Revenue maximization via machine learning with noisy data

Ellen Vitercik
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley
vitercik@berkeley.edu

Tom Yan
Department of Machine Learning
Carnegie Mellon University
tyyan@cmu.edu

Abstract

Increasingly, copious amounts of consumer data are used to learn high-revenue mechanisms via machine learning. Existing research on mechanism design via machine learning assumes that there is a distribution over the buyers' values for the items for sale and that the learning algorithm's input is a *training set* sampled from this distribution. This setup makes the strong assumption that no noise is introduced during data collection. In order to help place mechanism design via machine learning on firm foundations, we investigate the extent to which this learning process is robust to noise. Optimizing revenue using noisy data is challenging because revenue functions are extremely volatile: an infinitesimal change in the buyers' values can cause a steep drop in revenue. Nonetheless, we provide guarantees when arbitrarily correlated noise is added to the training set; we only require that the noise has bounded magnitude or is sub-Gaussian. We conclude with an application of our guarantees to *multi-task* mechanism design, where there are multiple distributions over buyers' values and the goal is to learn a high-revenue mechanism per distribution. To our knowledge, we are the first to study mechanism design via machine learning with noisy data as well as multi-task mechanism design.

1 Introduction

Revenue maximization in multi-item settings is one of the most important, long-standing open problems in mechanism design. In Bayesian Mechanism Design, there is a set of items for sale and an underlying distribution defining a set of agents' values for the items. A mechanism determines which buyers receive which items and what they pay. For decades, research in economics has assumed that the mechanism designer must know the exact distribution over buyers' values. An explosion of research [e.g., 1, 4, 19, 20, 25, 26, 28, 30, 31, 33, 37, 40, 44–48, 51] has relaxed this strong assumption: instead, the distribution is unknown and the mechanism designer only has a training set of i.i.d. samples. Using the training set, the goal is to learn a mechanism with high expected revenue. *Learning-based* mechanism design is on the verge of taking over as the main tool for designing high-revenue mechanisms for selling items—a cornerstone of many modern enterprises.

Motivated by recent literature on the brittleness of deep learning models in the face of *imperceptible* noise [32, 41, 52], an important question is whether adversarial noise has the same effect in learning-based mechanism design. Learning methods deployed in real-world settings to design mechanisms

35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia.

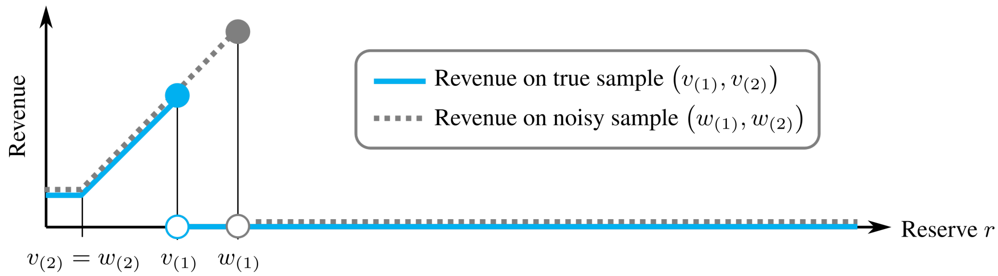

Figure 1: We illustrate revenue function volatility using the single-item second-price auction with a reserve r. Let $v_{(1)}$ be the highest bid and $v_{(2)}$ be the second-highest bid. Revenue as a function of the reserve r equals $v_{(2)}$ if $r \le v_{(2)}$, r if $v_{(2)} < r \le v_{(1)}$, and 0 if $r > v_{(1)}$, as illustrated by the blue solid line. Suppose we receive a noisy sample with highest bid $w_{(1)} > v_{(1)}$ and second-highest bid $w_{(2)} = v_{(2)}$. Revenue is illustrated by the dotted grey line. Setting the reserve equal to $w_{(1)}$ maximizes revenue on the noisy sample, but leads to zero revenue on the true values $(v_{(1)}, v_{(2)})$.

must be robust to minute noise in the data. We provide guarantees for learning mechanisms with high expected revenue in the face of an adversary that can add arbitrarily correlated noise to the training set; we only require that the noise has bounded magnitude or is sub-Gaussian. In contrast, classic learning-theoretic results typically rely on the stronger assumption that the training samples are independent.

1.1 Summary of main contributions and overview of techniques

We set out to determine whether imperceptible adversarial noise in the training set can cause a catastrophic loss in the learned mechanism's expected revenue. Our main contribution is a *sensitivity analysis* of revenue with respect to the noise's magnitude, which answers this question in the negative. Our results apply to *empirical revenue maximization*, the canonical approach to learning-based mechanism design. To our knowledge, we are the first to study learning-based mechanism design under adversarial noise. We provide guarantees for optimizing revenue over several classic mechanism classes: second-price auctions with non-anonymous reserves under additive buyers, single-priced lottery mechanisms under unit-demand buyers, and item-pricing mechanisms under unit-demand buyers. These classes have been studied extensively [e.g., 6, 18, 24, 46, 51] and can be viewed as *hypothesis classes*, just as DNNs correspond to a particular (brittle) type of hypothesis class. In some settings, mechanisms from these classes have been shown to provide approximately optimal revenue [22, 23, 35, 39].

The key challenge we face is that revenue functions are extremely sensitive to small perturbations of the buyers' values. In a second-price auction, for example, slightly shifting the highest bid from below to above the reserve price can cause an arbitrarily large drop in revenue, as illustrated in Figure 1. The set of bidders whose bids are above their reserve prices can completely change when even an infinitesimal amount of noise is added, radically altering the analytical form of the revenue function. This is unlike most functions that we understand well from a learning-theoretic perspective, which generally are smooth, continuous, or—more broadly—exhibit a straightforward connection between parameters and output value. Despite the volatility of these revenue functions, we nonetheless are able to prove bounds on the revenue loss incurred by optimizing over a noisy training set.

Our second main contribution is an application of our guarantees to *multi-task mechanism design*. To our knowledge, we are the first to study this problem. The existing literature on mechanism design via machine learning assumes that there is a *single* distribution defining the buyers' values. Often, however, the mechanism designer may be interested in designing high-revenue mechanisms for multiple related distributions. Each distribution thus defines a distinct learning *task*. The goal is to leverage the similarity between the distributions to learn a high-revenue mechanism per distribution. Multi-task learning has proven useful in fields such as Computer Vision and Natural Language Processing [53], and we demonstrate its value in mechanism design as well. Our main technical insight is that we can transform training instances from one task into slightly noisy training instances for another task and make use of our near-optimality guarantees under noised samples.

1.2 Related research

Adversarial Machine Learning: A prominent line of research on machine learning in the presence of noise studies the case where some fraction of the training data may be noisy (data poisoning). Research in this vein includes settings where every instance can be corrupted with probability $\eta < 1$ [e.g., 3] and where an η-fraction of the training data can be corrupted [e.g., 16, 27]. Meanwhile, in our setting, every sample may be adversarially perturbed with probability 1, not just an η-fraction, though we require that the perturbation be bounded. Another line of research, especially prominent in Computer Vision, studies the case where the test data can be adversarially perturbed within, for example, an ℓ_p-ball [32, 41, 52]. We, on the other hand, are concerned with noisy training data and face the unique challenges imposed by mechanism design.

Mechanism Design with Noised Distributions: While the robustness of mechanism design in the face of model uncertainty has been well studied (see for instance work by Bergemann and Morris [14] and Bergemann and Schlag [15]), to our knowledge, there has been relatively less work done on mechanism design via machine learning in the particular setting where samples are not necessarily drawn i.i.d. from the true distribution. The paper most closely related to ours is that by Cai and Daskalakis [20], who show that given access to a noisy distribution, it is possible to learn a mechanism with high expected revenue over the true distribution if the Kolmogorov distance between the noisy and true distributions are small. One might hope that the uniform distribution over the noisy training data could constitute the noisy distribution, but the Kolmogorov distance between this empirical distribution and the true distribution could be as large as 1, so these results do not apply to our setting. Moreover, we note that Cai and Daskalakis [20] makes the stronger assumption that the buyers' values follow a product distribution, whereas our assumed noise model allows for arbitrarily correlated distributions.

Multi- vs Single-item Mechanism Design: Further afield, Huang et al. [38] and Guo et al. [33] have also studied mechanism design under noisy settings, albeit under *single-item* settings. By contrast, we study a wider and different set of *multi-item* mechanism classes including multi-item lotteries and multi-item item-pricing mechanisms.

The past few decades of research on revenue-maximizing multi-item mechanism design has demonstrated that results and intuition from the single-item setting do not always carry over to the multi-item setting. And we see that this is the case in our paper as well. For example, an item-pricing mechanism for a single item would have a straight-forward stability analysis (similar to the analysis of a single-item anonymous second-price auction that we include for intuition in Appendix B.1). With multiple items and unit-demand buyers, however, stability does not even hold. Therefore, intuition from the single-item case does not carry over to the multi-item case, and we must use completely different analysis techniques.

Dispersion: The technique needed is that of *dispersion* (section 3.4). Our bounds improve based on how "nice" the distribution over buyers' values is, quantified by dispersion [5]. Dispersion has primarily been used to provide regret bounds [5, 8]. Balcan et al. [7] also use dispersion in mechanism design, though for a very different problem: estimating how much utility an agent can gain by misreporting their value in a manipulable mechanism.

The appendix summarizes additional related work on mechanism design with side information, multi-task learning, transfer learning, and approximate incentive compatibility.

2 Notation

There are n buyers and m items for sale. Each buyer $i \in [n]$ has a value for each item $j \in [m]$, denoted as $v_{ij} \geq 0$. We analyze unit-demand and additive buyers. If buyer i is unit-demand, he is only interested in obtaining one item, so his value for a bundle $b \subseteq [m]$ of goods is equal to the maximum value he has for any item in b, $\max_{j \in b} v_{ij}$. If buyer i is additive, his value for a bundle $b \subseteq [m]$ is $\sum_{j \in b} v_{ij}$. We use the notation $\boldsymbol{v}_i = (v_{i1}, \dots, v_{im})$ to denote buyer i's values for all m items and $\boldsymbol{v} = (\boldsymbol{v}_1, \dots, \boldsymbol{v}_n) \in \mathbb{R}^{nm}$ to denote all n buyers' values. When there is only one item, we use the notation $\boldsymbol{v} = (v_1, \dots, v_n) \in \mathbb{R}^n$, where $v_i \in \mathbb{R}$ is buyer i's value for the item.

We study mechanism classes parameterized by vectors $\boldsymbol{r} \in \mathbb{R}^d$ for some d (for example, the class of second-price auctions parameterized by non-anonymous reserves $\boldsymbol{r}$). The mechanisms we analyze are dominant-strategy incentive compatible, so we assume the bids equal the buyers' true values. We denote the revenue of the mechanism defined by $\boldsymbol{r}$ when the buyers' values equal $\boldsymbol{v}$ by $\text{rev}_{\boldsymbol{r}}(\boldsymbol{v})$. For simplicity of notation, we assume $\text{rev}_{\boldsymbol{r}}(\boldsymbol{v})$ is in $[0, 1]$ though our results can be extended to the case where $\text{rev}_{\boldsymbol{r}}(\boldsymbol{v}) \in [0, H]$ for some H (all bounds must simply be multiplied by H).

3 Learning under adversarial noise

In this section, we study mechanism design via machine learning when the data is corrupted. Let $\mathcal{S} = \left\{\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(L)}\right\} \subset \mathbb{R}^{nm}_{\geq 0}$ be a set of valuation vectors drawn from an unknown distribution $\mathcal{D}$. Our learning algorithm receives a poisoned training set $\mathcal{S}' = \left\{\boldsymbol{w}^{(1)}, \dots, \boldsymbol{w}^{(L)}\right\} \subset \mathbb{R}^{nm}_{\geq 0}$.

Bounded Noise Model: To model the *imperceptible* noise that corrupts the bids, we assume the bounded noise model that is commonly assumed in adversarial defense literature. That is, for some known $\epsilon > 0$, all samples $\ell \in [L]$, all bidders $i \in [n]$, and all items $j \in [m]$, $w_{ij}^{(\ell)} \in \left[v_{ij}^{(\ell)} - \epsilon, v_{ij}^{(\ell)}\right]$, or more succinctly, $\boldsymbol{v}^{(\ell)} - \boldsymbol{\epsilon} \leq \boldsymbol{w}^{(\ell)} \leq \boldsymbol{v}^{(\ell)}$. (If we only know that $\left\|\boldsymbol{v}^{(\ell)} - \boldsymbol{w}^{(\ell)}\right\|_\infty \leq \epsilon$, we can shift all bids in each vector $\boldsymbol{w}^{(\ell)}$ down by ϵ, in which case $\boldsymbol{v}^{(\ell)} - \boldsymbol{2\epsilon} \leq \boldsymbol{w}^{(\ell)} \leq \boldsymbol{v}^{(\ell)}$.). Our goal is to use the noisy set $\mathcal{S}'$ to learn a mechanism with high expected revenue over $\mathcal{D}$.

We provide some more motivating factors for our choice of the noise model in the context of mechanism design. Besides its prevalence in adversarial learning literature, (1) Such noise may arise when one is using estimated bids to learn auction parameters. We give a concrete example in Section 4 where we study multi-task mechanism design. (2) The noise may result from bounded rationality on the part of the bidders, leading to small, seemingly-innocuous differences between the buyers' true values and reported values. (3) Lastly, we note that our results can immediately be extended to cover unbounded standard noise models such as *sub-Gaussian* noise. In this case, suppose that each element of each vector $\boldsymbol{v}^{(i)} \in \mathcal{S}$ is perturbed by $\text{subG}(\sigma^2)$ to obtain the noisy vector $\boldsymbol{w}^{(i)}$. Then with probability $1 - \delta$, for all $i \in [L]$, $||\boldsymbol{w}^{(i)} - \boldsymbol{v}^{(i)}||_\infty \leq \sigma\sqrt{2 \log \frac{2Lnm}{\delta}}$. The noise need not be independent. Therefore, all of our results hold with (high) probability $1 - 2\delta$ over the draw of the true values $\mathcal{S}$ and the noise for $\epsilon = 2\sigma\sqrt{2 \log \frac{2Lnm}{\delta}}$.

Our Approach: We begin by fixing a mechanism class parameterized by vectors $\boldsymbol{r} \in \mathbb{R}^d$ for some d (for example, the class of second-price auctions parameterized by non-anonymous reserves $\boldsymbol{r}$). Given the training set $\mathcal{S}'$, the most widely used learning algorithm is *empirical revenue maximization (ERM)*, which returns the parameter setting $\hat{\boldsymbol{r}}'$ that maximizes average empirical revenue over $\mathcal{S}'$. This algorithm will be the subject of our study. Throughout this section, we analyze the following key question: what is the difference between the expected revenue of the mechanism defined by $\hat{\boldsymbol{r}}'$, $\mathbb{E}_{\boldsymbol{v}\sim\mathcal{D}}\left[\text{rev}_{\hat{\boldsymbol{r}}'}(\boldsymbol{v})\right]$, and that of the optimal mechanism in the class, $\max_{\boldsymbol{r}\in\mathbb{R}^d} \mathbb{E}\left[\text{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right]$?

Another key aspect of learning besides optimality is sample complexity. In machine learning beyond the context of mechanism design, prior research [e.g., 49] has shown that a large increase in sample complexity is needed to handle noise with bounded ℓ_∞-norm. By contrast, under the same noise assumption, we show that empirical revenue maximization can achieve near-optimal revenue *without significantly higher sample complexity*. This exposes a notable contrast between the two learning tasks. We provide lower bounds showing that ERM's loss has an optimal dependence on the noise ϵ.

3.1 Robustness

We begin by providing guarantees for any mechanism class that satisfies a notion of robustness, which helps us isolate exactly the properties we need to prove that robust revenue optimization is possible. A class is robust if it satisfies two properties. The first is a *stability* property. Let $\hat{\boldsymbol{r}}$ be the *empirically optimal* parameter setting over the set $\mathcal{S}$ of true samples, which means that of all $\boldsymbol{r} \in \mathbb{R}^d$, average revenue over $\mathcal{S}$ is maximized when $\boldsymbol{r} = \hat{\boldsymbol{r}}$. Let $\hat{\boldsymbol{r}}'$ be the empirically optimal parameter vector over the noisy set $\mathcal{S}'$. Stability is satisfied when the average revenues of $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{r}}'$ over $\mathcal{S}$ are close.

Our second property relies on a classic learning-theoretic notion of convergence. It is satisfied when, for every parameter vector $\boldsymbol{r}$, average revenue over the set $\mathcal{S}$ is close to expected revenue. Although convergence bounds have been derived in prior research for the mechanism classes we analyze [6, 46], a convergence bound alone *does not imply any guarantees* whatsoever for optimization with noisy

Mechanism class	**Buyers' values**	p**-stable**	**Error bound**
Single-item anonymous second-price auction	Additive	$p(\epsilon, n, m) = 2\epsilon$ (Theorem B.2)	$\tilde{O}(\epsilon + \sqrt{1/L})$ (Corollary B.6)
Multi-item non-anonymous second-price auction	Additive	$p(\epsilon, n, m) = 2m\epsilon$ (Theorem 3.3)	$\tilde{O}(m\epsilon + \sqrt{nm/L})$ (Corollary 3.7)
Multi-item non-anonymous lotteries	Unit-demand	$p(\epsilon, n, m) = n\epsilon$ (Theorem 3.9)	$\tilde{O}(n\epsilon + \sqrt{nm/L})$ (Corollary 3.10)
Item-pricing mechanisms	Unit-demand	Does not satisfy (Lemma B.11)	$\tilde{O}(nm^2(\kappa\epsilon + \sqrt{1/L}))^*$ (Theorem 3.13, Lemma 3.15)

* The distribution over buyers' values is κ-bounded (Definition 3.14).

Table 1: Our p-stability guarantees together with the resulting error bounds. Item-pricing mechanisms do not satisfy p-stability, so we use alternative techniques to provide an error bound (Section 3.4).

data. This is why we introduce the separate notion of stability, which has not been previously studied in the automated mechanism design literature. The challenge then lies in proving that a variety of mechanism classes satisfy stability. To begin, we define the two properties formally as follows:

Definition 3.1. Given two functions $p : [0,1] \times \mathbb{Z}^2 \to \mathbb{R}$ and $q : [0,1] \times \mathbb{Z}^3 \to \mathbb{R}$, we say that a mechanism class is *p-stable and q-convergent* if the following conditions hold:

1. *p-stable.* For any $L \geq 1$, let $\mathcal{S} = \left\{\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(L)}\right\} \subset \mathbb{R}^{nm}_{\geq 0}$ and $\mathcal{S}' = \left\{\boldsymbol{w}^{(1)}, \dots, \boldsymbol{w}^{(L)}\right\} \subset \mathbb{R}^{nm}_{\geq 0}$ be two arbitrary sets of vectors such that $\boldsymbol{v}^{(\ell)} - \boldsymbol{\epsilon} \leq \boldsymbol{w}^{(\ell)} \leq \boldsymbol{v}^{(\ell)}$ for all $\ell \in [L]$. Let $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{r}}'$ be the empirically optimal parameter vectors over $\mathcal{S}$ and $\mathcal{S}'$: $\hat{\boldsymbol{r}} = \operatorname{argmax}_{\boldsymbol{r} \in \mathbb{R}^d} \sum_{\ell=1}^L \operatorname{rev}_{\boldsymbol{r}}\left(\boldsymbol{v}^{(\ell)}\right)$ and $\hat{\boldsymbol{r}}' = \operatorname{argmax}_{\boldsymbol{r} \in \mathbb{R}^d} \sum_{\ell=1}^L \operatorname{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right)$. We require that on average over $\mathcal{S}$, the revenues of $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{r}}'$ are close: $\frac{1}{L}\sum_{\ell=1}^L \operatorname{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) - \operatorname{rev}_{\hat{\boldsymbol{r}}'}\left(\boldsymbol{v}^{(\ell)}\right) \leq p(\epsilon, n, m)$.
2. *q-convergent.* For any $L \geq 1$ and $\delta \in (0,1)$, with probability $1-\delta$ over the draw $\mathcal{S} \sim \mathcal{D}^L$, for every vector $\boldsymbol{r} \in \mathbb{R}^d$, the difference between the average revenue over $\mathcal{S}$ and the expected revenue is at most $q(\delta, L, n, m)$. In other words, $\left|\frac{1}{L}\sum_{\boldsymbol{v} \in \mathcal{S}} \operatorname{rev}_{\boldsymbol{r}}(\boldsymbol{v}) - \mathbb{E}\left[\operatorname{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right]\right| \leq q(\delta, L, n, m)$.

If a mechanism class is p-stable and q-convergent, then the expected revenue of the empirically optimal mechanism over the noisy samples $\mathcal{S}'$ is close to the expected revenue of the optimal mechanism in the class. For completeness, the proof is in Appendix B.

Fact 3.2. *Fix a p-stable and q-convergent mechanism class. Let $\mathcal{S} = \left\{\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(L)}\right\}$ and $\mathcal{S}' = \left\{\boldsymbol{w}^{(1)}, \dots, \boldsymbol{w}^{(L)}\right\}$ be two sets of valuation vectors such that for all $\ell \in [L]$, $\boldsymbol{v}^{(\ell)} - \boldsymbol{\epsilon} \leq \boldsymbol{w}^{(\ell)} \leq \boldsymbol{v}^{(\ell)}$. Let $\hat{\boldsymbol{r}}'$ be empirically optimal over $\mathcal{S}'$: $\hat{\boldsymbol{r}}' = \operatorname{argmax}_{\boldsymbol{r} \in \mathbb{R}^d} \sum_{\ell=1}^L \operatorname{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right)$. With probability $1-\delta$ over $\mathcal{S} \sim \mathcal{D}^L$, $\max_{\boldsymbol{r} \in \mathbb{R}^d} \mathbb{E}_{\boldsymbol{v} \sim \mathcal{D}}\left[\operatorname{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right] - \mathbb{E}_{\boldsymbol{v} \sim \mathcal{D}}\left[\operatorname{rev}_{\hat{\boldsymbol{r}}'}(\boldsymbol{v})\right] \leq p(\epsilon, n, m) + 2q(\delta, L, n, m)$.*

We now prove that several mechanism classes are stable: second-price auctions with non-anonymous reserves and lotteries. (For intuition, we also analyze the simpler class of second-price auctions with anonymous reserves in Appendix B.1.) Table 1 summarizes our results.

3.2 Non-anonymous second-price auctions

We prove that *second-price auctions with non-anonymous reserves* and additive bidders are robust. In the single-item setting, this auction is defined by a vector $\boldsymbol{r} \in \mathbb{R}^n$, where r_i is bidder i's *reserve price*. Each bidder submits a bid and the mechanism discards all bidders whose bids are smaller than their reserves. If there are bidders remaining, the highest bidder, say bidder i, wins and pays the maximum of the second-highest remaining bid and r_i (or r_i if there are no other remaining bids). In the multi-item case, there is only one copy of each item and there is a separate auction per item. The mechanism is defined by $\boldsymbol{r} = (\boldsymbol{r}_1, \dots, \boldsymbol{r}_m) \in \mathbb{R}^{nm}$, where $\boldsymbol{r}_j \in \mathbb{R}^n$ is the reserve vector for item j.

We begin by analyzing single-item auctions, which then implies guarantees for multiple items. The key challenge is that the set of bidders whose bids are above their reserves can completely change when even infinitesimal noise is added, radically altering the analytical form of the revenue function.

Theorem 3.3. *Single-item non-anonymous second-price auctions are p-stable with $p(\epsilon, n, m) = 2\epsilon$.*

Proof. For any $L \geq 1$, let $\mathcal{S} = \left\{\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(L)}\right\} \subset \mathbb{R}^n_{\geq 0}$ and $\mathcal{S}' = \left\{\boldsymbol{w}^{(1)}, \dots, \boldsymbol{w}^{(L)}\right\} \subset \mathbb{R}^n_{\geq 0}$ be two sets of valuation vectors such that $\boldsymbol{v}^{(\ell)} - \boldsymbol{\epsilon} \leq \boldsymbol{w}^{(\ell)} \leq \boldsymbol{v}^{(\ell)}$ for all $\ell \in [L]$. Let $\hat{\boldsymbol{r}}'$ be the empirically optimal reserve vector over $\mathcal{S}'$ and let $\hat{\boldsymbol{r}}$ be empirically optimal over $\mathcal{S}$. This proof relies on two key lemmas. The first states that if we shift the reserve vector $\hat{\boldsymbol{r}}$—which is empirically optimal over $\mathcal{S}$—down by an additive factor of $\boldsymbol{\epsilon} = (\epsilon, \dots, \epsilon) \in \mathbb{R}^n$ and apply it to the valuations in $\mathcal{S}'$, little revenue is lost. We use the standard notation $\langle x \rangle = \max\{x, 0\}$. The full proof is in Appendix B.2.

