| \textstyle\sum_j v_j\right] = \frac{1}{n}\textstyle\sum_{j=1}^n v_j,$$

so G is stochastically comonotonic with $\xi_i = 1/n$ for all i. Therefore, Theorem 3 nests Theorem 1 of Deb and Roesler (2023), which proves the same result under exchangeable prior. This generalization is relevant both from conceptual and practical perspectives. It is analytically important since it speaks to the essential feature of the prior that makes pure bundling informationally robust. The reader of Deb and Roesler (2023) may conclude that the exchangeability of the prior, with all the restrictions it involves, may be crucial for the result. In particular, one may wonder if symmetry is an important driver of what makes pure bundling robustly optimal.[17] Our analysis shows that the symmetry implied by an exchangeable prior is not crucial for pure bundling to be robustly optimal. As is clear, stochastic comonotonicity allows for arbitrary asymmetry in terms of the mean values. In this sense, our theorem uncovers the fundamental property of the prior that makes pure bundling informationally robust.

[17] Deb and Roesler (2023) study two extensions that permit asymmetry across items. They show that (i) pure bundling is still robustly optimal when the value of a proper subset $B \subsetneq N$ of items can be lower than $\sum_{i \in B} v_i$, and (ii) bundling a proper subset $B \subsetneq N$ is robustly optimal when the values are "exchangeable" only within B. Nevertheless, both extensions rely on the symmetry of all items within the considered bundle. Our Theorem 3 can be easily extended to obtain (i) analogously to Corollary 3 and to obtain (ii) as is explained in the subsequent paragraphs.

Second, the generality gained from stochastic comonotonicity is not just significant, it is also practically relevant. Stochastic comonotonicity is consistent with many simple models or heuristics used in various settings. For instance, consider an investment bank's problem in pricing the assets. There are n assets, all belonging to a sector N. The capital asset pricing model (CAPM) implies a prior distribution of the asset values consistent with stochastic comonotonicity. To be concrete, suppose the bank subscribes to the model that assesses the return of each of n assets as:

$$r_i = \beta_i \cdot r_m + e_i,$$

where the random variable r_m is the "market return" of the sector, a constant β_i is the "beta" of the asset and e_i is the idiosyncratic risk satisfying $\sum e_i = 0$ and $\mathbb{E}[e_i|r_m] = 0$. Since $\mathbb{E}[r_i|\sum_j r_j] = \beta_i \sum_j r_j$, such a model satisfies stochastic comonotonicity. If the investment bank is concerned with informational ambiguity, the max-min optimal policy would be to sell its bundle as an asset.

Finally, while Theorem 3 only covers the case of $\mathcal{K} = \{N\}$, with pure bundling as the optimal solution, it is not difficult to generalize the theorem to obtain partial bundling as an informationally robust policy by expanding the nature of ambiguity. For any arbitrary partition $\mathcal{K}$, suppose that the seller only knows the marginal distribution G^K for each (partial) bundle K of items, without knowing anything at all about the correlation of the values across different K's within $\mathcal{K}$. Further, the seller does not know the information the buyer may have about the item values. In this case, the seller faces both distributional and informational ambiguity. Applying Theorem 1, one can show that the seller facing such ambiguity will find $\mathcal{K}$-bundled sales as a robustly optimal strategy. See Che and Zhong (2022) for details. Cast in the investment banker example, if her prior is given by the sector-specific CAPM model, then an informationally robust strategy calls for bundling all assets in each sector K and selling the alternative bundles separately. In other words, the optimal menu of portfolios contains the market return (i.e. a market *index*) for each sector.

6 General Distributional Ambiguity

We have so far focused on ambiguity sets characterized by moment conditions. In this section, we go beyond moment conditions and identify a general structure of ambiguity sets that would give rise to the robust optimality of $\mathcal{K}$-bundled sales. Special cases will identify the conditions that justify separate sales and pure bundling.

To state the general condition on the ambiguity set, we first define an operator $\Upsilon_{\mathcal{K}} : F \mapsto \Delta(\mathbb{R}_+)^{|\mathcal{K}|}$:

$$\Upsilon_{\mathcal{K}}(F) := \left\{ (F_K)_{K \in \mathcal{K}} \in \Delta(\mathbb{R}_+)^{|\mathcal{K}|} : \forall K \in \mathcal{K}, \forall z \in \mathbb{R}_+, F_K(z) := \mathbb{P}_F\{\textstyle\sum_{j \in K} v_j \leqslant z\} \right\},$$

where $\mathbb{P}_F\{\cdot\} := \mathbb{E}_F[\mathbf{1}_{\{\cdot\}}]$. In words, $\Upsilon_{\mathcal{K}}(F)$ calculates the marginal distribution of the total value of each bundle $K \in \mathcal{K}$, given the initial distribution F. $\Upsilon_{\mathcal{K}}(F)$ is called the **$\mathcal{K}$-marginals** of F.

Definition 1 *Fix any arbitrary partition $\mathcal{K}$ of N. An ambiguity set $\mathcal{F} \subset \Delta(\mathbb{R}_+^n)$ exhibits* **$\mathcal{K}$-Knightian ambiguity** *if $\mathcal{F} = \Upsilon_{\mathcal{K}}^{-1} \circ \Upsilon_{\mathcal{K}}(\mathcal{F})$.*

The notion of $\mathcal{K}$-Knightian ambiguity assumes two types of ambiguity. First, the seller has arbitrary knowledge about the $\mathcal{K}$-marginals; thus, all $\mathcal{K}$-marginals (F_K) in $\Upsilon_{\mathcal{K}}(\mathcal{F})$ are considered possible. Second, for each tuple of $\mathcal{K}$-marginals that the seller considers possible, she faces full ambiguity about the joint distribution; thus, all joint distributions in $\Upsilon_{\mathcal{K}}^{-1}((F_K))$ are considered possible. In particular, this means she faces ambiguity on a) the correlation of total values of product groups K's across those in $\mathcal{K}$ and b) the distribution of values across items within each product group $K \in \mathcal{K}$.

A special case of $\mathcal{K}$-Knightian ambiguity is the case studied in Section 2 where the seller knows only moments of the $\mathcal{K}$-marginals. But there are many other examples. For instance, $\mathcal{K}$-marginals may be constrained such that $(F_K)_{K \in \mathcal{K}} \in \mathcal{G} \subset \Delta(\mathbb{R}_+)^{|\mathcal{K}|}$, for some arbitrary set $\mathcal{G}$. We list specific examples of $\mathcal{K}$-Knightian ambiguity:

- For each K, the ambiguity set may include every F_K within a distance, say $\delta_K > 0$, from some reference marginal distribution F_K^0 the seller finds plausible.[18] Bergemann and Schlag (2011) formulated ambiguity in this sense.

- For each K, the ambiguity set may require $\underline{F}_K \leqslant_{SO} F_K \leqslant_{SO} \overline{F}_K$ for some benchmark distributions $\underline{F}_K, \overline{F}_K$ and some arbitrary stochastic order $\leqslant_{SO}$ that is closed under convex combinations. Examples of such stochastic orders are

[18]The metric could be sup norm or Levy-Prokhorov, among others.

First-Order Stochastic Order, Second-Order Stochastic Dominance, Lehmann, Supermodularity, or combinations thereof.[19]

In addition to the knowledge specified by $\mathcal{F}$, we allow the seller to have arbitrary knowledge about the means of item values. Specifically, consider a set

$$\widehat{\mathcal{F}} := \{F \in \Delta(\mathbb{R}^n_+) : (\mu_1(F), ..., \mu_n(F)) \in \Omega\},$$

where Ω is an arbitrary nonempty subset of $\mathbb{R}^n_{++}$. We then assume that the seller's ambiguity set is given by $\mathcal{F} \bigcap \widehat{\mathcal{F}}$. Clearly, when $\Omega = \mathbb{R}^n_+$, the constraint specified by $\widehat{\mathcal{F}}$ has no bite at all.

The main result requires some technical assumptions. We say a set $\mathcal{F}'$ of distributions is **regular** if $\mathcal{F}'$ *is nonempty, convex, closed under weak topology, tight, and has bounded expectation.*[20] Our main theorem then follows:

Theorem 4 *Fix any partition $\mathcal{K}$ of N. Suppose the seller faces a regular ambiguity set $\mathcal{F} \bigcap \widehat{\mathcal{F}}$, where $\mathcal{F}$ exhibits $\mathcal{K}$-Knightian ambiguity. Then, a $\mathcal{K}$-bundled sales mechanism is robustly optimal in the sense that*

$$\sup_{M \in \mathcal{M}} \inf_{F \in \mathcal{F} \bigcap \widehat{\mathcal{F}}} R(M, F) = \sup_{M \in \mathcal{M}_{\mathcal{K}}} \inf_{F \in \mathcal{F} \bigcap \widehat{\mathcal{F}}} R(M, F).$$

PROOF: See Appendix A.4. ■

$\mathcal{K}$-Knightian ambiguity crystallizes the insight that gives rise to separation and bundling in the earlier section. Specifically, the concept captures the ambiguity about how the values of alternative bundles in $\mathcal{K}$ are correlated and the ambiguity about how a given value of a bundle $K \in \mathcal{K}$ is distributed across items within K. The former gives rise to the separation of sales across alternative bundles in $\mathcal{K}$ whereas the latter ambiguity gives rise to the bundled sales of items within each K.

