

Robustly Optimal Mechanisms for Selling Multiple Goods

YEON-KOO CHE and WEIJIE ZHONG*

January 2, 2023

Abstract

We study robustly optimal mechanisms for selling multiple items. The seller maximizes revenue *robustly* 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 a variety of moment conditions on the values of subsets of goods. We also identify general properties of the ambiguity set that lead to the robust optimality of partial bundling which includes separate sales and pure bundling as special cases.

1 INTRODUCTION

How should a seller sell multiple goods to a buyer? This seemingly simple problem has long baffled scholars. With only one good, it is optimal for the seller to charge a simple monopoly price (Myerson (1981) and Riley and Zeckhauser (1983)). With multiple goods, however, the optimal mechanism is not easy to characterize. Even when the buyer demands only two goods and values them additively according to a quasi-linear utility function, the optimal mechanism may involve a menu of lotteries whose exact forms are highly sensitive to the distribution of the buyer’s values (Daskalakis et al. (2017), and Manelli and Vincent (2007)). Unsettlingly still, simple mechanisms such as item pricing or bundled pricing can perform arbitrarily poorly compared to theoretical optima (Briest et al. (2010); Hart and Nisan (2013)).

*Che: Department of Economics, Columbia University (email: yeonkooche@gmail.com); Zhong: Graduate School of Business, Stanford University (email: wayne.zhongwj@gmail.com)

While complete analysis is still elusive, recent progress based on a linear programming dual approach (Daskalakis et al. (2017) and Manelli and Vincent (2006)) has helped the authors to find sufficient conditions for the optimality of simple, realistic mechanisms such as pure or mixed bundling.¹ These works constitute major advances in the subject matter and provide useful insights. Yet, these insights apply only to specific, and often limited, environments, and it is unclear what general lessons can be obtained regarding the design of optimal mechanisms or those employed in practice.²

In this regard, Carroll (2017)’s robust optimality approach offers a promising new lead. In his model, the seller knows the distribution of the buyer’s valuation for each good, but she faces ambiguity on how the item values are correlated with one another.³ The seller is ambiguity averse, so she maximizes revenue under the worst-case correlation structure. Strikingly, the mechanism that is robustly optimal in this sense “separates”—namely, the seller simply sells each good separately at a monopoly price. The optimality of separate sales is robust in the sense that it does not depend on the marginal distribution of item values.

As striking and robust as the result is, the model employs a stark assumption that the seller knows the *exact* marginal distributions of buyer’s item values while knowing *nothing at all* about their correlation. Intuitively, her knowledge about the marginals should make the seller confident, at least, about her revenue from selling each good separately, whereas her total ignorance about their correlation should make her nervous about linking the sale of alternative goods in a particular way. For instance, pure bundling would work for some correlation structures but may prove disastrous for others. For this reason, the robust optimality of a separate mechanism may not

¹Pure bundling is optimal when the virtual value of each individual good is nonnegative in the two good setting (Menicucci et al. (2015)), when the lower bound for a uniform distribution is sufficiently high (Pavlov (2011) and Daskalakis et al. (2017)), or when the value of a smaller bundle is relatively higher for a high type than for a low type (Haghpanah and Hartline, 2020). A “mixed bundling” which offers a menu of individual items and their bundle for sale is optimal for the two goods case if the buyers’ values are iid uniform on $[0, 1]$ (Manelli and Vincent, 2007), or if the values follow special distributions such as exponential or power distribution (Daskalakis et al., 2017). In a similar spirit, Hart and Reny (2019) show that separate sales attain certain lower bounds of optimal revenue, although the analysis does not compare separate sales with, say, pure bundling.

²Daskalakis et al. (2017) provide a general characterization for the optimality of pure bundling. The characterizing conditions are useful for verifying/certifying the optimality of pure bundling but are not easy to interpret in economic terms.

³The buyer is assumed to have a quasilinear utility function that is additively separable across goods.

appear so surprising given this particular knowledge structure.⁴ One may thus wonder what role, if any, the stark form of ambiguity he assumes plays in his model for the sweeping conclusion. Would the conclusion change if, more realistically, the seller does not know the precise marginal distributions of item values, but rather knows only some aspects or summary statistics about them? What happens if the seller also has (possibly limited) knowledge about the aggregate value of some or all goods?

In this paper, we consider general, and arguably more realistic, forms of ambiguity the seller may face given her limited knowledge about the buyer’s valuation. This knowledge may take the form of summary statistics such as moments of individual item values or of total values for several goods, or it may involve other restrictions on the buyer’s value distribution. We then study the robustly optimal selling mechanism given such a general form of ambiguity. Specifically, the seller maximizes the expected revenue by choosing any feasible—i.e., incentive compatible and individually rational—selling mechanism given the worst-case distribution consistent with the seller’s ambiguity. Formally, this is obtained by solving for a saddle point, or an equilibrium, of a zero-sum game played by the seller who seeks to maximize her revenue and the adversarial nature who seeks to minimize it. The saddle point identifies the optimal revenue guarantee for the seller.

In our model, a seller has $n \geq 2$ heterogeneous goods to sell to a buyer. In keeping with the literature, we assume that the buyer has a quasilinear utility function that is additive in $(v_1, \dots, v_n) \in \mathbb{R}_+^n$ for goods. The values are the buyer’s private information, and thus unobserved by the seller. The seller only knows that the joint distribution F of the values lies within some set $\mathcal{F} \subset \Delta(\mathbb{R}_+^n)$, called an *ambiguity set*. \mathcal{F} is characterized by a variety of moment conditions with a partition structure. Let \mathcal{K} be an arbitrary partition of goods $\{1, \dots, n\}$, and each element $K \in \mathcal{K}$ is a “bundle” of items. We assume that the seller possesses some knowledge of the mean value of each individual item and a dispersion moment of the total value of each bundle K , where a *dispersion moment* is the expectation of an arbitrary convex moment function. Formally, \mathcal{F} contains all value distributions whose means of each item and dispersion moments of each bundle in \mathcal{K} lie in an arbitrary convex and compact set. Special cases of the ambiguity set include the setting where the seller only knows the moments of

⁴As Carroll (2017) correctly points out, this insight is incomplete, for it is in principle possible for some form of bundling to outperform the optimal separate mechanism *regardless of the correlation structure*.

individual item values. An ambiguity set given by such item-wise moments captures the spirit of [Carroll \(2017\)](#) while relaxing the stark form of ambiguity: the seller has *some* knowledge about individual item values but not about their correlation. In another extreme case, the seller may know the dispersion moment of the *total value of all items*.

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} is the finest partition and the coarsest partition, respectively. These results are intuitive. In the former case, the seller has knowledge of the dispersion of individual item values, but lacks knowledge on the correlation across alternative item values; separate sales are thus robust against this form of ambiguity. By contrast, in the latter case, the seller has knowledge of the dispersion of total value of all items, but faces ambiguity about how the dispersion of total value is distributed across individual item values; this form of ambiguity can be best dealt with by selling all goods in a grand bundle. This intuition explains the robust optimal of partial bundling for a general partition \mathcal{K} : the ambiguity about the cross-item distribution within the value of each product group $K \in \mathcal{K}$ explains the bundling of items within that group, whereas the ambiguity on the correlation across the values of alternative product groups suggests the separation across the groups.

We next show in [Theorem 2](#) that the partial bundling structure is not only sufficient, but also necessary for the robust optimality of the mechanism: it is no longer robustly optimal either to separate items within any product group $K \in \mathcal{K}$ or to bundle across multiple product groups in \mathcal{K} . The counterfactual distribution that would make these alternative mechanisms suboptimal deepens our understanding about what necessitates the seller to choose the particular partial bundling mechanism identified in [Theorem 1](#). Separate sales of items (or product groups) are motivated by the fear that asymmetry in the buyer’s distribution may lead to screening inefficiency and revenue loss if items were bundled. By contrast, a bundled sales of items is motivated by the fear that a certain negative correlation across values would lead to revenue loss if items were sold separately. This latter finding harks back to the classic insight by [Adams and Yellen \(1976\)](#).

In [Section 5](#), we apply our framework to study “informational ambiguity,” wherein

the seller faces ambiguity about the possible signals the buyer might possess about his values. There has been a growing interest within the economics literature to identify a mechanism that is robustly optimal against such informational ambiguity. While the existing literature assumes that the ambiguity is purely informational, meaning the seller has an unambiguous prior on the value distribution, we consider a more general framework in which the seller faces both prior and informational ambiguity. Specifically, for an arbitrary partition \mathcal{K} , the seller in our model knows the (marginal) distributions of item values in each group $K \in \mathcal{K}$. But she faces ambiguity on the correlation of item values across product groups as well as on the signals that the buyer has about the item values consistent with the marginals. Theorem 3 identifies \mathcal{K} -bundled sales to be the robustly optimal sale mechanism, under the assumption that the marginal on each product group $K \in \mathcal{K}$ can be decomposed into a co-monotone common component and idiosyncratic components with zero mean. This latter condition contains a large class of distributions that nests as a special case the exchangeable prior required by [Deb and Roesler \(2021\)](#); in particular, it allows for a full range of distributional asymmetries across item values.

While moment conditions are natural and reasonable in many contexts, other forms of knowledge/ambiguity may emerge in other contexts. In Section 6, we identify a general sufficient condition that justifies the use of the \mathcal{K} -bundled sales as a robustly optimal sales mechanism for any arbitrary partition \mathcal{K} of goods. The key condition we develop is **\mathcal{K} -Knightian ambiguity**. To explain, fix any arbitrary partition \mathcal{K} of goods, and suppose that the seller finds some \mathcal{K} -marginals—the marginal distributions of the total values of alternative bundles in \mathcal{K} —to be compatible with her ambiguity set \mathcal{F} . If any other distribution, say F' , also gives rise to the same \mathcal{K} -marginals, then \mathcal{K} -Knightian ambiguity requires that such a distribution also belong to the ambiguity set \mathcal{F} . We show that \mathcal{K} -Knightian ambiguity encompasses not only the moment conditions studied in Section 3, but also the other types of distributional or informational ambiguities mentioned earlier.

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 \(2020\)](#). The current paper

departs from this literature by taking a robustness approach.

Second, the current paper contributes to the literature on robust mechanism design. A growing number of 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 the context of 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. \(2019\)](#), [Che \(2022\)](#), and [Brooks and Du \(2021\)](#) 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.⁵ We develop his framework further and provide a robustness-based rationale for general forms of partial bundling, which include separate sales and pure bundling as special cases.⁶

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 \(2019\)](#), and [Deb and Roesler \(2021\)](#). 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. 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](#)

⁵[He and Li \(2020\)](#) study the robust optimal mechanism for selling a single item to multiple buyers. Similar to [Carroll \(2017\)](#), the seller knows exact marginal distributions of buyers’ valuations without any knowledge of their correlation.

⁶Although 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 \(2019\)](#) 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 that is 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 \(2013\)](#)).

and Roesler (2021), which is concurrent and independent of the current paper, deals with the multi-item selling problem, showing that pure bundling is informationally robust when the prior belief is exchangeable across alternative items. As mentioned, we consider a more general environment in which the seller faces ambiguity on her prior as well as the buyer’s information, and present a condition for the robust optimality of a general \mathcal{K} -bundling, which nests Deb and Roesler (2021)’s pure bundling under exchangeable prior as a special case.⁷

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. Section 5 extends the optimality of \mathcal{K} -bundled sales to a setting with informational ambiguity. Section 6 develops general sufficient conditions justifying the \mathcal{K} -bundled sales mechanism. Section 7 concludes.

2 MODEL

A seller has n items for sale to a single buyer. The buyer has values $\mathbf{v} := (v_1, \dots, v_n)$ for the alternative items whose distribution is unknown to the seller.⁸ The seller simply knows that the distribution lies within some *ambiguity set* $\mathcal{F} \subset \Delta(\mathbb{R}_+^n)$.

Ambiguity Set: For the main analysis, we will consider a general set of moment conditions. These conditions are defined in terms of means and dispersion 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, we imagine 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 $K \in \mathcal{K}$ interpreted as a **product**

⁷Although the current paper is concurrent with Deb and Roesler (2021), Section 5 is subsequent to Deb and Roesler (2021). Our generalization in Section 5 should therefore be regarded as an extension of their contribution.

⁸To be precise, the “values” $\mathbf{v} := (v_1, \dots, v_n)$ 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 on what the buyer “knows.”

group. For each product group $K \in \mathcal{K}$, we let

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

be the **dispersion** of product group K 's value under F , where $\phi_K : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is a twice-differentiable convex function satisfying $\phi_K'' \geq \varepsilon$ for some $\varepsilon > 0$. We will refer to such a function as **convex moment function**. The convex moment function is quite flexible. If ϕ_K is quadratic, then $\sigma_K(F) - \mathbb{E}_F[\sum_{i \in K} v_i]^2$ reduces to the variance of $\sum_{i \in K} v_i$ under F . However, the convex function ϕ can be much more general and versatile than the conventional power moment functions considered, for example, by Carrasco et al. (2018). For instance, we show in Section 5 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. Finally, the case of *domain restriction* in which the value of each bundle $K \in \mathcal{K}$, $\sum_{i \in K} v_i$, lies in some interval $[0, \bar{v}_K]$, can be seen as a limiting case of convex moments. While the differentiability condition we impose does not make a domain restriction a special case of our model, Appendix B.6 shows that all our results carry over to that case.

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

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

The generality of this set Ω allows us to capture a wide range of scenarios in terms of the seller's ambiguity. For instance, Ω could be arbitrarily close to $\mathbb{R}_+^{n+|\mathcal{K}|}$. In this case, the moment condition entails almost no restriction on the ambiguity set.⁹ 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: $\psi_j(\mu_1(F), \dots, \mu_n(F)) \geq 0$, for some concave functions $\psi_j, j = 1, \dots, n$. This allows for the cases in which the seller knows the average values of subsets of items.¹⁰ In a similar vein, the size of Ω may correspond to the seller being more or less ambiguous about the dispersion of relevant bundle values.

⁹A truly unrestricted ambiguity set is uninteresting, however, since the worst case distribution for the seller will be degenerate at zero.

¹⁰For instance, we could have $\sum_{i \in K} \mathbb{E}_F[v_i] = m_K$ for each $K \in \mathcal{K}$.

Most important, the partition of goods encodes the structure of the seller’s dispersional knowledge about the value distribution. The partitional structure may arise from the characteristics of goods being sold. For instance, the items within each group K may be close substitutes while the relationships across items in distinct groups may not be clear cut or well understood. The partitional structure captures the seller’s ambiguity about (i) how the values of alternative bundles in \mathcal{K} are correlated and (ii) how the item values within each bundle K are distributed. We will argue that the partitional structure of the dispersional knowledge leads to a generalized partial bundling—a \mathcal{K} -bundled sales mechanism.

While we consider a more general ambiguity set in Section 6, there are a couple of reasons to study this class of ambiguity sets. First, moments are natural and salient summary statistics widely used for decision making. Second, even though our general analysis in Section 6 identifies the *type* of optimal mechanism, the analysis there does not identify its *exact form* or the corresponding worst-case distribution, which we are able to do under this constrained class of ambiguity sets. Further, the current framework allows us to handle informational ambiguity in Section 5, which is not subsumed by Section 6.