Lemma 3.4. *Let $\boldsymbol{r}_\epsilon = (\langle \hat{r}_1 - \epsilon \rangle, \dots, \langle \hat{r}_n - \epsilon \rangle)$. If $\text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) > 0$, then $\text{rev}_{\boldsymbol{r}_\epsilon}\left(\boldsymbol{w}^{(\ell)}\right) \geq \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) - 2\epsilon$.*

Proof sketch of Lemma 3.4. Suppose that $\text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) > 0$. Let i be the winner under the values $\boldsymbol{v}^{(\ell)}$ and reserve $\hat{\boldsymbol{r}}$. Since $v_i^{(\ell)} \geq \hat{r}_i$, we know that $w_i^{(\ell)} \geq \langle v_i^{(\ell)} - \epsilon \rangle \geq \langle \hat{r}_i - \epsilon \rangle$, so under the values $\boldsymbol{w}^{(\ell)}$ and reserve $\boldsymbol{r}_\epsilon$, there is at least one bidder whose bid is at least his reserve. Let i' be the winner under the valuation vector $\boldsymbol{w}^{(\ell)}$ and reserve $\boldsymbol{r}_\epsilon$. We split this proof into four cases that depend on:

1. Whether or not $i = i'$, and
2. Whether—under $\boldsymbol{v}^{(\ell)}$ and $\hat{\boldsymbol{r}}$—there is another bidder besides i whose bid is at least his reserve (in other words, whether or not $\{t : v_t^{(\ell)} \geq \hat{r}_t, t \neq i\} = \emptyset$).

In the first case, $i = i'$ and $\{t : v_t^{(\ell)} \geq \hat{r}_t, t \neq i\} \neq \emptyset$. Let k be the index of the second-highest bidder in $\boldsymbol{v}^{(\ell)}$ whose bid is above his reserve in $\hat{\boldsymbol{r}}$: $k = \text{argmax}_{t \neq i}\{v_t^{(\ell)} : v_t^{(\ell)} \geq \hat{r}_t\}$. This means that $\text{rev}_{\hat{\boldsymbol{r}}}(\boldsymbol{v}^{(\ell)}) = \max\{\hat{r}_i, v_k^{(\ell)}\}$. Since $v_k^{(\ell)} \geq \hat{r}_k$, we have that $w_k^{(\ell)} \geq \langle v_k^{(\ell)} - \epsilon \rangle \geq \langle \hat{r}_k - \epsilon \rangle$. Since $k \neq i$ and $i = i'$, it must be that $k \neq i'$, so $\{t : w_t^{(\ell)} \geq \langle \hat{r}_t - \epsilon \rangle, t \neq i'\} \neq \emptyset$ (in particular, the set contains k). Let k' be the index of the second-highest bidder in $\boldsymbol{w}^{(\ell)}$ whose bid is above his reserve in $\boldsymbol{r}_\epsilon$: $k' = \text{argmax}_{t \neq i'}\{w_t^{(\ell)} : w_t^{(\ell)} \geq \langle \hat{r}_t - \epsilon \rangle\}$, which means that $\text{rev}_{\boldsymbol{r}_\epsilon}(\boldsymbol{w}^{(\ell)}) = \max\{\hat{r}_{i'} - \epsilon, w_{k'}^{(\ell)}\}$. Since $k \in \{t : w_t^{(\ell)} \geq \langle \hat{r}_t - \epsilon \rangle, t \neq i'\}$, we know that $w_{k'}^{(\ell)} \geq w_k^{(\ell)} \geq v_k^{(\ell)} - \epsilon$. Putting this all together, we prove that $\text{rev}_{\boldsymbol{r}_\epsilon}(\boldsymbol{w}^{(\ell)}) \geq \max\{\hat{r}_i - \epsilon, w_{k'}^{(\ell)}\} \geq \max\{\hat{r}_i, v_k^{(\ell)}\} - 2\epsilon = \text{rev}_{\hat{\boldsymbol{r}}}(\boldsymbol{v}^{(\ell)}) - 2\epsilon$. We prove the other three cases in the appendix. □

We use Lemma 3.4 to continue Theorem 3.3's proof. Let I be the set of indices ℓ such that $\text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) > 0$. By Lemma 3.4, $\sum_{\ell=1}^L \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) = \sum_{\ell \in I} \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) \leq \sum_{\ell \in I} \text{rev}_{\boldsymbol{r}_\epsilon}\left(\boldsymbol{w}^{(\ell)}\right) + 2L\epsilon \leq \sum_{\ell=1}^L \text{rev}_{\boldsymbol{r}_\epsilon}\left(\boldsymbol{w}^{(\ell)}\right) + 2L\epsilon \leq \sum_{\ell=1}^L \text{rev}_{\hat{\boldsymbol{r}}'}\left(\boldsymbol{w}^{(\ell)}\right) + 2L\epsilon$.

Next, we prove that for any reserve vector $\boldsymbol{r}$, revenue under the samples $\boldsymbol{v}^{(\ell)}$ will only be higher than revenue under the samples $\boldsymbol{w}^{(\ell)}$, which intuitively makes sense since $\boldsymbol{w}^{(\ell)} \leq \boldsymbol{v}^{(\ell)}$. The proof, which has a similar structure as the proof of Lemma 3.4, is in Appendix B.2.

Lemma 3.5. *For all samples $\ell \in [L]$ and reserve vectors $\boldsymbol{r} \in \mathbb{R}^n$, $\text{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right) \leq \text{rev}_{\boldsymbol{r}}\left(\boldsymbol{v}^{(\ell)}\right)$.*

Since $\sum_{\ell=1}^L \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) \leq \sum \text{rev}_{\hat{\boldsymbol{r}}'}\left(\boldsymbol{w}^{(\ell)}\right) + 2L\epsilon$, Lemma 3.5 implies that the theorem holds. □

We next use Theorem 3.3 to prove stability guarantees in the case where there are multiple items and n additive buyers. It follows from the fact that the mechanism's revenue decomposes additively over the items. The proof is in Appendix B.2.

Theorem 3.6. *The set of multi-item non-anonymous second-price auctions under n additive buyers is p-stable with $p(\epsilon, n, m) = 2m\epsilon$.*

Theorem 3.6 and the fact that the class is q-convergent with $q(\delta, L, n, m) = O\left(\sqrt{\frac{1}{L}\left(nm\log(nm) + \log\frac{1}{\delta}\right)}\right)$ [46] implies that optimization under noise is possible:

Corollary 3.7. *Let $\mathcal{S} = \left\{\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(L)}\right\} \subset \mathbb{R}^{nm}_{\geq 0}$ and $\mathcal{S}' = \left\{\boldsymbol{w}^{(1)}, \dots, \boldsymbol{w}^{(L)}\right\} \subset \mathbb{R}^{nm}_{\geq 0}$ be two sets such that for all $\ell \in [L]$, $\boldsymbol{v}^{(\ell)} - \boldsymbol{\epsilon} \leq \boldsymbol{w}^{(\ell)} \leq \boldsymbol{v}^{(\ell)}$. Let $\hat{\boldsymbol{r}}' = \operatorname{argmax}_{\boldsymbol{r} \in \mathbb{R}^{nm}} \sum_{\ell=1}^{L} \operatorname{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right)$. With high probability over $\mathcal{S} \sim \mathcal{D}^L$, $\max_{\boldsymbol{r} \in \mathbb{R}^{nm}} \mathbb{E}_{\boldsymbol{v} \sim \mathcal{D}}\left[\operatorname{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right] - \mathbb{E}_{\boldsymbol{v} \sim \mathcal{D}}\left[\operatorname{rev}_{\hat{\boldsymbol{r}}'}(\boldsymbol{v})\right] = \tilde{O}\left(m\epsilon + \sqrt{\frac{nm}{L}}\right)$.*

This dependence on $m\epsilon$ is tight: no algorithm has better error than $m\epsilon$. Therefore, empirical revenue maximization provides an optimal dependence on the error term ϵ. The proof is in Appendix B.2.

Proposition 3.8. *Fix an arbitrary error term ϵ. For any deterministic algorithm $\mathcal{A}$ that takes as input a training set $\mathcal{S} \subseteq \mathbb{R}^{nm}$ and returns a vector of non-anonymous reserves $\mathcal{A}(\mathcal{S}) \in \mathbb{R}^{nm}$, there exists a distribution $\mathcal{D}$ such that with probability 1 over the draw $\mathcal{S} = \left\{\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(L)}\right\} \sim \mathcal{D}^L$, $\max_{\boldsymbol{r} \in \mathbb{R}^{nm}} \mathbb{E}\left[\operatorname{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right] - \mathbb{E}\left[\operatorname{rev}_{\mathcal{A}(\mathcal{S}')}(\boldsymbol{v})\right] = \Omega(m\epsilon)$ for some noisy training set $\mathcal{S}' = \left\{\boldsymbol{w}^{(1)}, \dots, \boldsymbol{w}^{(L)}\right\}$ such that $\left\|\boldsymbol{v}^{(\ell)} - \boldsymbol{w}^{(\ell)}\right\|_{\infty} \leq \epsilon$ for all $\ell \in [L]$.*

3.3 Lotteries

We next prove that single-priced lottery mechanisms satisfy our stability and convergence conditions. We analyze a setting where there are n unit-demand buyers with values for m items. Unlike the previous section, we assume there are at least n units of each good available. A single-priced lottery is defined by $n(m+1)$ parameters: for each buyer i, there is a price $r_{i0} \in \mathbb{R}_{\geq 0}$ and a set of probabilities $r_{i1}, \dots, r_{im} \in [0,1]$ with $\sum_{j=1}^{m} r_{ij} = 1$. If the buyer chooses to pay r_{i0}, she will receive one item $J \in [m]$, and $\Pr[J = j] = r_{ij}$. Therefore, her expected utility is $\sum_{j=1}^{m} v_{ij} r_{ij} - r_{i0}$. She will choose to participate in the lottery so long as her expected utility is at least 0. We prove that this mechanism class satisfies our stability and convergence conditions. The full proof is in Appendix B.3.

Theorem 3.9. *The set of lotteries with n unit-demand buyers is p-stable with $p(\epsilon, n, m) = n\epsilon$.*

Proof sketch. We sketch the proof that $p(\epsilon, n, m) = \epsilon$ in the single-buyer setting $(n = 1)$ and for the sake of generality, we prove the guarantee for n buyers in the appendix. For any $L \geq 1$, let $\mathcal{S} = \left\{\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(L)}\right\} \subset \mathbb{R}^{m}_{\geq 0}$ and $\mathcal{S}' = \left\{\boldsymbol{w}^{(1)}, \dots, \boldsymbol{w}^{(L)}\right\} \subset \mathbb{R}^{m}_{\geq 0}$ be two arbitrary sets of valuation vectors such that $\boldsymbol{v}^{(\ell)} - \boldsymbol{\epsilon} \leq \boldsymbol{w}^{(\ell)} \leq \boldsymbol{v}^{(\ell)}$ for all $\ell \in [L]$. Let $\hat{\boldsymbol{r}} = (\hat{r}_0, \hat{r}_1, \dots, \hat{r}_m)$ be the empirically optimal parameter vector over the set $\mathcal{S}$ and let $\hat{\boldsymbol{r}}' = (\hat{r}'_0, \hat{r}'_1, \dots, \hat{r}'_m)$ be the empirically optimal parameter vector over the set $\mathcal{S}'$. We prove that $\frac{1}{L} \sum_{\ell=1}^{L} \operatorname{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) - \operatorname{rev}_{\hat{\boldsymbol{r}}'}\left(\boldsymbol{v}^{(\ell)}\right) \leq \epsilon$.

We first prove that if we shift the price $\hat{r}_0$ down by ϵ and evaluate the resulting lottery over $\mathcal{S}'$, little revenue is lost. Specifically, letting $\boldsymbol{r}_\epsilon = (\max\{\hat{r}_0 - \epsilon, 0\}, \hat{r}_1, \dots, \hat{r}_n)$ we prove that if $\operatorname{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) > 0$, then $\operatorname{rev}_{\boldsymbol{r}_\epsilon}\left(\boldsymbol{w}^{(\ell)}\right) \geq \operatorname{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) - \epsilon$. This implies that $\sum_{\ell=1}^{L} \operatorname{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) \leq \sum_{\ell=1}^{L} \operatorname{rev}_{\boldsymbol{r}_\epsilon}\left(\boldsymbol{w}^{(\ell)}\right) + L\epsilon$. Since $\hat{\boldsymbol{r}}'$ is empirically optimal under $\mathcal{S}'$, $\sum_{\ell=1}^{L} \operatorname{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) \leq \sum_{\ell=1}^{L} \operatorname{rev}_{\hat{\boldsymbol{r}}'}\left(\boldsymbol{w}^{(\ell)}\right) + L\epsilon$.

Next, we prove that for any parameter vector $\boldsymbol{r}$, revenue under the samples $\boldsymbol{v}^{(\ell)}$ will only be higher than revenue under the samples $\boldsymbol{w}^{(\ell)}$. Specifically, for every $\ell \in [L]$ and any parameter vector $\boldsymbol{r} \in \mathbb{R}^{m+1}$, $\operatorname{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right) \leq \operatorname{rev}_{\boldsymbol{r}}\left(\boldsymbol{v}^{(\ell)}\right)$. This implies that $\sum_{\ell=1}^{L} \operatorname{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) \leq L\epsilon + \sum_{\ell=1}^{L} \operatorname{rev}_{\hat{\boldsymbol{r}}'}\left(\boldsymbol{v}^{(\ell)}\right)$. ☐

By a natural generalization of prior research [6], $q(\delta, L, n, m) = O\left(\sqrt{\frac{1}{L}\left(nm\log(nm) + \log\frac{1}{\delta}\right)}\right)$. Fact 3.2 and Theorem 3.9 imply our main result for this section—that the class of lotteries is robust:

Corollary 3.10. *Let $\mathcal{S} = \left\{\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(L)}\right\} \subset \mathbb{R}^{nm}_{\geq 0}$ and $\mathcal{S}' = \left\{\boldsymbol{w}^{(1)}, \dots, \boldsymbol{w}^{(L)}\right\} \subset \mathbb{R}^{nm}_{\geq 0}$ be two sets such that for all $\ell \in [L]$, $\boldsymbol{v}^{(\ell)} - \boldsymbol{\epsilon} \leq \boldsymbol{w}^{(\ell)} \leq \boldsymbol{v}^{(\ell)}$. Let $\hat{\boldsymbol{r}}' = \operatorname{argmax}_{\boldsymbol{r} \in \mathbb{R}^{n(m+1)}} \sum_{\ell=1}^{L} \operatorname{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right)$. With high probability over $\mathcal{S} \sim \mathcal{D}^L$, $\max_{\boldsymbol{r} \in \mathbb{R}^{n(m+1)}} \mathbb{E}_{\boldsymbol{v} \sim \mathcal{D}}\left[\operatorname{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right] - \mathbb{E}\left[\operatorname{rev}_{\hat{\boldsymbol{r}}'}(\boldsymbol{v})\right] = \tilde{O}\left(n\epsilon + \sqrt{\frac{nm}{L}}\right)$.*

This dependence on $n\epsilon$ is tight: no algorithm has better error than $n\epsilon$. The proof is in Appendix B.3.

Proposition 3.11. *Fix an arbitrary error term ϵ. For any deterministic algorithm $\mathcal{A}$ that takes as input a training set $\mathcal{S} \subseteq \mathbb{R}^{nm}$ and returns a vector of lottery parameters $\mathcal{A}(\mathcal{S}) \in \mathbb{R}^{n(m+1)}$, there exists a distribution $\mathcal{D}$ such that with probability 1 over the draw $\mathcal{S} = \left\{\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(L)}\right\} \sim \mathcal{D}^L$, $\max_{\boldsymbol{r} \in \mathbb{R}^{n(m+1)}} \mathbb{E}\left[\operatorname{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right] - \mathbb{E}\left[\operatorname{rev}_{\mathcal{A}(\mathcal{S}')}(\boldsymbol{v})\right] = \Omega(n\epsilon)$ for some noisy training set $\mathcal{S}' = \left\{\boldsymbol{w}^{(1)}, \dots, \boldsymbol{w}^{(L)}\right\}$ such that $\left\|\boldsymbol{v}^{(\ell)} - \boldsymbol{w}^{(\ell)}\right\|_{\infty} \leq \epsilon$ for all $\ell \in [L]$.*

3.4 Guarantees via dispersion

We now provide guarantees for the class of item-pricing mechanisms under unit-demand buyers, under which revenue is a particularly volatile function. In fact, the p-stability property does not hold for any non-trivial stability function p. We show that we can use another tool—called *dispersion* [5]—to obtain guarantees for learning with a noisy dataset. Therefore, we illustrate that p-stability is a sufficient but not necessary condition for obtaining robustness guarantees for a wide range of mechanism classes. We showcase this for item-pricing mechanisms.

Item-pricing mechanisms: Let there be n buyers and m items and as in Section 3.2, we assume there is only one unit of each item for sale. There is a non-anonymous price $r_{ij} \in \mathbb{R}$ for each buyer $i \in [n]$ and each item $j \in [m]$. First, buyer 1 arrives and buys the item $j \in [m]$ that maximizes his utility $v_{1j} - r_{1j}$ (or chooses not to buy if $v_{1j} < r_{1j}$ for all items $j \in [m]$). Next, buyer 2 arrives and selects the item among the remaining that maximizes his utility $v_{2j} - r_{2j}$ (or chooses not to buy). This process continues for each buyer $i \in [n]$. As in the previous section, we use the notation $\text{rev}_{\boldsymbol{r}}(\boldsymbol{v})$ to denote the revenue of the mechanism with prices $\boldsymbol{r} \in \mathbb{R}^{nm}$ when the buyers have values $\boldsymbol{v} \in \mathbb{R}^{nm}$.

Under item-pricing mechanisms, we face an immediate obstacle which is that p-stability does not necessarily hold for any non-trivial choice of the function p: the empirically optimal prices over the noisy set $\mathcal{S}'$ might have terrible revenue on average over the true set $\mathcal{S}$ (Lemma B.11 in Appendix B.4).

The primary challenge in proving a stability guarantee is that an agent may have very different values for two items but very similar utilities given a set of prices $\boldsymbol{r}$. If we add a little noise to the buyer's values, their preference ordering over the items may flip. In turn, this may cause the buyer to select a low-cost item instead of a high-cost item, thus triggering a sharp drop in revenue. As a result, average revenue over a noisy training set may be completely different from average revenue over the uncorrupted training set. If, however, given any price vector $\boldsymbol{r}$, the buyers' utilities across items are not too similar (they are "dispersed"), their preferences will not change if a bit of noise is added to their values. Below, we formally capture this notion of dispersion.

Definition 3.12 ((ϵ, k)-dispersion)**.** Let $\mathcal{S} = \left\{\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(L)}\right\} \subset \mathbb{R}^{nm}_{\geq 0}$ be a set of valuation vectors. We say that $\mathcal{S}$ is (ϵ, k)-dispersed if for any price vector $\boldsymbol{r} \in \mathbb{R}^m$, there are at most k valuation vectors in $\mathcal{S}$ such that for some buyer $i \in [n]$, either:
1. For some item pair $j, j' \in [m]$, buyer i's utility for item j is within ϵ of her utility for item j', or
2. Buyer i's utility for some item j is between 0 and ϵ.

The parameter k allows for some slack: for some—but not all—of the valuation vectors in $\mathcal{S}$, the buyers' utilities can concentrate. Later in this section, we demonstrate that dispersion holds under mild assumptions. First, we provide a guarantee based on dispersion for optimizing prices using a noisy training set. The full proof is in Appendix B.4.

Theorem 3.13. *Let $\mathcal{S} = \left\{\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(L)}\right\} \sim \mathcal{D}^L$ be a set of $(2\epsilon, k)$-dispersed vectors. Let $\mathcal{S}' = \left\{\boldsymbol{w}^{(1)}, \dots, \boldsymbol{w}^{(L)}\right\} \subset \mathbb{R}^{nm}_{\geq 0}$ be another set such that for all $\ell \in [L]$, $\left\|\boldsymbol{v}^{(\ell)} - \boldsymbol{w}^{(\ell)}\right\|_\infty \leq \epsilon$. Let $\hat{\boldsymbol{r}}'$ be empirically optimal over $\mathcal{S}'$: $\hat{\boldsymbol{r}}' = \text{argmax}_{\boldsymbol{r} \in \mathbb{R}^{nm}} \sum_{\ell=1}^{L} \text{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right)$. With probability $1-\delta$ over the draw of $\mathcal{S}$, $\max_{\boldsymbol{r} \in \mathbb{R}^{nm}} \mathbb{E}_{\boldsymbol{v} \sim \mathcal{D}}\left[\text{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right] - \mathbb{E}\left[\text{rev}_{\hat{\boldsymbol{r}}'}(\boldsymbol{v})\right] = O\left(\frac{k}{L} + \sqrt{\frac{1}{L}\left(nm\log(nm) + \log\frac{1}{\delta}\right)}\right)$.*

Proof sketch. We first prove that for any price vector $\boldsymbol{r}$, average revenue over $\mathcal{S}$ is within $\frac{k}{L}$ of average revenue over $\mathcal{S}'$, for the following reason. Due to dispersion, for most of the valuation vectors $\boldsymbol{v}^{(\ell)} \in \mathcal{S}$, the buyers will choose to buy the same set of goods regardless of whether their values are defined by $\boldsymbol{v}^{(\ell)}$ or $\boldsymbol{w}^{(\ell)}$, so the revenue remains constant. For at most k valuation vectors in $\mathcal{S}$, the allocation may change arbitrarily, but revenue is always bounded in the interval $[0, 1]$, which is why the bound contains the term $\frac{k}{L}$. Finally, to relate average and expected revenue, we use a generalization bound from prior research [46] which equals the second summand of our bound. $\square$

This result raises an important question: when will a set of valuations be dispersed? One example, also observed in prior research [5], is when the distribution over buyers' values is relatively "smooth" (*a la* smoothed analysis [50]), formalized as follows.

Definition 3.14. For any distribution over an abstract set $\mathcal{X}$ with probability density function $f : \mathcal{X} \to \mathbb{R}_{\geq 0}$, the density function is *κ-bounded* if $\max_{x \in \mathcal{X}} f(x) \leq \kappa$.

The proof of the following lemma is in Appendix B.4.

Lemma 3.15. *Suppose that for every buyer $i \in [n]$ and every pair of items $j, j' \in [m]$, buyer i's values for items j and j' have a κ-bounded joint density function. Then for any $\epsilon > 0$, with probability $1-\delta$ over the draw $\mathcal{S} = \left\{\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(L)}\right\} \sim \mathcal{D}^L$, the set $\mathcal{S}$ is $(2\epsilon, k)$-dispersed with $k = 4Lnm^2\kappa\epsilon + O\left(nm^2\sqrt{L\log\frac{nm}{\delta}}\right)$.*

Theorem 3.13 and Lemma 3.15 imply learning guarantees for optimizing with noisy data when dispersion holds, even for these extremely volatile revenue functions.

4 Multi-task learning

We now show how our results from the previous section apply to multi-task mechanism design.

Setup: In this setting, there are T tasks, and each task $t \in [T]$ is defined by a distribution $\mathcal{D}^{(t)}$ over buyers' values. Each distribution could represent, for example, buyers from different regions or market segments. The learning algorithm receives a training set sampled from each task's distribution. Given a mechanism class parameterized by vectors $\boldsymbol{r} \in \mathbb{R}^d$ (which equal, for example, reserve prices) and the T training sets, our goal is to learn a parameter setting $\boldsymbol{r}^{(t)}$ for each task t with high expected revenue over $\mathcal{D}^{(t)}$. In particular, $\max_{\boldsymbol{r}\in\mathbb{R}^d}\mathbb{E}_{\boldsymbol{v}\sim\mathcal{D}^{(t)}}\left[\text{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right] - \mathbb{E}_{\boldsymbol{v}\sim\mathcal{D}^{(t)}}\left[\text{rev}_{\boldsymbol{r}^{(t)}}(\boldsymbol{v})\right]$ should be small.