[19]Recall, however, from Section 5 that the Second-Order Stochastic Dominance Order, or equivalently the Mean Preserving Spread Order, can be handled by dispersion moment conditions involving particular (piece-wise linear) convex moment functions.

[20]A set of measures on $\mathbb{R}^n_+$ is tight if for any $\epsilon > 0$ there is a compact subset $S \subset \mathbb{R}^n_+$ whose measure is at least $1 - \epsilon$. All other notions are standard.

As special cases, the theorem provides conditions for the robust optimality of two canonical sales mechanisms:[21]

Corollary 5 *The seller's ambiguity set $\mathcal{F}$ exhibits $\mathcal{K}$-Knightian ambiguity.*

1. *If $\mathcal{K}$ is the finest partition of N and $\mathcal{F}$ is regular, then separate sales are robustly optimal.*
2. *If $\mathcal{K}$ is the coarsest partition of N and $\mathcal{F} \bigcap \widehat{\mathcal{F}}$ is regular, then pure bundling is robustly optimal.*

Theorem 4 identifies $\mathcal{K}$-Knightian ambiguity as a fundamental general condition for $\mathcal{K}$-bundled sales to be robustly optimal. To the best of our knowledge, this condition provides for the most general characterization of the extent to which items should be bundled or separated. Since $\mathcal{K}$-Knightian ambiguity holds under the moment restrictions considered in Section 3, this condition can be seen as responsible for the robust optimality found in that section.[22] Nevertheless, Theorem 4 does not make that section superfluous. Note that the current theorem does not identify the exact form of the optimal mechanism or the worst-case distribution, whereas the additional structure given by moment restrictions allowed us to identify them in Theorem 1. Not only is the exact identification of the mechanism and distribution important and useful of its own right, it enables us to go beyond $\mathcal{K}$-Knightian ambiguity, which is sufficient but not necessary for $\mathcal{K}$-bundled sales to be robustly optimal. For instance, as noted by Corollary 2 and Theorem 3, the exact solution of the joint distribution enables us to identify a robustly optimal mechanism—i.e., $\mathcal{K}$-bundled sales—even when the ambiguity set $\tilde{\mathcal{F}}$ fails $\mathcal{K}$-Knightian ambiguity. Finally, the worst-case distribution

[21] Corollary 5 part 1 formalizes the conjecture in p. 481 of Carroll (2017): when $\mathcal{G} = \Upsilon_{\mathcal{K}}(\mathcal{F})$ is "well-behaved enough to contain a single worst marginal distribution," then separate sale is robustly optimal. Indeed, even when the seller's ambiguity set contains a non-singleton set of marginal distributions, if it admits a unique saddle point, then the optimal mechanism in the saddle point must be robustly optimal against the exact marginal distributions associated with that saddle point, so the mechanism must be separating, following Theorem 1 of Carroll (2017). However, it is not easy to guarantee existence or uniqueness of a saddle point. For instance, our regularity condition does not necessarily lead to the existence of a saddle point or its uniqueness. We therefore did not follow Carroll (2017)'s conjectured recipe of establishing (unique) saddle points for each item. Our regularity condition, although not sufficient for existence of a saddle point, guarantees the robust optimality of $\mathcal{K}$-bundled sales mechanisms, following a minimax argument.

[22] The moment conditions required by $\mathcal{F}$ in Section 3 clearly satisfies $\mathcal{K}$-Knightian ambiguity. We prove in Appendix B.5 that $\mathcal{F}$ considered in Section 3 is regular.

found in Theorem 1 plays a crucial part in the proof of Theorem 4, which makes the former indispensable for obtaining the current generalization.

7 Concluding Remarks

The current paper has characterized robustly optimal mechanisms for selling multiple goods for a monopolist faced with ambiguity on the buyer's private valuations of the goods. The nature of the robustly optimal mechanism depends on the type of ambiguity facing the seller. We have identified moment conditions as well as general distributional conditions leading to the robust optimality of a $\mathcal{K}$-bundled sales mechanism, which includes the commonly used sales mechanisms of separate sales and pure bundling as two special cases. The distributional condition that we identify, namely, $\mathcal{K}$-Knightian ambiguity, is the most general kind known to date that rationalizes these sales mechanisms. More importantly, the concept captures the clear economic insights that give rise to separation and bundling of items in a (robustly) optimal sale. As argued in detail, ambiguity about the correlation of values across items/bundles leads to separation of items/bundles, whereas ambiguity about across-items value dispersion leads to the bundling of items in the sale. In particular, the latter ambiguity features the threat of negatively-correlated item values as a reason for favoring a bundled sales, thus connecting with the classic insight provided by Adams and Yellen (1976).

Carrying the theme of Carroll (2017) to its fruition, the current paper thus provides a general robustness perspective on the rationale for alternative canonical sales mechanisms. As such, it offers a complementary as well as an alternative perspective on the subject matter which has so far been approached almost exclusively from a Bayesian mechanism design perspective.

There are at least two avenues along which one could further extend the current paper. First, our model, like all other papers on the subject matter, assumes a single buyer, and naturally, one might consider introducing multiple buyers into the model. Two concurrent papers have made progress under such generality; however, there is still no general answer.[23] He et al. (2024) provide a limiting result that as the

[23] A few papers identify robustly optimal mechanisms in *single-item* auctions. Brooks and Du (2021a) finds a robustly optimal auction mechanism, when robustness is required with respect to value distributions with known means and common domain, buyers' high-order beliefs, and to equilibrium

number of ex-ante identical buyers grows large, it is *asymptotically* robustly optimal to auction off via a second-price format the robustly optimal bundle identified in the current paper. Brooks and Du (2023) provide a duality result that enables the calculation of the *informationally* robust revenue guarantee via linear programming and show that it is without loss to consider an auction format with one-dimensional message space.

Second, while the current paper offers a robustness-based rationale for separate sales and pure bundling as well as more general $\mathcal{K}$-bundling, we do not offer a rationale for so-called "mixed-bundling," i.e., a menu of options for buying goods both separately and a bundle. Although the nature of ambiguity that would justify such a mechanism remains unknown, we hope our current paper will offer useful insights for future inquiry into this topic.

References

Adams, W. J. and Yellen, J. L. (1976). Commodity bundling and the burden of monopoly. *The quarterly journal of economics*, pages 475–498.

Armstrong, M. (1996). Multiproduct nonlinear pricing. *Econometrica*, 64:51–76.

Armstrong, M. (1999). Price discrimination by a many-product firm. *The Review of Economic Studies*, 66:151–168.

Babaioff, M., Immorlica, N., Lucier, B., and Weinberg, S. M. (2014). A simple and approximately optimal mechanism for an additive buyer. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pages 21–30. IEEE.

Bergemann, D., Brooks, B., and Morris, S. (2016). Informationally robust optimal auction design. Cowles Foundation Discussion Paper No. 2065.

Bergemann, D., Brooks, B., and Morris, S. (2019). Revenue guarantee equivalence. *American Economic Review*, 109(5):1911–1929.

Bergemann, D. and Schlag, K. (2011). Robust monopoly pricing. *Journal of Economic Theory*, 146(6):2527–2543.

Bergemann, D. and Schlag, K. H. (2008). Pricing without priors. *Journal of the European Economic Association*, 6(2-3):560–569.

selection. Considering a similar mean constraint but restricting attention to the private-value setting, Che (2022) identifies a robustly optimal auction mechanism within a class of "competitive" mechanisms which encompass standard auctions. Similarly, Bergemann et al. (2019) identifies an informationally-robust optimal auction mechanism in the class of symmetric and standard auctions.

Blackwell, D. (1951). Comparison of experiments. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 93–102.

Briest, P., Chawla, S., Kleinberg, R., and Weinberg, M. (2010). Pricing randomized allocations. In *ACM Sympsium on Discrete Algorithms*, pages 585–597.

Brooks, B. and Du, S. (2021a). Maxmin auction design with known expected values. Technical report, Tech. rep., The University of Chicago and University of California-San Diego.

Brooks, B. and Du, S. (2021b). Optimal auction design with common values: An informationally robust approach. *Econometrica*, 89(3):1313–1360.

Brooks, B. and Du, S. (2023). On the structure of informationally robust optimal mechanisms. *Available at SSRN 3663721*.

Carrasco, V., Luz, V. F., Kos, N., Messner, M., Monteiro, P., and Moreira, H. (2018). Optimal selling mechanisms under moment conditions. *Journal of Economic Theory*.

Carroll, G. (2015). Robustness and linear contracts. *American Economic Review*, 105(2):536–563.

Carroll, G. (2017). Robustness and separation in multidimensional screening. *Econometrica*, 85(2):453–488.

Carroll, G. (2019). Robust incentives for information acquisition. *Journal of Economic Theory*, 181:382–420.

Che, E. (2022). Robustly optimal auction design under mean constraints. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 153–181.

Che, Y.-K. and Zhong, W. (2022). Robustly optimal mechanisms for selling multiple goods. *arXiv preprint arXiv:2105.02828*.

Daskalakis, C., Deckelbaum, A., and Tzamos, C. (2013). Mechanism design via optimal transport. In *Proceedings of the fourteenth ACM conference on Electronic commerce*, pages 269–286.

Daskalakis, C., Deckelbaum, A., and Tzamos, C. (2017). Strong duality for a multiple-good monopolist. *Econometrica*, 85(3):735–767.