Feasible Mechanisms: Facing an ambiguity set \mathcal{F} , 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 $M = (q(\mathbf{v}), t(\mathbf{v}))$, where the *allocation rule* $q : \mathbf{v} \mapsto [0, 1]^n$ specifies the probability of allocating each item to the buyer, and the *payment rule* $t : \mathbf{v} \mapsto \mathbb{R}^+$ specifies the expected payment received from the buyer, both as Borel measurable functions of the vector \mathbf{v} of values reported by the buyer.¹¹ The mechanism satisfies *incentive compatibility* and *individual rationality*:

$$\mathbf{v} \cdot q(\mathbf{v}) - t(\mathbf{v}) \geq \sup_{\mathbf{v}' \in \mathbb{R}_+^n} \mathbf{v} \cdot q(\mathbf{v}') - t(\mathbf{v}') \quad (\text{IC})$$

$$\mathbf{v} \cdot q(\mathbf{v}) - t(\mathbf{v}) \geq 0 \quad (\text{IR})$$

¹¹Note that ambiguity in our model does not invalidate the revelation principle. Interpreting $(q(v), t(v))$ as the outcome when the buyer has value v , the feasibility conditions (IC) and (IR) are necessary and sufficient for the outcome to be implementable.

for each $\mathbf{v} \in \mathbb{R}_+^n$. Let \mathcal{M} denote the set of all direct mechanisms satisfying the *(IC)* and *(IR)* constraints—called *feasible* mechanisms.¹²

Among the feasible mechanisms, certain types of mechanisms will be of special interest to us. Let \mathcal{K} be an arbitrary partition of the set $N := \{1, \dots, n\}$ of goods. Each element of the partition can be interpreted as a bundle of items; the partition \mathcal{K} then represents a particular collection of bundles. We may imagine that the seller sells each bundle K separately. Let $K(i)$ be $K \in \mathcal{K}$ such that $i \in K$. 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 $q_K : \mathbb{R}_+ \rightarrow [0, 1]$ and $t_K : \mathbb{R}_+ \rightarrow \mathbb{R}$ such that $t(\mathbf{v}) = \sum_{K \in \mathcal{K}} t_K(\sum_{j \in K} v_j)$ and $q_i(\mathbf{v}) = q_{K(i)}(\sum_{j \in K} v_j)$.¹³ 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.

\mathcal{K} -bundled sales include two canonical mechanisms as special cases. When \mathcal{K} is the *finest* partition, namely when $\mathcal{K} = \{\{1\}, \dots, \{n\}\}$, \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 $\mathcal{K} = \{\{1, \dots, n\}\}$, \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 $R(M, F) := \int t(\mathbf{v})F(d\mathbf{v})$.¹⁴ Let $R \in \mathbb{R}$ be a *revenue guarantee* if there exists a mechanism $M \in \mathcal{M}$ such that $R(M, F) \geq R$ for all $F \in \mathcal{F}$.

¹² It is without loss to require *(IC)* and *(IR)* for all types in \mathbb{R}_+^n , rather than only for $\mathbf{v} \in \bigcup_{F \in \mathcal{F}} \text{supp}(F)$. Proposition B.1 shows that, for any feasible mechanism defined on $\bigcup_{F \in \mathcal{F}} \text{supp}(F)$, 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.

¹³ The feasibility of these one-dimensional mechanisms is implied by the feasibility of $M = (q, t)$. Specifically, *(IC)* of (q_K, t_K) is implied by *(IC)* of (q, t) . Likewise, *(IR)* of (q, t) implies that t_K can be adjusted by a constant so that (q_K, t_K) satisfies *(IR)*.

¹⁴ Here, we implicitly assume that the (opportunity) cost to the seller of selling each item is zero. This is without loss. If there are unit costs $\mathbf{c} = (c_i) \geq 0$ for the items, then the problem facing the seller is exactly the same as in our model in which she faces $\mathbf{w} = \mathbf{v} - \mathbf{c}$ 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 \mathbf{w} in the new model.

The seller's objective is to maximize the revenue guarantee. Let

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

be the **optimal revenue guarantee**, and we say the mechanism attaining that guarantee is **robustly optimal**. Finding such a mechanism amounts to analyzing a zero-sum game in which the seller chooses an optimal mechanism $M \in \mathcal{M}$ against adversarial nature, who chooses F to minimize the seller's revenue. Its equilibrium $(M^*, F^*) \in (\mathcal{M}, \mathcal{F})$ is a **saddle point**; i.e., $\forall M \in \mathcal{M}, \forall F \in \mathcal{F}$,

$$R(M, F^*) \leq R(M^*, F^*) \leq R(M^*, F). \quad (2)$$

It is well-known that a saddle point gives rise to an optimal revenue guarantee (see [Osborne and Rubinstein \(1994\)](#), Proposition 22.2-b):

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

In fact, one can find a robustly optimal mechanism either by finding a saddle point or by solving the maxi-minimization problem directly:

LEMMA 1 ([OSBORNE AND RUBINSTEIN \(1994\)](#), PROPOSITION 22-2.A,C). *If (M^*, F^*) is a saddle point, then*

$$M^* \in \arg \max_{M \in \mathcal{M}} \min_{F \in \mathcal{F}} R(M, F) \text{ and } F^* \in \arg \min_{F \in \mathcal{F}} \max_{M \in \mathcal{M}} R(M, F). \quad (4)$$

Conversely, if (M^, F^*) satisfies (4), and a saddle point exists, then (M^*, F^*) is a saddle point.*

3 ROBUSTLY OPTIMAL SALE MECHANISMS

Here 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. To this end, we will explicitly construct a candidate saddle point (M^*, F^*) and verify that it satisfies (2). Importantly, the optimal mechanism will involve \mathcal{K} -bundling.

Construction of F^* . We first construct F^* , nature's choice of distribution. This involves two steps. We first fix an arbitrary pair $(\mathbf{m}, \mathbf{s}) \in \Omega$ and construct a distribution $F^{(\mathbf{m}, \mathbf{s})}$ with means and dispersion characterized by (\mathbf{m}, \mathbf{s}) . We later describe how (\mathbf{m}, \mathbf{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 .

The distribution $F^{(\mathbf{m}, \mathbf{s})}$ is supported on an one-dimensional ray emanating from the origin. Let X be a random variable distributed from $[1, \infty)$ according to a cdf H :

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

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:

$$\int_1^{\frac{\beta_K}{\alpha_K}} \frac{\alpha_K}{x} dx + \alpha_K = m_K; \quad (5)$$

$$\int_1^{\frac{\beta_K}{\alpha_K}} \frac{\phi_K(\alpha_K x)}{x^2} dx + \frac{\phi_K(\beta_K) \alpha_K}{\beta_K} = s_K. \quad (6)$$

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 2 guarantees that such a pair exists for any $(m_K, s_K) \gg (0, \phi_K(m_K))$.

LEMMA 2. *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 (5) and (6). The mapping $(m_K, s_K)_K \mapsto (\alpha_K)_K$ is continuous.*

PROOF: See Appendix A.1. ■

Figure 1 illustrates $F^{(\mathbf{m}, \mathbf{s})}$ for $\mathcal{K} = \{\{1\}, \{2, 3\}\}$. The red curve is the support of $F^{(\mathbf{m}, \mathbf{s})}$. The valuations of the three goods are clearly co-monotonic. The asymmetry between different goods can be seen by projecting the support onto each pair of goods, illustrated by the dashed red curves. The marginal distribution of (v_2, v_3) is supported on a straight line, while the marginal distribution of (v_1, v_2) is supported on a kinked line.

We now describe how (\mathbf{m}, \mathbf{s}) is chosen. By continuity of $(\alpha_K)_{K \in \mathcal{K}}$ and compactness of Ω , there exists

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

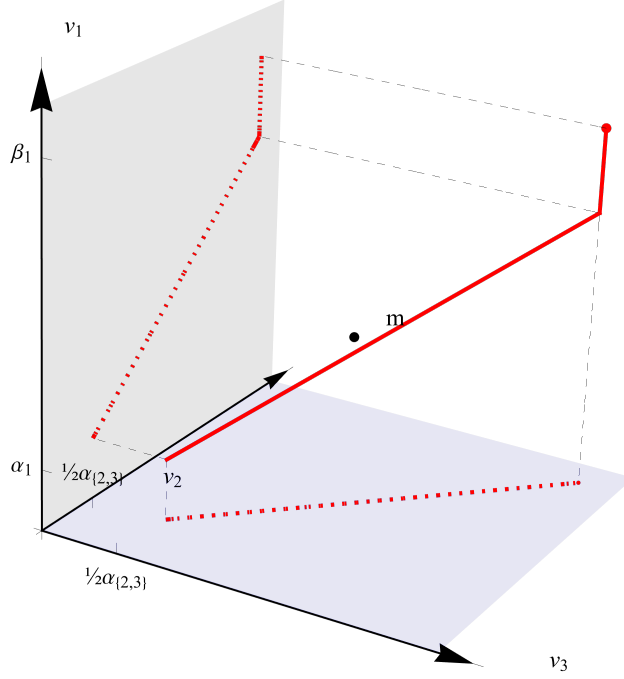


Fig. 1: Support of the joint distribution for $\mathcal{K} = \{\{1\}, \{2, 3\}\}$.

Note: $m_1 = 0.6$, $m_2 = m_3 = 0.25$, $\phi_{\{1\}}(v) = \phi_{\{2,3\}}(v) = v^2$, $s_{\{1\}} = 0.46$, $s_{\{2,3\}} = 0.1625$.

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

We provide some intuition behind the construction of F^* . Note that values of alternative items are generated via an common random variable X , which is distributed according to a Pareto distribution H . For the reason well-known by now, the Pareto distribution plays a crucial role in suppressing the seller's revenue. As shown by Carrasco et al. (2018) and Roesler and Szentes (2017) in the single-good case, this property of the distribution incentivizes the seller to charge a price equal to the lowest possible value type. This feature generalizes to the multi-item sale, but with interesting new wrinkles. To see this more clearly, one can write the virtual value of each item $i \in K$ for some $K \in \mathcal{K}$ as:

$$V_i(x) - V'_i(x) \frac{1 - H(x)}{h(x)} = \alpha_K \frac{m_i}{m_K} J_K(x).$$

Observe $J_K(x) = x - \frac{1-H(x)}{h(x)} = 0$ for $x < \beta_K/\alpha_K$ and $J_K(x) = \beta_K/\alpha_K > 0$ for $x \geq \beta_K/\alpha_K$. Following the standard Myersonian logic, facing such a distribution, it is optimal to allocate each $i \in K$ with any probability $q \in [0, 1]$ for $x < \beta_K/\alpha_K$ and with probability $q = 1$ for $x \geq \beta_K/\alpha_K$. Since this probability can be identical across

all items in the same group K , bundling that group is optimal.

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:

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

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 $\lambda_K := 1/\left[\int_{\alpha_K}^{\beta_K} \frac{\phi'_K(\beta_K) - \phi'_K(x)}{x} dx\right]$ normalizes the density so that it integrates to one.

We provide some intuition behind the construction of the mechanism. First, randomizing over selling prices can be seen as a “hedging” by the seller against nature’s adversarial choice of value distribution F^* .¹⁵ While any deterministic price is susceptible to nature putting a large mass just below that price (in which case no sale arises), a random price can immunize the seller from such a targeted “attack” and is therefore more robust. How should this randomization be structured? We can show that the density g should be chosen to “keep nature in line,” namely, to prevent it from deviating to a different $F \in \mathcal{F}$ from F^* , as required by (2).

Although the detailed proof is provided later, here we sketch the argument that the density $(g_K)_K$ of the bundle prices keeps nature from deviating to any F with the same moments $(\mathbf{m}^*, \mathbf{s}^*)$. To this end, write nature’s problem as:

$$\min_{F \in \Delta(\mathbb{R}_+^n)} \mathbb{E}_F[t^*(\mathbf{v})]$$

subject to

$$\mathbb{E}_F[v_i] = m_i^*, \forall i, \text{ and } \mathbb{E}_F[\phi_K(\sum_{i \in K} v_i)] = s_K^*, \forall K.$$

Consider the associated Lagrangian function, $L(\mathbf{v}) := t^*(\mathbf{v}) + \sum_K \lambda_K \phi_K(\sum_{i \in K} v_i)$. Our construction of $(g_K)_K$ ensures that for $(\lambda_K)_K$ chosen above, the Lagrangian $L(\mathbf{v})$ is

¹⁵As evident from the direct mechanism specification, the random pricing can be equivalently implemented via a menu of lotteries of selling each bundle at distinct prices.

linear in \mathbf{v} on the support of F^* and convex on \mathbb{R}_+^n . These features mean that the Lagrangian is maximized at F^* : for any arbitrary F with the same moments $(\mathbf{m}^*, \mathbf{s}^*)$ that nature may deviate to, we have

$$\int L(\mathbf{v})F^*(d\mathbf{v}) = L\left(\int \mathbf{v}F^*(d\mathbf{v})\right) = L\left(\int \mathbf{v}F(d\mathbf{v})\right) \leq \int L(\mathbf{v})F(d\mathbf{v}),$$

where the first equality follows from the linearity of L within the support of F^* , the second from F having the same item value means as F^* , and the last from the convexity of L . In words, a deviation by nature from F^* to any distribution F with the same means \mathbf{m}^* must be a mean-preserving redistribution of masses. The linearity of L on $\text{supp}(F^*)$ means that no such redistribution benefits nature, and its convexity outside the support means that nature has no incentive to put mass outside the support. As long as $(\lambda_K)_K$ reflect the accurate shadow cost of the dispersion constraints (which is established in our proof), maximization of the Lagrangian function means that nature will indeed never deviate from F^* .

The linearity of L on the support of F^* means that $t^*(\cdot)$ is actually concave in that support; specifically, $t^*(\cdot)$ must have the same curvature as $-\phi_K$, on its K^{th} coordinate within the support. The concavity feature can be readily seen in Figures 2 and 3, which plot the revenue function of mechanism M^* for the three-good example featured in Figure 1. Since bundles $\{1\}$ and $\{2, 3\}$ are sold separately, it is sufficient to plot the revenue functions on good 1 and bundle $\{2, 3\}$ side by side. Figure 2 plots the revenue function from selling good 1. Figure 3 plots the revenue function from selling the bundle $\{2, 3\}$. In both figures, the revenue functions t_K are proportional to $-\phi_K$ on the supports of the type distributions.

Our main result follows.

THEOREM 1. *The pair (M^*, F^*) is a saddle point satisfying (2). In the saddle point, the seller attains revenue $\sum_{K \in \mathcal{K}} \alpha_K$, and hence it is robustly optimal for the seller to sell each bundle K separately at a random price according to G_K .*

PROOF: See Appendix A.2. ■

This theorem provides 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*

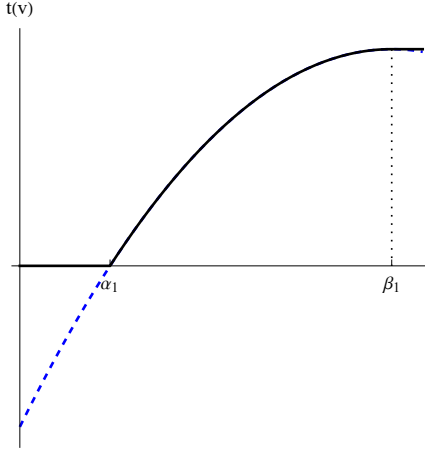


Fig. 2: Revenue from good 1.

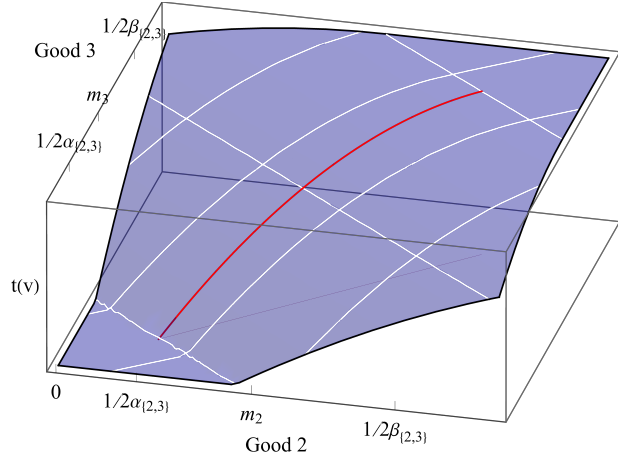


Fig. 3: Revenue from bundle $\{2, 3\}$.

Note: $m_1 = 0.6$, $m_2 = m_3 = 0.25$, $\phi_{\{1\}}(v) = \phi_{\{2,3\}}(v) = v^2$, $s_{\{1\}} = s_{\{2,3\}} = 0.1$.

ambiguity set in (1) where \mathcal{K} is the coarsest partition, then a sale of the grand bundle is robustly optimal.