Task relatedness: If the distributions are completely unrelated, there is no hope that sharing information across tasks could improve learning. Thus, we require some notion of "task-relatedness". Under our model of task-relatedness, buyers' values across tasks are similar up to additive shifts which represent, for example, differences in their income brackets. More formally, for each buyer $i \in [n]$ and item $j \in [m]$, there is an underlying distribution $\mathcal{D}_{ij}$ over $\mathbb{R}_{\geq 0}$. Let $\mathcal{D} = \times_{i,j}\mathcal{D}_{ij}$ denote the cross product of these nm distributions. Each task $t \in [T]$ is defined by an unknown vector of *common-base values* $\boldsymbol{b}^{(t)} \in \mathbb{R}^n$. Intuitively, when $b_i^{(t)}$ is large, buyer i tends to be willing to pay more for items, and when $b_i^{(t)}$ is small, buyer i tends to be frugal. This model is inspired by the well-studied *common-base-value* model [9, 24]. For each task $t \in [T]$, our learning algorithm receives L samples $\boldsymbol{v}^{(1,t)}, \dots, \boldsymbol{v}^{(L,t)} \in \mathbb{R}^{nm}$. Each sample $\boldsymbol{v}^{(\ell,t)}$ is generated by sampling $\boldsymbol{z}^{(\ell,t)} \sim \mathcal{D}$ and defining buyer i's value for item j to be $v_{ij}^{(\ell,t)} = b_i^{(t)} + z_{ij}^{(\ell,t)}$. This defines a distribution $\mathcal{D}_{ij}^{(t)}$ over buyer i's value for item j. We use the notation $\mathcal{D}^{(t)} = \times_{i,j}\mathcal{D}_{ij}^{(t)}$ to denote the cross product of these nm distributions for task t. The learning algorithm does not know the vectors $\boldsymbol{b}^{(t)}$ or $\boldsymbol{z}^{(\ell,t)}$; it only observes the valuation vectors $\boldsymbol{v}^{(\ell,t)}$. Throughout this section, we use the (slight abuse of) notation $\boldsymbol{b}^{(t)} + \boldsymbol{z}^{(\ell,t)}$ to denote $\boldsymbol{v}^{(\ell,t)}$. In Appendix D, we supply empirical evidence that demonstrates that this is a reasonable model of task-relatedness on real-world auction data.

Sample bootstrapping: Our multi-task approach follows from the observation that for each task $t \in [T]$, $\mathcal{S}_t = \left\{\boldsymbol{b}^{(t)} + \boldsymbol{z}^{(\ell,\tau)} : \tau \in [T], \ell \in [L]\right\}$ is a training set of LT samples from the t^{th} task's distribution $\mathcal{D}^{(t)}$ since each vector $\boldsymbol{z}^{(\ell,\tau)}$ is sampled from $\mathcal{D}$. However, we cannot compile this training set since we do not know the base values $\boldsymbol{b}^{(t)}$. Nonetheless, in Appendix C, we show that it is straightforward to compute a set of vectors $\mathcal{S}_t' = \left\{\boldsymbol{w}^{(\ell,t,\tau)} : \ell \in [L], \tau \in [T]\right\}$ that closely approximate* the vectors in $\mathcal{S}_t$: with high probability, for all pairs of tasks $t, \tau \in [T]$ and indices $\ell \in [L]$, $\left\|\boldsymbol{b}^{(t)} + \boldsymbol{z}^{(\ell,\tau)} - \boldsymbol{w}^{(\ell,t,\tau)}\right\|_\infty = \tilde{O}\left(\sqrt{\frac{1}{Lm}}\right)$. From our results in Section 3, we know that it is possible to learn a mechanism with high expected revenue over the distribution $\mathcal{D}^{(t)}$ using the noisy training set $\mathcal{S}_t'$. We summarize the resulting guarantees below. All proofs are in Appendix C.

Multi-item non-anonymous second-price auctions. Our results from Section 3.2 imply the following guarantee for learning a high-revenue multi-item second-price auction with non-anonymous reserves under additive bidders.

Theorem 4.1. *For each task $t \in [T]$, let $\hat{\boldsymbol{r}}_t'$ be the empirically optimal reserve vector over the set $\mathcal{S}_t'$: $\hat{\boldsymbol{r}}_t' = \operatorname{argmax}_{\boldsymbol{r}\in\mathbb{R}^{nm}}\sum_{\ell=1}^L\sum_{\tau=1}^T \text{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell,t,\tau)}\right)$. With probability $1-\delta$, for every task $t \in [T]$, $\max_{\boldsymbol{r}\in\mathbb{R}^{nm}}\mathbb{E}_{\boldsymbol{v}\sim\mathcal{D}^{(t)}}\left[\text{rev}_{\boldsymbol{r}}(\boldsymbol{v}) - \text{rev}_{\hat{\boldsymbol{r}}_t'}(\boldsymbol{v})\right] = \tilde{O}\left(\sqrt{\frac{1}{LT}\left(nm + \log\frac{T}{\delta}\right) + \frac{m}{L}\log\frac{nT}{\delta}}\right)$.*

*In some cases, we show that it is also possible to exactly recover the vectors in $\mathcal{S}_t$ (Appendix C.2).

This theorem implies that for any $\gamma \in (0,1)$, when $L = \tilde{\Omega}\left(m/\gamma^2\right)$ and $T = \tilde{\Omega}(n)$, the revenue of the auction defined by the reserves $\hat{\boldsymbol{r}}'_t$ is within γ of optimal: $\max \mathbb{E}_{\boldsymbol{v}\sim\mathcal{D}^{(t)}}\left[\text{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right] - \mathbb{E}_{\boldsymbol{v}\sim\mathcal{D}^{(t)}}\left[\text{rev}_{\hat{\boldsymbol{r}}'_t}(\boldsymbol{v})\right] \leq \gamma$. In contrast, the best-known single-task sample complexity guarantee requires $L = \tilde{\Omega}\left(nm/\gamma^2\right)$ [46][†]. Our per-task sample complexity is smaller by a multiplicative factor equalling the number of buyers.

Lotteries. Our results from Section 3.3 imply that under unit-demand buyers, optimizing lottery parameters using the noisy training sets $\mathcal{S}'_t$ results in nearly optimal revenue.

Theorem 4.2. *For each task t, let* $\hat{\boldsymbol{r}}'_t$ *be the empirically optimal lottery parameter vector over the set* $\mathcal{S}'_t$*:* $\hat{\boldsymbol{r}}'_t = \text{argmax}_{\boldsymbol{r}\in\mathbb{R}^{n(m+1)}} \sum_{\ell=1}^{L}\sum_{\tau=1}^{T} \text{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell,t,\tau)}\right)$. *With probability* $1-\delta$*, for every* $t \in [T]$*,* $\max_{\boldsymbol{r}\in\mathbb{R}^{n(m+1)}} \mathbb{E}_{\boldsymbol{v}\sim\mathcal{D}^{(t)}}\left[\text{rev}_{\boldsymbol{r}}(\boldsymbol{v}) - \text{rev}_{\hat{\boldsymbol{r}}'_t}(\boldsymbol{v})\right] = \tilde{O}\left(\sqrt{\frac{1}{LT}\left(nm + \log\frac{T}{\delta}\right) + \frac{n^2}{Lm}\log\frac{nT}{\delta}}\right).$

Theorem 4.2 implies that for any $\gamma \in (0,1)$, when $L = \tilde{\Omega}\left(n/\gamma^2\right)$, $T = \tilde{\Omega}(m)$, and there are more items than buyers $(m > n)$, the revenue of the lottery defined by $\hat{\boldsymbol{r}}'_t$ is within γ of optimal: $\max \mathbb{E}_{\boldsymbol{v}\sim\mathcal{D}^{(t)}}\left[\text{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right] - \mathbb{E}_{\boldsymbol{v}\sim\mathcal{D}^{(t)}}\left[\text{rev}_{\hat{\boldsymbol{r}}'_t}(\boldsymbol{v})\right] \leq \gamma$. Meanwhile, the best-known single-task sample complexity bound requires $L = \tilde{\Omega}\left(nm/\gamma^2\right)$ [6]. Our approach's per-task sample complexity is better by a multiplicative factor of m.

5 Conclusions

We provided guarantees for learning high-revenue mechanisms with noisy data. Learning in the presence of noise is particularly challenging in mechanism design because revenue functions are volatile, exhibiting many jump discontinuities. We were nonetheless able to provide revenue guarantees only under the assumption that the magnitude of the noise is bounded or sub-Gaussian. Our guarantees apply to the dominant approach to learning-based mechanism design: empirical revenue maximization (ERM). We demonstrated the application of our guarantees to multi-task mechanism design. We thus initiated the study of both learning with noisy data and multi-task learning in mechanism design.

Future directions: (1) We focused on learning with bounded noise rather than noise of arbitrary magnitude (the full adversarial game). This is an exciting future direction if any guarantees are indeed possible when bids may be arbitrarily altered, and may require the development of new learning algorithms beyond ERM. (2) A major bottleneck for empirical evaluations in learning-based mechanism design is the lack of public datasets—we are only aware of one, from eBay [44]. We focused on theoretical guarantees, but applied research on robust and multi-task learning in mechanism design is a great future direction. (3) Building on our results for multi-task learning, what other notions of task-relatedness permit strong multi-task sample complexity bounds?

Negative societal impacts: This paper falls into the broader line of research on using machine learning for mechanism design. In this direction, collecting individuals' data and using it to fine-tune prices could have negative privacy implications. The development of privacy-preserving approaches to mechanism design via machine learning is a great direction for future research.

References

[1] Noga Alon, Moshe Babaioff, Yannai A Gonczarowski, Yishay Mansour, Shay Moran, and Amir Yehudayoff. Submultiplicative Glivenko-Cantelli and uniform convergence of revenues. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2017.

[2] Rie Kubota Ando and Tong Zhang. A framework for learning predictive structures from multiple tasks and unlabeled data. *Journal of Machine Learning Research*, 6(Nov):1817–1853, 2005.

[†]The pseudo-dimension of the mechanism class is at least nm [6], which suggests that the dependence on n may be necessary in the single-task setting.

[3] Pranjal Awasthi, Maria-Florina Balcan, Nika Haghtalab, and Ruth Urner. Efficient learning of linear separators under bounded noise. In *Proceedings of the Conference on Learning Theory (COLT)*, pages 167–190, 2015.

[4] Maria-Florina Balcan, Avrim Blum, Jason Hartline, and Yishay Mansour. Reducing mechanism design to algorithm design via machine learning. *Journal of Computer and System Sciences*, 74: 78–89, December 2008.

[5] Maria-Florina Balcan, Travis Dick, and Ellen Vitercik. Dispersion for data-driven algorithm design, online learning, and private optimization. In *Proceedings of the IEEE Symposium on Foundations of Computer Science (FOCS)*, 2018.

[6] Maria-Florina Balcan, Tuomas Sandholm, and Ellen Vitercik. A general theory of sample complexity for multi-item profit maximization. In *Proceedings of the ACM Conference on Economics and Computation (EC)*, 2018.

[7] Maria-Florina Balcan, Tuomas Sandholm, and Ellen Vitercik. Estimating approximate incentive compatibility. In *Proceedings of the ACM Conference on Economics and Computation (EC)*, 2019.

[8] Maria-Florina Balcan, Travis Dick, and Wesley Pegden. Semi-bandit optimization in the dispersed setting. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI)*, 2020.

[9] MohammadHossein Bateni, Sina Dehghani, MohammadTaghi Hajiaghayi, and Saeed Seddighin. Revenue maximization for selling multiple correlated items. In *Proceedings of the European Symposium on Algorithms (ESA)*, pages 95–105. Springer, 2015.

[10] Jonathan Baxter. A model of inductive bias learning. *Journal of artificial intelligence research*, 12:149–198, 2000.

[11] Xiaohui Bei and Zhiyi Huang. Bayesian incentive compatibility via fractional assignments. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pages 720–733. SIAM, 2011.

[12] Shai Ben-David and Reba Schuller Borbely. A notion of task relatedness yielding provable multiple-task learning guarantees. *Machine learning*, 73(3):273–287, 2008.

[13] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine learning*, 79 (1-2):151–175, 2010.

[14] Dirk Bergemann and Stephen Morris. Robust mechanism design. *Econometrica*, pages 1771–1813, 2005.

[15] Dirk Bergemann and Karl Schlag. Robust monopoly pricing. *Journal of Economic Theory*, 146 (6):2527–2543, 2011.

[16] Kush Bhatia, Prateek Jain, and Purushottam Kar. Robust regression via hard thresholding. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, pages 721–729, 2015.

[17] Avrim Blum, Nika Haghtalab, Ariel D Procaccia, and Mingda Qiao. Collaborative pac learning. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, pages 2392–2401, 2017.

[18] Patrick Briest, Shuchi Chawla, Robert Kleinberg, and S Matthew Weinberg. Pricing randomized allocations. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 585–597. SIAM, 2010.

[19] Sébastien Bubeck, Nikhil R Devanur, Zhiyi Huang, and Rad Niazadeh. Online auctions and multi-scale online learning. In *Proceedings of the ACM Conference on Economics and Computation (EC)*, 2017.

[20] Yang Cai and Constantinos Daskalakis. Learning multi-item auctions with (or without) samples. In *Proceedings of the IEEE Symposium on Foundations of Computer Science (FOCS)*, 2017.

[21] Yang Cai, Argyris Oikonomou, Grigoris Velegkas, and Mingfei Zhao. An efficient epsilon-bic to bic transformation and its application to black-box reduction in revenue maximization. *arXiv preprint arXiv:1911.10172*, 2019.

[22] Shuchi Chawla, Jason D Hartline, and Robert Kleinberg. Algorithmic pricing via virtual valuations. In *Proceedings of the ACM Conference on Economics and Computation (EC)*, 2007.

[23] Shuchi Chawla, Jason D Hartline, David L Malec, and Balasubramanian Sivan. Multi-parameter mechanism design and sequential posted pricing. In *Proceedings of the Annual Symposium on Theory of Computing (STOC)*, 2010.

[24] Shuchi Chawla, David Malec, and Balasubramanian Sivan. The power of randomness in Bayesian optimal mechanism design. *Games and Economic Behavior*, 91:297–317, 2015.

[25] Richard Cole and Tim Roughgarden. The sample complexity of revenue maximization. In *Proceedings of the Annual Symposium on Theory of Computing (STOC)*, 2014.

[26] Nikhil R Devanur, Zhiyi Huang, and Christos-Alexandros Psomas. The sample complexity of auctions with side information. In *Proceedings of the Annual Symposium on Theory of Computing (STOC)*, 2016.

[27] Ilias Diakonikolas, Gautam Kamath, Daniel Kane, Jerry Li, Ankur Moitra, and Alistair Stewart. Robust estimators in high-dimensions without the computational intractability. *SIAM Journal on Computing*, 48(2):742–864, 2019.

[28] Edith Elkind. Designing and learning optimal finite support auctions. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2007.

[29] Tomer Galanti, Lior Wolf, and Tamir Hazan. A theoretical framework for deep transfer learning. *Information and Inference: A Journal of the IMA*, 5(2):159–209, 2016.

[30] Yannai A Gonczarowski and Noam Nisan. Efficient empirical revenue maximization in single-parameter auction environments. In *Proceedings of the Annual Symposium on Theory of Computing (STOC)*, pages 856–868, 2017.

[31] Yannai A Gonczarowski and S Matthew Weinberg. The sample complexity of up-to-ε multi-dimensional revenue maximization. In *Proceedings of the IEEE Symposium on Foundations of Computer Science (FOCS)*, 2018.

[32] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2014.

[33] Chenghao Guo, Zhiyi Huang, and Xinzhi Zhang. Settling the sample complexity of single-parameter revenue maximization. In *Proceedings of the Annual Symposium on Theory of Computing (STOC)*, 2019.

[34] Rishi Gupta and Tim Roughgarden. A PAC approach to application-specific algorithm selection. *SIAM Journal on Computing*, 46(3):992–1017, 2017.

[35] Sergiu Hart and Noam Nisan. Approximate revenue maximization with multiple items. In *Proceedings of the ACM Conference on Economics and Computation (EC)*, 2012.

[36] Jason D Hartline and Brendan Lucier. Bayesian algorithmic mechanism design. In *Proceedings of the Annual Symposium on Theory of Computing (STOC)*, pages 301–310, 2010.

[37] Zhiyi Huang, Yishay Mansour, and Tim Roughgarden. Making the most of your samples. In *Proceedings of the ACM Conference on Economics and Computation (EC)*, 2015.

[38] Zhiyi Huang, Jinyan Liu, and Xiangning Wang. Learning optimal reserve price against non-myopic bidders. *arXiv preprint arXiv:1804.11060*, 2018.

[39] Xinye Li and Andrew Chi-Chih Yao. On revenue maximization for selling multiple independently distributed items. *Proceedings of the National Academy of Sciences*, 110(28): 11232–11237, 2013.

[40] Anton Likhodedov and Tuomas Sandholm. Methods for boosting revenue in combinatorial auctions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2004.

[41] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2018.

[42] Yishay Mansour, Mehryar Mohri, and Afshin Rostamizadeh. Domain adaptation: Learning bounds and algorithms. In *Proceedings of the Conference on Learning Theory (COLT)*, 2009.

[43] Daniel McNamara and Maria-Florina Balcan. Risk bounds for transferring representations with and without fine-tuning. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 2373–2381, 2017.

[44] Mehryar Mohri and Andrés Muñoz Medina. Learning theory and algorithms for revenue optimization in second price auctions with reserve. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2014.

[45] Jamie Morgenstern and Tim Roughgarden. On the pseudo-dimension of nearly optimal auctions. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2015.

[46] Jamie Morgenstern and Tim Roughgarden. Learning simple auctions. In *Proceedings of the Conference on Learning Theory (COLT)*, 2016.

[47] Andrés Muñoz Medina and Sergei Vassilvitskii. Revenue optimization with approximate bid predictions. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2017.

[48] Tim Roughgarden and Okke Schrijvers. Ironing in the dark. In *Proceedings of the ACM Conference on Economics and Computation (EC)*, 2016.

[49] Ludwig Schmidt, Shibani Santurkar, Dimitris Tsipras, Kunal Talwar, and Aleksander Madry. Adversarially robust generalization requires more data. *arXiv preprint arXiv:1804.11285*, 2018.

[50] Daniel A Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM (JACM)*, 51(3):385–463, 2004.

[51] Vasilis Syrgkanis. A sample complexity measure with applications to learning optimal auctions. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2017.

[52] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2014.

[53] Yu Zhang and Qiang Yang. A survey on multi-task learning. *arXiv preprint arXiv:1707.08114*, 2017.

6 Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] We discussed the limitations of our work in the "Future directions" paragraph of Section 5.
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] We discussed any potential negative societal impacts in Section 5.

(d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]

2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes] We provide full, detailed proofs of all results. Due to space constraints, the main body includes proof sketches of some of the results, and the full proofs of these results are in the appendix.
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [N/A]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A Additional related research

Multi-task learning. Learning algorithms often require large training sets to find accurate models. *Multi-task learning* aims to leverage information across multiple related tasks to improve the learning outcomes per task. The majority of the existing literature on multi-task learning is empirical, whereas we focus on theoretical guarantees. Two early works on the sample complexity of multi-task learning [10, 12] study binary classification, whereas we study real-valued revenue functions. Blum et al. [17] also study multi-task binary classification. In their case, one can pool non-noisy data across different tasks to train the classifiers, which is the not the case in our setting. They also assume that there is a common optimal or close-to-optimal target function that performs well with respect to each task's distribution. We, on the other hand, do not assume the existence of a common set of parameters that yield nearly optimal revenue on each task. Moving beyond binary classification, Ando and Zhang [2] provide generalization guarantees that hold on average over the tasks, whereas we bound the error for each individual task.

Mechanism design with side information. Devanur et al. [26] study single-item mechanism design via machine learning when the seller can see some public information—or *signal*—about each buyer. The signal indicates, for example, their annual income. Each signal defines a distinct marginal distribution over buyers' values, thus characterizing—in essence—a distinct learning task. Devanur et al. [26] show how to learn a mechanism that has high revenue in expectation over the draw of a fresh task (or in other words, the draw of a fresh signal). For us, there is no distribution over tasks: the tasks are fixed up front and we aim to learn a high-revenue mechanism for every task. Moreover, we study the multi-item, not single-item, setting.

Transfer learning and domain adaptation. Transfer learning and domain adaptation are closely related to multi-task learning. In this setting, there is a *source domain* and a *target domain*. The learning algorithm receives training examples from the source domain (and perhaps a few training examples from the target domain), and aims to learn a hypothesis with low loss on the target domain. Research with provable guarantees on this topic include papers by Ben-David et al. [13], Mansour et al. [42], and McNamara and Balcan [43]. In our multi-task mechanism design setting, for any one task, we do not have enough data to provide strong generalization guarantees, let alone to provide strong guarantees if we were to apply the learned mechanism to another task, no matter how related it is. Galanti et al. [29] study a setting where there are a number of *source tasks* together with a *target task*. The source tasks are used to help in the target learning task, but optimizing the loss of the source tasks is not part of the learner's goal. This is in contrast to our goal, which is to learn a high-revenue mechanism per task.

Approximate incentive compatibility. Prior research on approximate *incentive compatibility (IC)* has studied the conversion of η-IC mechanisms into IC mechanisms [e.g., 11, 21, 36]. A manipulable mechanism is η-IC if each agent can improve his utility by at most η when he misreports his values by any arbitrary amount. In contrast, the mechanism classes we study in this paper are dominant strategy incentive compatible (DSIC). Our overall goal is to use the noisy training data to learn a mechanism that has high revenue in expectation over the fresh draw from the true distribution over buyers' values. Since the resulting mechanism is DSIC, we may assume that with respect to this fresh draw, the buyers' bids equal their true values, so no notion of approximate incentive compatibility is relevant. In our model, the adversary has an ℓ_∞-norm budget of ϵ. This ϵ should not be confused with η in the literature on η-IC mechanism design, which captures the maximum amount the buyer's utility can change when he misreports his bid.

B Additional proofs about learning under adversarial noise (Section 3)

Notation. Throughout this appendix, we use the notation $\langle x \rangle = \max\{x, 0\}$.

Fact 3.2. *Fix a p-stable and q-convergent mechanism class. Let $\mathcal{S} = \{\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(L)}\}$ and $\mathcal{S}' = \{\boldsymbol{w}^{(1)}, \dots, \boldsymbol{w}^{(L)}\}$ be two sets of valuation vectors such that for all $\ell \in [L]$, $\boldsymbol{v}^{(\ell)} - \boldsymbol{\epsilon} \leq \boldsymbol{w}^{(\ell)} \leq \boldsymbol{v}^{(\ell)}$. Let $\hat{\boldsymbol{r}}'$ be empirically optimal over $\mathcal{S}'$: $\hat{\boldsymbol{r}}' = \operatorname{argmax}_{\boldsymbol{r} \in \mathbb{R}^d} \sum_{\ell=1}^{L} \operatorname{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right)$. With probability $1 - \delta$ over $\mathcal{S} \sim \mathcal{D}^L$, $\max_{\boldsymbol{r} \in \mathbb{R}^d} \mathbb{E}_{\boldsymbol{v} \sim \mathcal{D}}\left[\operatorname{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right] - \mathbb{E}_{\boldsymbol{v} \sim \mathcal{D}}\left[\operatorname{rev}_{\hat{\boldsymbol{r}}'}(\boldsymbol{v})\right] \leq p(\epsilon, n, m) + 2q(\delta, L, n, m)$.*

Proof. Let $\hat{\boldsymbol{r}}$ be the parameter vector of the empirically optimal mechanism over the true samples $\mathcal{S}$: $\hat{\boldsymbol{r}} = \text{argmax}_{\boldsymbol{r} \in \mathbb{R}^d} \sum_{\ell=1}^{L} \text{rev}_{\boldsymbol{r}}\left(\boldsymbol{v}^{(\ell)}\right)$. Next, let $\boldsymbol{r}^*$ be the parameter vector of the mechanism with the highest expected revenue over the distribution $\mathcal{D}$: $\boldsymbol{r}^* = \text{argmax}_{\boldsymbol{r} \in \mathbb{R}^d} \mathbb{E}_{\boldsymbol{v} \sim \mathcal{D}}\left[\text{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right]$.

