Do Prices Coordinate Markets?

Justin Hsu^{*}
Department of Computer and
Information Science
University of Pennsylvania
USA
justhsu@cis.upenn.edu

Jamie Morgenstern[†]
Departments of Computer and
Information Science and
Economics
University of Pennsylvania
USA
jamiegmt@cs.cmu.edu

Ryan Rogers[‡]
Department of Applied
Mathematics and
Computational Science
University of Pennsylvania
USA
ryrogers@sas.upenn.edu

Aaron Roth[§]
Department of Computer and
Information Sciences
University of Pennsylvania
USA
aaroht@cis.upenn.edu

Rakesh Vohra
Economics Department
University of Pennsylvania
USA
rvohra@seas.upenn.edu

ABSTRACT

Walrasian equilibrium prices have a remarkable property: they allow each buyer to purchase a bundle of goods that she finds the most desirable, while guaranteeing that the induced allocation over all buyers will globally maximize social welfare. However, this clean story has two caveats: first, the prices may induce indifference. In fact, the minimal equilibrium prices *necessarily* induce indifference. Accordingly, buyers may need to coordinate with one another to arrive at a socially optimal outcome—the prices alone are *not* sufficient to coordinate the market; second, although natural procedures converge to Walrasian equilibrium prices on a fixed population, in practice buyers typically observe prices without participating in a price computation process. These prices cannot be perfect Walrasian equilibrium prices, but instead somehow reflect distributional information about the market.

To better understand the performance of Walrasian prices when facing these two problems, we give two results. First, we propose a mild genericity condition on valuations under which the minimal Walrasian equilibrium prices induce allocations which result in low over-demand, no matter how the buyers break ties. In fact, under genericity the over-demand of any good can be bounded by 1, which is the best possible at the minimal prices. We demonstrate our results for unit demand valuations and give an extension to ma-

troid based valuations (MBV), conjectured to be equivalent to gross substitute valuations (GS).

Second, we use techniques from learning theory to argue that the over-demand and welfare induced by a price vector converge to their expectations uniformly over the class of all price vectors, with respective sample complexity linear and quadratic in the number of goods in the market. These results make no assumption on the form of the valuation functions.

These two results imply that under a mild genericity condition, the exact Walrasian equilibrium prices computed in a market are guaranteed to induce both low over-demand and high welfare when used in a new market where agents are sampled independently from the same distribution, whenever the number of agents is larger than the number of commodities in the market.

Categories and Subject Descriptors

F.0 [Theory of Computation]: General

Keywords

Walrasian Equilibrium, Learning Theory

1. INTRODUCTION

The power of Walrasian equilibrium is often summarized with a pithy slogan: prices coordinate markets. However, this is not exactly true—a Walrasian equilibrium specifies a price for each good *and* an assignment of goods to buyers. The assignment is just as important as the prices, since there can be multiple bundles of goods which maximize a buyer's utility at given prices. If buyers select arbitrarily among these bundles, they may over-demand some goods.

One way to avoid this problem is to assume strictly concave valuation functions ensuring that each buyer always has a unique, utility maximizing bundle of goods. However, this idea does not apply in many natural economic settings; for instance, when goods are indivisible. For another possible solution, perhaps we could use Walrasian prices that eliminate the coordination problem. For example, with a single good and distinct valuations, a price *strictly* between the largest and second largest buyer valuations will eliminate indifference. However, how do we arrive at such prices? Relying on a coordinator with full knowledge of the market to solve the coordination problem defeats the purpose of markets.

^{*}Supported in part by NSF grants CCF-1101389, TWC-1513694 and CNS-1065060 and by a grant from the Simons Foundation (#360368).

[†]Supported in part by the Warren Center.

[‡]Supported in part by an NSF grant CNS-1253345

[§]Supported in part by an NSF grant CCF-1101389 and an Alfred P. Sloan Research Fellowship.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

Copyright is held by the owner/author(s). Publication rights licensed to ACM.

STOC'16, June 19–21, 2016, Cambridge, MA, USA
ACM. 978-1-4503-4132-5/16/06...\$15.00
<http://dx.doi.org/10.1145/2897518.2897559>

In this work, we consider a different approach: natural assumptions on buyers to simplify the coordination problem. We focus our investigation on *minimal* Walrasian equilibrium prices, which result from many natural market dynamics¹ [29]. We aim to allocate bundles of m types of indivisible goods g , each with some supply s_g , to n buyers who have matroid based Valuations² and quasi-linear preferences; the assignment (“unit demand”) model is an important special case. We detail our model in Section 2.

We begin our technical results in Section 3, by showing that indifference at the minimal Walrasian prices can be a serious problem—goods can be in the demand correspondence of *every* buyer, possibly leading to all buyers demanding the same good. Clever tie-breaking does not help; for any tie-breaking rule, the induced over-demand can be as large as $\Omega(n)$. (We provide a simple example in Section 3). Even worse, we observe that the minimal Walrasian equilibrium prices always induce indifference.