More generally, Theorem 1 rationalizes a form of partial bundling that is “aligned” with the structure of the seller’s dispersional knowledge, as represented by the partition \mathcal{K} . The rough intuition is as follows. On one hand, the ambiguity about the correlation between values of bundles in \mathcal{K} leads to separate sales of these alternative bundles. On the other, the ambiguity about how a given value of each bundle $K \in \mathcal{K}$ is distributed across items within K leads to the bundling of the items within K . This insight will be further generalized in the next section.

Before closing, we make an observation that will expand the applicability of Theorem 1 beyond the ambiguity set \mathcal{F} in (1).

COROLLARY 2. *Consider an ambiguity set $\tilde{\mathcal{F}} \subset \mathcal{F}$ such that $F^* \in \tilde{\mathcal{F}}$. Then, given this ambiguity set, (M^*, F^*) is a saddle point.*

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

This corollary states that F^* remains robustly optimal within any ambiguity set $\tilde{\mathcal{F}}$ 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. More interestingly, we can utilize the corollary to handle multiple moment conditions.

Multiple moment conditions: Consider a collection of twice-differentiable convex functions $(\phi_K^a)_{a \in A}$ satisfying $\phi_K^a(0) = \phi_K^{a'}(0) = 0$ and $\phi_K^{a''} \in [\epsilon, \bar{\phi}]$ for $\bar{\phi} \geq \epsilon > 0$, where A is an arbitrary index set. Define the ambiguity set

$$\tilde{\mathcal{F}} = \left\{ F \in \Delta(\mathbb{R}_+^n) \mid \mu_i(F) = m_i; \sigma_K^a(F) \leq s_K^a, \forall K \in \mathcal{K}, a \in A \right\}, \quad (7)$$

where $s_K^a \in [\delta, \bar{s}]$ for $\bar{s} \geq \delta > 0$.

PROPOSITION 1. *There exists a \mathcal{K} -bundled sales mechanism M^* and $F^* \in \tilde{\mathcal{F}}$ such that (M^*, F^*) is a saddle point of $R(M, F)$.*

Proposition 1 is proven by identifying the “most binding” moment condition, which corresponds to the index a^* that minimizes the support of $F^{(m, s^a)}$. $F^{(m, s^{a^*})}$ has two key properties: (i) it satisfies all the moment conditions and (ii) $F^{(m, s^{a^*})}$ is part of a saddle point under the single moment condition $\sigma_K^{a^*}(F) \leq s_K^{a^*}$. Given these observations, it now follows from Corollary 2 that any saddle point under the single moment condition $\sigma_K^{a^*}(F) \leq s_K^{a^*}$ remains a saddle point under $\tilde{\mathcal{F}}$.

4 NECESSITY OF THE ESSENTIAL FEATURES

Theorem 1 establishes the optimal revenue guarantee and identifies a mechanism that attains that guarantee. However, it leaves open the possibility that another mechanism may attain that same revenue guarantee. For a sharper characterization, one would like to know whether a different mechanism could offer the same revenue guarantee. Unfortunately, establishing the *necessity* of a robustly optimal mechanism in this sense is generally intractable. For each mechanism M differing from M^* —and there are infinitely many of them—one would need to find $F \in \mathcal{F}$ such that $R(M, F) < R(M^*, F^*)$. Not surprisingly, we are not aware of any paper that establishes the *necessity* of a robustly optimal mechanism.

Nevertheless, it is important to know whether the main qualitative feature of the mechanism is *essential* in a suitable sense. This is what we establish next in Theorem 2.

This analysis is crucial to understanding what motivates the choice of a mechanism. To illustrate, recall the three-good example featured in Figure 1. The robust-optimal mechanism sells good 1 and bundle $\{2, 3\}$ separately at random prices. However, it is not the unique best response by the seller to the worst case distribution F^* . Given F^* , both bundling all three goods at price $\alpha_1 + \alpha_{\{2,3\}}$ and selling each good separately at prices $(\alpha_1, \frac{\alpha_{\{2,3\}}}{2}, \frac{\alpha_{\{2,3\}}}{2})$ yield the same revenue $\alpha_1 + \alpha_{\{2,3\}}$ attained by the robustly optimal mechanism. Nevertheless and most importantly, these alternative mechanisms are *not* robustly optimal (even when the prices can be chosen arbitrarily). We will show that there exist some *off-path* distributions \tilde{F} that make their revenues strictly lower under these mechanisms.

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

1. *Suppose there are $K, K' \in \mathcal{K}$ such that $\beta_K/\alpha_K \neq \beta_{K'}/\alpha_{K'}$ (as defined in (5) and (6)). Then, no mechanism that bundles all items $i \in K \cup K'$ is robustly optimal.*
2. *Suppose there is $K \in \mathcal{K}$ with $|K| \geq 2$. Then, for any nonempty sets $J, J' \subset K$ with $J \cap J' = \emptyset$, no mechanism that separates J and J' is robustly optimal.*

It is important to understand the counterfactual value distributions that dissuade the seller from choosing the “wrong” sales formats, since they offer valuable insights as to *why* the seller may wish to choose the mechanism identified in Theorem 1. While the proof identifies the counterfactual distributions precisely, we illustrate them in the context of the three-good example. Recall that it is robustly optimal to sell good 1 and a bundle $\{2, 3\}$ separately. We first illustrate why it is not robustly optimal to bundle all three goods.

Why is pure bundling not robustly optimal? As observed earlier, a bundled sales of all three goods at price $\alpha_1 + \alpha_{\{2,3\}}$ yields the maxmin revenue against F^* . In fact, the same revenue is attained when the grand bundle is sold at any price p within $[\alpha_1 + \alpha_{\{2,3\}}, \beta_{\{2,3\}} \left(1 + \frac{\alpha_1}{\alpha_{\{2,3\}}}\right)]$; see the left panel of Figure 4. However, the same figure hints at why selling the grand bundle is not robustly optimal. Suppose the seller charges an even higher price $p > \beta_{\{2,3\}} \left(1 + \frac{\alpha_1}{\alpha_{\{2,3\}}}\right)$ 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 the right

panel of Figure 4: the seller may fail to sell goods 2 and 3 to the buyer *even when* he has the highest value $\beta_{\{2,3\}}$ for the bundle $\{2, 3\}$, if his value of good 1 is less than $p - \beta_{\{2,3\}}$. 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 $\beta_{\{2,3\}}$ for the bundle $\{2, 3\}$.

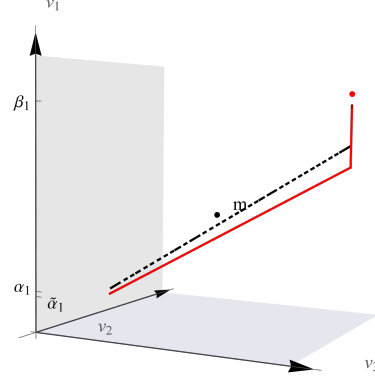
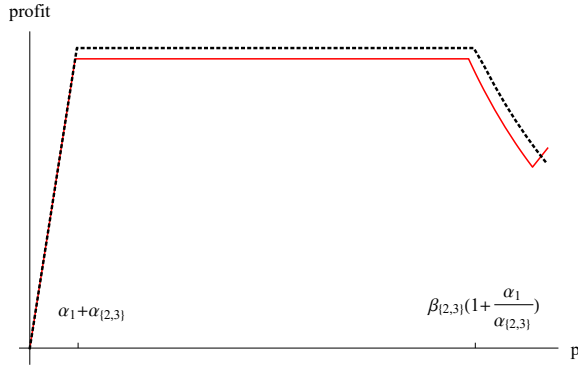


Fig. 4: Profit from pure bundling mechanism under F^* (dashed) and \tilde{F} (solid) and \tilde{F} (solid)

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 red curve in Figure 5. 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 $\tilde{\beta}_1$ of good 1. The seller cannot take advantage of this increased mass at $\tilde{\beta}_1$ under pure bundling since the profit at p in the neighborhood of $\beta_1 + \beta_{\{2,3\}}$ was strictly lower than $\alpha_1 + \alpha_{\{2,3\}}$, as can be seen in Figure 4.¹⁶ Hence, the new distribution keeps the seller’s revenue strictly below $\alpha_1 + \alpha_{\{2,3\}}$ 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.

¹⁶The robust optimality of M^* means that the seller would receive at least $\alpha_1 + \alpha_{\{2,3\}}$ from M^* given \tilde{F} .

Why is full separation not robustly optimal? Could separating the sales of *all three* goods be robustly optimal? The answer is no. Consider a distribution \tilde{F} , which is the same as F^* except that a small mass ε is transferred from $(\beta_1, \frac{1}{2}\beta_{\{2,3\}}, \frac{1}{2}\beta_{\{2,3\}})$ (the point mass at the top) to $(\beta_1, \beta_{\{2,3\}}, 0)$ and $(\beta_1, 0, \beta_{\{2,3\}})$, each with respective masses of $\frac{1}{2}\varepsilon$. See Figure 6.

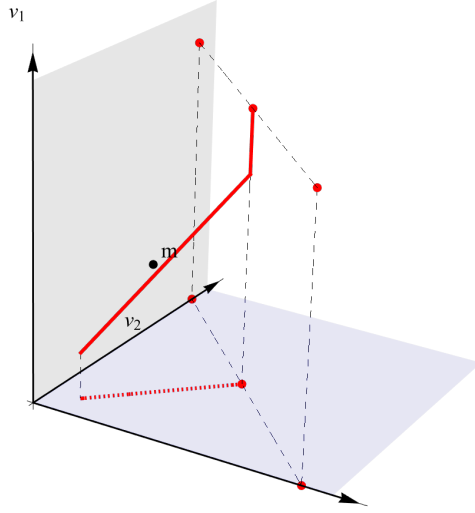


Fig. 6: The support of distribution \tilde{F}

This change keeps all constraints satisfied and does not alter the revenue of M^* since the distributions of v_1 and $v_2 + v_3$ remain the same. Yet the change has increased the dispersion of each individual item value in a way that makes separate sales less profitable. To see this, suppose the seller sells all three items separately, in particular, separating items 2 and 3. For ε sufficiently small, the seller will never wish to charge prices 0 or $\beta_{\{2,3\}}$ 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 $\alpha_1 + \alpha_{\{2,3\}}$ by selling the three items separately. In essence, the fear of this “negatively-correlated” counterfactual distribution motivates the seller to bundle the of goods 2 and 3.

As demonstrated above, the counterfactual distributions teach us a valuable lesson—in fact as valuable as the worst case distribution F^* —for understanding the chosen mechanism. It is well known that negatively-correlated item values make bundling desirable in the standard Bayesian context (see [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 clarifies this issue: it is a *possible* negative correlation “off the path” that motivates the seller to use bundling in the current environment.¹⁷

5 INFORMATIONAL ROBUSTNESS

A rather surprising application of our analysis is informational robustness, 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 (2019); Roesler and Szentes (2017); Ravid et al. (2019); Bergemann et al. (2019). Here, we will show how our approach based on convex moment functions can be used to study informational robustness.

The existing literature on this topic focuses solely on the informational ambiguity, assuming that the seller has *exact* knowledge about the prior distribution of the buyer’s valuations. Here, we take a more general approach by allowing for ambiguity about both the prior distribution of the buyer’s valuations *and* the buyer’s information about the realizations of those distributions. Fix any arbitrary partition \mathcal{K} . We assume that, for each product group $K \in \mathcal{K}$, the seller knows the exact distribution of $|K|$ -dimensional values of items within K , denoted by $G_K \in \Delta(\mathbb{R}_+^{|K|})$, but does not know how the item values are correlated *across* different groups K . In addition, the seller faces full ambiguity on the information the buyer possesses on his valuations.

Problem formulation. For each product group K , consider a prior distribution G_K of $|K|$ -dimensional vector of item values for K . For any such profile $(G_K)_{K \in \mathcal{K}}$, we first define the set of distributions $G \in \Delta(\mathbb{R}_+^n)$ of all item values that are compatible with $(G_K)_{K \in \mathcal{K}}$:

$$\mathcal{G} = \left\{ G \in \Delta(\mathbb{R}_+^n) \mid \mathbb{E}_G[h(v_k)_{k \in K}] = \mathbb{E}_{G_K}[h(v_k)_{k \in K}], \forall K \in \mathcal{K}, \forall h \in C(\mathbb{R}_+^{|K|}) \right\}, \quad (8)$$

where $C(\cdot)$ is the set of all continuous functions on $[\cdot]$. Then, the ambiguity set \mathcal{F} is defined as:

$$\mathcal{F} = \left\{ F \in \Delta(\mathbb{R}_+^n) \mid \exists G \in \mathcal{G} \text{ s.t. } \mathbb{E}_F[\phi(\mathbf{v})] \leq \mathbb{E}_G[\phi(\mathbf{v})], \forall \text{ convex } \phi \in C(\mathbb{R}_+^n) \right\}. \quad (9)$$

¹⁷When we state “off-the-path”, we are invoking the maximization characterization of robust optimality (recall Lemma 1): the seller acts first, knowing nature’s response.

In words, \mathcal{G} contains all distributions of valuations that the seller deems possible and \mathcal{F} contains all distributions of the buyer's estimated valuations of items consistent with priors in \mathcal{G} . A robustly optimal mechanism given the ambiguity set \mathcal{F} defined in (9) is then called **informationally robust**.

Assumption on the prior. We next introduce an assumption on the priors $(G_K)_K$. For any $G_K \in \Delta(\mathbb{R}_+^{|K|})$, let random vector \mathbf{v}^K be the values of the items in K distributed according to G_K . Define a random vector

$$\bar{\mathbf{v}}^K := \mathbf{m}^K \cdot \frac{\sum_{i \in K} v_i^K}{\sum_{i \in K} m_i^K},$$

where $\mathbf{m}^K := \mathbb{E}[\mathbf{v}^K]$ is the mean values of the items in K .

We call $\bar{\mathbf{v}}^K$ a *co-monotonic component* of the valuations since it varies co-monotonically by a scalar random variable $Y := \frac{\sum v_i^K}{\sum m_i^K}$. Its support is thus a ray emanating from the origin and passing through item value means \mathbf{m}^K . Note that the mean of Y is one, so $\mathbb{E}[\bar{\mathbf{v}}^K] = \mathbf{m}^K$. The remaining term $\mathbf{e} = \mathbf{v}^K - \bar{\mathbf{v}}^K$ captures the *idiosyncratic component* of the values. By construction, the idiosyncratic component has zero *ex ante* means. We require that its *conditional* means vanish:

ASSUMPTION 1. For each $K \in \mathcal{K}$, $\mathbb{E}[\mathbf{v}^K | \bar{\mathbf{v}}^K] \equiv \bar{\mathbf{v}}^K$.

Assumption 1 requires that the co-monotonic component is a profile of unbiased estimates of \mathbf{v}^K . Effectively, we are assuming \mathbf{v}^K as being generated by garbling some co-monotonic random vector $\bar{\mathbf{v}}^K$ by white noises. For each $K \in \mathcal{K}$, let $\Gamma_K \subset \Delta(\mathbb{R}^{|K|})$ be the set of all distributions satisfying Assumption 1. The set Γ_K is quite large. For instance, it can accommodate arbitrary means and variances of item values. In particular, both are allowed to be asymmetric across items within K .

Informationally robust mechanism. Our main result then follows.

THEOREM 3. Fix an arbitrary \mathcal{K} , if the ambiguity set \mathcal{F} is given by (9) with marginal distributions $(G_K) \in (\Gamma_K)$, then \mathcal{K} -bundled sales are informationally robust.

We briefly sketch the proof. Let $(G_K) \in (\Gamma_K)$ be the (marginal) prior known to the seller. Instead of considering the problem at hand, we consider an alternative problem in which the seller faces a more relaxed ambiguity set $\check{\mathcal{F}} \supset \mathcal{F}$ in which each

admitted distribution (F_K) is dominated in convex order by (G_K) *only with regard to total value* $\sum_{i \in K} v_i$ of items in K for each $K \in \mathcal{K}$.¹⁸ Note that $\check{\mathcal{F}}$ can be characterized by a continuum of convex moment conditions on $\sum_{i \in K} v_i$. Proposition 1 then identifies the worst-case distribution F^* and a robustly optimal mechanism M^* — \mathcal{K} -bundled sales. Invoking Corollary 2, it now remains to show that $F^* \in \mathcal{F}$. This follows from (1) recalling that F^* is a mean-preserving contraction of the co-monotonic component \bar{v}^K for each K and (2) noting that, by Assumption 1, \bar{v}^K is in turn a mean-preserving contraction of G_K for each $K \in \mathcal{K}$. By the standard argument due to Strassen (1965), F^* is a feasible signal compatible with the (marginal) prior (G_K) , completing the proof.