Deb, R. and Roesler, A.-K. (2023). Multi-Dimensional Screening: Buyer-Optimal Learning and Informational Robustness. *The Review of Economic Studies*, page rdad100.

Du, S. (2018). Robust mechanisms under common valuation. *Econometrica*, 86(5):1569–1588.

Guo, Y. and Shmaya, E. (2023). Robust monopoly regulation. Mimeo.

Haghpanah, N. and Hartline, J. (2021). When is pure bundling optimal? *The Review of Economic Studies*, 88(3):1127–1156.

Hart, S. and Nisan, N. (2012). Approximate revenue maximization with multiple items. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, EC '12, page 656, New York, NY, USA. Association for Computing Machinery.

Hart, S. and Nisan, N. (2013). The menu-size complexity of auctions. In *ACM Conference on Electronic Commerce*, pages 565–566.

Hart, S. and Reny, P. J. (2015). Maximal revenue with multiple goods: Nonmonotonicity and other observations. *Theoretical Economics*, 10(3):893–922.

Hart, S. and Reny, P. J. (2019). The better half of selling separately. *ACM Transactions on*

Economics and Computation (TEAC), 7(4):1–18.

He, W. and Li, J. (2022). Correlation-robust auction design. *Journal of Economic Theory*, 200:105403.

He, W., Li, J., and Zhong, W. (2024). Rank-guaranteed auctions. Mimeo.

Kartik, N. and Zhong, W. (2023). Lemonade from lemons: Information design and adverse selection. Mimeo.

Koçyiğit, Ç., Iyengar, G., Kuhn, D., and Wiesemann, W. (2019). Distributionally robust mechanism design. *Management Science*, Articles in Advance:1–31.

Koçyiğit, Ç., Rujeerapaiboon, N., and Kuhn, D. (2021). Robust multidimensional pricing: Separation without regret. *Mathematical Programming*, pages 1–34.

Li, X. and Yao, A. C.-C. (2013). On revenue maximization for selling multiple independently distributed items. *Proceedings of the National Academy of Sciences*, 110(28):11232–11237.

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

Manelli, A. M. and Vincent, D. R. (2006). Bundling as an optimal selling mechanism for a multiple-good monopolist. *Journal of Economic Theory*, 127(1):1–35.

McAfee, R. P. and McMillan, J. (1988). Multidimensional incentive compatibility and mechanism design. *Journal of Economic theory*, 46(2):335–354.

Menicucci, D., Hurkens, S., and Jeon, D.-S. (2015). On the optimality of pure bundling for a monopolist. *Journal of Mathematical Economics*, 60:33–42.

Osborne, M. J. and Rubinstein, A. (1994). *A Course in Game Theory*. The MIT Press; Cambridge, Massachusetts; London, England.

Ravid, D., Roesler, A.-K., and Szentes, B. (2022). Learning before trading: on the inefficiency of ignoring free information. *Journal of Political Economy*, 130(2):346–387.

Rochet, J.-C. and Chone, P. (1988). Ironing, sweeping, and multidimensional screening. *Econometrica*, 66:783–826.

Roesler, A.-K. and Szentes, B. (2017). Buyer-optimal learning and monopoly pricing. *American Economic Review*, 107(7):2072–80.

Scarf, H. E. (1957). A min-max solution of an inventory problem. Technical report P-910, The RAND Corporation.

Sion, M. (1958). On general minimax theorems. *Pacific Journal of mathematics*, 8(1):171–176.

A Appendix: Proofs

A.1 Proof of Theorem 1

Proof: We prove Theorem 1 by explicitly constructing a $\mathcal{K}-$bundled sales mechanism M^* and a valuation distribution F^* and verify that they constitute a saddle point.

Construction of F^*. We first construct F^*, nature's choice of distribution. This involves two steps. We first fix an arbitrary pair $(\boldsymbol{m}, \boldsymbol{s}) \in \Omega$ and construct a distribution $F^{(\boldsymbol{m},\boldsymbol{s})}$ with means and dispersion characterized by $(\boldsymbol{m}, \boldsymbol{s})$. We later describe how $(\boldsymbol{m}, \boldsymbol{s})$ is chosen. To begin, let $m_K := \sum_{j \in K} m_j$ for each $K \in \mathcal{K}$, and let $K(i) = \{K \in \mathcal{K} : i \in K\}$ denote the bundle containing item i.

Let X be a random variable distributed from $[1, \infty)$ according to a cdf H:

$$H(x) := \text{Prob}[X \leqslant x] = 1 - \frac{1}{x}.$$

Then, the value of item $i \in K$, for $K \in \mathcal{K}$, is given by:

$$V_i(X) := \min\{\alpha_K X, \beta_K\} \cdot \frac{m_i}{m_K},$$

where, for each $K \in \mathcal{K}$, the parameters $0 < \alpha_K < m_K < \beta_K$ satisfy:

$$\int_1^{\frac{\beta_K}{\alpha_K}} \frac{\alpha_K}{x} \mathrm{d}x + \alpha_K = m_K; \tag{6}$$

$$\int_1^{\frac{\beta_K}{\alpha_K}} \frac{\phi_K(\alpha_K x)}{x^2} \mathrm{d}x + \frac{\phi_K(\beta_K)\alpha_K}{\beta_K} = s_K. \tag{7}$$

In short, the total value of each product group $K \in \mathcal{K}$ rises co-monotonically and linearly with the common random variable X at rate α_K; the value of each item i is then determined in proportion to its mean m_i relative to the total mean of the group value. The parameters $(\alpha_K, \beta_K)_{K \in \mathcal{K}}$ are in turn determined to satisfy the moment conditions with respect to the sum of means m_K and dispersion s_K. Lemma A.1 guarantees that such a pair exists for any $(m_K, s_K) \gg (0, \phi_K(m_K))$.

By continuity of $(\alpha_K)_{K \in \mathcal{K}}$ and compactness of Ω, there exists

$$(\boldsymbol{m}, \boldsymbol{s}) = \arg \min_{(\tilde{\boldsymbol{m}}, \tilde{\boldsymbol{s}}) \in \Omega} \sum_{K \in \mathcal{K}} \alpha_K(\tilde{m}_K, \tilde{s}_K).$$

Setting $F^* := F^{(\boldsymbol{m},\boldsymbol{s})}$ completes the construction of nature's choice of distribution.

Construction of M^*. Next, we define the candidate optimal mechanism M^*. In a nutshell, the seller sells each bundle K separately at a random price distributed according to G_K. The corresponding direct mechanism is:

$$\begin{cases} q_i^*(\boldsymbol{v}) = G_{K(i)}\left(\sum_{j \in K(i)} v_j\right), \\ t^*(\boldsymbol{v}) = \sum_{K \in \mathcal{K}} \int_{p \leqslant \sum_{j \in K} v_j} p G_K(\mathrm{d}p). \end{cases}$$

The cdf G_K is defined via the density function:

$$g_K(v) := \lambda_K \cdot \frac{\phi_K'(\beta_K) - \phi_K'(v)}{v}$$

on $[\alpha_K, \beta_K]$ and zero elsewhere, where $\lambda_K := 1/[\int_{\alpha_K}^{\beta_K} \frac{\phi_K'(\beta_K)-\phi_K'(x)}{x}\mathrm{d}x]$ normalizes the density so that it integrates to one. As is illustrated in Example 1, g_K is constructed so that the revenue from selling bundling K is an affine transformation of the moment function ϕ_K.

Verification of saddle point We first compute the value $R(M^*, F^*)$. For any $\boldsymbol{v}$ in the support of F^*,

$$\begin{aligned} t^*(\boldsymbol{v}) &= \sum_{K\in\mathcal{K}} \int_{\alpha_K}^{\sum_{j\in K} v_j} p g_K(\mathrm{d}p) \\ &= \sum_{K\in\mathcal{K}} \lambda_K \left\{ \phi_K'(\beta_K) \left(\textstyle\sum_{j\in K} v_j - \alpha_K \right) - \phi_K \left(\textstyle\sum_{j\in K} v_j \right) + \phi_K(\alpha_K) \right\}. \end{aligned}$$

Hence,

$$\begin{aligned} R(M^*,F^*) &= \int t^*(\boldsymbol{v}) F^*(\mathrm{d}\boldsymbol{v}) \\ &= \sum_{K\in\mathcal{K}} \lambda_K \left\{ \phi_K'(\beta_K)(m_K - \alpha_K) + \phi_K(\alpha_K) - \int \left(\phi_K \left(\textstyle\sum_{j\in K} v_j \right) \right) F^*(\mathrm{d}\boldsymbol{v}) \right\} \\ &= \sum_{K\in\mathcal{K}} \lambda_K \{ \phi_K'(\beta_K)(m_K - \alpha_K) + \phi_K(\alpha_K) - s_K \} \\ &= \sum_{K\in\mathcal{K}} \frac{\phi_K'(\beta_K)\alpha_K \log(\beta_K/\alpha_K) - \alpha_K \int_{\alpha_K}^{\beta_K} \frac{\phi_K'(x)}{x}\mathrm{d}x}{\int_{\alpha_K}^{\beta_K} \frac{\phi_K'(\beta_K)-\phi_K'(x)}{x}\mathrm{d}x} = \sum_{K\in\mathcal{K}} \alpha_K. \end{aligned} \tag{8}$$

The first three equalities are straightforward. The fourth equality follows from (6) and (7) and from recalling that $\lambda_K = 1/[\int_{\alpha_K}^{\beta_K} \frac{\phi_K'(\beta_K)-\phi_K'(x)}{x}\mathrm{d}x]$.