Hence, we cannot hope to rule out all over-demand at the minimal prices. But, we give a “genericity” condition on buyer valuations that is the next-best thing: the over-demand for each good g will be at most 1, independent of its supply s_g and the tie-breaking strategy used by buyers. Therefore, as the supply grows, worst-case over-demand becomes negligible. We warm up with the assignment model in Section 4, where our condition is simple to state: buyer valuations for goods should be linearly independent over the coefficients $\{-1, 0, 1\}$. To follow, we generalize our techniques to the matroid based valuations case in Section 5. The situation is significantly more complicated, but the core genericity definition and proof strategy are in the same spirit.

After we show that exact minimal Walrasian prices generically induce low over-demand, a natural question is whether this property holds when the same prices are used on “similar buyers”; we turn to this question in Section 6. More formally, imagine a sample N_1 of n buyers drawn from an unknown distribution Π of buyer valuations. The goods are priced using the minimal Walrasian equilibrium prices, computed from the valuations of buyers in N_1 . Now, keeping these prices fixed, we draw a fresh sample N_2 of n buyers from Π , who each choose some bundle from their demand correspondence at the given prices (breaking ties arbitrarily). Will the over-demand and welfare on N_2 be close to the over-demand and welfare on N_1 ?

If the supply s_g of a good g is small, the difference in over-demand between N_1 and N_2 may be large when compared to s_g . However, we show that if $s_g \geq \tilde{O}(m/\epsilon^2)$, the demand for any good g on sample N_2 will be within a $1 \pm \epsilon$ factor of the supply s_g of good g . Note that the supply requirement is independent of the market size n . Similarly, if the optimal welfare for N_1, N_2 is large, the induced welfare of the two markets will be within a $1 \pm \epsilon$ of one another (and within a $1 - \epsilon$ factor of the optimal welfare for N_2).

Furthermore, we are able to prove these bounds *without* any assumption on the structure of the valuation functions. This lack of structure makes it difficult to argue directly about notions of combinatorial dimension like VC dimension, and so we take a different approach which may be of independent interest. Using a recent compression argument of Daniely and Shalev-Shwartz [15], we show that assuming fixed but unknown prices, the class of functions *predicting* a buyer’s demanded bundle at those prices is learn-

able using $\tilde{O}(m/\epsilon^2)$ many samples. Because this is a multi-class learning problem, learning does not imply uniform convergence. However, the binary problem of predicting demand for a *particular* good is a 1-dimensional projection of the bundle prediction problem, and hence is also learnable with the same number of samples. By a classical result of Ehrenfeucht et al. [21], learning and uniform convergence have the same sample complexity in binary prediction problems. So, we can bound the VC dimension, and thus the sample complexity for uniform convergence for demand. Moreover, our bound is tight—even for unit demand buyers, the VC dimension of the class of demand predictors is $\Omega(m)$.

Welfare, unlike demand, corresponds to a real-valued prediction problem, and so the sample complexity needed for uniform convergence cannot be bounded by bounding the sample complexity of learning. Instead, we directly bound the pseudo-dimension of the class of welfare predictors by $\tilde{O}(m^2)$, again without making any assumptions about the form of the valuation functions. We show that if the optimal welfare is $\tilde{\Omega}(m^4 H^3 \sqrt{n}/\epsilon^2)$, the welfare induced by the Walrasian prices p for N_1 when applied to N_2 is within a $1 - \epsilon$ factor of optimal.

Related work. We follow a long line of work on understanding how markets behave under limited coordination. If buyers’ valuations are strictly concave and items are divisible, Arrow and Debreu [2] show that there exist item prices p such that each agent has a unique utility-maximizing bundle at p , and that when each agent selects her unique utility-maximizing bundle the market clears. With indivisible items, anonymous equilibrium item prices may not exist; even when they do, finding such prices is NP-hard [16]). Mount and Reiter [33] consider the size of the message space needed to compute Walrasian equilibria, and Nisan and Segal [37] show that polynomial communication is sufficient. This stands in sharp contrast to the situation for submodular buyers, where exponential communication is needed to compute prices which support an efficient allocation [37].

Our work is also related to the growing area of learning for mechanism design, where a mechanism is selected from some class of mechanisms as a function of sampled buyers. Recent work measures the sample complexity of revenue maximization in the single-parameter [4–7, 10, 11, 13, 14, 19, 22, 24, 27, 28, 31, 32, 41] and multi-parameter settings [17, 20].