Applications of Theorem 3. We comment on the generality of Assumption 1 and its applicability.

- **Exchangeable prior distribution:** Suppose, as assumed by Deb and Roesler (2021), \mathcal{K} is the coarsest partition and the prior distribution G is *exchangeable*; namely, for all permutations (i_1, \dots, i_n) of $(1, \dots, n)$, $G(v_1, \dots, v_n) = G(v_{i_1}, \dots, v_{i_n})$.

We can verify that G satisfies Assumption 1. By exchangeability, for each i ,

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

which implies that

$$\mathbb{E}[v | \bar{v}] = \bar{v},$$

where we used the fact that $m_1 = \dots = m_n$ and thus that $m_i / (\sum_j m_j) = 1/n$. Hence, an exchangeable distribution G satisfies Assumption 1.

As a direct corollary of Theorem 3, pure bundling is informationally robust with respect to exchangeable prior (Theorem 3 of Deb and Roesler (2021)). Deb and Roesler (2021) may give the reader an impression that a strong symmetry across item values (as implied by exchangeability) is necessary for pure bundling to be informational robust. Recall, however, that Assumption 1 allows for arbitrary asymmetry both in the item value means and their variances. Hence, the more general result of Theorem 3 clarifies that symmetry is not essential for the pure bundling to be informationally robust

¹⁸It is easy to see that $\check{\mathcal{F}} \supset \mathcal{F}$ since if each F_K is convex-order dominated by G_K , surely the total value induced by F_K is convex-order dominated by the corresponding total value induced by G_K .

(when \mathcal{K} is the coarsest partition). What is essential is that G be a mean-preserving spread of its co-monotonic component.

• **Portfolio design and indexing:** Imagine an investment bank is designing its product(s). There are n assets, each belonging to a sector $K \in \mathcal{K}$. The investment bank knows the exact marginal distribution of the possible returns of assets within each sector K but is uncertain about the correlation across sectors. Within each sector, the return of each asset lies on the *capital market line* predicted by the capital asset pricing model (CAPM) under zero risk-free rate:

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

where r_m^K is the “market return” of the sector, β_i^K is the “beta” of the asset and e_i^K is the idiosyncratic risk satisfying $\sum e_i^K = 0$ and $\mathbb{E}[e_i^K | r_m^K] = 0$.

Then, the distribution of \mathbf{r}^K satisfies Assumption 1 by letting $\mathbf{v}^K = \mathbf{r}^K$ and $\bar{\mathbf{v}}^K = \boldsymbol{\beta}^K \cdot r_m^K$. The investment bank’s goal is to design a menu of portfolios to maximize the revenue guarantee against the worst case correlation across sectors and worst case private information obtained by potential buyers. Then the portfolio design problem exactly maps to the robust mechanism design problem introduced in Section 2, with an ambiguity set defined according to (9).

As a direct corollary of Theorem 3, the informationally robust strategy is to bundle all assets in each sector K and to sell the alternative bundles separately. In other words, the optimal menu of portfolios contains one *index* for each sector.

• **“Separable” informational ambiguity:** \mathcal{K} is the finest partition $\{\{1\}, \dots, \{n\}\}$. This case is comparable to the model of Carroll (2017). Since Assumption 1 has no bite in that case, our model reduces to the seller knowing an arbitrary marginal distribution G_i of each item $i = 1, \dots, n$. Our theorem, which “predicts” separate sales, can be seen as an extension of his main theorem into the informational ambiguity environment. This extension is of independent interest, and is not trivial. Unlike his model (as well as our model in Section 3), the seller faces ambiguity with respect to the buyer’s information in the current model. This latter feature means that the associated ambiguity set \mathcal{F} is not separable across (estimated) item values: the convex function ϕ characterizing \mathcal{F} in (9) is arbitrary and thus may non-trivially restrict the correlation across (estimated) item values.

6 GENERAL AMBIGUITY SETS

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 $\Upsilon_{\mathcal{K}} : F \mapsto \Delta(\mathbb{R}_+)^{|\mathcal{K}|}$:

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

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.¹⁹ [Bergemann and Schlag \(2011\)](#) formulated ambiguity in this sense.
- For each K , the ambiguity set may require $\underline{F}_K \leq_{SO} F_K \leq_{SO} \overline{F}_K$ for some benchmark distributions $\underline{F}_K, \overline{F}_K$ and some arbitrary stochastic order \leq_{SO} that

¹⁹The metric could be sup norm or Levy-Prokhorov, among others.

is closed under convex combinations. Examples of such stochastic orders are First-Order Stochastic Order, Second-Order Stochastic Dominance, Lehmann, Supermodularity, or combinations thereof.²⁰

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

$$\hat{\mathcal{F}} := \{F \in \Delta(\mathbb{R}_+^n) : (\mu_1(F), \dots, \mu_n(F)) \in \Omega\},$$

where Ω is an arbitrary nonempty subset of \mathbb{R}_+^n . We then assume that the seller's ambiguity set is given by $\mathcal{F} \cap \hat{\mathcal{F}}$. Clearly, when $\Omega = \mathbb{R}_+^n$, the constraint specified by $\hat{\mathcal{F}}$ 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.²¹ Our main theorem then follows:

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

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

PROOF: See Appendix A.6. ■

\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 .

²⁰Recall, 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.

²¹A 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:²²

COROLLARY 3. *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 $\mathcal{F} \cap \hat{\mathcal{F}}$ 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.²³ 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 found

²²Corollary 3 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.

²³ 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.

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 robustly optimal mechanism depends on the type of ambiguity facing the seller. We have identified moment conditions as well as general distributional conditions that lead 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. We are not aware of any successful generalization in this regard.²⁴ Nevertheless, our

²⁴A few papers identify robustly optimal mechanisms in *single-item* auctions. Brooks and Du (2021) finds an robustly optimal auction mechanism, when robustness is required for all value distributions with known means and common domain, buyers’ high-order beliefs, equilibrium 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.

current results extend to multiple buyers in one important sense. Following [He et al. \(2022\)](#), one can argue that as the 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. The reader is referred to [He et al. \(2022\)](#) for details.

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.
- Briest, P., Chawla, S., Kleinberg, R., and Weinberg, M. (2010). Pricing randomized allocations. In *ACM Symposium on Discrete Algorithms*, pages 585–597.
- Brooks, B. and Du, S. (2019). Optimal auction design with common values: An informationally-robust approach. Working paper.
- Brooks, B. and Du, S. (2021). Maxmin auction design with known expected values. Technical report, Tech. rep., The University of Chicago and University of California-San Diego.
- 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.
- 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. (2021). Multi-dimensional screening: buyer-optimal learning and informational robustness. Working paper.
- Du, S. (2018). Robust mechanisms under common valuation. *Econometrica*, 86(5):1569–1588.
- Guo, Y. and Shmaya, E. (2019). Robust monopoly regulation. *arXiv preprint arXiv:1910.04260*.
- Haghpanah, N. and Hartline, J. (2020). When is pure bundling optimal? *The Review of Economic Studies*, page forthcoming.
- 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. (2020). Correlation-robust auction design.
- He, W., Li, J., and Zhong, W. (2022). Order statistics of large samples: Theory and an application to robust auction design. Mimeo.
- Kartik, N. and Zhong, W. (2020). 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.
- Myerson, R. B. (1981). Optimal auction design. *Mathematics of Operations Research*, 6(1):58–73.

- Osborne, M. J. and Rubinstein, A. (1994). *A Course in Game Theory*. The MIT Press; Cambridge, Massachusetts; London, England.
- Pavlov, G. (2011). Optimal mechanism for selling two goods. *The B.E. Journal of Theoretical Economics*, 11:1–35.
- Ravid, D., Roesler, A.-K., and Szentes, B. (2019). Learning before trading: on the inefficiency of ignoring free information. *Available at SSRN 3317917*.
- Riley, J. and Zeckhauser, R. (1983). Optimal selling strategies: When to haggle, when to hold firm. *The Quarterly Journal of Economics*, 98:267–289.
- 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.
- Strassen, V. (1965). The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, 36(2):423–439.

A APPENDIX: PROOFS

A.1 Proof of Lemma 2

From (5), we can solve for $\beta_K = \alpha_K e^{\frac{m_K - \alpha_K}{\alpha_K}}$. Substituting this into (6), its LHS becomes a continuous function of α_K . It is strictly decreasing in α_K for any $\alpha_K < \beta_K$:

$$\begin{aligned}
\frac{d\text{LHS of (6)}}{d\alpha_K} &= \left(\frac{\phi_K(\beta_K)}{(\beta_K/\alpha_K)^2} - \frac{\phi_K(\beta_K)}{(\beta_K/\alpha_K)^2} \right) \cdot \frac{d(\beta_K/\alpha_K)}{d\alpha_K} + \int_1^{\frac{\beta_K}{\alpha_K}} \frac{\phi'_K(\alpha_K x)}{x} dx + \frac{\phi'_K(\beta_K)\alpha_K}{\beta_K} \cdot \frac{d\beta_K}{d\alpha_K} \\
&= \int_1^{\frac{\beta_K}{\alpha_K}} \frac{\phi'_K(\alpha_K x)}{x} dx - \phi'_K(\beta_K) \frac{m_K - \alpha_K}{\alpha_K} \\
&= \phi'_K(\beta_K) \left(\int_1^{\frac{\beta_K}{\alpha_K}} \frac{\phi'_K(\alpha_K x)}{x \phi'_K(\beta_K)} dx - \frac{m_K - \alpha_K}{\alpha_K} \right) \\
&< \phi'_K(\beta_K) \left(\log \beta_K - \log \alpha_K - \frac{m_K - \alpha_K}{\alpha_K} \right) \\
&= 0,
\end{aligned}$$

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 (6) 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

$$\int_1^{\frac{\beta_K}{\alpha_K}} \frac{\phi_K(\alpha_K x)}{x^2} dx \geq \int_1^{\frac{\beta_K}{\alpha_K}} \left(\frac{\phi_K(0)}{x^2} + \frac{\phi'_K(0)(\alpha_K x)}{x^2} + \frac{1}{2} \varepsilon \frac{(\alpha_K x)^2}{x^2} \right) dx$$

$$\begin{aligned}
&= \phi_K(0) \left(1 - \frac{\alpha_K}{\beta_K}\right) + \phi'_K(0) \alpha_K (\log \beta_K - \log \alpha_K) + \frac{1}{2} \varepsilon \alpha_K (\beta_K - \alpha_K) \\
&\geq -|\phi_K(0)| + \phi'_K(0) (m_K - \alpha_K) + \frac{1}{2} \varepsilon \alpha_K^2 \left(e^{\frac{m_K - \alpha_K}{\alpha_K}} - 1\right).
\end{aligned}$$

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

Collecting the observations so far, we conclude that there exists a unique pair (α_K, β_K) satisfying (5) and (6).

It is easy to see that both sides of (5) and (6) 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.2 Proof of Theorem 1

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

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

Hence,

$$\begin{aligned}
R(M^*, F^*) &= \int t^*(\mathbf{v}) F^*(d\mathbf{v}) \\
&= \sum_{K \in \mathcal{K}} \lambda_K \left\{ \phi'_K(\beta_K) (m_K - \alpha_K) + \phi_K(\alpha_K) - \int \left(\phi_K \left(\sum_{j \in K} v_j \right) \right) F^*(d\mathbf{v}) \right\} \\
&= \sum_{K \in \mathcal{K}} \lambda_K \{ \phi'_K(\beta_K) (m_K - \alpha_K) + \phi_K(\alpha_K) - s_K \} \\
&= \sum_{K \in \mathcal{K}} \frac{\phi'_K(\beta_K) \alpha_K \log(\beta_K / \alpha_K) - \alpha_K \int_{\alpha_K}^{\beta_K} \frac{\phi'_K(x)}{x} dx}{\int_{\alpha_K}^{\beta_K} \frac{\phi'_K(\beta_K) - \phi'_K(x)}{x} dx} = \sum_{K \in \mathcal{K}} \alpha_K. \tag{10}
\end{aligned}$$

The first three equalities are straightforward. The fourth equality follows from (5) and (6) and from recalling that $\lambda_K = 1 / [\int_{\alpha_K}^{\beta_K} \frac{\phi'_K(\beta_K) - \phi'_K(x)}{x} dx]$.

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 $\mathbf{V}(x)$, the mechanism M can be represented equivalently via $(\psi(x), \tau(x)) := (q(\mathbf{V}(x)), t(\mathbf{V}(x)))$. Since M satisfies (IC), it must satisfy the envelope condition:

$$\tau(x) = \psi(x) \cdot \mathbf{V}(x) - \int_1^x \psi(z) \cdot \mathbf{V}'(z) dz.$$

Hence,

$$R(M, F^*) = \int \tau(x) H(dx)$$

$$\begin{aligned}
&\leq \sup_{\psi(\cdot)} \int \psi(x) \cdot \left(\mathbf{V}(x) - \mathbf{V}'(x) \frac{1-H(x)}{h(x)} \right) H(dx) \\
&= \sup_{\psi} \sum_i \int_1^{\frac{\beta_{K(i)}}{\alpha_{K(i)}}} \psi_i(x) \cdot 0 H(dx) + \int_{\frac{\beta_{K(i)}}{\alpha_{K(i)}}}^{\infty} \psi_i(x) \cdot \gamma_i \cdot \beta_{K(i)} H(dx) \\
&\leq \sum_i \gamma_i \cdot \beta_{K(i)} \cdot \frac{\alpha_{K(i)}}{\beta_{K(i)}} = \sum_{K \in \mathcal{K}} \alpha_K = R(M^*, F^*), \tag{11}
\end{aligned}$$

where $\gamma_i := \frac{m_i}{\sum_{j \in N} m_j}$. The second inequality follows from $\psi_i \leq 1$. The third equality follows from $\sum_{i \in N} \gamma_i = 1$. The last equality follows from (10).

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

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

To see why this inequality holds, observe first that $t^*(\mathbf{v}) = RHS$ when $\sum_{j \in K} v_j \in [\alpha_K, \beta_K]$ (recall the very first displayed equation in the proof). Outside that region, $t^*(\mathbf{v})$ is constant in \mathbf{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 $F \in \mathcal{F}$,

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

Note that (5) and (6), together with $\lambda_K = 1 / [\int_{\alpha_K}^{\beta_K} \frac{\phi'_K(\beta_K) - \phi'_K(x)}{x} dx]$, imply that $A = \sum_{K \in \mathcal{K}} \alpha_K = R(M^*, F^*)$.

The above inequalities imply that $R(M^*, F) \geq R(M^*, F^*) - B$. 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 $\frac{dR(M^*, F^\delta)}{d\delta} \big|_{\delta=0} < 0$. In particular,

this implies that $\frac{dB}{d\delta}|_{\delta=0} > 0$. However, one can show that

$$\frac{dB}{d\delta}\Big|_{\delta=0} = - \frac{d \sum_{K \in \mathcal{K}} \alpha_K}{d\delta}\Big|_{\delta=0}. \quad (12)$$

(see Appendix B.1 for the details). Hence, $\frac{dB}{d\delta}|_{\delta=0} > 0$ 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 (\mathbf{m}, \mathbf{s}) . Therefore, we conclude that $R(M^*, F) \geq R(M^*, F^*)$.

A.3 Proof of Proposition 1

We prove the saddle point by construction. Consider distribution $F^{(\mathbf{m}, \mathbf{s}^a)}$, each defined by a parameter array $(\alpha_K^a)_{K \in \mathcal{K}}$ according to (5) and (6).²⁵ Let $\alpha_K^* = \sup_{a \in A} \alpha_K^a$, $\beta_K^* = \alpha_K^* e^{\frac{m_K - \alpha_K^*}{\alpha_K^*}}$ and F^* be defined by parameters α_K^*, β_K^* .