Next, we show that $M^* \in \arg\max_{M\in\mathcal{M}} R(M, F^*)$. Fix any $M = (q, t) \in \mathcal{M}$. Since the support of F^* is a parametric curve $\boldsymbol{V}(x)$, the mechanism M can be represented equivalently via $(\psi(x), \tau(x)) := (q(\boldsymbol{V}(x)), t(\boldsymbol{V}(x)))$. Since M satisfies (IC), it must satisfy the envelope condition:

$$\tau(x) = \psi(x) \cdot \boldsymbol{V}(x) - \int_1^x \psi(z) \cdot \boldsymbol{V}'(z)\mathrm{d}z.$$

Hence,

$$\begin{aligned}
R(M, F^*) &= \int \tau(x) H(\mathrm{d}x) \\
&\leqslant \sup_{\psi(\cdot)} \int \psi(x) \cdot \left(\boldsymbol{V}(x) - \boldsymbol{V}'(x) \frac{1 - H(x)}{h(x)} \right) H(\mathrm{d}x) \\
&= \sup_{\psi} \sum_i \int_1^{\frac{\beta_{K(i)}}{\alpha_{K(i)}}} \psi_i(x) \cdot 0 H(\mathrm{d}x) + \int_{\frac{\beta_{K(i)}}{\alpha_{K(i)}}}^{\infty} \psi_i(x) \cdot \gamma_i \cdot \beta_{K(i)} H(\mathrm{d}x) \\
&\leqslant \sum_i \gamma_i \cdot \beta_{K(i)} \cdot \frac{\alpha_{K(i)}}{\beta_{K(i)}} = \sum_{K \in \mathcal{K}} \alpha_K = R(M^*, F^*), \qquad (9)
\end{aligned}$$

where $\gamma_i := \frac{m_i}{\sum_{j \in N} m_j}$. The second inequality follows from $\psi_i \leqslant 1$. The third equality follows from $\sum_{i \in N} \gamma_i = 1$. The last equality follows from (8).

Finally, we show that $F^* \in \arg\min_{F \in \mathcal{F}} R(M^*, F)$. To this end, observe

$$t^*(\boldsymbol{v}) \geqslant \sum_{K \in \mathcal{K}} \lambda_K \left\{ \phi'_K(\beta_K) \left(\textstyle\sum_{j \in K} v_j - \alpha_K \right) - \phi_K \left(\textstyle\sum_{j \in K} v_j \right) + \phi_K(\alpha_K) \right\}.$$

To see why this inequality holds, observe first that $t^*(\boldsymbol{v}) = RHS$ when $\sum_{j \in K} v_j \in [\alpha_K, \beta_K]$ (recall the very first displayed equation in the proof). Outside that region, $t^*(\boldsymbol{v})$ is constant in $\boldsymbol{v}$, while the RHS is strictly decreasing in $\sum_{j \in K} v_j$ when $\sum_{j \in K} v_j > \beta_K$ and strictly increasing in $\sum_{j \in K} v_j$ when $\sum_{j \in K} v_j < \alpha_K$. It then follows that, for any $F \in \mathcal{F}$,

$$\begin{aligned}
R(M^*, F) &= \int t^*(\boldsymbol{v}) F(\mathrm{d}\boldsymbol{v}) \\
&\geqslant \int \sum_{K \in \mathcal{K}} \lambda_K \left\{ \phi'_K(\beta_K) \left(\textstyle\sum_{j \in K} v_j - \alpha_K \right) - \phi_K \left(\textstyle\sum_{j \in K} v_j \right) + \phi_K(\alpha_K) \right\} F(\mathrm{d}\boldsymbol{v}) \\
&= \sum_{K \in \mathcal{K}} \lambda_K \left\{ \phi'_K(\beta_K) \left(\textstyle\sum_{j \in K} \mathbb{E}_F[v_j] - \alpha_K \right) + \phi_K(\alpha_K) \right\} - \sum_{K \in \mathcal{K}} \lambda_K \int \phi_K \left(\textstyle\sum_{j \in K} v_j \right) F(\mathrm{d}\boldsymbol{v}) \\
&= \underbrace{\sum_{K \in \mathcal{K}} \lambda_K \{ \phi'_K(\beta_K)(m_K - \alpha_K) + \phi_K(\alpha_K) - s_K \}}_{A} \\
&\quad - \underbrace{\sum_{K \in \mathcal{K}} \lambda_K \left\{ \phi'_K(\beta_K)(m_K - \textstyle\sum_{j \in K} \mathbb{E}_F[v_j]) + \int \left(\phi_K \left(\textstyle\sum_{j \in K} v_j \right) - s_K \right) F(\mathrm{d}\boldsymbol{v}) \right\}}_{B}.
\end{aligned}$$

Note that (6) and (7), together with $\lambda_K = 1/[\int_{\alpha_K}^{\beta_K} \frac{\phi'_K(\beta_K) - \phi'_K(x)}{x} \mathrm{d}x]$, imply that $A = \sum_{K \in \mathcal{K}} \alpha_K = R(M^*, F^*)$.

The above inequalities imply that $R(M^*, F) \geqslant R(M^*, F^*) - B$. If F has the same moments as F^*, then $B = 0$, so we are done. Hence, assume F has different moments than F^*. Suppose for the sake of contradiction that $R(M^*, F) < R(M^*, F^*)$. Since $\mathcal{F}$ is a convex set, if we define $F^\delta = F^* + \delta(F - F^*)$, then $F^\delta \in \mathcal{F}$ for any $\delta \in [0, 1]$. Since $R(M^*, F)$ is linear in F, we must then have $\frac{\mathrm{d}R(M^*,F^\delta)}{\mathrm{d}\delta}\big|_{\delta=0} < 0$. In particular, this implies that $\frac{\mathrm{d}B}{\mathrm{d}\delta}\big|_{\delta=0} > 0$. However, one can show that

$$\frac{\mathrm{d}B}{\mathrm{d}\delta}\Big|_{\delta=0} = -\frac{\mathrm{d}\sum_{K\in\mathcal{K}} \alpha_K}{\mathrm{d}\delta}\Big|_{\delta=0}. \tag{10}$$

(see Appendix B.1 for the details). Hence, $\frac{\mathrm{d}B}{\mathrm{d}\delta}\big|_{\delta=0} > 0$ means that F^δ entails a smaller value of $\sum \alpha_K$ relative to F^*, and thus lower revenue, for sufficiently small δ. However, this contradicts the fact that $\sum \alpha_K$ is minimized at $(\boldsymbol{m}, \boldsymbol{s})$. Therefore, we conclude that $R(M^*, F) \geqslant R(M^*, F^*)$. ■

Lemma A.1 *For any $(m_K, s_K) \gg (0, \phi_K(m_K))$ for each $K \in \mathcal{K}$, there exists a unique pair (α_K, β_K) satisfying (6) and (7). The mapping $(m_K, s_K)_K \mapsto (\alpha_K)_K$ is continuous.*

PROOF: From (6), we can solve for $\beta_K = \alpha_K e^{\frac{m_K - \alpha_K}{\alpha_K}}$. Substituting this into (7), its LHS becomes a continuous function of α_K. It is strictly decreasing in α_K for any $\alpha_K < \beta_K$:

$$\begin{aligned}
\frac{\text{dLHS of (7)}}{\mathrm{d}\alpha_K} &= \left(\frac{\phi_K(\beta_K)}{(\beta_K/\alpha_K)^2} - \frac{\phi_K(\beta_K)}{(\beta_K/\alpha_K)^2}\right)\cdot\frac{\mathrm{d}(\beta_K/\alpha_K)}{\mathrm{d}\alpha_K} + \int_1^{\frac{\beta_K}{\alpha_K}} \frac{\phi_K'(\alpha_K x)}{x}\mathrm{d}x + \frac{\phi_K'(\beta_K)\alpha_K}{\beta_K}\cdot\frac{\mathrm{d}\beta_K}{\mathrm{d}\alpha_K} \\
&= \int_1^{\frac{\beta_K}{\alpha_K}} \frac{\phi_K'(\alpha_K x)}{x}\mathrm{d}x - \phi_K'(\beta_K)\frac{m_K - \alpha_K}{\alpha_K} \\
&= \phi_K'(\beta_K)\left(\int_1^{\frac{\beta_K}{\alpha_K}} \frac{\phi_K'(\alpha_K x)}{x\phi_K'(\beta_K)}\mathrm{d}x - \frac{m_K - \alpha_K}{\alpha_K}\right) \\
&< \phi_K'(\beta_K)\left(\log\beta_K - \log\alpha_K - \frac{m_K - \alpha_K}{\alpha_K}\right) \\
&= 0,
\end{aligned}$$

where the strict inequality follows from the convexity of ϕ, and the last equality is from substituting $\beta_K = \alpha_K e^{\frac{m_K - \alpha_K}{\alpha_K}}$.