Since the mechanism class is (p, q)-robust, we know that with probability $1 - \delta$, the average revenue of any parameter $\boldsymbol{r}$ over the true samples $\mathcal{S}$ is close to its expected revenue:

$$\left| \underset{\boldsymbol{v} \sim \mathcal{D}}{\mathbb{E}}\left[\text{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right] - \frac{1}{L} \sum_{\ell=1}^{L} \text{rev}_{\boldsymbol{r}}\left(\boldsymbol{v}^{(\ell)}\right) \right| \leq q(\delta, L, n, m). \tag{1}$$

Since $\hat{\boldsymbol{r}}$ is the empirically optimal parameter vector over the true samples $\mathcal{S}$, the average revenue of $\boldsymbol{r}^*$ over $\mathcal{S}$ is at most the average revenue of $\hat{\boldsymbol{r}}$ over $\mathcal{S}$. Combined with Equation (1) with $\boldsymbol{r} = \boldsymbol{r}^*$, this means that

$$\begin{aligned} \underset{\boldsymbol{v} \sim \mathcal{D}}{\mathbb{E}}\left[\text{rev}_{\boldsymbol{r}^*}(\boldsymbol{v})\right] &\leq q(\delta, L, n, m) + \frac{1}{L} \sum_{\ell=1}^{L} \text{rev}_{\boldsymbol{r}^*}\left(\boldsymbol{v}^{(\ell)}\right) \\ &\quad - \frac{1}{L} \sum_{\ell=1}^{L} \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) + \frac{1}{L} \sum_{\ell=1}^{L} \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) \\ &\leq q(\delta, L, n, m) + \frac{1}{L} \sum_{\ell=1}^{L} \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right). \end{aligned}$$

Next, we combine this inequality and Equation (1) with $\boldsymbol{r} = \hat{\boldsymbol{r}}'$:

$$\begin{aligned} &\underset{\boldsymbol{v} \sim \mathcal{D}}{\mathbb{E}}\left[\text{rev}_{\boldsymbol{r}^*}(\boldsymbol{v})\right] - \underset{\boldsymbol{v} \sim \mathcal{D}}{\mathbb{E}}\left[\text{rev}_{\hat{\boldsymbol{r}}'}(\boldsymbol{v})\right] \\ &\leq \frac{1}{L} \sum_{\ell=1}^{L} \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) - \text{rev}_{\hat{\boldsymbol{r}}'}\left(\boldsymbol{v}^{(\ell)}\right) + 2q(\delta, L, n, m) \end{aligned}$$

Since the class is (p, q)-robust, $\frac{1}{L} \sum \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) - \text{rev}_{\hat{\boldsymbol{r}}'}\left(\boldsymbol{v}^{(\ell)}\right) \leq p(\epsilon, n, m)$, so the fact holds. □

Lemma B.1. *Let $f_1, \ldots, f_m$ be m functions mapping $\mathbb{R}^n$ to $\mathbb{R}$. Then*

$$\max_{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_m \in \mathbb{R}^n} \sum_{j=1}^{m} f_j\left(\boldsymbol{r}_j\right) = \sum_{j=1}^{m} \max_{\boldsymbol{r}_j \in \mathbb{R}^n} f_j\left(\boldsymbol{r}_j\right).$$

Proof. Clearly,

$$\max_{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_m \in \mathbb{R}^n} \sum_{j=1}^{m} f_j\left(\boldsymbol{r}_j\right) \leq \sum_{j=1}^{m} \max_{\boldsymbol{r}_j \in \mathbb{R}^n} f_j\left(\boldsymbol{r}_j\right).$$

Suppose

$$\max_{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_m \in \mathbb{R}^n} \sum_{j=1}^{m} f_j\left(\boldsymbol{r}_j\right) < \sum_{j=1}^{m} \max_{\boldsymbol{r}_j \in \mathbb{R}^n} f_j\left(\boldsymbol{r}_j\right).$$

Let $(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_m)$ be a set of vectors in $\text{argmax}_{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_m \in \mathbb{R}^n} \sum_{j=1}^{m} f_j\left(\boldsymbol{r}_j\right)$ and for each $j \in [m]$, let $\boldsymbol{q}_j'$ be a vector in $\text{argmax}_{\boldsymbol{r}_j \in \mathbb{R}^n} f_j\left(\boldsymbol{r}_j\right)$. By assumption $\sum_{j=1}^{m} f_j\left(\boldsymbol{q}_j\right) < \sum_{j=1}^{m} f_j\left(\boldsymbol{q}_j'\right)$. However, this contradicts the fact that

$$\sum_{j=1}^{m} f_j\left(\boldsymbol{q}_j\right) = \max_{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_m \in \mathbb{R}^n} \sum_{j=1}^{m} f_j\left(\boldsymbol{r}_j\right).$$

Therefore, the lemma statement holds. □

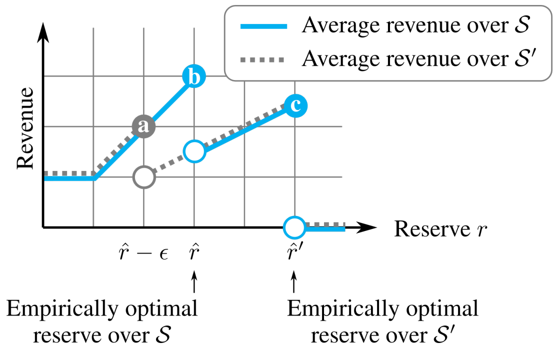

Figure 2: Illustration of Example B.3, which exemplifies Theorem B.2 when $\epsilon = 1$, $\boldsymbol{v}^{(1)} = (1,3)$, $\boldsymbol{w}^{(1)} = (1,2)$, and $\boldsymbol{v}^{(2)} = \boldsymbol{w}^{(2)} = (1,5)$. The blue solid line is average revenue over $\mathcal{S} = \left\{\boldsymbol{v}^{(1)}, \boldsymbol{v}^{(2)}\right\}$ as a function of the reserve r and the grey dotted line is average revenue over $\mathcal{S}' = \left\{\boldsymbol{w}^{(1)}, \boldsymbol{w}^{(2)}\right\}$. The lemma shows that the difference between points (b) and (c) is small.

B.1 Second-price auctions with anonymous reserves

In this section, we analyze second-price auctions with anonymous reserves, which is a bit less involved than our analysis of multi-item *non-anonymous* reserves. We assume there is only one unit of the item for sale. A single-item second-price auction with an anonymous reserve is defined by a reserve price $r \in \mathbb{R}$. Each bidder submits a real-valued bid to the auctioneer. The highest bidder wins so long as her bid is above the reserve and pays the maximum of the second-highest bid and the reserve. We prove that this mechanism class is p-stable.

Theorem B.2. *The class of single-item second-price auctions with anonymous reserves is p-stable with $p(\epsilon, n, m) = 2\epsilon$.*

We first provide an example of Theorem B.2's implications and a proof overview.

Example B.3. Figure 2 llustrates Theorem B.2 when $\epsilon = 1$, $\boldsymbol{v}^{(1)} = (1,3)$, $\boldsymbol{w}^{(1)} = (1,2)$, and $\boldsymbol{v}^{(2)} = \boldsymbol{w}^{(2)} = (1,5)$. The blue solid line is average revenue over $\mathcal{S} = \left\{\boldsymbol{v}^{(1)}, \boldsymbol{v}^{(2)}\right\}$, which equals $\frac{1}{2}\sum_{\ell=1}^{2} \text{rev}_r\left(\boldsymbol{v}^{(\ell)}\right)$, as a function of the reserve r. The grey dotted line is average revenue over $\mathcal{S}' = \left\{\boldsymbol{w}^{(1)}, \boldsymbol{w}^{(2)}\right\}$. The reserves that maximize average revenue over $\mathcal{S}$ and $\mathcal{S}'$ are $\hat{r} = 3$ and $\hat{r}' = 5$, respectively. Theorem B.2 states that the difference between average revenue over $\mathcal{S}$ when $r = \hat{r}'$ and $r = \hat{r}$ is small. One might hope that $\boldsymbol{v}^{(\ell)} - \boldsymbol{\epsilon} \leq \boldsymbol{w}^{(\ell)} \leq \boldsymbol{v}^{(\ell)}$ would imply that $\hat{r} - \epsilon \leq \hat{r}' \leq \hat{r}$, which might imply the lemma holds. Figure 2 shows this is not the case. We first show that the difference between points (a) and (b) is small, where (a) is the average revenue of $\hat{r} - \epsilon$ over $\mathcal{S}'$ and (b) is the average revenue of $\hat{r}$ over $\mathcal{S}$. We then show that this implies that Theorem B.2 holds: the difference between points (b) and (c) is small, where (c) is the average revenue of $\hat{r}'$ over $\mathcal{S}$.

Proof of Theorem B.2. For any $L \geq 1$, let $\mathcal{S} = \left\{\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(L)}\right\} \subset [0,1]^n$ and $\mathcal{S}' = \left\{\boldsymbol{w}^{(1)}, \dots, \boldsymbol{w}^{(L)}\right\} \subset [0,1]^n$ be two arbitrary sets of valuation vectors such that $\boldsymbol{v}^{(\ell)} - \boldsymbol{\epsilon} \leq \boldsymbol{w}^{(\ell)} \leq \boldsymbol{v}^{(\ell)}$ for all $\ell \in [L]$. Let $\hat{r}'$ be the empirically optimal reserve over the set $\mathcal{S}'$ and $\hat{r}$ be the empirically optimal reserve over the set $\mathcal{S}$. We prove that $\frac{1}{L}\sum_{\ell=1}^{L} \text{rev}_{\hat{r}}\left(\boldsymbol{v}^{(\ell)}\right) - \text{rev}_{\hat{r}'}\left(\boldsymbol{v}^{(\ell)}\right) \leq \epsilon$.

This proof relies on two key claims. The first, Claim B.4 states that if we shift the reserve $\hat{r}$—which is empirically optimal for the samples in $\mathcal{S}$—down by an additive factor of ϵ and apply this reserve to the samples in $\mathcal{S}'$, little revenue is lost. Specifically, Claim B.4 guarantees that for all $\ell \in [L]$, if $\text{rev}_{\hat{r}}\left(\boldsymbol{v}^{(\ell)}\right) > 0$, then $\text{rev}_{\langle \hat{r}-\epsilon \rangle}\left(\boldsymbol{w}^{(\ell)}\right) \geq \text{rev}_{\hat{r}}\left(\boldsymbol{v}^{(\ell)}\right) - 2\epsilon$.

Claim B.4. *For all $\ell \in [L]$, if $\text{rev}_{\hat{r}}\left(\boldsymbol{v}^{(\ell)}\right) > 0$, then $\text{rev}_{\langle \hat{r}-\epsilon \rangle}\left(\boldsymbol{w}^{(\ell)}\right) \geq \text{rev}_{\hat{r}}\left(\boldsymbol{v}^{(\ell)}\right) - 2\epsilon$.*

Proof of Claim B.4. Fix an arbitrary sample $\ell \in [L]$ such that $\text{rev}_{\hat{r}}\left(\boldsymbol{v}^{(\ell)}\right) > 0$. Let i be the bidder with the highest valuation in $\boldsymbol{v}^{(\ell)}$. Since $\text{rev}_{\hat{r}}\left(\boldsymbol{v}^{(\ell)}\right) > 0$, it must be that bidder i's bid is at least the reserve $\hat{r}$. This implies that $w_i^{(\ell)} \geq \left\langle v_i^{(\ell)} - \epsilon \right\rangle \geq \langle \hat{r} - \epsilon \rangle$, so under the valuation vector $\boldsymbol{w}^{(\ell)}$ and reserve $\langle \hat{r}-\epsilon \rangle$, there is at least one bidder whose bid is at least the reserve. Let i' be the highest bidder (or in other words, the winner) under the valuation vector $\boldsymbol{w}^{(\ell)}$. There are two cases, depending on whether or not $i = i'$. In both cases, we show that $\text{rev}_{\langle \hat{r}-\epsilon \rangle}\left(\boldsymbol{w}^{(\ell)}\right) \geq \text{rev}_{\hat{r}}\left(\boldsymbol{v}^{(\ell)}\right) - 2\epsilon$, so the claim holds. We use the notation k to denote the second-highest bidder under $\boldsymbol{v}^{(\ell)}$ $\left(k = \text{argmax}_{t \neq i} v_t^{(\ell)}\right)$ and k' to denote the second-highest bidder under $\boldsymbol{w}^{(\ell)}$ $\left(k' = \text{argmax}_{t \neq i'} w_t^{(\ell)}\right)$. Using this notation, $\text{rev}_{\hat{r}}\left(\boldsymbol{v}^{(\ell)}\right) = \max\left\{\hat{r}, v_k^{(\ell)}\right\}$ and $\text{rev}_{\langle \hat{r}-\epsilon \rangle}\left(\boldsymbol{w}^{(\ell)}\right) = \max\left\{\langle \hat{r}-\epsilon \rangle, w_{k'}^{(\ell)}\right\} = \max\left\{\hat{r}-\epsilon, w_{k'}^{(\ell)}\right\}$.

In the first case, $i = i'$. Since $k \neq i$, it must be that $k \neq i'$. Bidder i' is the highest bidder under $\boldsymbol{w}^{(\ell)}$, so it must be that bidder k is the t^{th} highest bidder under $\boldsymbol{w}^{(\ell)}$ for some $t \geq 2$. Since bidder k' is the second-highest bidder under $\boldsymbol{w}^{(\ell)}$, this means that $w_{k'}^{(\ell)} \geq w_k^{(\ell)} \geq v_k^{(\ell)} - \epsilon$. Therefore, $\text{rev}_{\langle \hat{r}-\epsilon \rangle}\left(\boldsymbol{w}^{(\ell)}\right) = \max\left\{\hat{r}-\epsilon, w_{k'}^{(\ell)}\right\} \geq \max\left\{\hat{r}, w_{k'}^{(\ell)}\right\} - \epsilon \geq \max\left\{\hat{r}, v_k^{(\ell)} - \epsilon\right\} - \epsilon \geq \max\left\{\hat{r}, v_k^{(\ell)}\right\} - 2\epsilon = \text{rev}_{\hat{r}}\left(\boldsymbol{v}^{(\ell)}\right) - 2\epsilon$.

In the second case, $i \neq i'$. Bidder i' is the highest bidder under $\boldsymbol{w}^{(\ell)}$, so it must be that bidder i is the t^{th} highest bidder under $\boldsymbol{w}^{(\ell)}$ for some $t \geq 2$. Since bidder k' is the second-highest bidder under $\boldsymbol{w}^{(\ell)}$, this means that $w_{k'}^{(\ell)} \geq w_i^{(\ell)} \geq v_i^{(\ell)} - \epsilon$. Therefore, $\text{rev}_{\langle \hat{r}-\epsilon \rangle}\left(\boldsymbol{w}^{(\ell)}\right) \geq w_{k'}^{(\ell)} \geq v_i^{(\ell)} - \epsilon \geq \text{rev}_{\hat{r}}\left(\boldsymbol{v}^{(\ell)}\right) - \epsilon$, where the final inequality holds because the auction's revenue is always at most the highest bid. ☐

Let $I \subseteq [L]$ be the set of indices ℓ such that $\text{rev}_{\hat{r}}\left(\boldsymbol{v}^{(\ell)}\right) > 0$ (for all other indices, $\text{rev}_{\hat{r}}\left(\boldsymbol{v}^{(\ell)}\right) = 0$). By Claim B.4, for all $\ell \in I$, $\text{rev}_{\hat{r}}\left(\boldsymbol{v}^{(\ell)}\right) \leq \text{rev}_{\langle \hat{r}-\epsilon \rangle}\left(\boldsymbol{w}^{(\ell)}\right) + 2\epsilon$. Therefore,

$$\sum_{\ell=1}^{L} \text{rev}_{\hat{r}}\left(\boldsymbol{v}^{(\ell)}\right) = \sum_{\ell \in I} \text{rev}_{\hat{r}}\left(\boldsymbol{v}^{(\ell)}\right) \leq \sum_{\ell \in I} \text{rev}_{\langle \hat{r}-\epsilon \rangle}\left(\boldsymbol{w}^{(\ell)}\right) + 2L\epsilon \leq \sum_{\ell=1}^{L} \text{rev}_{\langle \hat{r}-\epsilon \rangle}\left(\boldsymbol{w}^{(\ell)}\right) + 2L\epsilon. \tag{2}$$

(In Figure 2, this proves the claim that the difference between points (a) and (b) is small.) Since $\hat{r}'$ is the empirically optimal reserve under the perturbed training instances $\mathcal{S}'$ (mathematically, $\hat{r}' = \text{argmax}_{r \in [0,1]} \sum_{\ell=1}^{L} \text{rev}_r\left(\boldsymbol{w}^{(\ell)}\right)$), Equation (2) implies that

$$\sum_{\ell=1}^{L} \text{rev}_{\hat{r}}\left(\boldsymbol{v}^{(\ell)}\right) \leq 2L\epsilon + \sum_{\ell=1}^{L} \text{rev}_{\hat{r}'}\left(\boldsymbol{w}^{(\ell)}\right). \tag{3}$$

Next, we prove that for any reserve r, revenue under the samples $\boldsymbol{v}^{(\ell)}$ will only be higher than revenue under the samples $\boldsymbol{w}^{(\ell)}$. This claim intuitively makes sense since $\boldsymbol{w}^{(\ell)} \leq \boldsymbol{v}^{(\ell)}$.

Claim B.5. *For every sample $\ell \in [L]$ and any reserve $r \in [0,1]$,* $\text{rev}_r\left(\boldsymbol{w}^{(\ell)}\right) \leq \text{rev}_r\left(\boldsymbol{v}^{(\ell)}\right)$.

Proof of Claim B.5. Fix an arbitrary index $\ell \in [L]$ such that $\text{rev}_r\left(\boldsymbol{w}^{(\ell)}\right) > 0$ (if $\text{rev}_r\left(\boldsymbol{w}^{(\ell)}\right) = 0$, then the claim clearly holds). Let i' be the highest bidder under the valuation vector $\boldsymbol{w}^{(\ell)}$. Since $\text{rev}_r\left(\boldsymbol{w}^{(\ell)}\right) > 0$, it must be that the bid of bidder i' is at least the reserve r. Since $v_{i'}^{(\ell)} \geq w_{i'}^{(\ell)}$, we know that $v_{i'}^{(\ell)} \geq r$ as well. Therefore, under the valuation vector $\boldsymbol{v}^{(\ell)}$ and reserve r, there is at least one bidder whose bid is at least the reserve. Let i be the highest bidder (or in other words, the winner) under the valuation vector $\boldsymbol{v}^{(\ell)}$. As in the proof of Claim B.4, there are two cases, depending on whether or not $i = i'$. In both cases, we show that $\text{rev}_r\left(\boldsymbol{w}^{(\ell)}\right) \leq \text{rev}_r\left(\boldsymbol{v}^{(\ell)}\right)$, so the claim holds. Again, we use the notation k to denote the second-highest bidder under $\boldsymbol{v}^{(\ell)}$ $\left(k = \text{argmax}_{t \neq i} v_t^{(\ell)}\right)$ and k' to denote the second-highest bidder under $\boldsymbol{w}^{(\ell)}$ $\left(k' = \text{argmax}_{t \neq i'} w_t^{(\ell)}\right)$. Using this notation, $\text{rev}_{\hat{r}}\left(\boldsymbol{v}^{(\ell)}\right) = \max\left\{\hat{r}, v_k^{(\ell)}\right\}$ and $\text{rev}_{\hat{r}}\left(\boldsymbol{w}^{(\ell)}\right) = \max\left\{\hat{r}, w_{k'}^{(\ell)}\right\}$.

In the first case, $i = i'$. Since $k' \neq i'$, it must be that $k' \neq i$. Therefore, bidder k' is the t^{th} highest bidder under the valuation vector $\boldsymbol{v}^{(\ell)}$ for some $t \geq 2$. Since bidder k is the second-highest bidder under $\boldsymbol{v}^{(\ell)}$, we know that $v_k^{(\ell)} \geq v_{k'}^{(\ell)} \geq w_{k'}^{(\ell)}$. Therefore, $\text{rev}_r\left(\boldsymbol{w}^{(\ell)}\right) = \max\left\{r, w_{k'}^{(\ell)}\right\} \leq \max\left\{r, v_k^{(\ell)}\right\} = \text{rev}_r\left(\boldsymbol{v}^{(\ell)}\right)$.

In the second case, $i \neq i'$, so it must be that bidder i' is the t^{th} highest bidder under the valuation vector $\boldsymbol{v}^{(\ell)}$ for some $t \geq 2$. Since bidder k is the second-highest bidder under $\boldsymbol{v}^{(\ell)}$, we know that $v_k^{(\ell)} \geq v_{i'}^{(\ell)}$. Therefore, $\text{rev}_r\left(\boldsymbol{v}^{(\ell)}\right) \geq v_k^{(\ell)} \geq v_{i'}^{(\ell)} \geq w_{i'}^{(\ell)} \geq \text{rev}_r\left(\boldsymbol{w}^{(\ell)}\right)$, where the final inequality follows from the fact that the revenue of the auction is always at most the valuation of the highest bidder. □

Equation (3) and Claim B.5 imply $\sum_{\ell=1}^{L} \text{rev}_{\hat{r}}\left(\boldsymbol{v}^{(\ell)}\right) \leq 2L\epsilon + \sum_{\ell=1}^{L} \text{rev}_{\hat{r}'}\left(\boldsymbol{v}^{(\ell)}\right)$, as claimed. □

Morgenstern and Roughgarden [46] prove the mechanism class is q-convergent with $q(\delta, L, n, m) = O\left(\sqrt{\frac{1}{L}\log\frac{1}{\delta}}\right)$. This fact together with Fact 3.2 and Theorem B.2 implies the following guarantee:

Corollary B.6. *Let $\mathcal{S} = \left\{\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(L)}\right\} \subset \mathbb{R}^n_{\geq 0}$ and $\mathcal{S}' = \left\{\boldsymbol{w}^{(1)}, \ldots, \boldsymbol{w}^{(L)}\right\} \subset \mathbb{R}^n_{\geq 0}$ be two sets such that for all $\ell \in [L]$, $\boldsymbol{v}^{(\ell)} - \boldsymbol{\epsilon} \leq \boldsymbol{w}^{(\ell)} \leq \boldsymbol{v}^{(\ell)}$. Let $\hat{r}' = \text{argmax} \sum_{\ell=1}^{L} \text{rev}_r\left(\boldsymbol{w}^{(\ell)}\right)$. With high probability,* $\max \mathbb{E}_{\boldsymbol{v}\sim\mathcal{D}}\left[\text{rev}_r(\boldsymbol{v})\right] - \mathbb{E}_{\boldsymbol{v}\sim\mathcal{D}}\left[\text{rev}_{\hat{r}'}(\boldsymbol{v})\right] = \tilde{O}\left(\epsilon + \sqrt{\frac{1}{L}}\right)$.

B.2 Additional proofs about second-price auctions with non-anonymous reserves

Lemma 3.4. *Let $\boldsymbol{r}_\epsilon = \left(\langle \hat{r}_1 - \epsilon\rangle, \ldots, \langle \hat{r}_n - \epsilon\rangle\right)$. If $\text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) > 0$, then $\text{rev}_{\boldsymbol{r}_\epsilon}\left(\boldsymbol{w}^{(\ell)}\right) \geq \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) - 2\epsilon$.*

Proof. Fix an arbitrary index $\ell \in [L]$ such that $\text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) > 0$. Let i be the winner under the valuation vector $\boldsymbol{v}^{(\ell)}$ and reserve vector $\hat{\boldsymbol{r}}$. Since $v_i^{(\ell)} \geq \hat{r}_i$, we know that $w_i^{(\ell)} \geq \left\langle v_i^{(\ell)} - \epsilon\right\rangle \geq \langle \hat{r}_i - \epsilon\rangle$, so under the valuation vector $\boldsymbol{w}^{(\ell)}$ and reserve vector $\boldsymbol{r}_\epsilon$, there is at least one bidder whose bid is at least his reserve. Let i' be the winner under the valuation vector $\boldsymbol{w}^{(\ell)}$ and reserve vector $\boldsymbol{r}_\epsilon$.