Note that if the set of $\{a\}$ is finite, then the proof is already done because the constraint corresponding to the largest α_K^a is the “most binding” constraint — F^* satisfies the constraint while keeping all other constraints slack. The rest follows from Corollary 2. The rest of the proof deals with the general case where no single constraint is “most binding”; hence, we will construct such a constraint by taking the limit.

First, we argue that $0 < \alpha_K^* < m_K$. This is because when $\alpha_K^a \rightarrow m_K$, $\sigma_K^a(F^{(\mathbf{m}, \mathbf{s}^a)}) \leq \bar{\phi} \text{Var}(F^{(\mathbf{m}, \mathbf{s}^a)}) \rightarrow 0 < \delta$ and when $\alpha_K^a \rightarrow 0$, $\sigma_K^a(F^{(\mathbf{m}, \mathbf{s}^a)}) \geq \epsilon \text{Var}(F^{(\mathbf{m}, \mathbf{s}^a)}) \rightarrow \infty > \bar{s}$. Therefore, $\beta_K^* < m_K < \alpha_K^*$ and F^* is well-defined.

For each K , since $\phi_K^{a''} \in [\epsilon, \bar{\phi}]$, $(\phi_K^{a'})$ is a totally bounded and equicontinuous collection of functions on $[0, \beta_K^*]$. Therefore, there exists a countable set (a_n) s.t. $\alpha_K^{a_n} \rightarrow \alpha_K^*$ and $\phi_K^{a_n'}$ converges uniformly to some function $h(\cdot)$. WLOG, pick a subsequence of (a_n) s.t. $s_K^{a_n} \rightarrow s_K$. Extend the definition of h from $[0, \beta_K^*]$ to \mathbb{R} by letting $h(x) \equiv h(\beta_K^*)$ when $x > \beta_K^*$. Define:

$$\phi_K(x) = \int_0^x h(s) ds$$

In word, ϕ_K is the limit of $\phi_K^{a_n'}$ within $[0, \beta_K^*]$ and extends linearly to \mathbb{R} . By Fatou’s lemma, $\phi_K(x)$ is the limit of $\phi_K^{a_n}(x)$ for all $x \in [\alpha_K^*, \beta_K^*]$. Therefore,

$$\begin{aligned} \mathbb{E}_{F^*} \left[\phi_K \left(\sum_{i \in K} v_i \right) \right] &= \lim \mathbb{E}_{F^*} \left[\phi_K^{a_n} \left(\sum_{i \in K} v_i \right) \right] \\ &= \lim \mathbb{E}_{F^{(\mathbf{m}, \mathbf{s}^{a_n})}} \left[\phi_K^{a_n} \left(\sum_{i \in K} v_i \right) \right] \end{aligned}$$

²⁵ β_K^a is uniquely pinned down by $\alpha_K^a e^{\frac{m_K - \alpha_K^a}{\alpha_K^a}}$ for each α_K^a .

$$\begin{aligned}
&= \lim s_K^{a_n} \\
&= s_K.
\end{aligned}$$

Now, construct M^* according to Theorem 1 for moment functions (ϕ_K) . Obviously, by construction, (M^*, F^*) is a saddle point given the ambiguity set \mathcal{F} defined as:

$$\mathcal{F} = \left\{ F \in \Delta(\mathbb{R}_+^n) : \mu_i(F) = m_i; \mathbb{E}_F \left[\phi_K \left(\sum_{i \in K} v_i \right) \right] \leq s_K \right\}.$$

Finally, we verify that \mathcal{F} and $\tilde{\mathcal{F}}$ satisfies the conditions of Corollary 2. Since $\forall a \in A$, $\alpha_K^* \geq \alpha_K^a$, $\sigma_K^a(F^*) \leq \sigma_K^a(F^a) = s_K^a$. Therefore, $F \in \tilde{\mathcal{F}}$. We verify that $\mathcal{F} \subset \tilde{\mathcal{F}}$: $\forall F \notin \tilde{\mathcal{F}}$, if $\mu_i(F) \neq m_i$, then $F \notin \mathcal{F}$. Else if $\mathbb{E}_F[\phi_K(\sum_{i \in K} v_i)] > s_K$, then we prove that for some a , $\sigma_K^a(F) > s_K^a$. $\forall a$, define

$$\tilde{\phi}_K^a(x) = \begin{cases} \phi_K^a(x) & \text{when } x \in [0, \beta_K^*] \\ \phi_K^a(\beta_K^*) + \phi_K^{a'}(\beta_K^*)(x - \beta_K^*) & \text{when } x > \beta_K^* \end{cases}.$$

$\tilde{\phi}_K^a$ extends ϕ_K^a outside of $[0, \beta_K^*]$ linearly. Since ϕ_K^a is strictly convex, $\tilde{\phi}_K^a \leq \phi_K^a$.

$$\begin{aligned}
&\lim_{n \rightarrow \infty} \mathbb{E}_F[\phi_K^{a_n}(\sum_{i \in K} v_i)] \\
&\geq \lim_{n \rightarrow \infty} \mathbb{E}_{F, \sum v_i \in [0, \beta_K^*]} \left[\phi_K^{a_n}(\sum_{i \in K} v_i) \right] + \phi_K^{a_n} \left(\mathbb{E}_F \left[\sum_{i \in K} v_i \mid \sum v_i > \beta_K^* \right] \right) F(\sum v_i > \beta_K^*) \\
&\geq \lim_{n \rightarrow \infty} \mathbb{E}_{F, \sum v_i \in [0, \beta_K^*]} \left[\phi_K^{a_n}(\sum_{i \in K} v_i) \right] + \tilde{\phi}_K^{a_n} \left(\mathbb{E}_F \left[\sum_{i \in K} v_i \mid \sum v_i > \beta_K^* \right] \right) F(\sum v_i > \beta_K^*) \\
&= \mathbb{E}_{F, \sum v_i \in [0, \beta_K^*]} \left[\phi_K(\sum_{i \in K} v_i) \right] + \phi_K \left(\mathbb{E}_F \left[\sum_{i \in K} v_i \mid \sum v_i > \beta_K^* \right] \right) F(\sum v_i > \beta_K^*) \\
&> \lim_{n \rightarrow \infty} s_K^{a_n}.
\end{aligned}$$

The first inequality is Jensen's inequality. The second inequality is from $\tilde{\phi}_K^a \leq \phi_K^a$. The equality is from the convergence of $\phi_K^{a_n}$. The last inequality is from the linearity of ϕ_K outside of $[0, \beta_K^*]$.

To sum up, we prove that $\mathcal{F} \subset \tilde{\mathcal{F}}$. Corollary 2 implies that (M^*, F^*) is also a saddle point given $\tilde{\mathcal{F}}$.

A.4 Proof of Theorem 2

(Part 1) 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 $\tilde{F} \in \mathcal{F}$ such that $\sup_{M \in \mathcal{M}_{K'}} R(M, \tilde{F}) < R(M^*, F^*)$, for all 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 \tilde{F} as follows. Without loss, assume $\ell_K > \ell_{K'}$ and let $\ell \in (\ell_{K'}, \ell_K)$. Let H_ε be given by:

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

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

$$\int_1^{\ell^\varepsilon} (\alpha^\varepsilon x) H_\varepsilon(dx) + (\alpha^\varepsilon \ell^\varepsilon)(1 - H_\varepsilon(\ell^\varepsilon)) = m_K; \quad (13)$$

$$\int_1^{\ell^\varepsilon} \phi_K(\alpha^\varepsilon x) H_\varepsilon(dx) + \phi_K(\alpha^\varepsilon \ell^\varepsilon)(1 - H_\varepsilon(\ell^\varepsilon)) = s_K. \quad (14)$$

Denote the LHS of (13) and (14) 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 $\mathbf{f}_K := (f_K^1, f_K^2)$ with respect to $(\alpha^\varepsilon, \ell^\varepsilon)$:

$$\begin{aligned} & \mathbf{J}_{\alpha^\varepsilon, \ell^\varepsilon} \mathbf{f}_K(\alpha^\varepsilon, \ell^\varepsilon, \varepsilon) \\ &= \begin{bmatrix} \int_1^{\ell^\varepsilon} x H_\varepsilon(dx) + \ell^\varepsilon(1 - H_\varepsilon(\ell^\varepsilon)) & \alpha^\varepsilon(1 - H_\varepsilon(\ell^\varepsilon)) \\ \int_1^{\ell^\varepsilon} \phi'_K(\alpha^\varepsilon x) x H_\varepsilon(dx) + \phi'_K(\alpha^\varepsilon \ell^\varepsilon) \ell^\varepsilon(1 - H_\varepsilon(\ell^\varepsilon)) & \phi'_K(\alpha^\varepsilon \ell^\varepsilon) \alpha^\varepsilon(1 - H_\varepsilon(\ell^\varepsilon)) \end{bmatrix} \\ &\implies \mathbf{J}_{\alpha^\varepsilon, \ell^\varepsilon} \mathbf{f}_K(\alpha^\varepsilon, \ell^\varepsilon, \varepsilon) \Big|_{\varepsilon=0} \\ &= \begin{bmatrix} \int_1^{\ell_K} x H(dx) + \ell_K(1 - H(\ell_K)) & \alpha_K(1 - H(\ell_K)) \\ \int_1^{\ell_K} \phi'_K(\alpha_K x) x H(dx) + \phi'_K(\alpha_K \ell_K) \ell_K(1 - H(\ell_K)) & \phi'_K(\alpha_K \ell_K) \alpha_K(1 - H(\ell_K)) \end{bmatrix}; \end{aligned}$$

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

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

By the inverse function theorem,

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

Therefore, for ε sufficiently close to 0, $\alpha^\varepsilon < \alpha_K$. Let X be the random variable

distributed according to CDF H . Define $\tilde{\mathbf{V}} := (\tilde{V}_1, \dots, \tilde{V}_n)$, where

$$\tilde{V}_i = \begin{cases} \frac{m_i}{m_J} \min \{\alpha_J X, \alpha_J \ell_J\} & \text{if } i \in J \neq K \\ \frac{m_i}{m_K} \min \{\alpha^\varepsilon (X + \varepsilon \mathbf{1}_{\{X > \ell - \varepsilon\}}), \alpha^\varepsilon, \ell^\varepsilon\} & \text{if } i \in J = K. \end{cases}$$

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

Now, consider any mechanism M that bundles $K \cup K'$. M can be written as $(\psi(x), \tau(x), \tilde{\psi}(x), \tilde{\tau}(x))$, where $(\psi(x), \tau(x))$ is the allocation and the payment for items $i \notin K \cup K'$ and $(\tilde{\psi}(x), \tilde{\tau}(x))$ 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 $i \in K \cup K'$. The envelope theorem implies

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

The second inequality is from the definition of α_J 's and the fact that $\tilde{\psi}$ equivalently characterizes a mechanism that bundles $K \cup K'$. For $x \leq \ell - \varepsilon$, $H_\varepsilon(x) = H(x)$, 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 $\varepsilon \rightarrow 0$, 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, $R(M, \tilde{F}) < \sum \alpha_J = R(M^*, F^*)$. Therefore,

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

as was to be shown.

(Part 2) Fix any nonempty $J, J' \subset K$ for some $K \in \mathcal{K}$ such that $J \cap J' = \emptyset$. We show that it is never robustly optimal to separate J and J' . To this end, it suffices to find $\tilde{F} \in \mathcal{F}$ such that $\sup_{M \in \mathcal{M}_{\mathcal{K}'}} R(M, \tilde{F}) < R(M^*, F^*)$, for any partition \mathcal{K}' such that $\{J, J'\} \subset \mathcal{K}'$.

We construct \tilde{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 \tilde{F} is then defined by the item values:

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

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, $\tilde{F} \in \mathcal{F}$.

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 \mathbf{V} is effectively uni-dimensional for $i \notin J \cup J'$, the envelope

condition implies:

$$\begin{aligned} \int t^{-J \cup J'}(\mathbf{v}) F(d\mathbf{v}) &\leq \sup_{\psi} \sum_{i \notin J \cup J'} \int_0^{\frac{\beta_{K(i)}}{\alpha_{K(i)}}} \psi_i(x) \cdot 0 H(dx) + \int_{\frac{\beta_{K(i)}}{\alpha_{K(i)}}}^{\infty} \psi_i(x) \phi_i \beta_{K(i)} H(dx) \\ &\leq \sum_{i \notin J \cup J'} \alpha_{K(i)} \frac{m_i}{m_{K(i)}}. \end{aligned}$$

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,

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

A.5 Proof of Theorem 3

Fix any $(G_K) \in (\Gamma_K)$. We prove by construction and verification. Recall that Proposition 1 applies to a setting with moment conditions defined on the total value of each product group, while the convex order defining \mathcal{F} consists of a continuum of arbitrary dispersion moment conditions. To invoke Proposition 1 and construct the saddle point, we begin with identifying a relaxed ambiguity set. For all $\mathbf{v} \in \mathbb{R}_+^n$, let $\mathbf{v}_K = (v_i)_{i \in K}$. Fix any vector $\mathbf{z} \in \mathbb{R}_{++}^{|\mathcal{K}|}$. Consider an ambiguity set

$\mathcal{F}_{\mathbf{z}} = \{F \in \Delta(\mathbb{R}_+^n) \mid \forall K \in \mathcal{K}, \mathbb{E}_F[\mathbf{v}_K] = \mathbb{E}_{G_K}[\mathbf{v}_K] \text{ and } \mathbb{E}_F[\phi_{z_K}(\mathbf{v}_K)] \leq \mathbb{E}_{G_K}[\phi_{z_K}(\mathbf{v}_K)]\}$, indexed by \mathbf{z} , where each $\phi_{z_K}(\mathbf{v}_K) := \max\{z_K - \sum_{i \in K} v_i, 0\}$. Observe that ϕ_{z_K} is convex and continuous everywhere. More importantly, the associated condition captures the second-order stochastic dominance order: namely, the random variable $\sum_{i \in K} v_i$ distributed according to F second-order stochastically dominates (SOSD) the corresponding random variable distributed according to G if $\mathbb{E}_F[\phi_{z_K}(\mathbf{v}_K)] \leq \mathbb{E}_{G_K}[\phi_{z_K}(\mathbf{v}_K)]$

for all $z_K \geq 0$.²⁶

Proposition 1 implies that there exists a \mathcal{K} -bundled sales mechanism M^* and $F^* \in \bigcap \mathcal{F}_z$ that form a saddle point (with a minor modification to accommodate non-differentiable moment functions).²⁷ We are now ready to prove our statement: (M^*, F^*) is a saddle point under ambiguity set \mathcal{F} , defined in (9). To this end, note first that $\mathcal{F} \subset \bigcap_z \mathcal{F}_z$. Hence, in light of Corollary 2, it suffices to prove that $F^* \in \mathcal{F}$.

$$\begin{aligned}
\mathbb{E}_{G_K}[\phi(\mathbf{v}_K)] &= \mathbb{E}_{G_K} \left[\mathbb{E} \left[\phi(\mathbf{v}^K) \mid \bar{\mathbf{v}}^K \right] \right] \\
&\geq \mathbb{E}_{G_K} \left[\mathbb{E} \left[\phi(\bar{\mathbf{v}}^K) \mid \bar{\mathbf{v}}^K \right] \right] \\
&= \mathbb{E}_{G_K} [\phi(\bar{\mathbf{v}}^K)] \\
&= \mathbb{E}_{G_K} \left[\phi \left(\mathbf{m}^K \frac{\sum v_i}{\sum m_i^K} \right) \right] \\
&\geq \mathbb{E}_{F^*} \left[\phi \left(\mathbf{m}^K \frac{\sum v_i}{\sum m_i^K} \right) \right] \\
&= \mathbb{E}_H \left[\phi \left(\min \left\{ \alpha_K x, \alpha_K e^{\frac{\sum m_i^K - \alpha_K}{\alpha_K}} \right\} \cdot \frac{\mathbf{m}^K}{\sum m_i^K} \right) \right] \\
&= \mathbb{E}_{F^*} [\phi(\mathbf{v}_K)].
\end{aligned}$$

The first inequality follows from convexity of ϕ and $\mathbb{E}[\mathbf{v}^K | \bar{\mathbf{v}}^K] = \bar{\mathbf{v}}^K$. The second inequality holds since the total value of F^* on components K SOSD that of G_K , a fact implied by $F^* \in \bigcap_z \mathcal{F}_z$. The last two equalities follow from the definition of F^* .