Observe next that the LHS of (7) is strictly less than its RHS when $\alpha_K = m_K$. It is strictly greater than the RHS when α_K is sufficiently low. To see this, note

$$\int_1^{\frac{\beta_K}{\alpha_K}} \frac{\phi_K(\alpha_K x)}{x^2}\mathrm{d}x \geqslant \int_1^{\frac{\beta_K}{\alpha_K}} \left(\frac{\phi_K(0)}{x^2} + \frac{\phi_K'(0)(\alpha_K x)}{x^2} + \frac{1}{2}\varepsilon\frac{(\alpha_K x)^2}{x^2}\right)\mathrm{d}x$$

$$=\phi_K(0)\left(1-\frac{\alpha_K}{\beta_K}\right)+\phi_K'(0)\alpha_K(\log\beta_K-\log\alpha_K)+\frac{1}{2}\varepsilon\alpha_K(\beta_K-\alpha_K)$$
$$\geqslant-|\phi_K(0)|+\phi_K'(0)(m_K-\alpha_K)+\frac{1}{2}\varepsilon\alpha_K^2\left(e^{\frac{m_K-\alpha_K}{\alpha_K}}-1\right).$$

The last line tends to ∞ as $\alpha_K \to 0$.

Collecting the observations so far, we conclude that there exists a unique pair (α_K, β_K) satisfying (6) and (7).

It is easy to see that both sides of (6) and (7) are continuous in $(\alpha_K, \beta_K, m_K, s_K)$. Therefore, α_K and β_K each as a correspondence of (m_K, s_K) has a closed graph. Since we have shown that $\alpha_K(m_K, s_K)$ is a function, it is continuous. ■

A.1.1 Proof of Corollary 3

Let (M^*, F^*) be the saddle point pinned down in Theorem 1. It is straightforward that F^* remains optimal within $\mathcal{F}$ because M^* remains incentive compatible as the values for all bundles within $\mathcal{K}$ remain unchanged. Then, the revenue from M^* does not change with u_K for any F.

Now, we prove that M^* remains optimal. Note that since buyer's value is no longer additive, the definition of an allocation should be generalized to $\psi : x \to [0,1]^{2^N}$, i.e. probability of allocating each bundle, subject to $\forall i, \sum_{K\ni i}\psi_K(x) \leqslant 1$. Then, given $M = (\psi(x), \tau(x))$ the envelope condition implies:

$$\tau(x) = \sum_{K\in 2^N} u_K \cdot \psi_K(x) \cdot \sum_{i\in K} V_i(x) - \int_{\underline{x}}^{x} \sum_{K\in 2^N} u_K \cdot \psi_K(z) \cdot \sum_{i\in K} V_i(z)\mathrm{d}z$$
$$\Longrightarrow \quad R(M, F^*) \leqslant \sup_{\psi(\cdot)} \int \sum_{K\in 2^N} u_K \cdot \psi_K(x) \cdot \sum_{i\in K}\left(V_i(x) - V_i'(x)\frac{1-H(x)}{h(x)}\right) H(\mathrm{d}x)$$
$$= \sup_{\psi(\cdot)} \sum_i \int_{\frac{\beta_{K(i)}}{\alpha_{K(i)}}}^{\infty} \sum_{K\ni i} \psi_K(x) u_K \gamma_i \beta_{K(i)} H(\mathrm{d}x)$$
$$\leqslant \sum_i \gamma_i \alpha_{K(i)} = R(M^*, F^*).$$

The second inequality is from $\sum_{K\ni i}\psi_K(x) \leqslant 1$ and $u_K \leqslant 1$.

A.2 Proof of Theorem 2

(Part 1) Fix any nonempty $J, J' \subset K$ for some $K \in \mathcal{K}$ such that $J \cap J' = \varnothing$. We show that it is never robustly optimal to separate J and J'. To this end, it suffices to find $\widetilde{F} \in \mathcal{F}$ such that $\sup_{M \in \mathcal{M}_{\mathcal{K}'}} R(M, \widetilde{F}) < R(M^*, F^*)$, for any partition $\mathcal{K}'$ such that $\{J, J'\} \subset \mathcal{K}'$.

We construct $\widetilde{F}$ as follows. Define CDF H and (α_K, β_K) as in Theorem 1. Recall $X \sim H$. Define a new binomial random variable Y whose value is zero with probability $\frac{m_{J'}}{m_{J \cup J'}}$ and one with probability $\frac{m_J}{m_{J \cup J'}}$. Let $0 < \varepsilon < \min_K \frac{\alpha_K}{\beta_K}$. The distribution $\widetilde{F}$ is then defined by the item values:

$$V_i(X, Y) = \begin{cases} \min\left\{\alpha_{K(i)} X, \beta_{K(i)}\right\} \cdot \dfrac{m_i}{m_{K(i)}} & \text{if } i \notin J \cup J' \\ \min\left\{\alpha_{K(i)} X, \beta_{K(i)}\right\} \cdot \dfrac{m_i}{m_{K(i)}} & \text{if } i \in J \cup J' \text{ and } X \leqslant 1/\varepsilon \\ \beta_{K(i)} \cdot \dfrac{m_i}{m_{K(i)}} \cdot \dfrac{m_{J \cup J'}}{m_J} \cdot Y & \text{if } i \in J \text{ and } X > 1/\varepsilon \\ \beta_{K(i)} \cdot \dfrac{m_i}{m_{K(i)}} \cdot \dfrac{m_{J \cup J'}}{m_{J'}} \cdot (1 - Y) & \text{if } i \in J' \text{ and } X > 1/\varepsilon, \end{cases}$$

where recall $K(i) := K \in \mathcal{K}$ such that $i \in K$. In words, the values of items $i \notin J \cup J'$ are distributed same as F^*. The values of $j \in J \cup J'$ are also distributed same as F^* *conditional on* $X < 1/\varepsilon$, an event that occurs with probability $H(1/\varepsilon) = 1 - \varepsilon$. In the complementary event, the value of good $j \in J$ becomes either $\beta_{K(i)} \cdot \frac{m_i}{m_{K(i)}} \cdot \frac{m_{J \cup J'}}{m_{J'}}$ or zero. Effectively, mass ε of value $\beta_{K(i)} \cdot \frac{m_i}{m_{K(i)}}$ is split into a higher value and zero so that the expected value remains the same. Note that $j \in J'$ is split in the same fashion but in a way perfectly negatively correlated as the value of item $i \in J$. The negative correlation means that the dispersion of values of group K remains the same; recall both J and J' are in K. Hence, all moment conditions of (1) continue to be satisfied (since F^* satisfies them). Therefore, $\widetilde{F} \in \mathcal{F}$.

Since the mechanism M separates J and J', one can write:

$$M = \left(q^{-J \cup J'}(v_{i, i \notin J \cup J'}), t^{-J \cup J'}(v_{i, i \notin J \cup J'}), q^J(\sum_{i \in J} v_i), t^J(\sum_{i \in J} v_i), q^{J'}(\sum_{i \in J'} v_i), t^{J'}(\sum_{i \in J'} v_i) \right).$$

In words, groups J and J' are each bundled separately, and the mechanism can be arbitrarily defined on all other items. The IC and IR conditions imply that $(q^{-J \cup J'}, t^{-J \cup J'})$, (q^J, t^J) and $(q^{J'}, t^{J'})$ should each satisfy IC and IR. For $(q^{-J \cup J'}, t^{-J \cup J'})$, since the random vector $\boldsymbol{V}$ is effectively uni-dimensional for $i \notin J \cup J'$, the envelope condition implies:

$$\int t^{-J \cup J'}(\boldsymbol{v}) F(\mathrm{d}\boldsymbol{v}) \leqslant \sup_{\psi} \sum_{i \notin J \cup J'} \int_0^{\frac{\beta_{K(i)}}{\alpha_{K(i)}}} \psi_i(x) \cdot 0 H(\mathrm{d}x) + \int_{\frac{\beta_{K(i)}}{\alpha_{K(i)}}}^{\infty} \psi_i(x) \phi_i \beta_{K(i)} H(\mathrm{d}x)$$

$$\leqslant \sum_{i \notin J \cup J'} \alpha_{K(i)} \frac{m_i}{m_{K(i)}}.$$

The sub-mechanisms (q^J, t^J) and $(q^{J'}, t^{J'})$ sells bundles J and J' separately. For $\varepsilon > 0$ sufficiently small, it is suboptimal to charge price 0 for each bundle. However, charging any other price that leads to positive probability of sales generates revenue of

$$\alpha_K \frac{m_J}{m_K} \left(1 - \frac{m_{J'}}{m_{J \cup J'}} \varepsilon\right)$$

from the sale of bundle J (where $J \subset K$). Likewise, the sale of bundle J' results in the revenue at most of

$$\alpha_K \frac{m_{J'}}{m_K} \left(1 - \frac{m_J}{m_{J \cup J'}} \varepsilon\right).$$

Therefore,

$$\begin{aligned} R(M, \widetilde{F}) &< \sum_{i \notin J \cup J'} \alpha_{K(i)} \frac{m_i}{m_{K(i)}} + \alpha_K \frac{m_J}{m_K} + \alpha_K \frac{m_{J'}}{m_K} \\ &= R(M^*, F^*). \end{aligned}$$

(Part 2) For each $K \in \mathcal{K}$, let $\ell_K := \frac{\beta_K}{\alpha_K}$. Suppose there are $K, K' \in \mathcal{K}$ such that $\ell_K \neq \ell_{K'}$. We will show that it is never robustly optimal for the seller to bundle goods in $K \cup K'$. It suffices to find $\widetilde{F} \in \mathcal{F}$ such that $\sup_{M \in \mathcal{M}_{\mathcal{K}'}} R(M, \widetilde{F}) < R(M^*, F^*)$, for all $\mathcal{K}'$ such that $K \cup K' \in \mathcal{K}'$. This will imply that the revenue guarantee will be strictly lower for any selling mechanism that bundles the groups K and K'.