As in the proof of Claim B.4, we split this proof into cases. This proof, however, is a bit more involved because there are four cases, instead of two. These four cases depend on whether or not $i = i'$ (as in Claim B.4) and whether or not revenue depends on a second-highest bidder under the valuation vector $\boldsymbol{v}^{(\ell)}$ and reserve vector $\hat{\boldsymbol{r}}$ (in other words, whether or not $\left\{t : v_t^{(\ell)} \geq \hat{r}_t, t \neq i\right\} = \emptyset$). In all four cases, we show that $\text{rev}_{\boldsymbol{r}_\epsilon}\left(\boldsymbol{w}^{(\ell)}\right) \geq \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) - 2\epsilon$, so the claim holds.

Case 1: $i = i'$ and $\left\{t : v_t^{(\ell)} \geq \hat{r}_t, t \neq i\right\} \neq \emptyset$. Let k be the index of the second-highest bidder in $\boldsymbol{v}^{(\ell)}$ whose bid is above his reserve in $\hat{\boldsymbol{r}}$: $k = \text{argmax}_{t\neq i}\left\{v_t^{(\ell)} : v_t^{(\ell)} \geq \hat{r}_t\right\}$. This means that $\text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) = \max\left\{\hat{r}_i, v_k^{(\ell)}\right\}$. Since $v_k^{(\ell)} \geq \hat{r}_k$, we have that $w_k^{(\ell)} \geq \left\langle v_k^{(\ell)} - \epsilon\right\rangle \geq \langle \hat{r}_k - \epsilon\rangle$. Since $k \neq i$ and $i = i'$, it must be that $k \neq i'$, so $\left\{t : w_t^{(\ell)} \geq \langle \hat{r}_t - \epsilon\rangle, t \neq i'\right\} \neq \emptyset$ (in particular, the set contains k). Let k' be the index of the second-highest bidder in $\boldsymbol{w}^{(\ell)}$ whose bid is above his reserve in $\boldsymbol{r}_\epsilon$: $k' = \text{argmax}_{t\neq i'}\left\{w_t^{(\ell)} : w_t^{(\ell)} \geq \langle \hat{r}_t - \epsilon\rangle\right\}$, which means that $\text{rev}_{\boldsymbol{r}_\epsilon}\left(\boldsymbol{w}^{(\ell)}\right) = \max\left\{\langle \hat{r}_{i'} - \epsilon\rangle, w_{k'}^{(\ell)}\right\} = \max\left\{\hat{r}_{i'} - \epsilon, w_{k'}^{(\ell)}\right\}$. Since $k \in \left\{t : w_t^{(\ell)} \geq \langle \hat{r}_t - \epsilon\rangle, t \neq i'\right\}$, we know that $w_{k'}^{(\ell)} \geq w_k^{(\ell)} \geq v_k^{(\ell)} - \epsilon$, which means that $\text{rev}_{\boldsymbol{r}_\epsilon}\left(\boldsymbol{w}^{(\ell)}\right) = \max\left\{\hat{r}_{i'} - \epsilon, w_{k'}^{(\ell)}\right\} = \max\left\{\hat{r}_i - \epsilon, w_{k'}^{(\ell)}\right\} \geq \max\left\{\hat{r}_i, w_{k'}^{(\ell)}\right\} - \epsilon \geq \max\left\{\hat{r}_i, v_k^{(\ell)} - \epsilon\right\} - \epsilon \geq \max\left\{\hat{r}_i, v_k^{(\ell)}\right\} - 2\epsilon = \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) - 2\epsilon$.

Case 2: $i = i'$ **and** $\left\{t : v_t^{(\ell)} \geq \hat{r}_t, t \neq i\right\} = \emptyset$**.** In this case, $\text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) = \hat{r}_i$, so $\text{rev}_{\boldsymbol{r}_\epsilon}\left(\boldsymbol{w}^{(\ell)}\right) \geq \hat{r}_{i'} - \epsilon = \hat{r}_i - \epsilon = \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) - \epsilon$.

Case 3: $i \neq i'$ **and** $\left\{t : v_t^{(\ell)} \geq \hat{r}_t, t \neq i\right\} \neq \emptyset$**.** We know that $v_i^{(\ell)} \geq \hat{r}_i$, which means that $w_i^{(\ell)} \geq \left\langle v_i^{(\ell)} - \epsilon \right\rangle \geq \langle \hat{r}_i - \epsilon \rangle$. Since $i \neq i'$, it must be that $\left\{t : w_t^{(\ell)} \geq \langle \hat{r}_t - \epsilon \rangle, t \neq i'\right\} \neq \emptyset$ (in particular, the set contains i). Let k' be the index of the second-highest bidder in $\boldsymbol{w}^{(\ell)}$ whose bid is above his reserve in $\boldsymbol{r}_\epsilon$: $k' = \text{argmax}_{t \neq i'}\left\{w_t^{(\ell)} : w_t^{(\ell)} \geq \langle \hat{r}_t - \epsilon \rangle\right\}$, which means that $\text{rev}_{\boldsymbol{r}_\epsilon}\left(\boldsymbol{w}^{(\ell)}\right) = \max\left\{\langle \hat{r}_{i'} - \epsilon \rangle, w_{k'}^{(\ell)}\right\}$. Since $i \in \left\{t : w_t^{(\ell)} \geq \langle \hat{r}_t - \epsilon \rangle, t \neq i'\right\}$, it must be that $w_{k'}^{(\ell)} \geq w_i^{(\ell)} \geq v_i^{(\ell)} - \epsilon$. Therefore, $\text{rev}_{\boldsymbol{r}_\epsilon}\left(\boldsymbol{w}^{(\ell)}\right) \geq w_{k'}^{(\ell)} \geq v_i^{(\ell)} - \epsilon \geq \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) - \epsilon$, where the final inequality holds because the revenue of the auction is never more than the highest bid.

Case 4: $i \neq i'$ **and** $\left\{t : v_t^{(\ell)} \geq \hat{r}_t, t \neq i\right\} = \emptyset$**.** Suppose that $\left\{t : w_t^{(\ell)} \geq \langle \hat{r}_t - \epsilon \rangle, t \neq i'\right\} = \emptyset$. Since $i \neq i'$, this means that $\langle \hat{r}_i - \epsilon \rangle > w_i^{(\ell)}$. Since $w_i^{(\ell)} \geq 0$, this inequality implies that $\hat{r}_i - \epsilon > w_i^{(\ell)} \geq v_i^{(\ell)} - \epsilon$. However, this means that $\hat{r}_i > v_i^{(\ell)}$, which is a contradiction. Therefore, $\left\{t : w_t^{(\ell)} \geq \langle \hat{r}_t - \epsilon \rangle, t \neq i'\right\} \neq \emptyset$. Let k' be the index of the second-highest bidder in $\boldsymbol{w}^{(\ell)}$ whose bid is above his reserve in $\boldsymbol{r}_\epsilon$, or in other words, $k' = \text{argmax}_{t \neq i'}\left\{w_t^{(\ell)} : w_t^{(\ell)} \geq \langle \hat{r}_t - \epsilon \rangle\right\}$. Since $i \in \left\{t : w_t^{(\ell)} \geq \langle \hat{r}_t - \epsilon \rangle, t \neq i'\right\}$, we know that $w_{k'}^{(\ell)} \geq w_i^{(\ell)} \geq v_i^{(\ell)} - \epsilon$, so

$$\text{rev}_{\boldsymbol{r}_\epsilon}\left(\boldsymbol{w}^{(\ell)}\right) = \max\left\{\langle \hat{r}_{i'} - \epsilon \rangle, w_{k'}^{(\ell)}\right\} \geq w_{k'}^{(\ell)} \geq v_i^{(\ell)} - \epsilon \geq \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) - \epsilon,$$

where the final inequality holds because the revenue of the auction is never more than the highest bid. □

Lemma 3.5. *For all samples* $\ell \in [L]$ *and reserve vectors* $\boldsymbol{r} \in \mathbb{R}^n$, $\text{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right) \leq \text{rev}_{\boldsymbol{r}}\left(\boldsymbol{v}^{(\ell)}\right)$.

Proof. Fix an arbitrary index $\ell \in [L]$ such that $\text{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right) > 0$ (if $\text{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right) = 0$, then the claim clearly holds). Let i' be the winner under the valuation vector $\boldsymbol{w}^{(\ell)}$ and reserve vector $\boldsymbol{r}$. Since $w_{i'}^{(\ell)} \geq r_{i'}$ and $v_{i'}^{(\ell)} \geq w_{i'}^{(\ell)}$, we know that $v_{i'}^{(\ell)} \geq r_{i'}$. Therefore, under the valuation vector $\boldsymbol{v}^{(\ell)}$ and reserve vector $\boldsymbol{r}$, there is at least one bidder whose bid is at least his reserve. Let i be the winner under the valuation vector $\boldsymbol{v}^{(\ell)}$ and reserve vector $\boldsymbol{r}$.

There are four cases, depending on whether or not $i = i'$ and whether or not revenue depends on a second-highest bidder under the valuation vector $\boldsymbol{w}^{(\ell)}$ and reserve vector $\boldsymbol{r}$ (in other words, whether or not $\left\{t : w_t^{(\ell)} \geq r_t, t \neq i'\right\} = \emptyset$). As in the proof of Lemma 3.4, splitting the analysis into these four cases helps us deal with the challenge summarized by the proof sketch in the main body. In all four cases, we show that $\text{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right) \leq \text{rev}_{\boldsymbol{r}}\left(\boldsymbol{v}^{(\ell)}\right)$, so the claim holds.

Case 1: $i = i'$ **and** $\left\{t : w_t^{(\ell)} \geq r_t, t \neq i'\right\} \neq \emptyset$**.** Let k' be the index of the second-highest bidder in $\boldsymbol{w}^{(\ell)}$ whose bid is above his reserve: $k' = \text{argmax}_{t \neq i'}\left\{w_t^{(\ell)} : w_t^{(\ell)} \geq r_t\right\}$, so $\text{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right) = \max\left\{r_{i'}, w_{k'}^{(\ell)}\right\}$. Since $r_{k'} \leq w_{k'}^{(\ell)} \leq v_{k'}^{(\ell)}$, $k' \neq i'$, and $i = i'$, we know that $\left\{t : v_t^{(\ell)} \geq r_t, t \neq i\right\} \neq \emptyset$ (in particular, the set contains k'). Let k be the index of the second-highest bidder in $\boldsymbol{v}^{(\ell)}$ whose bid is above his reserve: $k = \text{argmax}_{t \neq i}\left\{v_t^{(\ell)} : v_t^{(\ell)} \geq r_t\right\}$. Since $k' \in \left\{t : v_t^{(\ell)} \geq r_t, t \neq i\right\}$, it must be that $v_k^{(\ell)} \geq v_{k'}^{(\ell)} \geq w_{k'}^{(\ell)}$. Therefore, $\text{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right) = \max\left\{r_{i'}, w_{k'}^{(\ell)}\right\} = \max\left\{r_i, w_{k'}^{(\ell)}\right\} \leq \max\left\{r_i, v_k^{(\ell)}\right\} = \text{rev}_{\boldsymbol{r}}\left(\boldsymbol{v}^{(\ell)}\right)$.

Case 2: $i = i'$ **and** $\left\{t : w_t^{(\ell)} \geq r_t, t \neq i'\right\} = \emptyset$**.** In this case, $\text{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right) = r_{i'} = r_i \leq \text{rev}_{\boldsymbol{r}}\left(\boldsymbol{v}^{(\ell)}\right)$.

Case 3: $i \neq i'$ **and** $\left\{t : w_t^{(\ell)} \geq r_t, t \neq i'\right\} \neq \emptyset$**.** Since $r_{i'} \leq w_{i'}^{(\ell)} \leq v_{i'}^{(\ell)}$ and $i' \neq i$, we know that $\left\{t : v_t^{(\ell)} \geq r_t, t \neq i\right\} \neq \emptyset$ (in particular, the set contains i'). Let k be the index of the second-highest bidder in $\boldsymbol{v}^{(\ell)}$ whose bid is above his reserve: $k = \text{argmax}_{t \neq i}\left\{v_t^{(\ell)} : v_t^{(\ell)} \geq r_t\right\}$, so $\text{rev}_{\boldsymbol{r}}\left(\boldsymbol{v}^{(\ell)}\right) = \max\left\{r_i, v_k^{(\ell)}\right\}$. Since $i' \in \left\{t : v_t^{(\ell)} \geq r_t, t \neq i\right\}$, it must be that $v_k^{(\ell)} \geq v_{i'}^{(\ell)}$. Moreover, by assumption, $v_{i'}^{(\ell)} \geq w_{i'}^{(\ell)}$. Therefore, $\text{rev}_{\boldsymbol{r}}\left(\boldsymbol{v}^{(\ell)}\right) \geq v_k^{(\ell)} \geq v_{i'}^{(\ell)} \geq w_{i'}^{(\ell)} \geq \text{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right)$, where the final inequality holds because the revenue of a second-price auction is never more than the highest bid.

Case 4: $i \neq i'$ **and** $\left\{t : w_t^{(\ell)} \geq r_t, t \neq i'\right\} = \emptyset$**.** Suppose that $\left\{t : v_t^{(\ell)} \geq r_t, t \neq i\right\} = \emptyset$. Since $i \neq i'$, this means that $r_{i'} > v_{i'}^{(\ell)} \geq w_{i'}^{(\ell)}$, which is a contradiction. Therefore, $\left\{t : v_t^{(\ell)} \geq r_t, t \neq i\right\} \neq \emptyset$. Let k be the index of the second-highest bidder in $\boldsymbol{v}^{(\ell)}$ whose bid is above his reserve: $k = \text{argmax}_{t \neq i}\left\{v_t^{(\ell)} : v_t^{(\ell)} \geq r_t\right\}$, so $\text{rev}_{\boldsymbol{r}}\left(\boldsymbol{v}^{(\ell)}\right) = \max\left\{r_i, v_k^{(\ell)}\right\}$. Since $i' \in \left\{t : v_t^{(\ell)} \geq r_t, t \neq i\right\}$, it must be that $v_k^{(\ell)} \geq v_{i'}^{(\ell)}$. Moreover, by assumption, $v_{i'}^{(\ell)} \geq w_{i'}^{(\ell)}$. Therefore, $\text{rev}_{\boldsymbol{r}}\left(\boldsymbol{v}^{(\ell)}\right) \geq v_k^{(\ell)} \geq v_{i'}^{(\ell)} \geq w_{i'}^{(\ell)} \geq \text{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right)$. □

Theorem 3.6. *The set of multi-item non-anonymous second-price auctions under n additive buyers is p-stable with $p(\epsilon, n, m) = 2m\epsilon$.*

Proof. Let $\mathcal{S} = \left\{\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(L)}\right\} \subset [0,1]^{nm}$ and $\mathcal{S}' = \left\{\boldsymbol{w}^{(1)}, \ldots, \boldsymbol{w}^{(L)}\right\} \subset [0,1]^{nm}$ be two arbitrary sets of valuation vectors such that $\boldsymbol{v}^{(\ell)} - \boldsymbol{\epsilon} \leq \boldsymbol{w}^{(\ell)} \leq \boldsymbol{v}^{(\ell)}$ for all $\ell \in [L]$. Let $\hat{\boldsymbol{r}}'$ be the empirically optimal reserve vector over the set $\mathcal{S}'$ and $\hat{\boldsymbol{r}}$ be the empirically optimal reserve vector over the set $\mathcal{S}$.

For a valuation vector $\boldsymbol{v}^{(\ell)}$, let $\boldsymbol{v}^{(\ell)}(j) \in \mathbb{R}^n$ denote all bidders' bids for item j. Given a reserve vector $\boldsymbol{r} = (\boldsymbol{r}_1, \ldots, \boldsymbol{r}_m) \in [0,1]^{nm}$, the revenue obtained from selling item j is not a function of $\boldsymbol{v}^{(\ell)}(j')$ or $\boldsymbol{r}_{j'}$ for any $j' \neq j$, so we can define $\text{rev}_{\boldsymbol{r}_j}\left(\boldsymbol{v}^{(\ell)}(j)\right)$ to be the revenue obtained from selling item j using the reserves $\boldsymbol{r}_j$. Of course, the overall revenue equals the sum of the revenues obtained from selling each item, so $\text{rev}_{\boldsymbol{r}}\left(\boldsymbol{v}^{(\ell)}\right) = \sum_{j=1}^m \text{rev}_{\boldsymbol{r}_j}\left(\boldsymbol{v}^{(\ell)}(j)\right)$. Using this notation, we can write

$$\sum_{\ell=1}^L \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) = \max_{\boldsymbol{r} \in [0,1]^{nm}} \sum_{\ell=1}^L \sum_{j=1}^m \text{rev}_{\boldsymbol{r}_j}\left(\boldsymbol{v}^{(\ell)}(j)\right) = \max_{\boldsymbol{r} \in [0,1]^{nm}} \sum_{j=1}^m \sum_{\ell=1}^L \text{rev}_{\boldsymbol{r}_j}\left(\boldsymbol{v}^{(\ell)}(j)\right). \quad (4)$$

In Lemma B.1, we prove that we can flip the order of the max and the sum in Equation (4):

$$\sum_{\ell=1}^L \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) = \sum_{j=1}^m \max_{\boldsymbol{r}_j \in [0,1]^n} \sum_{\ell=1}^L \text{rev}_{\boldsymbol{r}_j}\left(\boldsymbol{v}^{(\ell)}(j)\right) \quad (5)$$

and

$$\sum_{\ell=1}^L \text{rev}_{\hat{\boldsymbol{r}}'}\left(\boldsymbol{w}^{(\ell)}\right) = \sum_{j=1}^m \max_{\boldsymbol{r}_j \in [0,1]^n} \sum_{\ell=1}^L \text{rev}_{\boldsymbol{r}_j}\left(\boldsymbol{w}^{(\ell)}(j)\right).$$

Fix an arbitrary item $j \in [m]$. Let $\hat{\boldsymbol{q}}_j \in \mathbb{R}^n$ be the vector of non-anonymous reserves that maximizes average revenue over $\boldsymbol{v}^{(1)}(j), \ldots, \boldsymbol{v}^{(L)}(j)$:

$$\hat{\boldsymbol{q}}_j = \text{argmax}_{\boldsymbol{r}_j \in [0,1]^n} \sum_{\ell=1}^L \text{rev}_{\boldsymbol{r}_j}\left(\boldsymbol{v}^{(\ell)}(j)\right).$$

Similarly, let $\hat{\boldsymbol{q}}'_j \in \mathbb{R}^n$ be the vector of non-anonymous reserves that maximizes average revenue over $\boldsymbol{w}^{(1)}(j), \dots, \boldsymbol{w}^{(L)}(j)$:

$$\hat{\boldsymbol{q}}'_j = \text{argmax}_{\boldsymbol{r}_j \in [0,1]^n} \sum_{\ell=1}^{L} \text{rev}_{\boldsymbol{r}_j}\left(\boldsymbol{w}^{(\ell)}(j)\right).$$

Equation (5) implies that the empirically optimal reserve vector $\hat{\boldsymbol{r}}$ over $\mathcal{S}$ is simply the concatenation of the empirically optimal reserves per item: $\hat{\boldsymbol{r}} = (\hat{\boldsymbol{q}}_1, \dots, \hat{\boldsymbol{q}}_m)$. Similarly, $\hat{\boldsymbol{r}}' = (\hat{\boldsymbol{q}}'_1, \dots, \hat{\boldsymbol{q}}'_m)$. From Theorem 3.3, we know that the revenues of $\hat{\boldsymbol{q}}_j$ and $\hat{\boldsymbol{q}}'_j$ are close on average over $\mathcal{S}$. In other words, $\sum_{\ell=1}^{L} \text{rev}_{\hat{\boldsymbol{q}}_j}\left(\boldsymbol{v}^{(\ell)}(j)\right) - \text{rev}_{\hat{\boldsymbol{q}}'_j}\left(\boldsymbol{v}^{(\ell)}(j)\right) \leq 2L\epsilon$. Therefore,

$$\sum_{\ell=1}^{L} \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) - \text{rev}_{\hat{\boldsymbol{r}}'}\left(\boldsymbol{v}^{(\ell)}\right) = \sum_{j=1}^{m}\sum_{\ell=1}^{L} \text{rev}_{\hat{\boldsymbol{q}}_j}\left(\boldsymbol{v}^{(\ell)}(j)\right) - \text{rev}_{\hat{\boldsymbol{q}}'_j}\left(\boldsymbol{v}^{(\ell)}(j)\right) \leq 2Lm\epsilon,$$

so the lemma statement holds. □

Proposition 3.8. *Fix an arbitrary error term ϵ. For any deterministic algorithm $\mathcal{A}$ that takes as input a training set $\mathcal{S} \subseteq \mathbb{R}^{nm}$ and returns a vector of non-anonymous reserves $\mathcal{A}(\mathcal{S}) \in \mathbb{R}^{nm}$, there exists a distribution $\mathcal{D}$ such that with probability 1 over the draw $\mathcal{S} = \left\{\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(L)}\right\} \sim \mathcal{D}^L$, $\max_{\boldsymbol{r}\in\mathbb{R}^{nm}} \mathbb{E}\left[\text{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right] - \mathbb{E}\left[\text{rev}_{\mathcal{A}(\mathcal{S}')}(\boldsymbol{v})\right] = \Omega(m\epsilon)$ for some noisy training set $\mathcal{S}' = \left\{\boldsymbol{w}^{(1)}, \dots, \boldsymbol{w}^{(L)}\right\}$ such that $\left\|\boldsymbol{v}^{(\ell)} - \boldsymbol{w}^{(\ell)}\right\|_\infty \leq \epsilon$ for all $\ell \in [L]$.*

Proof. Fix a particular algorithm $\mathcal{A}$ and consider the $\mathcal{A}$'s choice of mechanism for a particular item i:

Case 1 There exists constant $\bar{s} > 3\epsilon$ such that for $\mathcal{S}' = \{\bar{s}\mathbf{1}, ..., \bar{s}\mathbf{1}\}$, $\mathcal{A}(\mathcal{S}')_i > \bar{s} - \epsilon$, where $\mathcal{A}(\mathcal{S}')_i$ denotes the mechanism parameter for the ith item. We will argue that there exists a distribution $\mathcal{D}$ such that algorithm $\mathcal{A}$ attains zero revenue in the worst case.

Consider buyer distribution $\mathcal{D}$ where each buyer-item distribution is a point mass distribution at $\bar{s} - \epsilon$. Then, with probability 1, $\mathcal{S} \sim \mathcal{D}$ is such that $\mathcal{S} = \{(\bar{s} - \epsilon)\mathbf{1}, ..., \{(\bar{s} - \epsilon)\mathbf{1}\}$ by definition. And so, for the ith item, since $\mathcal{A}(\mathcal{S}')_i > \bar{s} - \epsilon$:

$$\mathbb{E}_{\boldsymbol{v}\sim\mathcal{D}}[\text{rev}_{\mathcal{A}(\mathcal{S}')}(\boldsymbol{v})] = 0$$

And the constructed $\mathcal{S}'$ satisfies $|\mathcal{S} - \mathcal{S}'| \leq \epsilon$.

To conclude, it remains to note that since $\mathcal{D}$ is a point mass at the same point, the optimal revenue for the ith item is:

$$\max_{M\in\mathcal{M}} \mathbb{E}_{\boldsymbol{v}\sim\mathcal{D}}[\text{rev}_M(\boldsymbol{v})] = \bar{s} - \epsilon$$

And the statement follows from applying the same reasoning for all m items and summing.

Case 2 In the other case, $\mathcal{A}$ is such that for all constant $\bar{s} > 3\epsilon$ and $\mathcal{S}' = \{\bar{s}\mathbf{1}, ..., \bar{s}\mathbf{1}\}$, $\mathcal{A}(\mathcal{S}')_i \leq \bar{s} - \epsilon$.