We have so far shown that for each K , the marginal distribution of F on its components K is a mean-preserving contraction of G_K . Then, by Theorem 2 of Strassen (1965), there exists a dilatation G_K such that for any Borel-measurable set $O \subset \Delta(\mathbb{R}_+^{|K|})$,

$$G_K(O) = \int G_K(O | \mathbf{v}_K) F^*(d\mathbf{v}).$$

Define $G \in \Delta(\mathbb{R}_+^n)$,

$$G(\times O_K) = \int \prod_{K \in \mathcal{K}} G_K(O_K | \mathbf{v}_K) F^*(d\mathbf{v}),$$

²⁶This can be seen upon integration by parts:

$$\int_{\sum_{i \in K} v_i \leq z} F(\mathbf{v}_K) d\mathbf{v} = \int_{\sum_{i \in K} v_i \leq z_K} (z_K - \sum v_i) F(d\mathbf{v}) = \mathbb{E}_F[\phi_{z_K}(\mathbf{v}_K)] \leq \mathbb{E}_{G_K}[\phi_{z_K}(\mathbf{v})] = \int_{\sum_{i \in K} v_i \leq z} G_K(\mathbf{v}) d\mathbf{v},$$

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

²⁷The differentiability of ϕ_K is used to pin down (\mathbf{m}, \mathbf{s}) within a general Σ in Theorem 1. However, in the current problem, \mathbf{m} is given and \mathbf{s} trivially equals $(\mathbb{E}_{G_K}[\phi_{z_K}(\mathbf{v}_K)])$ for a given z_K . The full proof is relegated to Appendix B.4.

for each Borel measurable set $O_K \subset \mathbb{R}_+^{|K|}$. Clearly, $G \in \mathcal{G}$, since the marginal G on components K is G_K . Since $\prod_{K \in \mathcal{K}} G_K(\cdot | \mathbf{v}_K)$ is a dilatation on space \mathbb{R}_+^n , Theorem 2 of [Strassen \(1965\)](#) implies that for each convex and continuous ϕ , $\mathbb{E}_G[\phi(\mathbf{v})] \geq \mathbb{E}_{F^*}[\phi(\mathbf{v})]$. We have thus proven that $F^* \in \mathcal{F}$, and the proof is complete.

A.6 Proof of Theorem 4

PROOF: Observe

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

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

LEMMA A.1. *Fix any \mathcal{K} -marginals $(F_K)_{K \in \mathcal{K}}$ and any $\hat{\mathcal{F}}$, and let $\mathcal{F} := \Upsilon_{\mathcal{K}}^{-1}((F_K)_{K \in \mathcal{K}}) \cap \hat{\mathcal{F}} \neq \emptyset$. 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

$$\sup_{M \in \mathcal{M}_{\mathcal{H}}^k} R^k(M, G^*) = \sup_{M \in \mathcal{M}^k} R^k(M, G^*). \quad (15)$$

Now we construct $F^* \in \Delta(\mathbb{R}_+^n)$ using $G^* \in \Delta(\mathbb{R}_+^k)$. Choose any $F' \in \mathcal{F} = \Upsilon_K^{-1}((F_K)) \cap \hat{\mathcal{F}}$ (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 \mathbf{X} be a k -dimensional random vector distributed according to G^* , and let $\mathbf{V} = (V_i)$ be a n -dimensional random vector defined by

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

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

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

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

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, $F^* \in \hat{\mathcal{F}}$. We thus conclude that $F^* \in \mathcal{F}$.

We next prove that

$$\sup_{M \in \mathcal{M}^k} R^k(M, G^*) \geq \sup_{M \in \mathcal{M}} R(M, F^*). \quad (16)$$

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:

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

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

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

Therefore, (IC) and (IR) of M on $\text{supp}(F^*)$ imply (IC) and (IR) of \tilde{M} on $\text{supp}(G^*)$.

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 $M \in \mathcal{M}$, (16) follows.

We next prove

$$\sup_{M \in \mathcal{M}_K} R(M, F^*) \geq \sup_{M \in \mathcal{M}_H^k} R^k(M, G^*). \quad (17)$$

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

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

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

Combining (17), (15), and (16), we obtain

$$\sup_{M \in \mathcal{M}_K} R(M, F^*) \geq \sup_{M \in \mathcal{M}_H^k} R^k(M, G^*) = \sup_{M \in \mathcal{M}^k} R^k(M, G^*) \geq \sup_{M \in \mathcal{M}} R(M, F^*). \quad (18)$$

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

$$R(M, F) = \int t(\mathbf{v}) F(d\mathbf{v}) = \sum_{K \in \mathcal{K}} \int t(\sum_{j \in K} v_j) F(d\mathbf{v}) = \sum_{K \in \mathcal{K}} \int t(x) F_K(dx),$$

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

$$\sup_{M \in \mathcal{M}_K} \inf_{F \in \mathcal{F}} R(M, F) = \sup_{M \in \mathcal{M}_K} R(M, F^*). \quad (19)$$

Combining (19) with (18), we get

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

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

B ONLINE APPENDIX: NOT FOR PUBLICATION

B.1 Supplemental Results for Theorem 1

Here, we verify (12). Let

$$\begin{aligned} \Rightarrow \begin{cases} m'_K = \sum_{j \in K} \mathbb{E}_F[v_j] \\ s'_K = \int \phi_K(\sum_{j \in K} v_j) F(d\mathbf{v}) \\ \left. \frac{dm'_K}{d\delta} \right|_{\delta=0} = \sum_{j \in K} (\mathbb{E}_F[v_j] - \mathbb{E}_{F^*}[v_j]) \\ \left. \frac{ds'_K}{d\delta} \right|_{\delta=0} = \int \phi_K(\sum_{j \in K} v_j) F(d\mathbf{v}) - s_K \end{cases} \end{aligned} \quad (20)$$

Straightforward algebra yields:

$$\begin{aligned} \left. \frac{dB}{d\delta} \right|_{\delta=0} &= \sum_{K \in \mathcal{K}} \lambda_K \left(-\phi'_K(\beta_K) \left(\sum_{j \in K} (\mathbb{E}_F[v_j] - \mathbb{E}_{F^*}[v_j]) \right) + \int \phi_K(\sum_{j \in K} v_j) F(d\mathbf{v}) - \int \phi_K(\sum_{j \in K} v_j) F^*(d\mathbf{v}) \right) \\ &= - \sum_{K \in \mathcal{K}} \lambda_K \phi'_K(\beta_K) \cdot \left. \frac{dm'_K}{d\delta} \right|_{\delta=0} + \sum_{K \in \mathcal{K}} \lambda_K \cdot \left. \frac{ds'_K}{d\delta} \right|_{\delta=0}. \end{aligned} \quad (21)$$

We show that (21) coincides with $-\left. \frac{d \sum_{K \in \mathcal{K}} \alpha_K}{d\delta} \right|_{\delta=0}$. To this end, we first apply the implicit function theorem to (5) and (6). To begin, define function ψ :

$$\psi(\alpha, \beta, m, s) = \begin{cases} \int_{\alpha}^{\beta} \frac{\alpha}{x} dx + \alpha - m \\ \int_{\alpha}^{\beta} \frac{\alpha \phi(x)}{x^2} dx + \frac{\phi(\beta)\alpha}{\beta} - s \end{cases}.$$

Then by Lemma 2, (α_K, β_K) is the unique solution to $\psi(\alpha_K, \beta_K, m_K, s_K) = 0$.

$$\frac{d\psi}{d(\alpha, \beta)} = \begin{bmatrix} \log(\frac{\beta}{\alpha}) & \frac{\alpha}{\beta} \\ \int_{\alpha}^{\beta} \frac{\phi'(x)}{x} dx & \frac{\alpha \phi'(\beta)}{\beta} \end{bmatrix},$$

and

$$\det \left(\frac{d\psi}{d(\alpha, \beta)} \right) = \frac{\alpha}{\beta} \int \frac{\phi'(\beta) - \phi'(x)}{x} dx > 0,$$

so $\frac{d\psi}{d(\alpha, \beta)}$ is invertible at (α_K, β_K) . The implicit function theorem then implies that (α_K, β_K) is locally differentiable in m . With the subscript K suppressed for notational convenience, we have:

$$\begin{aligned} \frac{d(\alpha, \beta)^T}{d(m, s)} &= - \left[\frac{d\psi}{d(\alpha, \beta)} \right]^{-1} \cdot \frac{d\psi}{d(m, s)} \\ &= - \begin{bmatrix} \log(\frac{\beta}{\alpha}) & \frac{\alpha}{\beta} \\ \int_{\alpha}^{\beta} \frac{\phi'(x)}{x} dx & \frac{\alpha \phi'(\beta)}{\beta} \end{bmatrix}^{-1} \cdot \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}. \end{aligned}$$

It follows that:

$$\begin{cases} \left. \frac{d\alpha_K}{dm'_K} \right|_{m'_K=m_K} = \lambda_K \phi'_K(\beta_K); \\ \left. \frac{d\alpha_K}{ds'_K} \right|_{s'_K=s_K} = -\lambda_K. \end{cases} \quad (22)$$

Combining (21) and (22), we conclude that

$$\left. \frac{d \sum_{K \in \mathcal{K}} \alpha_K}{d\delta} \right|_{\delta=0} = \sum_{K \in \mathcal{K}} \left(\left. \frac{d\alpha_K}{dm'_K} \cdot \frac{dm'_K}{d\delta} \right|_{\delta=0} + \left. \frac{d\alpha_K}{ds'_K} \cdot \frac{ds'_K}{d\delta} \right|_{\delta=0} \right) = -\left. \frac{dB}{d\delta} \right|_{\delta=0},$$

which yields (12).

B.2 Technical Lemma used in the proof of Theorem 4

LEMMA B.1. *If \mathcal{G} is regular, then*

$$\inf_{F \in \mathcal{G}} \sup_{M \in \mathcal{M}_{\mathcal{K}}} R(M, F) = \sup_{M \in \mathcal{M}_{\mathcal{K}}} \inf_{F \in \mathcal{G}} R(M, F). \quad (23)$$

PROOF: In this proof, we still adopt the notation in the proof of Lemma A.1 whereby \mathcal{M}^1 denotes the set of feasible mechanisms for selling a single item. By the definition of $\mathcal{M}_{\mathcal{K}}$, there exists $(q_K, t_K) \in \mathcal{M}^1$ such that $t(\mathbf{v}) \equiv \sum_{K \in \mathcal{K}} t_K(\sum_{j \in K} v_j)$. Let $\mathcal{T} = \{t \in \mathbb{R}_+^{\mathcal{K}} : \exists M = (q, t) \in \mathcal{M}^1\}$ denote the projection of \mathcal{M}^1 onto the payment dimension. Then, $M \in \mathcal{M}_{\mathcal{K}}$ if and only if $(t_K) \in \mathcal{T}^{|\mathcal{K}|}$ such that $\forall F \in \Delta(\mathbb{R}_+^n)$:

$$R(M, F) = \int \sum_{K \in \mathcal{K}} t_K(\sum_{j \in K} v_j) F(d\mathbf{v}).$$

Therefore, (23) is equivalent to

$$\inf_{F \in \mathcal{G}} \sup_{(t_K) \in \mathcal{T}^{|\mathcal{K}|}} \int \sum_{K \in \mathcal{K}} t_K(\sum_{j \in K} v_j) F(d\mathbf{v}) = \sup_{(t_K) \in \mathcal{T}^{|\mathcal{K}|}} \inf_{F \in \mathcal{G}} \int \sum_{K \in \mathcal{K}} t_K(\sum_{j \in K} v_j) F(d\mathbf{v}). \quad (24)$$

Let $\mathcal{T}_c := \mathcal{T} \cap \mathcal{BC}(\mathbb{R}_+)$, where $\mathcal{BC}(\mathbb{R}_+)$ is the set of all bounded and continuous functions defined on \mathbb{R}_+ . We first establish the minimax theorem within $\mathcal{T}_c^{|\mathcal{K}|}$:

$$\text{CLAIM B.1. } \min_{F \in \mathcal{G}} \sup_{(t_K) \in \mathcal{T}_c^{|\mathcal{K}|}} \int \sum_{K \in \mathcal{K}} t_K(\sum_{j \in K} v_j) F(d\mathbf{v}) = \sup_{(t_K) \in \mathcal{T}_c^{|\mathcal{K}|}} \min_{F \in \mathcal{G}} \int \sum_{K \in \mathcal{K}} t_K(\sum_{j \in K} v_j) F(d\mathbf{v}).$$

PROOF: We observe that

1. $\mathcal{T}_c^{|\mathcal{K}|}$ is a convex subset of linear topological space $\mathcal{C}(\mathbb{R}_+)^{|\mathcal{K}|}$ (equipped with sup norm);
2. \mathcal{G} is a compact and convex subset of a linear topological space $\Delta(\mathbb{R}_+^n)$ (equipped with Lévy-Prokhorov metric);
3. $R(M, G) = \int \sum_{K \in \mathcal{K}} t_K(\sum_{j \in K} v_j) F(d\mathbf{v})$ is linear and continuous in both (t_K) and F .

Since every mechanism in \mathcal{M}^1 is incentive compatible, for any $t_K \in \mathcal{T}_c$, there exists a nondecreasing function $q_K : \mathbb{R}_+ \rightarrow [0, 1]$, such that $t_K(v) = vq_K(v) - t_K(0) - \int_0^v q_K(x)dx$ for each $v \in \mathbb{R}_+$. The convexity of \mathcal{T}_c then follows easily: since any convex combination of nondecreasing functions q_K and q'_K is still nondecreasing, a convex combination of t and t' in \mathcal{T}_c is still an element of \mathcal{T}_c . Then $\mathcal{T}_c^{|\mathcal{K}|}$ is convex since it is the product space. Next, the compactness of \mathcal{F} follows from Prokhorov theorem, since \mathcal{F} is tight and closed. The linearity of $\int \sum_{K \in \mathcal{K}} t_K(\sum_{j \in K} v_j) F(d\mathbf{v})$ is straightforward. The continuity follows since t_K is bounded and continuous and F is a probability measure (Portmanteau theorem). The three observations ensure that the Sion's minmax theorem (Sion (1958)) holds, from which the claim follows. \blacksquare

CLAIM B.2. Fix any $F \in \mathcal{G}$ and $(t_K) \in \mathcal{T}^{|\mathcal{K}|}$. For any $\varepsilon > 0$, there exists $(\tilde{t}_K) \in \mathcal{T}_c^{|\mathcal{K}|}$ such that

$$\int \sum_{K \in \mathcal{K}} \tilde{t}_K(\sum_{j \in K} v_j) F(d\mathbf{v}) \geq \int \sum_{K \in \mathcal{K}} t_K(\sum_{j \in K} v_j) F(d\mathbf{v}) - \varepsilon. \quad (25)$$

PROOF: For each $t \in \mathcal{T}$, there exists a nondecreasing function q , such that $t(v) = vq(v) - t(0) - \int_0^v q(s)ds$ for each $v \in \mathbb{R}^+$. Fix any such $t \in \mathcal{T}$ and the associated q . For each ε , we construct a continuous function $\hat{q} : \mathbb{R}_+ \rightarrow [0, 1]$ such that $\hat{q} \geq q - \varepsilon$ and $\int_0^v \hat{q}(x)dx - \int_0^v q(x)dx \leq \varepsilon$ for each $v \in \mathbb{R}_+$.

Since q is nondecreasing, there exist countably many discrete points, $\{v^m\}$, at which q jumps up. Let $d^m > 0$ be the size of the jump at v^m , respectively. Define $r^m(v) = \mathbf{1}_{\{q(v) \geq q(v^m+)\}} d^m$. r^m is a step function with step size d^m at v^m . Then, $w(v) = q(v) - \sum_m r^m(v)$ is a nondecreasing and continuous function. Recall $q(\cdot) \leq 1$, so $\sum_m \|r^m\| = \sum_m d^m < \infty$, and for any $\varepsilon > 0$, there exists N such that $\sum_{m > N} \|r^m(v)\| \leq \varepsilon$, where $\|\cdot\|$ is the supnorm.