We construct $\widetilde{F}$ as follows. Without loss, assume $\ell_K > \ell_{K'}$ and let $\ell \in (\ell_{K'}, \ell_K)$. Let H_ε be given by:

$$H_\varepsilon(x) := \begin{cases} H(x) & x \leqslant \ell - \varepsilon \\ H(\ell - \varepsilon) & x \in (\ell - \varepsilon, \ell), \\ H(x - \varepsilon) & x \geqslant \ell. \end{cases}$$

First, we define two parameters α^ε and ℓ^ε based on ℓ and ε:

$$\int_1^{\ell^\varepsilon} (\alpha^\varepsilon x) H_\varepsilon(\mathrm{d}x) + (\alpha^\varepsilon \ell^\varepsilon)(1 - H_\varepsilon(\ell^\varepsilon)) = m_K; \tag{11}$$

$$\int_1^{\ell^\varepsilon} \phi_K(\alpha^\varepsilon x) H_\varepsilon(\mathrm{d}x) + \phi_K(\alpha^\varepsilon \ell^\varepsilon)(1 - H_\varepsilon(\ell^\varepsilon)) = s_K. \tag{12}$$

Denote the LHS of (11) and (12) by $f_K^1(\alpha^\varepsilon, \ell^\varepsilon, \varepsilon)$ and $f_K^2(\alpha^\varepsilon, \ell^\varepsilon, \varepsilon)$, respectively. When ε is sufficiently small, ℓ^ε is close to ℓ_K and is thus strictly larger than ℓ.

Therefore, we compute the Jacobian matrix of the functions $\boldsymbol{f}_K := (f_K^1, f_K^2)$ with respect to $(\alpha^\varepsilon, \ell^\varepsilon)$:

$$\begin{aligned}
&\boldsymbol{J}_{\alpha^\varepsilon, \ell^\varepsilon} \boldsymbol{f}_K(\alpha^\varepsilon, \ell^\varepsilon, \varepsilon) \\
&= \begin{bmatrix} \int_1^{\ell^\varepsilon} x H_\varepsilon(\mathrm{d}x) + \ell^\varepsilon(1 - H_\varepsilon(\ell^\varepsilon)) & \alpha^\varepsilon(1 - H_\varepsilon(\ell^\varepsilon)) \\ \int_1^{\ell^\varepsilon} \phi_K'(\alpha^\varepsilon x) x H_\varepsilon(\mathrm{d}x) + \phi_K'(\alpha^\varepsilon \ell^\varepsilon)\ell^\varepsilon(1 - H_\varepsilon(\ell^\varepsilon)) & \phi_K'(\alpha^\varepsilon \ell^\varepsilon)\alpha^\varepsilon(1 - H_\varepsilon(\ell^\varepsilon)) \end{bmatrix} \\
&\implies \boldsymbol{J}_{\alpha^\varepsilon, \ell^\varepsilon} \boldsymbol{f}_K(\alpha^\varepsilon, \ell^\varepsilon, \varepsilon)\big|_{\varepsilon=0} \\
&= \begin{bmatrix} \int_1^{\ell_K} x H(\mathrm{d}x) + \ell_K(1 - H(\ell_K)) & \alpha_K(1 - H(\ell_K)) \\ \int_1^{\ell_K} \phi_K'(\alpha_K x) x H(\mathrm{d}x) + \phi_K'(\alpha_K \ell_K)\ell_K(1 - H(\ell_K)) & \phi_K'(\alpha_K \ell_K)\alpha_K(1 - H(\ell_K)) \end{bmatrix};
\end{aligned}$$

Meanwhile, the partial derivative of $\boldsymbol{f}_K$ with respect to ε is:

$$\begin{aligned}
&\boldsymbol{J}_\varepsilon \boldsymbol{f}_K(\alpha^\varepsilon, \ell^\varepsilon, \varepsilon) \\
&= \begin{bmatrix} \alpha^\varepsilon \varepsilon h(\ell - \varepsilon) + \alpha^\varepsilon \int_1^{\ell^\varepsilon} h(x + \varepsilon)\mathrm{d}x \\ (\phi_K(\alpha^\varepsilon \ell) - \phi_K(\alpha^\varepsilon(\ell - \varepsilon)))h(\ell - \varepsilon) + \alpha^\varepsilon \int_1^{\ell^\varepsilon} \phi_K'(\alpha^\varepsilon x) h(x + \varepsilon)\mathrm{d}x \end{bmatrix} \\
&\implies \boldsymbol{J}_\varepsilon \boldsymbol{f}_K(\alpha^\varepsilon, \ell^\varepsilon, \varepsilon)\big|_{\varepsilon=0} \\
&= \begin{bmatrix} \alpha_K(H(\ell_K) - H(1)) \\ \alpha_K \int_1^{\ell_K} \phi_K'(\alpha_K x) H(\mathrm{d}x) \end{bmatrix}.
\end{aligned}$$

By the inverse function theorem,

$$\frac{\mathrm{d}\alpha^\varepsilon}{\mathrm{d}\varepsilon}\bigg|_{\varepsilon=0} = -\boldsymbol{J}_{\alpha^\varepsilon, \ell^\varepsilon} \boldsymbol{f}_K^{-1} \cdot \boldsymbol{J}_\varepsilon \boldsymbol{f}_K\bigg|_{\varepsilon=0} = -\frac{\int_1^{\ell_K} (\phi'(\alpha_K x) - \phi'(\alpha_K \ell_K)) H(\mathrm{d}x)}{\int_1^{\ell_K} (\phi_K'(\alpha_K x) - \phi_K'(\alpha_K \ell_K)) x H(\mathrm{d}x)} < 0,$$

Therefore, for ε sufficiently close to 0, $\alpha^\varepsilon < \alpha_K$. Let X be the random variable distributed according to CDF H. Define $\widetilde{\boldsymbol{V}} := (\widetilde{V}_1, ..., \widetilde{V}_n)$, where

$$\widetilde{V}_i = \begin{cases} \dfrac{m_i}{m_J} \min\{\alpha_J X, \alpha_J \ell_J\} & \text{if } i \in J \neq K \\[2ex] \dfrac{m_i}{m_K} \min\left\{\alpha^\varepsilon(X + \varepsilon \mathbf{1}_{\{X > \ell - \varepsilon\}}), \alpha^\varepsilon, \ell^\varepsilon\right\} & \text{if } i \in J = K. \end{cases}$$

Note that by definition, $X + \varepsilon \mathbf{1}_{\{X > \ell - \varepsilon\}}$ is distributed according to H_ε. Let $\widetilde{F}$ be the distribution of $\widetilde{\boldsymbol{V}}$.

Now, consider any mechanism M that bundles $K \cup K'$. M can be written as $(\psi(x), \tau(x), \widetilde{\psi}(x), \widetilde{\tau}(x))$, where $(\psi(x), \tau(x))$ is the allocation and the payment for items $i \notin K \cup K'$ and $(\widetilde{\psi}(x), \widetilde{\tau}(x))$ is the allocation and the payment for items $i \in K \cup K'$, all as functions of x, the report of X. Note this formalism does not imply that the sales of items $i \notin K \cup K'$ is separated from those of $i \notin K \cup K'$. The envelope theorem implies

$$
\begin{aligned}
R(M, \widetilde{F}) \leqslant \sup_{\psi(\cdot), \widetilde{\psi}(\cdot)} & \left(\sum_{i \notin K \cup K'} \int \psi_i(x) (\widetilde{V}_i(x) - \widetilde{V}_i'(x) \frac{1 - H(x)}{h(x)}) H(\mathrm{d}x) \right. \\
& \left. + \int \widetilde{\psi}(x) \sum_{i \in K \cup K'} (\widetilde{V}_i(x) - \widetilde{V}_i'(x) \frac{1 - H(x)}{h(x)}) H(\mathrm{d}x) \right) \\
\leqslant & \sum_{J \in \mathcal{K}, J \neq K, K'} \alpha_J + \sup_x \left(\sum_{i \in K \cup K'} \widetilde{V}_i(x)(1 - H(x)) \right) \\
= & \sum_{J \in \mathcal{K}, J \neq K, K'} \alpha_J + \sup_x \underbrace{\left(\alpha_{K'} \min\{x, \ell_{K'}\}(1 - H(x)) + \alpha^\varepsilon \min\{x, \ell^\varepsilon\}(1 - H_\varepsilon(x)) \right)}_{A(x)}.
\end{aligned}
$$

The second inequality is from the definition of α_J's and the fact that $\widetilde{\psi}$ equivalently characterizes a mechanism that bundles $K \cup K'$. For $x \leqslant \ell - \varepsilon$, $H_\varepsilon(x) = H(x)$, but $\alpha^\varepsilon < \alpha_K$, so

$$
A(x) = \alpha_{K'} + \alpha^\varepsilon < \alpha_{K'} + \alpha_K.
$$

For $x \in (\ell - \varepsilon, \ell^\varepsilon]$, $x > \ell_{K'}$ when ε is chosen sufficiently small. Therefore,

$$
\begin{aligned}
A(x) = & \alpha_{K'} \frac{\ell_{K'}}{x} + \alpha^\varepsilon \frac{x}{x - \varepsilon} \\
< & \alpha_{K'} \frac{\ell_K}{\ell - \varepsilon} + \alpha^\varepsilon \frac{\ell - \varepsilon}{\ell - 2\varepsilon}.
\end{aligned}
$$

As $\varepsilon \to 0$, the latter expression tends to $\alpha_K + \alpha_{K'} \frac{\ell_K}{\ell} < \alpha_K + \alpha_{K'}$. Combining both case proves that when ε is sufficiently small, $R(M, \widetilde{F}) < \sum \alpha_J = R(M^*, F^*)$. Therefore,

$$
\sup_{M \in \mathcal{M}_{\mathcal{K}}} R(M, \tilde{F}) < R(M^*, F^*),
$$

as was to be shown.