To prove the statement, we may construct such a $\mathcal{D}$ by again setting it to be the same distribution as in Case 1 where each buyer-item distribution is a point mass at a particular value ρ. If $\mathcal{S} \sim \mathcal{D}$, then $\mathcal{S} = \{\rho\mathbf{1}, ..., \rho\mathbf{1}\}$.

Now, set $\mathcal{S}' = \{(\rho - \epsilon)\mathbf{1}, ..., (\rho - \epsilon)\mathbf{1}\}$, which satisfies $|\mathcal{S}' - \mathcal{S}| \leq \epsilon$. Then, $\mathcal{A}(\mathcal{S}')_i \leq (\rho - \epsilon) - \epsilon$. Hence, for the ith item:

$$\mathbb{E}_{\boldsymbol{v}\sim\mathcal{D}}[\text{rev}_{\mathcal{A}(\hat{S})}(\boldsymbol{v})] \leq \rho - 2\epsilon$$

As argued before, the optimal revenue attainable for the ith item is:

$$\max_{M\in\mathcal{M}} \mathbb{E}_{\boldsymbol{v}\sim\mathcal{D}}[\text{rev}_M(\boldsymbol{v})] = \rho$$

Summing across all m items yield the result.

$\square$

B.3 Additional proofs about lottery mechanisms

Single buyer. We begin by studying lotteries in the case where there is a single unit-demand buyer with values $(v_{11}, \dots, v_{1m}) \in [0,1]^m$ for m items. Lotteries are defined by a price $r_0 \in [0,1]$ and a set of probabilities $r_1, \dots, r_m \in [0,1]$ with $\sum_{j=1}^m r_j = 1$. If the buyer chooses to pay the price r_0, she will receive one item $J \in [m]$, and $\Pr[J = j] = r_j$. Therefore, her expected utility is $\sum_{j=1}^m v_{1j} r_j - r_0$. She will choose to participate in the lottery so long as her expected utility is at least 0. We now prove that this class of mechanisms satisfies our robustness condition (Definition 3.1).

Lemma B.7. *The set of lottery mechanisms with a single unit-demand buyer is (p,q)-robust with $p(\epsilon, n, m) = \epsilon$ and $q(\delta, L, n, m) = O\left(\sqrt{\frac{m \log m}{L}} + \sqrt{\frac{1}{L}\log\frac{1}{\delta}}\right)$.*

Proof. For any $L \geq 1$, let $\mathcal{S} = \left\{\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(L)}\right\} \subset [0,1]^m$ and $\mathcal{S}' = \left\{\boldsymbol{w}^{(1)}, \dots, \boldsymbol{w}^{(L)}\right\} \subset [0,1]^m$ be two arbitrary sets of valuation vectors such that $\boldsymbol{v}^{(\ell)} - \boldsymbol{\epsilon} \leq \boldsymbol{w}^{(\ell)} \leq \boldsymbol{v}^{(\ell)}$ for all $\ell \in [L]$. Let $\hat{\boldsymbol{r}} = (\hat{r}_0, \hat{r}_1, \dots, \hat{r}_m)$ be the empirically optimal parameter vector over the set $\mathcal{S}$ and let $\hat{\boldsymbol{r}}' = (\hat{r}'_0, \hat{r}'_1, \dots, \hat{r}'_m)$ be the empirically optimal parameter vector over the set $\mathcal{S}'$. We prove that $\frac{1}{L}\sum_{\ell=1}^L \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) - \text{rev}_{\hat{\boldsymbol{r}}'}\left(\boldsymbol{v}^{(\ell)}\right) \leq \epsilon$.

This proof relies on two claims. The first states that if we shift the price $\hat{r}_0$ down by ϵ and evaluate the resulting lottery over $\mathcal{S}'$, little revenue is lost. Again, we use the notation $\langle x \rangle = \max\{x, 0\}$.

Claim B.8. *Let $\boldsymbol{r}_\epsilon = (\langle \hat{r}_0 - \epsilon\rangle, \hat{r}_1, \dots, \hat{r}_n)$. If $\text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) > 0$, then $\text{rev}_{\boldsymbol{r}_\epsilon}\left(\boldsymbol{w}^{(\ell)}\right) \geq \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) - \epsilon$.*

Proof of Claim B.8. Fix an arbitrary $\boldsymbol{v}^{(\ell)}$ such that $\text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) > 0$, which means the buyer's utility must be at least 0: $\sum_{j=1}^m v_{1j}^{(\ell)} \hat{r}_j - \hat{r}_0 \geq 0$. As we show, this implies the buyer with values $\boldsymbol{w}^{(\ell)}$ has non-negative utility for the lottery defined by $\boldsymbol{r}_\epsilon$. To see why, first suppose that $\hat{r}_0 < \epsilon$. Since $\text{rev}_{\boldsymbol{r}_\epsilon}\left(\boldsymbol{w}^{(\ell)}\right) \geq 0$, it must be that $\text{rev}_{\boldsymbol{r}_\epsilon}\left(\boldsymbol{w}^{(\ell)}\right) \geq \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) - \epsilon = \hat{r}_0 - \epsilon$. Otherwise, $\hat{r}_0 \geq \epsilon$, so $\langle \hat{r}_0 - \epsilon \rangle = \hat{r}_0 - \epsilon$, which means that

$$
\begin{aligned}
0 \leq \sum_{j=1}^m v_{1j}^{(\ell)} \hat{r}_j - \hat{r}_0 &= \sum_{j=1}^m v_{1j}^{(\ell)} \hat{r}_j - \epsilon - (\hat{r}_0 - \epsilon) = \sum_{j=1}^m v_{1j}^{(\ell)} \hat{r}_j - \sum_{j=1}^m \hat{r}_j \epsilon - (\hat{r}_0 - \epsilon) && \left(\sum_{j=1}^m \hat{r}_j = 1\right) \\
&= \sum_{j=1}^m \left(v_{1j}^{(\ell)} - \epsilon\right)\hat{r}_j - (\hat{r}_0 - \epsilon) \leq \sum_{j=1}^m w_{1j}^{(\ell)} \hat{r}_j - (\hat{r}_0 - \epsilon). && \left(\boldsymbol{v}^{(\ell)} - \boldsymbol{\epsilon} \leq \boldsymbol{w}^{(\ell)}\right)
\end{aligned}
$$

As a result, the buyer with valuations $\boldsymbol{w}^{(\ell)}$ will participate in the lottery defined by $\boldsymbol{r}_\epsilon$ and pay a price of $\hat{r}_0 - \epsilon$. Therefore, $\text{rev}_{\boldsymbol{r}_\epsilon}\left(\boldsymbol{w}^{(\ell)}\right) = \hat{r}_0 - \epsilon = \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) - \epsilon$. $\square$

Let $I \subseteq [L]$ be the set of indices ℓ such that $\text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) > 0$ (for all other indices, $\text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) = 0$). By Claim B.8, for all $\ell \in I$, $\text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) \leq \text{rev}_{\boldsymbol{r}_\epsilon}\left(\boldsymbol{w}^{(\ell)}\right) + \epsilon$. Therefore,

$$
\sum_{\ell=1}^L \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) = \sum_{\ell \in I} \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) \leq \sum_{\ell \in I} \text{rev}_{\boldsymbol{r}_\epsilon}\left(\boldsymbol{w}^{(\ell)}\right) + L\epsilon \leq \sum_{\ell=1}^L \text{rev}_{\boldsymbol{r}_\epsilon}\left(\boldsymbol{w}^{(\ell)}\right) + L\epsilon. \tag{6}
$$

Since $\hat{\boldsymbol{r}}'$ is the empirically optimal reserve under the perturbed training instances $\mathcal{S}'$ (meaning that $\hat{\boldsymbol{r}}' = \text{argmax}_{\boldsymbol{r} \in [0,1]^{m+1}} \sum_{\ell=1}^L \text{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right)$), Equation (6) implies that

$$
\sum_{\ell=1}^L \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) \leq \sum_{\ell=1}^L \text{rev}_{\hat{\boldsymbol{r}}'}\left(\boldsymbol{w}^{(\ell)}\right) + L\epsilon. \tag{7}
$$

Next, we prove that for any parameter vector $\boldsymbol{r}$, revenue under the samples $\boldsymbol{v}^{(\ell)}$ will only be higher than revenue under the samples $\boldsymbol{w}^{(\ell)}$, which intuitively makes sense since $\boldsymbol{w}^{(\ell)} \leq \boldsymbol{v}^{(\ell)}$.

Claim B.9. *For every $\ell \in [L]$ and any parameter vector $\boldsymbol{r} \in [0,1]^{m+1}$, $\text{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right) \leq \text{rev}_{\boldsymbol{r}}\left(\boldsymbol{v}^{(\ell)}\right)$.*

Proof of Claim B.9. Fix an arbitrary index $\ell \in [L]$ such that $\text{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right) > 0$ (if $\text{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right) = 0$, then the claim clearly holds). Since the revenue is non-zero, it must be that the buyer's utility for the lottery is at least 0: $\sum_{j=1}^{m} w_{1j}^{(\ell)} r_j - r_0 \geq 0$. Because $\boldsymbol{v}^{(\ell)} \geq \boldsymbol{w}^{(\ell)}$, it must be that $\sum_{j=1}^{m} v_{1j}^{(\ell)} r_j - r_0 \geq \sum_{j=1}^{m} w_{1j}^{(\ell)} r_j - r_0 \geq 0$, so the buyer with valuations $\boldsymbol{v}^{(\ell)}$ will also participate in the lottery and pay a price of r_0. Therefore, $\text{rev}_r\left(\boldsymbol{w}^{(\ell)}\right) = \text{rev}_{\boldsymbol{r}}\left(\boldsymbol{v}^{(\ell)}\right) = r_0$. $\square$

Finally, Equation (7) and Claim B.9 imply that $\sum_{\ell=1}^{L} \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) \leq L\epsilon + \sum_{\ell=1}^{L} \text{rev}_{\hat{\boldsymbol{r}}'}\left(\boldsymbol{v}^{(\ell)}\right)$. $\square$

B.3.1 Multiple buyers.

Next, we move on to the case where there are n unit-demand buyers.

Theorem 3.9. *The set of lotteries with n unit-demand buyers is p-stable with $p(\epsilon, n, m) = n\epsilon$.*

Proof. Let $\mathcal{S} = \left\{\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(L)}\right\} \subset [0,1]^{nm}$ and $\mathcal{S}' = \left\{\boldsymbol{w}^{(1)}, \ldots, \boldsymbol{w}^{(L)}\right\} \subset [0,1]^{nm}$ be two arbitrary sets of valuation vectors such that $\boldsymbol{v}^{(\ell)} - \boldsymbol{\epsilon} \leq \boldsymbol{w}^{(\ell)} \leq \boldsymbol{v}^{(\ell)}$ for all $\ell \in [L]$. Let $\hat{\boldsymbol{r}}'$ be the empirically optimal reserve vector over the set $\mathcal{S}'$ and $\hat{\boldsymbol{r}}$ be the empirically optimal reserve vector over the set $\mathcal{S}$.

For a valuation vector $\boldsymbol{v}^{(\ell)}$, let $\boldsymbol{v}_i^{(\ell)} \in \mathbb{R}^m$ denote buyer i's values for all m items. Let $\boldsymbol{r} = (\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n) \in \mathbb{R}^{n(m+1)}$ be a lottery parameter vector, where $\boldsymbol{r}_i = (r_{i0}, r_{i1}, \ldots, r_{im})$ is the lottery offered to buyer i. The revenue obtained from buyer i is not a function of $\boldsymbol{v}_{i'}^{(\ell)}$ or $\boldsymbol{r}_{i'}$ for any $i' \neq i$, so we can define $\text{rev}_{\boldsymbol{r}_i}\left(\boldsymbol{v}_i^{(\ell)}\right)$ to be the revenue obtained from buyer i under the lottery defined by $\boldsymbol{r}$. Of course, the overall revenue equals the sum of the revenues obtained from each buyer, so $\text{rev}_{\boldsymbol{r}}\left(\boldsymbol{v}^{(\ell)}\right) = \sum_{i=1}^{n} \text{rev}_{\boldsymbol{r}_i}\left(\boldsymbol{v}_i^{(\ell)}\right)$. Using this notation, we can write

$$\sum_{\ell=1}^{L} \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) = \max_{\boldsymbol{r} \in [0,1]^{n(m+1)}} \sum_{\ell=1}^{L} \sum_{i=1}^{n} \text{rev}_{\boldsymbol{r}_i}\left(\boldsymbol{v}_i^{(\ell)}\right) = \max_{\boldsymbol{r} \in [0,1]^{n(m+1)}} \sum_{i=1}^{n} \sum_{\ell=1}^{L} \text{rev}_{\boldsymbol{r}_i}\left(\boldsymbol{v}_i^{(\ell)}\right). \quad (8)$$

In Lemma B.1 in Appendix B, we prove that we can flip the order of the max and the sum in Equation (8):

$$\sum_{\ell=1}^{L} \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) = \sum_{i=1}^{n} \max_{\boldsymbol{r}_i \in [0,1]^{m+1}} \sum_{\ell=1}^{L} \text{rev}_{\boldsymbol{r}_i}\left(\boldsymbol{v}_i^{(\ell)}\right) \quad (9)$$

and

$$\sum_{\ell=1}^{L} \text{rev}_{\hat{\boldsymbol{r}}'}\left(\boldsymbol{w}^{(\ell)}\right) = \sum_{i=1}^{n} \max_{\boldsymbol{r}_i \in [0,1]^{m+1}} \sum_{\ell=1}^{L} \text{rev}_{\boldsymbol{r}_i}\left(\boldsymbol{w}_i^{(\ell)}\right).$$

Fix an arbitrary bidder $i \in [n]$. Let $\hat{\boldsymbol{q}}_i \in \mathbb{R}^{m+1}$ be the lottery parameter vector that maximizes average revenue over $\boldsymbol{v}_i^{(1)}, \ldots, \boldsymbol{v}_i^{(L)}$:

$$\hat{\boldsymbol{q}}_i = \text{argmax}_{\boldsymbol{r}_i \in [0,1]^{m+1}} \sum_{\ell=1}^{L} \text{rev}_{\boldsymbol{r}_i}\left(\boldsymbol{v}_i^{(\ell)}\right).$$

Similarly, let $\hat{\boldsymbol{q}}_i' \in \mathbb{R}^{m+1}$ be the lottery parameter vector that maximizes average revenue over $\boldsymbol{w}_i^{(1)}, \ldots, \boldsymbol{w}_i^{(L)}$:

$$\hat{\boldsymbol{q}}_i' = \text{argmax}_{\boldsymbol{r}_i \in [0,1]^{m+1}} \sum_{\ell=1}^{L} \text{rev}_{\boldsymbol{r}_i}\left(\boldsymbol{w}_i^{(\ell)}\right).$$

Equation (9) implies that the empirically optimal parameter vector $\hat{\boldsymbol{r}}$ over $\mathcal{S}$ is simply the concatenation of the empirically optimal parameter vectors per bidder: $\hat{\boldsymbol{r}} = (\hat{\boldsymbol{q}}_1, \ldots, \hat{\boldsymbol{q}}_n)$. Similarly,

$\hat{\boldsymbol{r}}' = (\hat{\boldsymbol{q}}'_1, \dots, \hat{\boldsymbol{q}}'_n)$. From Lemma B.7, we know that the revenues of $\hat{\boldsymbol{q}}_i$ and $\hat{\boldsymbol{q}}'_i$ are close on average over $\mathcal{S}$. In other words, $\sum_{\ell=1}^L \text{rev}_{\hat{\boldsymbol{q}}_i}\left(\boldsymbol{v}_i^{(\ell)}\right) - \text{rev}_{\hat{\boldsymbol{q}}'_i}\left(\boldsymbol{v}_i^{(\ell)}\right) \leq L\epsilon$. Therefore,

$$\sum_{\ell=1}^L \text{rev}_{\hat{\boldsymbol{r}}}\left(\boldsymbol{v}^{(\ell)}\right) - \text{rev}_{\hat{\boldsymbol{r}}'}\left(\boldsymbol{v}^{(\ell)}\right) = \sum_{i=1}^n \sum_{\ell=1}^L \text{rev}_{\hat{\boldsymbol{q}}_i}\left(\boldsymbol{v}_i^{(\ell)}\right) - \text{rev}_{\hat{\boldsymbol{q}}'_i}\left(\boldsymbol{v}_i^{(\ell)}\right) \leq Ln\epsilon,$$

so the lemma statement holds. □

Lemma B.10. *For any $L \geq 1$ and $\delta \in (0,1)$, with probability $1-\delta$ over the draw of a set $\mathcal{S} \sim \mathcal{D}^L$, for every lottery parameter vector $\boldsymbol{r} \in \mathbb{R}^{n(m+1)}$,*

$$\left|\frac{1}{L}\sum_{\boldsymbol{v}\in\mathcal{S}} \text{rev}_{\boldsymbol{r}}(\boldsymbol{v}) - \mathop{\mathbb{E}}_{\boldsymbol{v}\sim\mathcal{D}}\left[\text{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right]\right| = O\left(\sqrt{\frac{nm\log(nm)}{L}} + \sqrt{\frac{1}{L}\log\frac{1}{\delta}}\right).$$

Proof. Given a valuation vector $\boldsymbol{v} \in \mathbb{R}^{nm}$, let $\text{rev}_{\boldsymbol{v}}(\boldsymbol{r})$ equal revenue as a function of the lottery parameter vector $\boldsymbol{r} \in \mathbb{R}^{n(m+1)}$. We prove this result by relying on the notion of *delineability*, which was introduced by Balcan et al. [6]. A mechanism class is (d,t)-*delineable* if:

1. It is parameterized by vectors $\boldsymbol{r} \in \mathbb{R}^d$, and
2. For any valuation vector $\boldsymbol{v}$, there is a set H of t hyperplanes in $\mathbb{R}^d$ such that in any connected component C of $\mathbb{R}^d \setminus H$, the function $\text{rev}_{\boldsymbol{v}}(\boldsymbol{r})$ is linear over C.

Balcan et al. [6] prove that for any (d,t)-delineable mechanism class, with probability $1-\delta$ over the draw of a set $\mathcal{S} \sim \mathcal{D}^L$, for every mechanism parameter vector $\boldsymbol{r} \in \mathbb{R}^d$,

$$\left|\frac{1}{L}\sum_{\boldsymbol{v}\in\mathcal{S}} \text{rev}_{\boldsymbol{r}}(\boldsymbol{v}) - \mathop{\mathbb{E}}_{\boldsymbol{v}\sim\mathcal{D}}\left[\text{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right]\right| = O\left(\sqrt{\frac{d\log(dt)}{L}} + \sqrt{\frac{1}{L}\log\frac{1}{\delta}}\right). \tag{10}$$

Clearly, in the case of lottery mechanisms, $d = n(m+1)$. We claim that $t = n$. To see why, for any valuation vector $\boldsymbol{v}$, bidder i will choose to participate in the lottery if and only if $\sum_{j=1}^m v_{ij}r_{ij} - r_{i0} \geq 0$, which is a hyperplane in $\mathbb{R}^{n(m+1)}$. On one side of this hyperplane, the bidder will pay r_{i0}, and on the other side, he will pay nothing. Let H be the set of n such hyperplanes—one per bidder. In any connected component C of $\mathbb{R}^{n(m+1)} \setminus H$, the function $\text{rev}_{\boldsymbol{v}}(\boldsymbol{r})$ is a linear function of the parameters $r_{10}, \dots, r_{n0}$. Therefore, the class is $(n(m+1), n)$-delineable, so the lemma statement follows from Equation (10). □

Proposition 3.11. *Fix an arbitrary error term ϵ. For any deterministic algorithm $\mathcal{A}$ that takes as input a training set $\mathcal{S} \subseteq \mathbb{R}^{nm}$ and returns a vector of lottery parameters $\mathcal{A}(\mathcal{S}) \in \mathbb{R}^{n(m+1)}$, there exists a distribution $\mathcal{D}$ such that with probability 1 over the draw $\mathcal{S} = \left\{\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(L)}\right\} \sim \mathcal{D}^L$, $\max_{\boldsymbol{r}\in\mathbb{R}^{n(m+1)}} \mathbb{E}\left[\text{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right] - \mathbb{E}\left[\text{rev}_{\mathcal{A}(\mathcal{S}')}(\boldsymbol{v})\right] = \Omega(n\epsilon)$ for some noisy training set $\mathcal{S}' = \left\{\boldsymbol{w}^{(1)}, \dots, \boldsymbol{w}^{(L)}\right\}$ such that $\left\|\boldsymbol{v}^{(\ell)} - \boldsymbol{w}^{(\ell)}\right\|_\infty \leq \epsilon$ for all $\ell \in [L]$.*

Proof. We first prove this guarantee for a single buyer and a single item. These mechanisms are defined by a single price r_{10} and probability $r_{11} \in [0,1]$. We will work with a simple distribution where with probability 1, $v_{11} = \rho$ for some $\rho \geq 0$ (v_{11} is the buyer's value for the item).

Under this distribution, the revenue-maximizing lottery sets $r_{10} = \rho$ and $r_{11} = 1$. This is because the buyer will choose to participate in the lottery so long as their expected utility $v_{11}r_{11} = \rho r_{11}$ is at least the price r_{10}. Therefore, the optimal revenue is $\max\left\{r_{10} : \rho r_{11} \geq r_{10}\right\}$, which is maximized when $r_{11} = 1$ and $r_{10} = \rho$.

We now split our analysis into two cases.

Case 1: In the first case, there exists a constant $\bar{s} > 2\epsilon$ such that when $\mathcal{A}$ receives the training set $\mathcal{S}' = \{\bar{s}, \dots, \bar{s}\}$, it returns a lottery $\mathcal{A}(\mathcal{S}') = (r_{10}, r_{11})$ with $(\bar{s} - \epsilon)\, r_{11} < r_{10}$. In this case, we define $\rho = \bar{s} - \epsilon$. Then the lottery that $\mathcal{A}$ returns has revenue 0, but the optimal lottery has revenue $\bar{s} - \epsilon > \epsilon$, so the proposition holds.

Case 2: Otherwise, in the second case, for every constant $\bar{s} > 2\epsilon$, when $\mathcal{A}$ receives the training set $\mathcal{S}' = \{\bar{s}, \dots, \bar{s}\}$, it returns a lottery $\mathcal{A}(\mathcal{S}') = (r_{10}, r_{11})$ with $(\bar{s} - \epsilon) r_{11} \geq r_{10}$. In this case, we define $\rho = \bar{s}$. The optimal lottery has revenue $\bar{s}$ and the revenue of the lottery that $\mathcal{A}$ returns is at most $r_{10} \leq \bar{s} - \epsilon$, so the proposition holds. □

B.4 Additional proofs about item-price mechanisms with unit-demand buyers under dispersion

Lemma B.11. *Suppose there are two items for sale and a single unit-demand buyer with values* $\boldsymbol{v} = (v_{11}, v_{12}) \in [0,1]^2$*. For any* $\epsilon > 0$*, there exists a vector* $\boldsymbol{w} \in [0,1]^2$ *such that* $\boldsymbol{v} - \boldsymbol{\epsilon} \leq \boldsymbol{w} \leq \boldsymbol{v}$ *and* $\max_{\boldsymbol{r}} \text{rev}_{\boldsymbol{r}}(\boldsymbol{v}) - \text{rev}_{\hat{\boldsymbol{r}}'}(\boldsymbol{v}) > |v_{11} - v_{12}|$*, where* $\hat{\boldsymbol{r}}' = \text{argmax}_{\boldsymbol{r}} \text{rev}_{\boldsymbol{r}}(\boldsymbol{w})$.

Since $|v_{11} - v_{12}|$ can be as large as 1, Lemma B.11 implies that we are not able to obtain a non-trivial bound on the stability function p.

Proof of Lemma B.11. Without loss of generality, suppose that $v_{11} \geq v_{12} > 0$. Also, suppose that we break ties in favor of item 1: if $v_{11} - r_{11} = v_{12} - r_{12} \geq 0$, then the buyer will buy item 1. In this case, the optimal set of prices are $\hat{\boldsymbol{r}} = (v_{11}, v_{12})$ and $\text{rev}_{\hat{\boldsymbol{r}}}(\boldsymbol{v}) = v_{11}$. Let $\boldsymbol{w} = (v_{11}, \langle v_{12} - \epsilon \rangle)$. In this case, $(v_{11}, \langle v_{12} - \epsilon \rangle) = \text{argmax}_{\boldsymbol{r}} \text{rev}_{\boldsymbol{r}}(\boldsymbol{w}) = \hat{\boldsymbol{r}}'$. Under this set of prices, the buyer with the values $\boldsymbol{v}$ will choose to buy item 2, so $\text{rev}_{\hat{\boldsymbol{r}}'}(\boldsymbol{v}) < v_{12}$. Therefore, the lemma statement holds. □

Before proving Theorem 3.13, we begin by restating Definition 3.12 more formally using mathematical notation.