Define:

$$\hat{r}^m(v) := \begin{cases} 0 & \text{when } v \leq v^m - \frac{\varepsilon}{N}, \\ \frac{Nd^m}{\varepsilon} \left(v - v^m + \frac{\varepsilon}{N} \right) & \text{when } v \in \left(v^m - \frac{\varepsilon}{N}, v^m \right), \\ d^m & \text{when } v \geq v^m. \end{cases}$$

In words, $\hat{r}^m(v)$ approximates the step function $r^m(v)$ using a continuous piecewise linear function. By definition, $\hat{r}^m(v)$ is nondecreasing and continuous. Moreover, $\hat{r}^m(v) \geq r^m(v)$ and $\int_0^v |r^m(s) - \hat{r}^m(s)|ds \leq \frac{\varepsilon d^m}{N}$. Now define $\hat{q}(v) := w(v) + \sum_{m=1}^N \hat{r}^m(v)$. Then, by definition,

1. $\hat{q}(v)$ is nondecreasing and continuous in v ;
2. $\hat{q}(v) - q(v) = \sum_{m=1}^N (\hat{r}^m(v) - r^m(v)) - \sum_{m=N+1}^{\infty} r^m(v) \geq -\varepsilon$;
3. $\int_0^v [\hat{q}(x) - q(x)] dx = \int_0^v [\sum_{m=1}^N (\hat{r}^m(x) - r^m(x)) - \sum_{m=N+1}^{\infty} r^m(x)] dx \leq N \times \frac{\varepsilon}{N} = \varepsilon$.

Next, since $\hat{q}(v)$ is nondecreasing and bounded, there exists v^* such that $\forall v \geq v^*$, $\hat{q}(v) \leq \hat{q}(v^*) + \varepsilon$. Truncate \hat{q} by defining $\tilde{q}(v) = \hat{q}(\min\{v, v^*\})$. We have

$$\begin{aligned}
\tilde{t}(v) &:= v\tilde{q}(v) - t(0) - \int_0^v \tilde{q}(x) dx \\
&\geq v(\hat{q}(v) - \varepsilon) - t(0) - \int_0^v \hat{q}(x) dx \\
&\geq v(q(v) - 2\varepsilon) - t(0) - \left(\int_0^v q(x) dx + \varepsilon \right) \\
&\geq t(v) - (1 + 2v)\varepsilon,
\end{aligned}$$

where the first inequality follows from the definition of \tilde{q} and the second follows from observations 2 and 3 above. Since $\tilde{q}(v) \equiv \tilde{q}(v^*)$ for $v \geq v^*$, $\tilde{t}(v) - \tilde{t}(v^*) = (v - v^*)\tilde{q}(v^*) - \int_{v^*}^v \tilde{q}(v^*) ds = 0$. So \tilde{t} is continuous and bounded and hence $\tilde{t} \in \mathcal{T}_c$.

For all $(t_K) \in \mathcal{T}^{|\mathcal{K}|}$, let (\tilde{t}_K) be constructed as above with parameter ε . Then, $\forall F \in \mathcal{F}$,

$$\int \sum_{K \in \mathcal{K}} \tilde{t}_K(\sum_{j \in K} v_j) F(d\mathbf{v}) \geq \int \sum_{K \in \mathcal{K}} t_K(\sum_{j \in K} v_j) F(d\mathbf{v}) - (1 + 2\mathbb{E}_F[\sum v_i]) \varepsilon$$

Recall that F has finite means for item values—an implication of \mathcal{F} being regular. Hence, by setting $\varepsilon := \epsilon / (1 + 2\mathbb{E}_F[\sum v_i])$, the claim is proven. \blacksquare

It follows from Claim B.2 that

$$\sup_{(t_K) \in \mathcal{T}_c^{|\mathcal{K}|}} \int \sum_{K \in \mathcal{K}} t_K(\sum_{j \in K} v_j) F(d\mathbf{v}) \geq \sup_{(t_K) \in \mathcal{T}^{|\mathcal{K}|}} \int \sum_{K \in \mathcal{K}} t_K(\sum_{j \in K} v_j) F(d\mathbf{v}).$$

Since $\mathcal{T}_c^{|\mathcal{K}|} \subset \mathcal{T}^{|\mathcal{K}|}$, this in turn implies that

$$\sup_{(t_K) \in \mathcal{T}_c^{|\mathcal{K}|}} \int \sum_{K \in \mathcal{K}} t_K(\sum_{j \in K} v_j) F(d\mathbf{v}) = \sup_{(t_K) \in \mathcal{T}^{|\mathcal{K}|}} \int \sum_{K \in \mathcal{K}} t_K(\sum_{j \in K} v_j) F(d\mathbf{v}). \quad (26)$$

Applying (26) to Claim B.1 establishes (24), which is in turn equivalent to (23). \blacksquare

B.3 Proposition B.1 referred to in Footnote 12

PROPOSITION B.1. *Let $V \subset \mathbb{R}^n$ be a closed set, and $(q, t) : V \rightarrow [0, 1]^n \times \mathbb{R}_+$ be Borel measurable function that satisfy (IC) and (IR) on V . Then there exists Borel*

measurable function (q^*, t^*) s.t. (i) $(q^*, t^*) \equiv (q, t)$ on V and (ii) (q^*, t^*) satisfy (IC) and (IR) on \mathbb{R}_+^n .

PROOF: Let $U := \mathbb{R}_+^n \setminus V$. Let $W := \text{cl}(\{(q(v), t(v)) : v \in V\})$, so W is closed. For each $v \in U$, define:

$$(Q(v), T(v)) = \left\{ (x, y) \in W : v \cdot x - y = \sup_{v' \in V} v \cdot q(v') - t(v') \right\}.$$

In words, (Q, T) is a correspondence that “maximally” extends (q, t) to all types in \mathbb{R}_+^n . Clearly, a mechanism using any selection from $(Q(v), T(v))$ will satisfy (IC) and (IR) for all $v \in U$. Now we verify that the correspondence $(Q(v), T(v))$ is Borel measurable on U . \forall closed set $D \in [0, 1]^n \times \mathbb{R}_+$:

$$\begin{aligned} & \left\{ v \in U : (Q(v), T(v)) \cap D \neq \emptyset \right\} \\ &= \left\{ v \in U : \max_{x, y \in D \cap W} v \cdot x - y = \sup_{v' \in V} v \cdot q(v') - t(v') \right\} \\ &= \left\{ v \in U : \sup_{x, y \in D \cap W} v \cdot x - y = \sup_{v' \in V} v \cdot q(v') - t(v') \right\} \\ &= U \setminus \left\{ v \in U : \sup_{x, y \in D \cap W} v \cdot x - y < \sup_{v' \in V} v \cdot q(v') - t(v') \right\} \end{aligned}$$

The first and third equality are identities. The second equality is by $D \cap W$ being closed and $\forall v$, the set of (x, y) s.t. $v \cdot x - y \geq 0$ is bounded. For any $v \in U$ such that $\sup_{x, y \in D \cap W} v \cdot x - y < \sup_{v' \in V} v \cdot q(v') - t(v')$, let ε be the gap. Then, for any $v'' \in B_{\frac{\varepsilon}{4n}}(v)$, the inequality holds for v'' , since x and q are bounded in $[0, 1]^n$. Therefore, we verified that $\{v \in U : (Q(v), T(v)) \cap D \neq \emptyset\}$ is a relatively closed subset of U , hence is Borel measurable. Then by then by the Kuratowski–Ryll–Nardzewski measurable selection theorem, there exists a selection (q^*, t^*) from (Q, T) on U that is Borel measurable. Now, extend (q^*, t^*) to V by letting $(q^*, t^*) = (q, t)$ on V . Since (q^*, t^*) and (q, t) are Borel measurable on open set U and closed set V , respectively, the extension (q^*, t^*) is also Borel measurable. It is easy to verify that (q^*, t^*) satisfies (IC) and (IR) for the entire domain \mathbb{R}_+^n . \blacksquare

B.4 Construction of saddle point in Section 5

PROOF: Fix any $z \in \mathbb{R}_+^{|K|}$. Consider a worst-case value distribution takes the form constructed in Section 3. Specifically, the item values $\mathbf{V}(X)$ are given by:

$$V_i(X) = \min \left\{ \alpha_K X, \alpha_K e^{\frac{m_K - \alpha_K}{\alpha_K}} \right\} \cdot \frac{m_i}{m_K},$$

for all $i \in K \in \mathcal{K}$, where $m_i = \mathbb{E}_{G_K}[v_i]$ and $m_K = \sum_{j \in K} m_j$, and X is distributed according to $H(x) = 1 - \frac{1}{x}$. (This is precisely the same before, except that $\beta_K = \alpha_K e^{\frac{m_K - \alpha_K}{\alpha_K}}$ has been substituted here.) Let F_α denote the distribution of random variable $\mathbf{V}(X)$. As before, α_K 's must be such that the constraint must hold relative to the test functions ϕ_{z_K} 's; namely, for each $K \in \mathcal{K}$, $\mathbb{E}_{F_\alpha}[\phi_{z_K}(\mathbf{v}_K)] \leq \mathbb{E}_{G_K}[\phi_{z_K}(\mathbf{v}_K)]$. The following argument slightly modifies the argument in Theorem 1 and identifies the worst cast distribution F_α :

Define the mechanism M_α as a \mathcal{K} -bundled sales mechanism with random prices for each bundle K distributed according to cdf:

$$\gamma_K(p) = \frac{\log(p) - \log(\alpha_K)}{\log(z_K) - \log(\alpha_K)}$$

for $p \in [\alpha_K, z_K]$. Then, the revenue function

$$\begin{aligned} t^*(\mathbf{v}) &= \sum_{K \in \mathcal{K}} \int_{\alpha_K}^{\sum_{i \in K} v_i} p \gamma_K(dp) \\ &= \sum_{K \in \mathcal{K}} \frac{\min\{\sum_{i \in K} v_i, z_K\} - \alpha_K}{\log(z_K) - \log(\alpha_K)}. \end{aligned}$$

Hence, $R(M_\alpha, F_\alpha) = \sum_{K \in \mathcal{K}} \frac{1}{\log(z_K) - \log(\alpha_K)} \left(\alpha_K \int_{\alpha_K}^{z_K} \frac{x - \alpha_K}{x^2} dx + \alpha_K \frac{z_K - \alpha_K}{z_K} \right) = \sum \alpha_K$. $\forall M \in \mathcal{M}$:

$$\begin{aligned} R(M, F_\alpha) &\leq \sup_{\psi(\cdot)} \int \psi(x) \cdot (\mathbf{V}(x) - \mathbf{V}'(x) \cdot x) \frac{1}{x^2} dx \\ &= \sup_{\gamma} \sum_i \gamma_i \cdot \frac{m_i}{m_{K(i)}} \cdot \alpha_{K(i)} \leq \sum \alpha_K. \end{aligned}$$

$\forall F \in \mathcal{F}_{z^*}$:

$$\begin{aligned} R(M_\alpha, F) &= \int t^*(\mathbf{v}) F(d\mathbf{v}) \\ &\geq \int \sum_{K \in \mathcal{K}} \frac{\min(\sum v_i, z_K) - \alpha_K}{\log(z_K) - \log(\alpha_K)} F(d\mathbf{v}) \\ &= \int \sum_{K \in \mathcal{K}} \frac{-\phi_{z_K}(\sum v_i) + z_K - \alpha_K}{\log(z_K) - \log(\alpha_K)} F(d\mathbf{v}) \\ &\geq \sum_{K \in \mathcal{K}} \frac{-\mathbb{E}_{\Gamma_K}[\phi_{z_K}(\mathbf{v}_K)] + z_K - \alpha_K}{\log(z_K) - \log(\alpha_K)} \\ &= \sum_{K \in \mathcal{K}} \alpha_K. \end{aligned}$$

The last equality is from $\mathbb{E}_{\Gamma_K}[\phi_{z_K}(\mathbf{v}_K)] = \theta_K(\alpha_K, z_K)$. Therefore, (M_α, F_α) forms a saddle point under ambiguity set \mathcal{F}_z .

We now vary \mathbf{z} and correspondingly vary α so that F_α is the worst case distribution

within $\bigcap \mathcal{F}_z$. Write

$$\mathbb{E}_{F_\alpha}[\phi_{z_K}(\mathbf{v}_K)] = \begin{cases} 0 & \text{when } z \leq \alpha \\ \int_1^{\frac{z}{\alpha}} (z - \alpha x) \frac{1}{x^2} dx & \text{when } z \in (\alpha, \alpha e^{\frac{m_K - \alpha}{\alpha}}) \\ z - m_K & \text{when } z \geq \alpha e^{\frac{m_K - \alpha}{\alpha}}. \end{cases} \quad (27)$$

Note the function $\theta_K(\alpha, z) := \mathbb{E}_{F_\alpha}[\phi_{z_K}(\mathbf{v}_K)]$ is jointly continuous in (α, z) and strictly decreasing in α when $z \in (\alpha, \alpha e^{\frac{m_K - \alpha}{\alpha}})$. Consider the set

$$A_K := \{\alpha \in (0, m_K] : \theta(\alpha, z) \leq \mathbb{E}_{G_K}[\phi_z(\mathbf{v}_K)], \forall z \in \mathbb{R}_+\}.$$

A_K is non-empty since $m_K \in A_K$. Let $\alpha_K^* = \inf A_K$. Then by the continuity of $\theta_K, \forall z$, $\theta(\alpha^*, z) \leq \mathbb{E}_{G_K}[\phi_z(\mathbf{v}_K)]$. Equality must hold at some $z_K^* \in (0, \infty)$, because otherwise in the region $[\alpha_K^*, \alpha_K^* e^{\frac{m_K - \alpha_K^*}{\alpha_K^*}}]$ (where θ_K depends on α_K), $\mathbb{E}_{G_K}[\phi_z(\mathbf{v}_K)]$ is bounded away from $\theta_K(\alpha_K^*, z)$ and α_K^* could have been chosen strictly smaller, contradicting $\alpha_K^* = \inf A_K$.²⁸

Now, let $F^* := F_{\alpha^*}$ and $M^* := M_{\alpha^*}$. By construction, F^* belongs to \mathcal{F}_z for all $z \in \mathbb{R}_+^{|\mathcal{K}|}$. Since F^* is the worst-case distribution given the ambiguity set \mathcal{F}_{z^*} , by Corollary 2, F^* is the worst case distribution and M^* , a \mathcal{K} -bundled sales mechanism, is robustly optimal, given the ambiguity set $\bigcap_z \mathcal{F}_z \subset \mathcal{F}_{z^*}$. ■

B.5 Proof of the statement in Footnote 23

PROPOSITION B.2. *Let \mathcal{F} be defined as in (1). Suppose further Ω is downward closed; i.e.,*

$$(\mathbf{m}, \mathbf{s}) \in \Omega \text{ and } \mathbf{0} \leq \mathbf{s}' < \mathbf{s} \implies (\mathbf{m}, \mathbf{s}') \in \Omega. \quad (28)$$

Then, \mathcal{F} is regular.

REMARK B.1. *We argue in words that imposing (28) on Ω is without loss of generality. This is because in the max-min problem, nature will never choose any $F' \in \mathcal{F}$ with $(\sigma_K(F'))_K = \mathbf{s}'$ if there exists $F \in \mathcal{F}$ such that $(\sigma_K(F))_K = \mathbf{s} > \mathbf{s}'$. For each marginal K , nature can always choose a mean-preserving spread, say F'' , of F' with a very large dispersion such that the probability of a zero value is close to one. Hence for this hypothetical distribution, the revenue can be arbitrarily close to 0. Mixing F'' with F'*

²⁸When \mathcal{K} is the coarsest partition, the parameters (α_K^*, z_K^*) are exactly (π^*, \bar{s}^*) in Theorem 3 of Deb and Roesler (2021). Such connection is observed in Du (2018) Proposition 1 and Ravid et al. (2019) Lemma 3 for Pareto distribution with tail index 1, and Kartik and Zhong (2020) Proposition 2 for Pareto distributions with general tail indices.

with sufficiently small probability strictly reduces revenue while keeping the moments below \mathbf{s} .