A.3 Proof of Theorem 3

PROOF: Fix any prior $G \in \Delta(\mathbb{R}_+^n)$ that is stochastically monotonic with sharing parameters $\boldsymbol{\xi} = (\xi_1, ..., \xi_n) \in \mathbb{R}_+^n$ with $\sum_i \xi_1 = 1$. We prove the theorem by construction and verification. We would like to invoke Theorem 1 for the case of $\mathcal{K} = \{N\}$. We begin by identifying a relaxed ambiguity set. For all $\boldsymbol{v} \in \mathbb{R}_+^n$, let $\boldsymbol{v} = (v_1, ..., v_n)$. For any $z > 0$, consider an ambiguity set

$$\mathcal{F}_z = \left\{ F \in \Delta(\mathbb{R}_+^n) | \mathbb{E}_F[\boldsymbol{v}] = \mathbb{E}_G[\boldsymbol{v}] \text{ and } \mathbb{E}_F[\phi_z(\boldsymbol{v})] \leqslant \mathbb{E}_G[\phi_z(\boldsymbol{v})] \right\},$$

indexed by z, where $\phi_z(\boldsymbol{v}) := \max\{z - \sum_i v_i, 0\}$. More importantly, the associated condition captures the second-order stochastic dominance order with respect to the total sum of item values: namely, the random variable $\sum_i v_i$ distributed according to F second-order stochastically dominates (SOSD) the corresponding random variable distributed according to G if $\mathbb{E}_F[\phi_z(\boldsymbol{v})] \leqslant \mathbb{E}_G[\phi_z(\boldsymbol{v})]$ for all $z > 0$.[24]

Observe that ϕ_z is not twice-differentiable, as is required by Theorem 1. Nevertheless, it is piecewise linear with only one kink; hence, the saddle point can be explicitly constructed following the same procedure as in Theorem 1, denoted by (F_z, M_z), where M_z is a pure bundling mechanism. Let α_z be the parameter defining F_z, indexed by z. Then, let z^* minimizes α_z while maintaining

$$\mathbb{E}_{F_{\alpha_z}}[\phi_{z'}(\boldsymbol{v})] \leqslant \mathbb{E}_G[\phi_{z'}(\boldsymbol{v})]$$

holds for all $z' \in \mathbb{R}_+$. Define $F^* := F_{z^*},\ M^* := M_{z^*}$.[25]

We are now ready to prove our statement: (M^*, F^*) is a saddle point under ambiguity set $\mathcal{F}_G$, defined in (5). To this end, note first that $\mathcal{F}_G \subset \bigcap_{z>0} \mathcal{F}_z \subset \mathcal{F}_{z^*}$ (since $(\phi_z)_{z>0}$ comprises only a subset of all possible convex functions). In light of Corollary 2, it suffices to show $F^* \in \mathcal{F}_G$, since that will imply that M^* is maxmin optimal given ambiguity set $\mathcal{F}_G$.

[24] This can be seen upon integration by parts:

$$\int_{\sum_i v_i \leqslant z} F(\boldsymbol{v})\mathrm{d}\boldsymbol{v} = \int_{\sum_i v_i \leqslant z} (z - \sum v_i) F(\mathrm{d}\boldsymbol{v}) = \mathbb{E}_F[\phi_z(\boldsymbol{v})] \leqslant \mathbb{E}_G[\phi_z(\boldsymbol{v})] = \int_{\sum_i v_i \leqslant z} G(\boldsymbol{v})\mathrm{d}\boldsymbol{v},$$

where the inequality is from $F \in \mathcal{F}_z$.

[25] The formal construction of the saddle point and the identification of z^* is relegated to Appendix B.4.

To show $F^* \in \mathcal{F}_G$, fix any arbitrary convex function $\phi : \mathbb{R}^n_+ \to \mathbb{R}$. Let $\boldsymbol{m} = (m_1, ..., m_n)$, where $m_i := \mathbb{E}_G[v_i] = \mathbb{E}_G[\xi_i(\sum_j v_j)]$.

$$\begin{aligned}
\mathbb{E}_G[\phi(\boldsymbol{v})] &= \mathbb{E}_G\Big[\mathbb{E}[\phi(\boldsymbol{v}) \,|{\textstyle\sum_i} v_i]\Big] \\
&\geqslant \mathbb{E}_G\left[\phi\big(\mathbb{E}[\boldsymbol{v}|{\textstyle\sum_i} v_i]\big)\right] \\
&= \mathbb{E}_G\big[\phi(\boldsymbol{\xi} \cdot \textstyle\sum v_i)\big] \\
&\geqslant \mathbb{E}_{F^*}\big[\phi(\boldsymbol{\xi} \cdot \textstyle\sum v_i)\big] \\
&= \mathbb{E}_{x\sim H}\left[\phi\left(\min\left\{\alpha_N x, \alpha_K e^{\frac{\sum m_i - \alpha_N}{\alpha_N}}\right\} \cdot \frac{\boldsymbol{m}}{\sum m_i}\right)\right] \\
&= \mathbb{E}_{F^*}\left[\phi(\boldsymbol{v})\right].
\end{aligned}$$

The first inequality follows from the convexity of ϕ. The second equality follows from stochastic comonotonicity of G; i.e., G satisfies Assumption 1 for some $\boldsymbol{\xi} = (\xi_1, ..., \xi_n) \in \mathbb{R}^n_+$ with $\sum_i \xi_1 = 1$. To understand the second inequality, observe first that the composition function $\phi \circ \boldsymbol{\xi}$ is convex in $\sum_i v_i$, the total sum of all item values. Next, recall that the total sum of all item values distributed according to F^* SOSD the total sum of item values distributed according to G—a fact implied by $F^* \in \bigcap_z \mathcal{F}_z$. Hence, the second inequality follows. The last two equalities follow from the definition of F^* given in Section 3. Since ϕ is an arbitrary convex function, we have shown that $F^* \in \mathcal{F}_G$, and the proof of Theorem 3 is complete. ■

A.4 Proof of Theorem 4

PROOF: Observe

$$\begin{aligned}
\sup_{M\in\mathcal{M}_\mathcal{K}} \inf_{F\in\mathcal{F}\bigcap\hat{\mathcal{F}}} R(M,F) &= \inf_{F\in\mathcal{F}\bigcap\hat{\mathcal{F}}} \sup_{M\in\mathcal{M}_\mathcal{K}} R(M,F) \\
&= \inf_{(F_K)\in\Upsilon_\mathcal{K}(\mathcal{F})} \inf_{F\in\Upsilon_\mathcal{K}^{-1}((F_K))\bigcap\hat{\mathcal{F}}} \sup_{M\in\mathcal{M}_\mathcal{K}} R(M,F) \\
&\geqslant \inf_{(F_K)\in\Upsilon_\mathcal{K}(\mathcal{F})} \sup_{M\in\mathcal{M}_\mathcal{K}} \inf_{F\in\Upsilon_\mathcal{K}^{-1}((F_K))\bigcap\hat{\mathcal{F}}} R(M,F) \\
&= \inf_{(F_K)\in\Upsilon_\mathcal{K}(\mathcal{F})} \sup_{M\in\mathcal{M}} \inf_{F\in\Upsilon_\mathcal{K}^{-1}((F_K))\bigcap\hat{\mathcal{F}}} R(M,F) \\
&\geqslant \sup_{M\in\mathcal{M}} \inf_{(F_K)\in\Upsilon_\mathcal{K}(\mathcal{F})} \inf_{F\in\Upsilon_\mathcal{K}^{-1}((F_K))\bigcap\hat{\mathcal{F}}} R(M,F) \\
&= \sup_{M\in\mathcal{M}} \inf_{F\in\mathcal{F}\bigcap\hat{\mathcal{F}}} R(M,F).
\end{aligned}$$

The first equality follows from Lemma B.1 proven in Appendix B.2, where $\mathcal{G} = \mathcal{F} \bigcap \widehat{\mathcal{F}}$. The two inequalities are min-max inequalities. The third equality follows from Lemma A.2. Since $\mathcal{M}_{\mathcal{K}} \subset \mathcal{M}$, the above inequalities yields the desired statement. ■

Lemma A.2 *Fix any $\mathcal{K}$-marginals $(F_K)_{K\in\mathcal{K}}$ and any $\widehat{\mathcal{F}}$, and let $\mathcal{F} := \Upsilon_{\mathcal{K}}^{-1}((F_K)_{K\in\mathcal{K}}) \bigcap \widehat{\mathcal{F}} \neq \varnothing$. Then, $\mathcal{K}$-bundled sales is robustly optimal in the sense that*

$$\sup_{M\in\mathcal{M}_{\mathcal{K}}} \inf_{F\in\mathcal{F}} R(M,F) = \sup_{M\in\mathcal{M}} \inf_{F\in\mathcal{F}} R(M,F).$$

PROOF: We first construct the worst case $F^* \in \Delta(\mathbb{R}_+^n)$. To this end, we imagine a hypothetical problem in which the seller sells $k = |\mathcal{K}|$ goods and faces full ambiguity given the knowledge of the marginal distributions F_K of the values of each item $K \in \mathcal{K}$. (That is, we interpret bundle K as a single item in this hypothetical problem.) This is precisely what Carroll (2017) analyzed. To recast his result in the current setup for this hypothetical problem, let $\mathcal{M}^k$ be the set of feasible mechanisms in this hypothetical problem with k items, and $R^k(M,G)$ denote the revenue the seller collects from a mechanism $M \in \mathcal{M}^k$ facing distribution $G \in \Delta(\mathbb{R}_+^k)$. (The corresponding notations for our original problem would then have superscript n, which we suppress for convenience.) Consider $\mathcal{M}_{\mathcal{H}}^k$, where $\mathcal{H}$ is the finest partition of $\mathcal{K}$. Then $\mathcal{M}_{\mathcal{H}}^k$ is the set of all "separate sales" mechanisms in this hypothetical problem.