Definition B.12 ((ϵ, k)-dispersion [5])**.** Let $\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(L)}$ be L valuation vectors. We say this set of vectors is (ϵ, k)-dispersed if for any price vector $\boldsymbol{r} \in \mathbb{R}^m$, there are at most k samples $\boldsymbol{v}^{(\ell)}$ such that for some buyer $i \in [n]$, either:

1. For some pair of items $j, j' \in [m]$, buyer i's utility for item j is within ϵ of her utility for item j', i.e.,
$$\left| v_{i,j}^{(\ell)} - r_j - \left(v_{i,j'}^{(\ell)} - r_{j'} \right) \right| \leq \epsilon,$$
or
2. Buyer i's utility for some item j is between 0 and ϵ, i.e., $0 \leq v_{i,j}^{(\ell)} \leq \epsilon$.

We begin with the following helpful lemma which allows us to prove Theorem 3.13.

Lemma B.13. *Let* $\mathcal{S} = \left\{ \boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(L)} \right\}$ *and* $\mathcal{S}' = \left\{ \boldsymbol{w}^{(1)}, \dots, \boldsymbol{w}^{(L)} \right\}$ *be two sets of valuation vectors such that for all* $\ell \in [L]$*,* $\left\| \boldsymbol{v}^{(\ell)} - \boldsymbol{w}^{(\ell)} \right\|_\infty \leq \epsilon$ *and all* $\boldsymbol{r} \in \mathbb{R}^m$*,* $\text{rev}_{\boldsymbol{r}}\left(\boldsymbol{v}^{(\ell)}\right)$ *and* $\text{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right)$ *are contained in the interval* $[0,1]$*. Suppose that the set* $\mathcal{S}$ *is* $(2\epsilon, k)$*-dispersed. Then for every parameter vector* $\boldsymbol{r} \in \mathbb{R}^m$*,* $\sum_{\ell=1}^L \left| \text{rev}_{\boldsymbol{r}}\left(\boldsymbol{v}^{(\ell)}\right) - \text{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right) \right| \leq k$.

Proof. For each valuation vector $\boldsymbol{v}^{(\ell)} \in \mathcal{S}$, every pair of items $j, j' \in [m]$, and every bidder $i \in [n]$, let $h_{i,j,j'}^{(\ell)} : \mathbb{R}^m \to \{0,1\}$ be a hyperplane indicator function such that

$$h_{i,j,j'}^{(\ell)}(\boldsymbol{r}) = \begin{cases} 1 & \text{if } v_{i,j}^{(\ell)} - r_j \geq v_{i,j'}^{(\ell)} - r_{j'} \\ 0 & \text{otherwise.} \end{cases}$$

If $h_{i,j,j'}^{(\ell)}(\boldsymbol{r}) = 1$, then buyer i with valuations defined by the vector $\boldsymbol{v}^{(\ell)}$ prefers item j to j' when presented with the prices $\boldsymbol{r}$. Otherwise, when $h_{i,j,j'}^{(\ell)}(\boldsymbol{r}) = 0$, buyer i prefers item j' to j. Similarly, for the set $\mathcal{S}'$, we define the same set of hyperplane indicator functions, which we denote as $\tilde{h}_{i,j,j'}^{(\ell)}$:

$$\tilde{h}_{i,j,j'}^{(\ell)}(\boldsymbol{r}) = \begin{cases} 1 & \text{if } w_{i,j}^{(\ell)} - r_j \geq w_{i,j'}^{(\ell)} - r_{j'} \\ 0 & \text{otherwise.} \end{cases}$$

Fix an arbitrary price vector $\boldsymbol{r} \in \mathbb{R}^m$. For any index $\ell \in [L]$, suppose that for all buyers $i \in [n]$ and item pairs $j, j' \in [m]$, $h_{i,j,j'}^{(\ell)}(\boldsymbol{r}) = \tilde{h}_{i,j,j'}^{(\ell)}(\boldsymbol{r})$. Then every buyer's preference ordering over items is the same under the valuations $\boldsymbol{v}^{(\ell)}$ as it is under the valuations $\boldsymbol{w}^{(\ell)}$, so the items each buyer buys will be the same under both valuation vectors. Therefore, $\text{rev}_{\boldsymbol{r}}\left(\boldsymbol{v}^{(\ell)}\right) = \text{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right)$.

We claim that due to dispersion, for at most k of the indices $\ell \in [L]$, $h^{(\ell)}_{i,j,j'}(\boldsymbol{r}) \neq \tilde{h}^{(\ell)}_{i,j,j'}(\boldsymbol{r})$ for some buyer $i \in [n]$ and item pair $j, j' \in [m]$. To see why, suppose $h^{(\ell)}_{i,j,j'}(\boldsymbol{r}) \neq \tilde{h}^{(\ell)}_{i,j,j'}(\boldsymbol{r})$ for some $\ell \in [n]$ and $j, j' \in [m]$. There are two cases, either $h^{(\ell)}_{i,j,j'}(\boldsymbol{r}) = 0$ or $h^{(\ell)}_{i,j,j'}(\boldsymbol{r}) = 1$:

1. In the first case, $h^{(\ell)}_{i,j,j'}(\boldsymbol{r}) = 0$, so $v^{(\ell)}_{i,j} - r_j < v^{(\ell)}_{i,j'} - r_{j'}$. Since $h^{(\ell)}_{i,j,j'}(\boldsymbol{r}) \neq \tilde{h}^{(\ell)}_{i,j,j'}(\boldsymbol{r})$, it must be that $w^{(\ell)}_{i,j} - r_j \geq w^{(\ell)}_{i,j'} - r_{j'}$. Therefore,

$$v^{(\ell)}_{i,j'} - r_{j'} \leq w^{(\ell)}_{i,j'} + \epsilon - r_{j'} \leq w^{(\ell)}_{i,j} + \epsilon - r_j \leq v^{(\ell)}_{i,j} + 2\epsilon - r_j < v^{(\ell)}_{i,j'} + 2\epsilon - r_{j'}.$$

   Rearranging terms, we have that

$$r_j - r_{j'} - 2\epsilon \leq v^{(\ell)}_{i,j} - v^{(\ell)}_{i,j'} < r_j - r_{j'}. \tag{11}$$

2. In the second case, $h^{(\ell)}_{i,j,j'}(\boldsymbol{r}) = 1$, so $v^{(\ell)}_{i,j} - r_j \geq v^{(\ell)}_{i,j'} - r_{j'}$. Since $h^{(\ell)}_{i,j,j'}(\boldsymbol{r}) \neq \tilde{h}^{(\ell)}_{i,j,j'}(\boldsymbol{r})$, it must be that $w^{(\ell)}_{i,j} - r_j < w^{(\ell)}_{i,j'} - r_{j'}$. Therefore,

$$v^{(\ell)}_{i,j} - r_j \leq w^{(\ell)}_{i,j} + \epsilon - r_j < w^{(\ell)}_{i,j'} + \epsilon - r_{j'} \leq v^{(\ell)}_{i,j'} + 2\epsilon - r_{j'} < v^{(\ell)}_{i,j} + 2\epsilon - r_j.$$

   Rearranging terms, we have that

$$r_j - r_{j'} \leq v^{(\ell)}_{i,j} - v^{(\ell')}_{i,j'} < r_j - r_{j'} + 2\epsilon. \tag{12}$$

From the fact that the set $\mathcal{S}$ is $(2\epsilon, k)$-dispersed, we know that Equations (11) and (12) can hold for at most k of the valuation vectors $\boldsymbol{v}^{(\ell)} \in \mathcal{S}$. Therefore, for at most k indices $\ell \in [L]$, $\text{rev}_{\boldsymbol{r}}\left(\boldsymbol{v}^{(\ell)}\right) \neq \text{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right)$. For these indices $\ell \in [L]$, $\left|\text{rev}_{\boldsymbol{r}}\left(\boldsymbol{v}^{(\ell)}\right) - \text{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right)\right| \leq 1$ because $\text{rev}_{\boldsymbol{r}}\left(\boldsymbol{v}^{(\ell)}\right), \text{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right) \in [0,1]$. Therefore, the lemma statement holds. $\square$

Theorem 3.13. *Let $\mathcal{S} = \left\{\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(L)}\right\} \sim \mathcal{D}^L$ be a set of $(2\epsilon, k)$-dispersed vectors. Let $\mathcal{S}' = \left\{\boldsymbol{w}^{(1)}, \ldots, \boldsymbol{w}^{(L)}\right\} \subset \mathbb{R}^{nm}_{\geq 0}$ be another set such that for all $\ell \in [L]$, $\left\|\boldsymbol{v}^{(\ell)} - \boldsymbol{w}^{(\ell)}\right\|_\infty \leq \epsilon$. Let $\hat{\boldsymbol{r}}'$ be empirically optimal over $\mathcal{S}'$: $\hat{\boldsymbol{r}}' = \operatorname{argmax}_{\boldsymbol{r} \in \mathbb{R}^{nm}} \sum_{\ell=1}^{L} \text{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right)$. With probability $1-\delta$ over the draw of $\mathcal{S}$, $\max_{\boldsymbol{r} \in \mathbb{R}^{nm}} \mathbb{E}_{\boldsymbol{v} \sim \mathcal{D}}\left[\text{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right] - \mathbb{E}\left[\text{rev}_{\hat{\boldsymbol{r}}'}(\boldsymbol{v})\right] = O\left(\frac{k}{L} + \sqrt{\frac{1}{L}\left(nm\log(nm) + \log\frac{1}{\delta}\right)}\right)$.*

Proof. From prior research by Morgenstern and Roughgarden [46], we know that with probability $1-\delta$, for all price vectors $\boldsymbol{r} \in \mathbb{R}^{nm}$, the average revenue of $\boldsymbol{r}$ over $\mathcal{S}$ is close to its expected revenue:

$$\left|\frac{1}{L}\sum_{\ell=1}^{L} \text{rev}_{\boldsymbol{r}}\left(\boldsymbol{v}^{(\ell)}\right) - \mathop{\mathbb{E}}_{\boldsymbol{v} \sim \mathcal{D}}\left[\text{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right]\right| = O\left(\sqrt{\frac{1}{L}\left(nm\log(nm) + \log\frac{1}{\delta}\right)}\right). \tag{13}$$

Let $\boldsymbol{r}^* = \operatorname{argmax}_{\boldsymbol{r}\in\mathbb{R}^m} \mathbb{E}_{\boldsymbol{v}\sim\mathcal{D}}\left[\operatorname{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right]$. Then

$$\begin{aligned}
&\max_{\boldsymbol{r}\in\mathbb{R}^m} \mathbb{E}_{\boldsymbol{v}\sim\mathcal{D}}\left[\operatorname{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right] - \mathbb{E}_{\boldsymbol{v}\sim\mathcal{D}}\left[\operatorname{rev}_{\hat{\boldsymbol{r}}'}(\boldsymbol{v})\right]\\
=\ & \mathbb{E}_{\boldsymbol{v}\sim\mathcal{D}}\left[\operatorname{rev}_{\boldsymbol{r}^*}(\boldsymbol{v})\right] - \frac{1}{L}\sum_{\ell=1}^L \operatorname{rev}_{\boldsymbol{r}^*}\left(\boldsymbol{v}^{(\ell)}\right) + \frac{1}{L}\sum_{\ell=1}^L \operatorname{rev}_{\boldsymbol{r}^*}\left(\boldsymbol{v}^{(\ell)}\right) - \mathbb{E}_{\boldsymbol{v}\sim\mathcal{D}}\left[\operatorname{rev}_{\hat{\boldsymbol{r}}'}(\boldsymbol{v})\right]\\
\leq\ & \frac{1}{L}\sum_{\ell=1}^L \operatorname{rev}_{\boldsymbol{r}^*}\left(\boldsymbol{v}^{(\ell)}\right) - \mathbb{E}_{\boldsymbol{v}\sim\mathcal{D}}\left[\operatorname{rev}_{\hat{\boldsymbol{r}}'}(\boldsymbol{v})\right] + O\left(\sqrt{\frac{1}{L}\left(nm\log(nm) + \log\frac{1}{\delta}\right)}\right) && (14)\\
\leq\ & \frac{1}{L}\sum_{\ell=1}^L \operatorname{rev}_{\boldsymbol{r}^*}\left(\boldsymbol{w}^{(\ell)}\right) + \frac{k}{L} - \mathbb{E}_{\boldsymbol{v}\sim\mathcal{D}}\left[\operatorname{rev}_{\hat{\boldsymbol{r}}'}(\boldsymbol{v})\right] + O\left(\sqrt{\frac{1}{L}\left(nm\log(nm) + \log\frac{1}{\delta}\right)}\right) && (15)\\
\leq\ & \frac{1}{L}\sum_{\ell=1}^L \operatorname{rev}_{\hat{\boldsymbol{r}}'}\left(\boldsymbol{w}^{(\ell)}\right) + \frac{k}{L} - \mathbb{E}_{\boldsymbol{v}\sim\mathcal{D}}\left[\operatorname{rev}_{\hat{\boldsymbol{r}}'}(\boldsymbol{v})\right] + O\left(\sqrt{\frac{1}{L}\left(nm\log(nm) + \log\frac{1}{\delta}\right)}\right) && (16)\\
\leq\ & \frac{1}{L}\sum_{\ell=1}^L \operatorname{rev}_{\hat{\boldsymbol{r}}'}\left(\boldsymbol{v}^{(\ell)}\right) + \frac{2k}{L} - \mathbb{E}_{\boldsymbol{v}\sim\mathcal{D}}\left[\operatorname{rev}_{\hat{\boldsymbol{r}}'}(\boldsymbol{v})\right] + O\left(\sqrt{\frac{1}{L}\left(nm\log(nm) + \log\frac{1}{\delta}\right)}\right) && (17)\\
\leq\ & \frac{2k}{L} + O\left(\sqrt{\frac{1}{L}\left(nm\log(nm) + \log\frac{1}{\delta}\right)}\right), && (18)
\end{aligned}$$

where Equations (14) and (18) follow from Equation (13), Equations (15) and (17) follow from Lemma B.13, and Equation (16) follows from the fact that $\hat{\boldsymbol{r}}' = \operatorname{argmax} \sum_{\ell=1}^L \operatorname{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right)$. □

When does dispersion hold? One example, also observed in prior research [5, 7, 34], is when the distribution over buyers' values is relatively "smooth." More formally, for any distribution over $[0,1]$ with probability density function $f : [0,1] \to \mathbb{R}_{\geq 0}$, we say the density function is *κ-bounded* if $\max_{x\in[0,1]} f(x) \leq \kappa$.

Lemma 3.15. *Suppose that for every buyer $i \in [n]$ and every pair of items $j, j' \in [m]$, buyer i's values for items j and j' have a κ-bounded joint density function. Then for any $\epsilon > 0$, with probability $1-\delta$ over the draw $\mathcal{S} = \left\{\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(L)}\right\} \sim \mathcal{D}^L$, the set $\mathcal{S}$ is $(2\epsilon, k)$-dispersed with $k = 4Lnm^2\kappa\epsilon + O\left(nm^2\sqrt{L\log\frac{nm}{\delta}}\right)$.*

Proof. Fix a buyer $i \in [n]$ and pair of items $j, j' \in [m]$. For any price vector $\boldsymbol{r} \in \mathbb{R}^m$, we want to bound the number of samples $\boldsymbol{v}^{(\ell)}$ such that buyer i's utility for item j is within 2ϵ of her utility for item j': $\left|v_{i,j}^{(\ell)} - r_j - \left(v_{i,j'}^{(\ell)} - r_{j'}\right)\right| \leq 2\epsilon$. Since buyer i's values for items j and j' have a κ-bounded joint density function, we know that $\Pr\left[\left|v_{i,j}^{(\ell)} - r_j - \left(v_{i,j'}^{(\ell)} - r_{j'}\right)\right| \leq 2\epsilon\right] \leq 4\kappa\epsilon$. Therefore,

$$\mathbb{E}\left[\left|\left\{\ell : \left|v_{i,j}^{(\ell)} - r_j - \left(v_{i,j'}^{(\ell)} - r_{j'}\right)\right| \leq 2\epsilon\right\}\right|\right] \leq 4N\kappa\epsilon.$$

This is because since $v_{i,j}$ and $v_{i,j'}$ are both random variables in $[0,1]$ with a κ-bounded joint density function, the difference $v_{i,j} - v_{i,j'}$ also has a κ-bounded density function [5].

Next, we can write

$$\mathbb{E}\left[\left|\left\{\ell : \left|v_{i,j}^{(\ell)} - r_j - \left(v_{i,j'}^{(\ell)} - r_{j'}\right)\right| \leq 2\epsilon\right\}\right|\right] = \mathbb{E}\left[\sum_{\ell=1}^L \mathbf{1}_{\left\{\left|v_{i,j}^{(\ell)} - r_j - \left(v_{i,j'}^{(\ell)} - r_{j'}\right)\right| \leq 2\epsilon\right\}}\right].$$

Since the VC dimension on intervals is 2, we know that with probability $1 - \frac{\delta}{nm^2}$, for any price vector $\boldsymbol{r} \in \mathbb{R}^m$,

$$\begin{aligned}\left|\left\{\ell : \left|v_{i,j}^{(\ell)} - r_j - \left(v_{i,j'}^{(\ell)} - r_{j'}\right)\right| \leq 2\epsilon\right\}\right| &= \sum_{\ell=1}^{L} \mathbf{1}_{\left\{\left|v_{i,j}^{(\ell)} - r_j - \left(v_{i,j'}^{(\ell)} - r_{j'}\right)\right| \leq 2\epsilon\right\}} \\ &\leq \mathbb{E}\left[\sum_{\ell=1}^{L} \mathbf{1}_{\left\{\left|v_{i,j}^{(\ell)} - r_j - \left(v_{i,j'}^{(\ell)} - r_{j'}\right)\right| \leq 2\epsilon\right\}}\right] + O\left(\sqrt{L \log \frac{nm}{\delta}}\right) \\ &\leq 4L\kappa\epsilon + O\left(\sqrt{L \log \frac{nm}{\delta}}\right). \qquad (19)\end{aligned}$$

Applying a union bound, we have show that with probability $1 - \delta$, for any price vector $\boldsymbol{r} \in \mathbb{R}^m$, any buyer $i \in [n]$, and any pair of items $j, j' \in [m]$ there are at most $4L\kappa\epsilon + O\left(\sqrt{L \log \frac{nm}{\delta}}\right)$ samples $\boldsymbol{v}^{(\ell)}$ such that buyer i's utility for item j is within 2ϵ of her utility for item j', i.e.:

$$\left|\left\{\ell : \left|v_{i,j}^{(\ell)} - r_j - \left(v_{i,j'}^{(\ell)} - r_{j'}\right)\right| \leq 2\epsilon\right\}\right| \leq 4L\kappa\epsilon + O\left(\sqrt{L \log \frac{nm}{\delta}}\right).$$

If we union these sets over all n buyers and all m^2 pairs of items, we have that with probability $1 - \delta$, for all price vector $\boldsymbol{r} \in \mathbb{R}^m$,

$$\left|\left\{\ell : \left|v_{i,j}^{(\ell)} - r_j - \left(v_{i,j'}^{(\ell)} - r_{j'}\right)\right| \leq \epsilon \text{ for some } i \in [n] \text{ and } j, j' \in [m]\right\}\right| = \tilde{O}\left(Lnm^2\kappa\epsilon\right). \qquad (20)$$

□

Corollary B.14. *Suppose that for every buyer $i \in [n]$ and every pair of items $j, j' \in [m]$, buyer i's values for items j and j' have a κ-bounded joint density function. For any $\delta \in (0, 1)$, let $\mathcal{S} = \left\{\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(L)}\right\} \sim \mathcal{D}^L$ be a set of L valuation vectors and let $\mathcal{S}' = \left\{\boldsymbol{w}^{(1)}, \ldots, \boldsymbol{w}^{(L)}\right\}$ be another set of vectors such that for all $\ell \in [L]$, $\left\|\boldsymbol{v}^{(\ell)} - \boldsymbol{w}^{(\ell)}\right\|_\infty \leq \epsilon$. Finally, let $\hat{\boldsymbol{r}} = \operatorname{argmax}_{\boldsymbol{r} \in \mathbb{R}^m} \sum_{\ell=1}^{L} \operatorname{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell)}\right)$. With probability $1 - \delta$,*

$$\max_{\boldsymbol{r} \in \mathbb{R}^m} \mathop{\mathbb{E}}_{\boldsymbol{v} \sim \mathcal{D}}\left[\operatorname{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right] - \mathop{\mathbb{E}}_{\boldsymbol{v} \sim \mathcal{D}}\left[\operatorname{rev}_{\hat{\boldsymbol{r}}}(\boldsymbol{v})\right] \leq O\left(nm^2\kappa\epsilon + nm^2\sqrt{\frac{1}{L} \log \frac{nm}{\delta}}\right).$$

C Additional results about multi-task mechanism design (Section 4)

Our multi-task learning approach is inspired by the following two observations, Observations 1 and 2.

Observation 1 (known vectors $\boldsymbol{b}^{(t)}$ and $\boldsymbol{z}^{(\ell,t)}$)). Suppose that rather than only being able to observe the valuation vectors $\boldsymbol{v}^{(\ell,t)}$, we were also able to observe the vectors $\boldsymbol{b}^{(t)}$ and $\boldsymbol{z}^{(\ell,t)}$, where $\boldsymbol{v}^{(\ell,t)} = \boldsymbol{b}^{(t)} + \boldsymbol{z}^{(\ell,t)}$. This would allow us to generate a total of LT samples from each distribution $\mathcal{D}^{(t)}$, namely, $\left\{\boldsymbol{b}^{(t)} + \boldsymbol{z}^{(\ell,\tau)} : \ell \in [L], \tau \in [T]\right\}$. If the mechanism class we optimize over is (p, q)-robust, we would be able to provide strong learnability guarantees, as we summarize below.

Theorem C.1. *Fix a (p, q)-robust mechanism class. With probability $1 - \delta$ over the draw of LT vectors $\left\{\boldsymbol{z}^{(\ell,\tau)} : \ell \in [L], \tau \in [T]\right\} \sim \mathcal{D}^{LT}$, for every task $t \in [T]$ and every parameter vector $\boldsymbol{r}$,*

$$\left|\frac{1}{LT} \sum_{\tau=1}^{T} \sum_{\ell=1}^{L} \operatorname{rev}_{\boldsymbol{r}}\left(\boldsymbol{b}^{(t)} + \boldsymbol{z}^{(\ell,\tau)}\right) - \mathop{\mathbb{E}}_{\boldsymbol{v} \sim \mathcal{D}^{(t)}}\left[\operatorname{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right]\right| \leq q\left(\frac{\delta}{T}, LT, n, m\right).$$

Before proving Theorem C.1, we highlight the following implication: if $\boldsymbol{r}^{(t)} \in \mathbb{R}^d$ maximizes average revenue over the LT samples from $\mathcal{D}^{(t)}$, then the expected revenue of $\boldsymbol{r}^{(t)}$ over $\mathcal{D}^{(t)}$ is nearly optimal. Formally, if $\boldsymbol{r}^{(t)} = \operatorname{argmax}_{\boldsymbol{r} \in \mathbb{R}^d} \sum_{\tau=1}^{T} \sum_{\ell=1}^{L} \operatorname{rev}_{\boldsymbol{r}}\left(\boldsymbol{b}^{(t)} + \boldsymbol{z}^{(\ell,\tau)}\right)$, then with probability $1 - \delta$, for every task t, $\max_{\boldsymbol{r} \in \mathbb{R}^d} \mathbb{E}_{\boldsymbol{v} \sim \mathcal{D}^{(t)}}\left[\operatorname{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right] - \mathbb{E}_{\boldsymbol{v} \sim \mathcal{D}^{(t)}}\left[\operatorname{rev}_{\boldsymbol{r}^{(t)}}(\boldsymbol{v})\right] \leq 2 \cdot q\left(\frac{\delta}{T}, LT, n, m\right)$. For example, in the case of multi-item non-anonymous second-price auctions, Corollary 3.7 implies that

$$\max_{\boldsymbol{r} \in \mathbb{R}^d} \mathop{\mathbb{E}}_{\boldsymbol{v} \sim \mathcal{D}^{(t)}}\left[\operatorname{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right] - \mathop{\mathbb{E}}_{\boldsymbol{v} \sim \mathcal{D}^{(t)}}\left[\operatorname{rev}_{\boldsymbol{r}^{(t)}}(\boldsymbol{v})\right] = O\left(\sqrt{\frac{nm \log(nm)}{LT}} + \sqrt{\frac{1}{LT} \log \frac{T}{\delta}}\right). \qquad (21)$$

Meanwhile, if we did not bootstrap the $L(T-1)$ additional samples per task and only used the L samples $\boldsymbol{v}^{(1,t)}, \dots, \boldsymbol{v}^{(L,t)}$ to select a reserve vector $\boldsymbol{r}^{(t)}$, then T would not appear in the denominator of Equation (21), so it would be worse by a multiplicative factor of $\sqrt{T}$. (The same would be true in the case of lottery mechanisms, as is evident from our bound on q from Theorem 3.9.)