PROOF: First, \mathcal{F} is clearly convex since all the constraints are linear in F and both Ω and \mathbf{s} is convex.

Next, we prove that it is closed under the weak topology. Consider any sequence $\{F_n\} \subset \mathcal{F}$ and $F_n \xrightarrow{w} F$. Let $m_i^n = \mu_i(F_n)$ and $s_K^n = \sigma_K(F_n)$. Without loss of generality, we pick a subsequence that $\lim_{n \rightarrow \infty} m_i^n = m_i$ and $\lim_{n \rightarrow \infty} s_K^n = s_K$.

Let $\bar{s}_K := \sup_{n \in \mathbb{Z}_+} s_K^n$, for each $K \in \mathcal{K}$. Next, apply Lemma B.2 below with $\bar{h}(\mathbf{v}) = \phi_K(\sum_{j \in K} v_j)$, $h(\mathbf{v}) = v_i$, and $C = \bar{s}_K$. Since Ω is compact, \bar{s}_K is finite. Lemma B.2 implies that:

$$\begin{aligned} \mathbb{E}_F[v_i] &= \lim_{n \rightarrow \infty} \mathbb{E}_{F_n}[v_i] \\ &= \lim_{n \rightarrow \infty} \mu_i(F_n) = m_i. \end{aligned}$$

Since Ω is compact, we must have $(\mu_i(F)) = \mathbf{m} \in \Omega$.

Now we verify that $(\sigma_K(F)) \in \Omega$.

$$\begin{aligned} \sigma_K(F) &= \mathbb{E}_F[\phi_K(\sum_{j \in K} v_j)] \\ &\leq \liminf_{n \rightarrow \infty} \mathbb{E}_{F_n} \left[\phi_K(\sum_{j \in K} v_j) \right] \\ &= \liminf_{n \rightarrow \infty} s_K^n \\ &= s_K. \end{aligned}$$

The inequality is implied by ϕ_K being continuous and bounded below (Portmanteau theorem). Therefore, since Ω is compact and satisfies (28), $(\sigma_K(F)) \leq \mathbf{s}$ implies that it is in Ω .

Finally, we prove that \mathcal{F} is tight. Since Ω is bounded, there exists $L > 0$ such that $\sum_{i \in N} m_i < L$ for all $\mathbf{m} \in \Omega$. For each $k > 0$, consider a set

$$U(k) := \{\mathbf{v} \in \mathbb{R}_+^n : \sum_{i=1}^n v_i \leq k\}.$$

The set $U(k)$ is compact. By Markov's inequality, we have

$$\mathbb{P}_F \{\mathbf{v} \notin U(k)\} \leq \frac{\sum_{i=1}^n \mathbb{E}_F[v_i]}{k} < \frac{L}{k},$$

for all $F \in \mathcal{F}$. Hence, for any $\varepsilon > 0$, one can take k large enough so that $\mathbb{P}_F \{\mathbf{v} \notin U(k)\} < \varepsilon$, as was to be shown. \blacksquare

LEMMA B.2. $\forall l \in \mathbb{N}$, let $\{F_n\} \subset \Delta(\mathbb{R}^l)$, $F_n \xrightarrow{w} F$, and $\bar{h} \in \mathcal{C}(\mathbb{R}^l)$ is a nonnegative function. If $\int \bar{h}(x) F_n(dx) \leq C$ for all n , then for all nonnegative function $h \in \mathcal{C}(\mathbb{R}^l)$

such that $\lim_{|x| \rightarrow \infty} \left| \frac{h(x)}{\bar{h}(x)} \right| = 0$,

$$\int h(x)F(dx) = \lim_{n \rightarrow \infty} \int h(x)F_n(dx).$$

PROOF: First, the Portmanteau theorem implies $\int h(x)F(dx) \leq \liminf_{n \rightarrow \infty} \int h(x)F_n(dx)$ for any nonnegative function h . Hence, it suffices to prove that $\int h(x)F(dx) \geq \limsup_{n \rightarrow \infty} \int h(x)F_n(dx)$. Suppose for the sake of contradiction that $\int h(x)F(dx) < \limsup_{n \rightarrow \infty} \int h(x)F_n(dx)$ (which is bounded by C). Without loss of generality, we pick a subsequence such that $\int \bar{h}(x)F_n(dx)$ converges. Along that subsequence,

$$\begin{aligned} & \int h(x)F(dx) < \limsup_{n \rightarrow \infty} \int h(x)F_n(dx) \\ \iff & \exists A > 0 \text{ s.t. } \int [\bar{h}(x) - Ah(x)]F(dx) > \liminf_{n \rightarrow \infty} \int [\bar{h}(x) - Ah(x)]F_n(dx) \\ \iff & \int [\bar{h}(x) - Ah(x) + B]F(dx) > \liminf_{n \rightarrow \infty} \int [\bar{h}(x) - Ah(x) + B]F_n(dx), \end{aligned}$$

for any $B \in \mathbb{R}$. Since $\lim_{|x| \rightarrow \infty} \left| \frac{h(x)}{\bar{h}(x)} \right| = 0$, there exists sufficiently large B such that $\bar{h}(x) - Ah(x) + B \geq 0$. The last part contradicts the Portmanteau theorem: $\int [\bar{h}(x) - Ah(x) + B]F(dx) \leq \liminf_{n \rightarrow \infty} \int [\bar{h}(x) - Ah(x) + B]F_n(dx)$. \blacksquare

B.6 Ambiguity Sets with Domain Restrictions

Here, we consider the ambiguity set \mathcal{F} which satisfies mean conditions (a special case of S being a singleton) but must satisfy domain restrictions instead of dispersion moment conditions. Specifically, fix any partition \mathcal{K} of N , with each element $K \in \mathcal{K}$ interpreted as a bundle of goods.

The seller now knows that the buyers' values lie within the domain $D := \{\mathbf{v} \in \mathbb{R}_+^n \mid \forall K \in \mathcal{K}, \sum_{i \in K} v_i \in [0, \bar{v}_K]\}$. The ambiguity set is now:

$$\mathcal{F} = \{F \in \Delta(D) : \mathbb{E}_F[v_i] = m_i, \forall i\},$$

where $0 < \sum_{i \in K} m_i < \bar{v}_K$ for each $K \in \mathcal{K}$.

As before, we exhibit a saddle point $(M^*, F^*) \in (\mathcal{M}, \mathcal{F})$ and prove that it satisfies the requirement (2).²⁹

²⁹Just like the dispersion moment conditions, it is easy to see that the ambiguity set \mathcal{F} exhibits \mathcal{K} -Knightian ambiguity as defined in Section 6. Therefore, Theorem 4 suggests that \mathcal{K} -bundled sales is robustly optimal. (Regularity is easy to verify in this case.)

Construction of F^* . Let the support of F^* be defined as a parametric curve $\mathbf{V}(x) : [1, \infty) \rightarrow D$ with the value of item i given by:

$$V_i(s) = \min \{ \alpha_{K(i)} \cdot x, \bar{v}_{K(i)} \} \cdot \frac{m_i}{\sum_{j \in K(i)} m_j},$$

where s is a scalar distributed from $[1, \infty)$ according to cdf H :

$$\text{Prob}(x \leq y) = H(y) = 1 - \frac{1}{y},$$

and $0 < \alpha_K < \sum_{j \in K} m_j$ satisfy:

$$\alpha_K(1 + \log(\bar{v}_K/\alpha_K)) = \sum_{j \in K} m_j. \quad (29)$$

The choice of (α_K) guarantees that the mean conditions are satisfied.

Construction of M^* . The construction of M^* is exactly the same as in Section 3. The seller sells each bundle K separately at independent random prices distributed according to G_K :

$$\begin{cases} q_i^*(\mathbf{v}) = G_{K(i)}(\sum_{j \in K(i)} v_j), \\ t^*(\mathbf{v}) = \sum_{K \in \mathcal{K}} \int_{p \leq \sum_{j \in K} v_j} p G_K(dp). \end{cases}$$

The cdf G_K is defined via the density function:

$$g_K(v) := \frac{1}{\log(\bar{v}_K/\alpha_K)v}$$

on $[\alpha_K, \bar{v}_K]$ and zero elsewhere.

THEOREM B.1. *The pair (M^*, F^*) is a saddle point satisfying (2). In the saddle point, seller attains revenue $\sum_{K \in \mathcal{K}} \alpha_K$ by selling each bundle K separately at a random price according to G_K .*

PROOF:. We first compute the value $R(M^*, F^*)$. On the support of F^* ,

$$t^*(\mathbf{v}) = \sum_{K \in \mathcal{K}} \frac{\sum_{j \in K} v_j - \alpha_K}{\log(\bar{v}_K/\alpha_K)}.$$

Hence,

$$\begin{aligned} R(M^*, F^*) &= \int t^*(\mathbf{v}) F^*(d\mathbf{v}) \\ &= \sum_{K \in \mathcal{K}} \frac{\sum_{j \in K} m_j - \alpha_K}{\log(\bar{v}_K/\alpha_K)} \\ &= \sum_{K \in \mathcal{K}} \alpha_K. \end{aligned} \quad (30)$$

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

$$\tau(x) = \psi(x) \cdot \mathbf{V}(x) - \int_1^x \psi(z) \cdot \mathbf{V}'(z) dz.$$

Hence,

$$\begin{aligned} R(M, F^*) &\leq \sup_{\psi} \int \psi \cdot \left(\mathbf{V}(x) - \mathbf{V}'(x) \frac{1 - H(x)}{h(x)} \right) H(dx) \\ &= \sup_{\psi} \sum_i \int_1^{\frac{\bar{v}_{K(i)}}{\alpha_{K(i)}}} \psi_i(x) \cdot 0 H(dx) + \int_{\frac{\bar{v}_{K(i)}}{\alpha_{K(i)}}}^{\infty} \psi_i(x) \cdot \gamma_i \cdot \bar{v}_{K(i)} H(dx) \\ &\leq \sum_i \gamma_i \cdot \bar{v}_{K(i)} \cdot \frac{\alpha_{K(i)}}{\bar{v}_{K(i)}} = \sum_{K \in \mathcal{K}} \alpha_K = R(M^*, F^*), \end{aligned} \quad (31)$$

where $\gamma_i := \frac{m_i}{\sum_{j \in N} m_j}$. The second inequality is from $\psi_i \leq 1$. The second equality is from $\sum_{i \in N} \gamma_i = 1$. The last equality is from (30).

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

$$t^*(\mathbf{v}) \geq \sum_{K \in \mathcal{K}} \frac{\sum_{j \in K} v_j - \alpha_K}{\log(\bar{v}_K / \alpha_K)}.$$

To see why the inequality holds, note that $t^*(\mathbf{v}) = RHS$ when $\sum_{j \in K} v_j \in [\alpha_K, \bar{v}_K]$. Outside that region, $t^*(\mathbf{v})$ is flat whereas the RHS is strictly increasing in $\sum_{j \in K} v_j$ whenever it is below α_K . It then follows that

$$\begin{aligned} R(M^*, F) &\geq \int \sum_{K \in \mathcal{K}} \frac{\sum_{j \in K} v_j - \alpha_K}{\log(\bar{v}_K / \alpha_K)} F(d\mathbf{v}) \\ &= \sum \alpha_K = R(M^*, F^*). \end{aligned} \quad (32)$$

Combining (31) and (32), the desired result follows. \blacksquare

REMARK B.2. *The saddle point here bears uncanny resemblance to that presented in Theorem 1. In particular, the worst-case distributions F^* are remarkably similar to each other in the two cases. In fact, they are identical if one were to replace \bar{v}_K by β_K in the formula. For any $(\bar{v}_K)_{K \in \mathcal{K}} > (m_K)_{K \in \mathcal{K}}$, one can find $(\phi_K)_{K \in \mathcal{K}}$ such that the optimal distribution for nature is identical. In this sense, the dispersion moment conditions play similar roles to upper bounds of bundle values. Intuitively, facing a dispersion moment for $\sum_{i \in K} v_K$, there is a largest bundle value $\sum_{i \in K} v_K$ beyond which nature finds it too costly to load any probability mass. At the same time, the resemblance is less than exact for the optimal selling mechanism. Given the absence*

of dispersion moment conditions, the revenue function t^* is linear (instead of concave) within the support of F^* (see the figures below). Despite these differences, the optimal mechanisms M^* are qualitatively similar between the two cases.

REMARK B.3. As with Section 3, Theorem B.1 specializes to two canonical cases. When \mathcal{K} is the finest partition of N , M^* involves a separate sales mechanism, and when \mathcal{K} is the coarsest partition, M^* involves a pure bundling mechanism. The support of F^* and the revenue from the optimal mechanism are depicted in each of these two cases in Figures 7 and 8 and Figures 9 and 10. Compared with Section 3, the only differences are that the revenue functions are linear (rather than concave) within the support of F^* , as noted above.

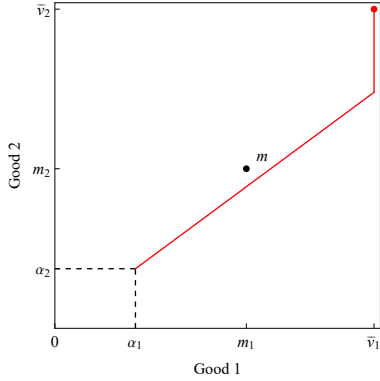


Fig. 7: Valuation distribution F^* when \mathcal{K} is the finest partition

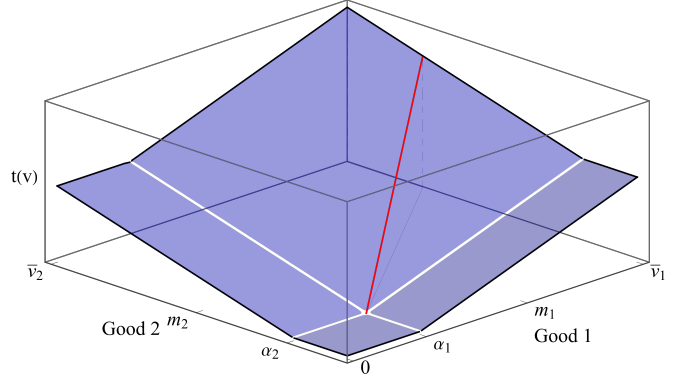


Fig. 8: Revenue from mechanism M^* when \mathcal{K} is the finest partition

Note: $m_1 = 0.6$, $m_2 = 0.5$, $\bar{v}_1 = \bar{v}_2 = 1$.

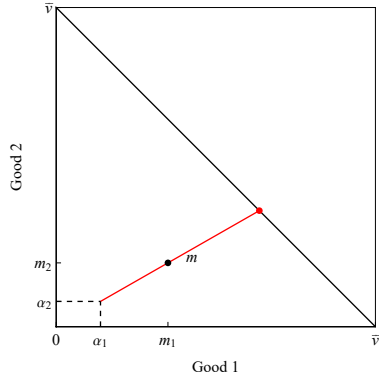


Fig. 9: Valuation distribution F^* when \mathcal{K} is the coarsest partition

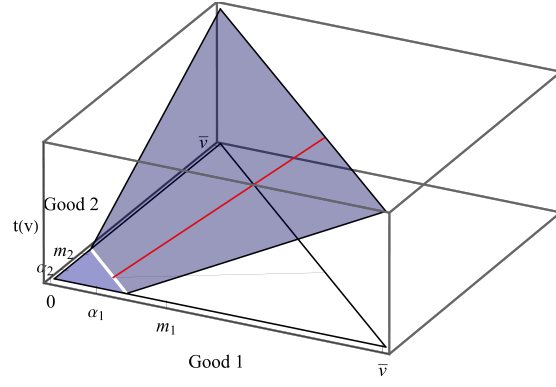


Fig. 10: Revenue from mechanism M^* when \mathcal{K} is the coarsest partition

Note: $m_1 = 0.7$, $m_2 = 0.4$, $\bar{v}_{\{1,2\}} = 2$.