Theorem 1 of Carroll (2017) then proves that there exists $G^* \in \Upsilon_{\mathcal{H}}^{-1}((F_K)_{K\in\mathcal{K}}) \subset \Delta(\mathbb{R}_+^k)$ and

$$\sup_{M\in\mathcal{M}_{\mathcal{H}}^k} R^k(M,G^*) = \sup_{M\in\mathcal{M}^k} R^k(M,G^*). \tag{13}$$

Now we construct $F^* \in \Delta(\mathbb{R}_+^n)$ using $G^* \in \Delta(\mathbb{R}_+^k)$. Choose any $F' \in \mathcal{F} = \Upsilon_K^{-1}((F_K)) \bigcap \widehat{\mathcal{F}}$ (which we assumed to be nonempty). For each i, let

$$\alpha_i := \frac{\mathbb{E}_{F'}[v_i]}{\sum_{j\in K(i)} \mathbb{E}_{F'}[v_j]}.$$

Let $\boldsymbol{X}$ be a k-dimensional random vector distributed according to G^*, and let $\boldsymbol{V} = (V_i)$ be a n-dimensional random vector defined by

$$V_i(\boldsymbol{X}) := \alpha_i X_{K(i)},$$

for each i. Let F^* be the distribution of $\boldsymbol{V}$.

We now prove $F^* \in \Upsilon_K^{-1}((F_K)) \bigcap \widehat{\mathcal{F}}$. Since $G^* \in \Upsilon_{\mathcal{H}}^{-1}((F_K))$, by construction, $F^* \in \Upsilon_K^{-1}((F_K))$. Since $F' \in \Upsilon_K^{-1}((F_K))$,

$$\mathbb{P}_{F'}\left\{\textstyle\sum_{j\in K} v_j \leqslant y\right\} = \mathbb{P}_{G^*}\left\{X_K \leqslant y\right\} = \mathbb{P}_{F^*}\left\{\textstyle\sum_{j\in K} v_j \leqslant y\right\},$$

for each $K \in \mathcal{K}$ and $y \in \mathbb{R}_+$. Hence, $\mathbb{E}_{F'}[\sum_{j\in K} v_j] = \mathbb{E}_{G^*}[X_K] = \mathbb{E}_{F^*}[\sum_{j\in K} v_j]$. It further follows that $\mathbb{E}_{F'}[v_j] = \mathbb{E}_{F^*}[v_j]$. Hence, $F^* \in \widehat{\mathcal{F}}$. We thus conclude that $F^* \in \mathcal{F}$.

We next prove that

$$\sup_{M\in\mathcal{M}^k} R^k(M, G^*) \geqslant \sup_{M\in\mathcal{M}} R(M, F^*). \tag{14}$$

To see this, fix any mechanism $M = (q, t) \in \mathcal{M}$ in our original problem. We now construct another mechanism $\tilde{M} = (\tilde{q}, \tilde{t}) \in \mathcal{M}^k$ for the hypothetical k-item problem as follows:

$$\begin{cases} \tilde{q}_K(\boldsymbol{x}) = \sum\limits_{j\in K} \alpha_j q_j(\boldsymbol{V}(\boldsymbol{x})), \\ \tilde{t}(\boldsymbol{x}) = t(\boldsymbol{V}(\boldsymbol{x})). \end{cases}$$

Observe that $\forall \boldsymbol{x}, \boldsymbol{x}'$,

$$\begin{aligned} \boldsymbol{x} \cdot \tilde{q}(\boldsymbol{x}') - \tilde{t}(\boldsymbol{x}') &= \sum_{K\in\mathcal{K}} x_K \sum_{j\in K} \alpha_j q_j(\boldsymbol{V}(\boldsymbol{x}')) - t(\boldsymbol{V}(\boldsymbol{x}')) \\ &= \sum_i \frac{V_i(\boldsymbol{x})}{\alpha_i} \alpha_i q_i(\boldsymbol{V}(\boldsymbol{x}')) - t(\boldsymbol{V}(\boldsymbol{x}')) = \boldsymbol{V}(\boldsymbol{x}) \cdot q(\boldsymbol{V}(\boldsymbol{x}')) - t(\boldsymbol{V}(\boldsymbol{x}')). \end{aligned}$$

Therefore, (IC) and (IR) of M on $\operatorname{supp}(F^*)$ imply (IC) and (IR) of $\tilde{M}$ on $\operatorname{supp}(G^*)$. Hence, $\tilde{M} \in \mathcal{M}^k$. Moreover, given G^*, $\tilde{M}$ yields the same expected revenue as M given F^*. Since one can find such $\tilde{M}$ for each $M \in \mathcal{M}$, (14) follows.

We next prove

$$\sup_{M\in\mathcal{M}_{\mathcal{K}}} R(M, F^*) \geqslant \sup_{M\in\mathcal{M}^k_{\mathcal{H}}} R^k(M, G^*). \tag{15}$$

Indeed, for any $\tilde{M} \in \mathcal{M}^k_{\mathcal{H}}$, we can construct a $\mathcal{K}$-bundled sales mechanism $\widehat{M} = (\widehat{q}, \widehat{t}) \in \mathcal{M}_{\mathcal{K}}$, where

$$\widehat{q}_i(\boldsymbol{v}) := \tilde{q}_{K(i)}(\textstyle\sum_{j\in K(i)} v_j) \text{ and } \widehat{t}(\boldsymbol{v}) := \displaystyle\sum_{K\in\mathcal{K}} \tilde{t}_K(\textstyle\sum_{j\in K} v_j).$$

Whenever $\boldsymbol{v} = \boldsymbol{V}(\boldsymbol{x})$, $\sum_{j\in K} v_j = \sum_{j\in K} \alpha_j x_K = x_K$. So, $\widehat{M}$ given F^* is payoff equivalent to $\tilde{M}$ given G^*. Since $\tilde{M} \in \mathcal{M}^k_{\mathcal{H}}$, satisfying (IR) and (IC) on supp(G^*), $\widehat{M}$ satisfies (IC) and (IR) on supp(F^*).

Combining (15), (13), and (14), we obtain

$$\sup_{M\in\mathcal{M}_{\mathcal{K}}} R(M,F^*) \geqslant \sup_{M\in\mathcal{M}^k_{\mathcal{H}}} R^k(M,G^*) = \sup_{M\in\mathcal{M}^k} R^k(M,G^*) \geqslant \sup_{M\in\mathcal{M}} R(M,F^*). \tag{16}$$

Finally, fix any mechanism $M \in \mathcal{M}_{\mathcal{K}}$. For any $F \in \mathcal{F} = \Upsilon^{-1}_K((F_K)) \bigcap \widehat{\mathcal{F}}$,

$$R(M,F) = \int t(\boldsymbol{v})F(\mathrm{d}\boldsymbol{v}) = \sum_{K\in\mathcal{K}} \int t(\textstyle\sum_{j\in K} v_j)F(\mathrm{d}\boldsymbol{v}) = \displaystyle\sum_{K\in\mathcal{K}} \int t(x)F_K(\mathrm{d}x),$$

where the first equality follows from the fact that $M \in \mathcal{M}_{\mathcal{K}}$ and the second follows from $F \in \Upsilon^{-1}_K((F_K))$. In other words, $R(M,F) = R(M,F')$ for any $F, F' \in \Upsilon^{-1}_K((F_K)) \bigcap \widehat{\mathcal{F}} = \mathcal{F}$, as long as $M \in M_{\mathcal{K}}$. Hence, it follows that

$$\sup_{M\in\mathcal{M}_{\mathcal{K}}} \inf_{F\in\mathcal{F}} R(M,F) = \sup_{M\in\mathcal{M}_{\mathcal{K}}} R(M,F^*). \tag{17}$$

Combining (17) with (16), we get

$$\sup_{M\in\mathcal{M}_{\mathcal{K}}} \inf_{F\in\mathcal{F}} R(M,F) = \sup_{M\in\mathcal{M}_{\mathcal{K}}} R(M,F^*) \geqslant \sup_{M\in\mathcal{M}} R(M,F^*) \geqslant \sup_{M\in\mathcal{M}} \inf_{F\in\mathcal{F}} R(M,F).$$

Since $\mathcal{M}_{\mathcal{K}} \subset \mathcal{M}$, the reverse inequality also holds, so we have the desired conclusion.

■