Proof of Theorem C.1. Fix a task t. By definition, $\left\{\boldsymbol{b}^{(t)} + \boldsymbol{z}^{(\ell,\tau)} : \ell \in [L], \tau \in [T]\right\}$ are samples from the distribution $\mathcal{D}^{(t)}$. Since the mechanism class is (p,q)-robust, we know that with probability $1 - \frac{\delta}{T}$ over the draw of the vectors $\left\{\boldsymbol{z}^{(\ell,\tau)} : \ell \in [L], \tau \in [T]\right\} \sim \mathcal{D}^{LT}$, for every parameter vector $\boldsymbol{r}$,

$$\left|\frac{1}{LT}\sum_{\tau=1}^{T}\sum_{\ell=1}^{L} \operatorname{rev}_{\boldsymbol{r}}\left(\boldsymbol{b}^{(t)} + \boldsymbol{z}^{(\ell,\tau)}\right) - \mathop{\mathbb{E}}_{\boldsymbol{v}\sim\mathcal{D}^{(t)}}\left[\operatorname{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right]\right| \leq q\left(\frac{\delta}{T}, LT, n, m\right).$$

The theorem statement therefore holds by a union bound over the T tasks. □

Observation 2 (known vectors $\boldsymbol{b}^{(t)} - \boldsymbol{b}^{(\tau)}$)). Observation 1 applied to the hypothetical scenario where we knew all vectors $\boldsymbol{b}^{(t)}$ and $\boldsymbol{z}^{(\ell,t)}$. Instead, suppose we knew only the differences $\boldsymbol{b}^{(t)} - \boldsymbol{b}^{(\tau)}$ for every pair of tasks $t, \tau \in [T]$, along with the true training set $\left\{\boldsymbol{v}^{(\ell,t)} : \ell \in [L], t \in [T]\right\}$. In this case, we could recover the vectors $\boldsymbol{b}^{(t)} + \boldsymbol{z}^{(\ell,\tau)}$ for every index $\ell \in [L]$ and pair of tasks $t, \tau \in [T]$, which is all we need to apply Theorem C.1. After all,

$$\boldsymbol{b}^{(t)} + \boldsymbol{z}^{(\ell,\tau)} = \boldsymbol{b}^{(t)} - \boldsymbol{b}^{(\tau)} + \boldsymbol{b}^{(\tau)} + \boldsymbol{z}^{(\ell,\tau)} = \boldsymbol{b}^{(t)} - \boldsymbol{b}^{(\tau)} + \boldsymbol{v}^{(\ell,\tau)}. \tag{22}$$

In reality, we do not know the differences $\boldsymbol{b}^{(t)} - \boldsymbol{b}^{(\tau)}$, but we can estimate them from data. We then use these estimates to bootstrap training instances as in Observation 2. We may not be able to estimate these differences exactly—there may be additive noise in our estimates—but from our results from the previous section (in particular, Fact 3.2), we know that this noise is not problematic in our final revenue guarantees so long as its magnitude is not too large.

We provide two results in this vein. In Appendix C.1, we show that the differences $\boldsymbol{b}^{(t)} - \boldsymbol{b}^{(\tau)}$ can be estimated approximately, and we show exactly how the resulting noise appears in our learning guarantees. For the mechanism classes we studied in the previous section, our per-task sample complexity guarantees are significantly better than the best-known single-task sample complexity bounds. In Appendix C.2, we show that in some cases, the differences $\boldsymbol{b}^{(t)} - \boldsymbol{b}^{(\tau)}$ can be estimated exactly, in which case our learning guarantees follow directly from Theorem C.1.

C.1 Approximate estimation

For each pair of tasks t and τ and bidder i, we define the following natural estimate of $b_i^{(t)} - b_i^{(\tau)}$:

$$\hat{b}_i^{(t,\tau)} = \frac{1}{Lm}\sum_{\ell=1}^{L}\sum_{j=1}^{m} v_{ij}^{(\ell,t)} - v_{ij}^{(\ell,\tau)},$$

where $v_{ij}^{(\ell,t)}$ is bidder i's value for item j under the ℓ^{th} sample $\boldsymbol{v}^{(\ell,t)}$ of task t. This estimate makes sense because in expectation, each summand equals $b_i^{(t)} - b_i^{(\tau)}$. A Hoeffding bound implies that this estimate's error is small, as we prove in the following lemma.

Lemma C.2. *With probability $1 - \delta$ over the draw of the LT samples $\left\{\boldsymbol{v}^{(\ell,t)} : \ell \in [L], t \in [T]\right\}$, for all bidders $i \in [n]$ and pairs of tasks $t, \tau \in [T]$, $\left|\hat{b}_i^{(t,\tau)} - \left(b_i^{(t)} - b_i^{(\tau)}\right)\right| \leq \sqrt{\frac{1}{2Lm}\log\frac{2nT^2}{\delta}}$.*

Proof. By definition, $v_{ij}^{(\ell,t)} - v_{ij}^{(\ell,\tau)} = b_i^{(t)} + z_{ij}^{(\ell,t)} - \left(b_i^{(\tau)} + z_{ij}^{(\ell,\tau)}\right)$, where $z_{ij}^{(\ell,t)}$ and $z_{ij}^{(\ell,\tau)}$ are both i.i.d. draws from the distribution $\mathcal{D}_{ij}$. Therefore, $\mathbb{E}\left[v_{ij}^{(\ell,t)} - v_{ij}^{(\ell,\tau)}\right] = b_i^{(t)} - b_i^{(\tau)}$, so the lemma follows from a Hoeffding bound and a union bound over all n bidders and $\binom{T}{2}$ pairs of tasks. □

We now show how to combine this estimate with Fact 3.2 and Theorem C.1 to provide strong multi-task learning guarantees. We use the notation $\hat{\boldsymbol{b}}^{(t,\tau)} = \left(\hat{b}_1^{(t,\tau)}, \dots, \hat{b}_n^{(t,\tau)}\right)$ and $\hat{\boldsymbol{b}}^{(t,\tau)} + \boldsymbol{v}^{(\ell,\tau)}$ to denote the vector in $\mathbb{R}^{nm}$ whose components equal $\hat{b}_i^{(t,\tau)} + v_{ij}^{(\ell,\tau)}$ for all bidders $i \in [n]$ and items $j \in [m]$. Fix an arbitrary task $t \in [T]$. Let $\mathcal{S}_t = \left\{\boldsymbol{b}^{(t)} + \boldsymbol{z}^{(\ell,\tau)} : \ell \in [L], \tau \in [T]\right\}$. This is a set of LT samples drawn from the distribution $\mathcal{D}^{(t)}$. We do not have access to this set of samples, but we do have access to a noisy version $\hat{\boldsymbol{b}}^{(t,\tau)} + \boldsymbol{v}^{(\ell,\tau)}$ of each sample $\boldsymbol{b}^{(t)} + \boldsymbol{z}^{(\ell,\tau)}$. From Lemma C.2 and Equation (22), we know that with probability $1 - \delta$, for all pairs of tasks $t, \tau \in [T]$ and all $\ell \in [L]$, our estimate $\hat{\boldsymbol{b}}^{(t,\tau)} + \boldsymbol{v}^{(\ell,\tau)}$ is close to the true sample $\boldsymbol{b}^{(t)} + \boldsymbol{z}^{(\ell,\tau)}$: $\left\|\boldsymbol{b}^{(t)} + \boldsymbol{z}^{(\ell,\tau)} - \left(\hat{\boldsymbol{b}}^{(t,\tau)} + \boldsymbol{v}^{(\ell,\tau)}\right)\right\|_\infty \leq \sqrt{\frac{1}{2Lm}\log\frac{2nT^2}{\delta}}$. To apply Fact 3.2, we must underestimate the buyers' values. Therefore, we define our noisy samples as $\mathcal{S}_t' = \left\{\boldsymbol{w}^{(\ell,t,\tau)} : \ell \in [L], \tau \in [T]\right\}$ where for each bidder i, each item j, and each $\ell \in [L]$, $w_{ij}^{(\ell,t,\tau)} = \hat{b}_i^{(t,\tau)} + v_{ij}^{(\ell,\tau)} - \sqrt{\frac{1}{2Lm}\log\frac{2nT^2}{\delta}}$. Lemma C.2 and Equation (22) imply the following corollary.

Corollary C.3. *Let $\epsilon = \sqrt{\frac{2}{Lm}\log\frac{2nT^2}{\delta}}$. With probability $1 - \delta$ over the draw of the LT samples $\left\{\boldsymbol{v}^{(\ell,t)} : \ell \in [L], t \in [T]\right\}$, for every bidder $i \in [n]$, every item $j \in [m]$, every pair of tasks $t, \tau \in [T]$, and every index $\ell \in [L]$, $\boldsymbol{b}^{(t)} + \boldsymbol{z}^{(\ell,\tau)} - \boldsymbol{\epsilon} \leq \boldsymbol{w}^{(\ell,t,\tau)} \leq \boldsymbol{b}^{(t)} + \boldsymbol{z}^{(\ell,\tau)}$.*

We now provide strong multi-task learning guarantees for lotteries and second-price auctions.

C.1.1 Lotteries

Corollaries 3.10 and C.3 together with Theorem C.1 imply that optimizing the reserves over the set of perturbed training instances $\mathcal{S}_t'$ results in a nearly optimal lottery.

Theorem 4.2. *For each task t, let $\hat{\boldsymbol{r}}_t'$ be the empirically optimal lottery parameter vector over the set $\mathcal{S}_t'$: $\hat{\boldsymbol{r}}_t' = \operatorname{argmax}_{\boldsymbol{r}\in\mathbb{R}^{n(m+1)}} \sum_{\ell=1}^L \sum_{\tau=1}^T \operatorname{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell,t,\tau)}\right)$. With probability $1 - \delta$, for every $t \in [T]$, $\max_{\boldsymbol{r}\in\mathbb{R}^{n(m+1)}} \mathbb{E}_{\boldsymbol{v}\sim\mathcal{D}^{(t)}}\left[\operatorname{rev}_{\boldsymbol{r}}(\boldsymbol{v}) - \operatorname{rev}_{\hat{\boldsymbol{r}}_t'}(\boldsymbol{v})\right] = \tilde{O}\left(\sqrt{\frac{1}{LT}\left(nm + \log\frac{T}{\delta}\right) + \frac{n^2}{Lm}\log\frac{nT}{\delta}}\right)$.*

Proof. This follows from Corollary 3.10 with $\epsilon = O\left(\sqrt{\frac{1}{Lm}\log\frac{nT}{\delta}}\right)$, as per Corollary C.3. □

C.1.2 Multi-item second-price auctions with non-anonymous reserves

Corollaries 3.7 and C.3 together with Theorem C.1 imply the following guarantee for second-price auctions with non-anonymous reserves.

Theorem 4.1. *For each task $t \in [T]$, let $\hat{\boldsymbol{r}}_t'$ be the empirically optimal reserve vector over the set $\mathcal{S}_t'$: $\hat{\boldsymbol{r}}_t' = \operatorname{argmax}_{\boldsymbol{r}\in\mathbb{R}^{nm}} \sum_{\ell=1}^L \sum_{\tau=1}^T \operatorname{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell,t,\tau)}\right)$. With probability $1 - \delta$, for every task $t \in [T]$, $\max_{\boldsymbol{r}\in\mathbb{R}^{nm}} \mathbb{E}_{\boldsymbol{v}\sim\mathcal{D}^{(t)}}\left[\operatorname{rev}_{\boldsymbol{r}}(\boldsymbol{v}) - \operatorname{rev}_{\hat{\boldsymbol{r}}_t'}(\boldsymbol{v})\right] = \tilde{O}\left(\sqrt{\frac{1}{LT}\left(nm + \log\frac{T}{\delta}\right) + \frac{m}{L}\log\frac{nT}{\delta}}\right)$.*

Proof. This result follows from Corollary 3.7 by the same logic as Theorem 4.2. □

C.2 Exact estimation

We now show that in some settings, we can, in fact, estimate the differences $\boldsymbol{b}^{(t)} - \boldsymbol{b}^{(\tau)}$ exactly, thus reducing the error in our revenue guarantees. This exact estimation is possible when there exists a quantal $\kappa \in \mathbb{R}$ such that for every bidder i and every pair of tasks $t, \tau \in [T]$, $b_i^{(t)} - b_i^{(\tau)} = s\kappa$ for some $s \in \mathbb{Z}$. In this case, we use the following estimate $\hat{b}_i^{(t,\tau)}$ for the difference $b_i^{(t)} - b_i^{(\tau)}$:

$$\hat{b}_i^{(t,\tau)} = \left[\frac{1}{Lm}\sum_{\ell=1}^L\sum_{j=1}^m v_{ij}^{(\ell,t)} - v_{ij}^{(\ell,\tau)}\right]_\kappa,$$

where the notation $[x]_\kappa$ denotes the value x rounded to the nearest multiple of κ. We prove that this estimate is exact so long as the number of samples L per task is sufficiently large.

Lemma C.4. *Suppose that $L = \Omega\left(\frac{1}{\kappa^2 m}\ln\frac{Tn}{\delta}\right)$. With probability $1-\delta$ over the draw of the LT samples $\left\{\boldsymbol{v}^{(\ell,t)} : \ell \in [L], t \in [T]\right\}$, for every bidder i and pair of tasks $t, \tau \in [T]$, $\hat{b}_i^{(t,\tau)} = b_i^{(t)} - b_i^{(\tau)}$.*

Proof. From Lemma C.2, we know with probability $1-\delta$, for every bidder i and pair $t, \tau \in [T]$,

$$\left|\frac{1}{Lm}\sum_{\ell=1}^{L}\sum_{j=1}^{m} v_{ij}^{(\ell,t)} - v_{ij}^{(\ell,\tau)} - \left(b_i^{(t)} - b_i^{(\tau)}\right)\right| \leq \sqrt{\frac{1}{2Lm}\ln\frac{2T^2 n}{\delta}}. \tag{23}$$

Therefore, when $L = \Omega\left(\frac{1}{\kappa^2 m}\ln\frac{Tn}{\delta}\right)$, the right-hand-side of Equation (23) is at most $\frac{\kappa}{2}$. As a result, when we round $\frac{1}{Lm}\sum_{\ell=1}^{L}\sum_{j=1}^{m} v_{ij}^{(\ell,t)} - v_{ij}^{(\ell,\tau)}$ to the nearest multiple of κ in order to formulate the estimate $\hat{b}_i^{(t,\tau)}$, we round it to exactly $b_i^{(t)} - b_i^{(\tau)}$, so the theorem statement holds. □

From Lemma C.4 and Equation (22), we know that with probability $1-\delta$, for every pair of tasks $t, \tau \in [T]$ and every index $\ell \in [L]$, $\hat{\boldsymbol{b}}^{(t,\tau)} + \boldsymbol{v}^{(\ell,\tau)} = \boldsymbol{b}^{(t)} + \boldsymbol{z}^{(\ell,\tau)}$, which is a sample from the distribution $\mathcal{D}^{(t)}$. As in Appendix C.1, we use the notation $\boldsymbol{w}^{(\ell,t,\tau)} = \hat{\boldsymbol{b}}^{(t,\tau)} + \boldsymbol{v}^{(\ell,\tau)}$ and $\mathcal{S}_t' = \left\{\boldsymbol{w}^{(\ell,t,\tau)} : \ell \in [L], \tau \in [T]\right\}$. We now provide learning guarantees similar to those in Appendix C.1. In the case of lottery mechanisms, Corollary 3.10, Theorem C.1, and Lemma C.4 imply that the empirically optimal lottery over the bootstrapped samples $\mathcal{S}_t'$ is nearly optimal overall.

Theorem C.5. *Suppose that $L = \Omega\left(\frac{1}{\kappa^2 m}\ln\frac{Tn}{\delta}\right)$. For each task $t \in [T]$, let $\hat{\boldsymbol{r}}_t'$ be the empirically optimal lottery parameter vector over the set $\mathcal{S}_t'$: $\hat{\boldsymbol{r}}_t' = \operatorname{argmax}_{\boldsymbol{r}\in[0,1]^{n(m+1)}}\sum_{\ell=1}^{L}\sum_{\tau=1}^{T}\operatorname{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell,t,\tau)}\right)$. With probability $1-\delta$, for every task $t \in [T]$,*

$$\max_{\boldsymbol{r}\in[0,1]^{n(m+1)}} \mathop{\mathbb{E}}_{\boldsymbol{v}\sim\mathcal{D}^{(t)}}\left[\operatorname{rev}_{\boldsymbol{r}}(\boldsymbol{v})\right] - \mathop{\mathbb{E}}_{\boldsymbol{v}\sim\mathcal{D}^{(t)}}\left[\operatorname{rev}_{\hat{\boldsymbol{r}}_t'}(\boldsymbol{v})\right] = O\left(\sqrt{\frac{nm\log(nm)}{LT}} + \sqrt{\frac{1}{LT}\log\frac{T}{\delta}}\right).$$

Proof. This theorem follows directly from Theorem C.1 because with probability $1-\delta$, for every index $\ell \in [L]$ and ever pair of tasks $t, \tau \in [T]$, $\boldsymbol{w}^{(\ell,t,\tau)} = \boldsymbol{b}^{(t)} + \boldsymbol{z}^{(\ell,\tau)}$. □

Theorem C.5 implies that for any $\gamma \in (0,1)$, when $LT = \tilde{\Omega}\left(\frac{nm}{\gamma^2}\right)$, the expected revenue of the lottery defined by the parameter vector $\hat{\boldsymbol{r}}_t'$ is within γ of optimal. Meanwhile, as we mentioned in Section 4, the best-known single-task sample complexity guarantee requires $L = \tilde{\Omega}\left(\frac{nm}{\gamma^2}\right)$. Therefore, using our approach, the multi-task sample complexity LT is equal to the best-known single-task sample complexity.

In the case of multi-item second-price auctions with non-anonymous reserves, Theorem C.1, Lemma C.4, and Corollary 3.7 imply the following guarantee.

Theorem C.6. *Suppose that $L = \Omega\left(\frac{1}{\kappa^2 m}\ln\frac{Tn}{\delta}\right)$. For each task $t \in [T]$, let $\hat{\boldsymbol{r}}_t'$ be the empirically optimal reserve vector over the set $\mathcal{S}_t'$: $\hat{\boldsymbol{r}}_t' = \operatorname{argmax}_{\boldsymbol{r}\in[0,1]^{nm}}\sum_{\ell=1}^{L}\sum_{\tau=1}^{T}\operatorname{rev}_{\boldsymbol{r}}\left(\boldsymbol{w}^{(\ell,t,\tau)}\right)$. With probability $1-\delta$, for every task $t \in [T]$,*

$$\max_{\boldsymbol{r}\in[0,1]^{nm}} \mathop{\mathbb{E}}_{\boldsymbol{v}\sim\mathcal{D}^{(t)}}\left[\operatorname{rev}_{\boldsymbol{r}}(\boldsymbol{v}) - \operatorname{rev}_{\hat{\boldsymbol{r}}_t'}(\boldsymbol{v})\right] = O\left(\sqrt{\frac{nm\log(nm)}{LT}} + \sqrt{\frac{1}{LT}\log\frac{T}{\delta}}\right).$$

Proof. This theorem follows by the same logic as Theorem C.5. □

This theorem implies that for any $\gamma \in (0,1)$, when $LT = \tilde{\Omega}\left(\frac{nm}{\gamma^2}\right)$, the revenues of the reserve vectors $\hat{\boldsymbol{r}}_t'$ are within γ of optimal. On the other hand, as we mentioned in Appendix C.1.2, the best-known single-task sample complexity guarantee is $L = \tilde{\Omega}\left(\frac{nm}{\gamma^2}\right)$ [46], which is equal to our multi-task sample complexity LT.

Category pairs	KS test statistic	Estimated base value difference
(27261, 27264)	0.06	-17.672
(27261, 27268)	0.072	1.757
(27261, 27269)	0.072	-30.594
(27264, 27268)	0.065	19.429
(27264, 27269)	0.052	-12.922
(27268, 27269)	0.06	-32.351

Table 2: For several category pairs (i, j), we report the KS test statistic measuring the similarity between the original dataset $\widehat{S}_i$ and the shifted dataset $\widehat{S}'_j$. We also report the estimated base-value difference mean $\left(\widehat{S}_j\right)$ − mean $\left(\widehat{S}_i\right)$.

D Testing the common base value assumption

Since the preceding theory relies on the common value assumption, it is important verify that this assumption holds up in real world setting. To verify this, we analyze the Ebay bidding data for sports merchandise [44] and test whether the common-base-value model [9, 24] reflects real-world value distributions for similar items.

Setup: Each element of the Ebay sports merchandise dataset describes the price an item was sold for on the platform, which equals the maximum of the reserve and the second-highest bid. We use this value as a proxy for a bid, since the individual bids are not reported in the dataset. The items are sorted into a number of categories including: baseball, basketball, football trading cards, and so on. Let $\widehat{S}_i \subset \mathbb{R}$ denote the set of all prices an item in category i was sold for in the dataset. We assume that each set $\widehat{S}_i$ is drawn i.i.d. from an unknown distribution $\mathcal{D}_i$. Intuitively, the distributions defined by similar categories should be similar.

Procedure: If distributions $\mathcal{D}_i$ and $\mathcal{D}_j$ fall under the common-base-value model, then the two distributions are related by an additive shift. This shift can be estimated by taking the difference of the estimated means. That is, we calculate shifted bids $\widehat{S}'_i = \left\{s + \text{mean}\left(\widehat{S}_j\right) - \text{mean}\left(\widehat{S}_i\right) \mid s \in \widehat{S}_i\right\}$. If the common-base-value model holds, then $\widehat{S}'_i$ and $\widehat{S}_j$ should be roughly from the same distribution. To test this, we use the well-known two-sample Kolmogorov–Smirnov test (KS) to test if $\widehat{S}'_i$ and $\widehat{S}_j$ are from the same distribution. KS tests the null hypothesis that the two sets of samples are from the same distribution, returning a test statistic that is the maximum absolute distance between the empirical CDFs of the two sets of samples. If this statistic is very small, it tells us that the two sets of samples come from very similar distributions.

Results: We observe that the KS test statistic is small for several distributions corresponding to different categories. As we show in Table 2, we find that the empirical buyer value distribution for the cluster of categories $\{27261, 27264, 27268, 27269\}$ falls closely inline with the common-base-value model: the KS test statistic computed for any pair of those distributions lie between 0.05 to 0.07. These are inline with the actual eBay categories that the IDs correspond to, which are baseball and basketball merchandise. Intuitively, it does make sense that buyer distributions for different sports merchandises have similar distributions.

We also observe sizable variance in the estimated base-value difference mean $\left(\widehat{S}_j\right)$ − mean $\left(\widehat{S}_i\right)$. These results suggest that the common-base-value model can be useful for understanding similarities between buyers' value distributions. Lastly, the average bids in each category range from 8 to 109 with a high standard deviation of 18 to 60. This high variance rules out the possibility that the distributions are very concentrated, in which case the CBV assumption would be trivially satisfied.