In particular, sample complexity results for revenue and welfare maximization using item pricings are known. Most papers consider buyers that make decisions sequentially, avoiding over-demand from uncoordinated resolution of indifference. In the unlimited supply setting (in which over-demand cannot arise), Balcan et al. [7] show how to learn approximately revenue-optimal prices with polynomial sample complexity using a covering argument. Extending this work, Balcan et al. [6] show how to handle large but limited supply settings and to handle welfare maximization. These papers imply that demand concentration will have convergence rates of $\tilde{O}(\frac{k^2}{\epsilon} \ln \frac{1}{\delta})$; our result improves this by a factor of k (where k is the number of distinct types of goods). This percolates into the guarantees regarding welfare concentration, in that our concentration result requires a weaker lower bound on the supply of each good. Our precise convergence rate for welfare concentration is looser than theirs in several parameters, in part because we allow buyers to break indifference in a worst-case way. Using similar techniques to Balcan et al. [6], we could recover similar convergence rates for welfare. Devanur and Hayes [18] also consider prior-independent pricing for revenue maximization with unit demand buyers in an online setting.

¹Minimal Walrasian prices are focal in other ways: e.g., in matching markets they correspond to VCG prices.

²“Matroid based valuations” are a structured subclass of *gross substitutes* valuations. In fact, Ostrovsky and Paes Leme [38] conjecture that the class of matroid based valuations is *equal* to the class of gross substitutes valuations, the largest class of valuations for which Walrasian equilibrium prices are guaranteed to exist.

Finally, we use compression schemes to derive several uniform convergence results; this compression tool was first used in the context of game theory by Balcan et al. [8], who use it directly to upper bound the PAC complexity of a learning problem rather than to imply uniform convergence over a class.

Our definition of genericity for unit demand valuations has also been called *independence* by Ashlagi et al. [3], Braverman et al. [12]. In fact, our bound of over-demand for generic unit demand valuations can be obtained from results due to Braverman et al. [12], although over-demand was not the focus of their work. However, our analysis extends to more general valuation functions.

Our genericity results rely on a novel *swap graph* construction, which bears a slight resemblance to the exchangeability graph of Murota [35]. The exchangeability graph also has nodes defined by goods, but without considering a Walrasian allocation and pricing. In contrast, our swap graph is models indifference at equilibrium.

2. MODEL

We consider a market with m indivisible goods, where good g has supply $s_g \geq 1$. We will write the set of goods as $[m] = \{1, \dots, m\}$ and denote the bundles of the m goods as $\mathcal{G} = 2^{[m]}$. The market will also have a set N of n buyers, where each buyer demands at most one copy of each good. For simplicity, we consider valuations defined over subsets of goods rather than arbitrary sets of copies; see the extended version for a formal treatment. For each buyer $q \in N$, let $v_q : \mathcal{G} \rightarrow [0, H]$ denote q 's valuation function. We will assume v_q is monotone and $v_q(\emptyset) = 0$. Our goal is to produce a feasible assignments of goods to buyers.

Definition 2.1 (Allocation). *An allocation $\mu : N \rightarrow \mathcal{G}$ assigns each buyer a bundle such that the whole assignment is feasible:*

$$\sum_{q \in N} \mathbf{1}\{g \in \mu(q)\} \leq s_g \text{ for all } g \in [m].$$

As is typical, we consider quasi-linear utility functions $u_q : \mathcal{G} \times \mathbb{R}_{\geq 0}^m \rightarrow \mathbb{R}_{\geq 0}$ defined by

$$u_q(S; p) = v_q(S) - \sum_{g \in S} p_g \text{ for all } q \in N \text{ and } S \in \mathcal{G}.$$

We will consider prices assigned to goods—each *copy* of a good has the same price. We will write prices as vectors $p = (p_g)_{g \in [m]} \in \mathbb{R}_{\geq 0}^m$ or as functions over bundles $p : \mathcal{G} \rightarrow \mathbb{R}_{\geq 0}$ such that $p(S) = \sum_{g \in S} p_g$.

Definition 2.2 (Demand correspondence). *The demand correspondence for buyer $q \in N$ at prices p is $\mathcal{D}_q(p) = \arg\max_{S \in \mathcal{G}} \{u_q(S; p)\}$. We call bundles $S \in \mathcal{D}_q(p)$ demand bundles. Note that $\mathcal{D}_q(p)$ contains only bundles with non-negative utility, since $u_q(\emptyset; p) = 0$ for every p .*

We focus our investigation on *Walrasian equilibria*, defined by a pricing and an allocation.

Definition 2.3 (Walrasian equilibrium). *For valuations $\{v_q\}_{q \in N}$, we say that a pair (p, μ) of prices p and allocation $\mu = (\mu_q)_{q \in N}$ is a Walrasian equilibrium (WE) if both:*

- $\mu_q \in \mathcal{D}_q(p)$ for all $q \in N$; and
- $p_g = 0$ implies $\sum_{q \in N} \mathbf{1}\{g \in \mu(q)\} < s_g$.

We call the price vector p a *Walrasian equilibrium price vector*. Note that there may be many distinct Walrasian equilibrium price vectors: in fact, the set of all Walrasian prices forms a lattice. The

minimum *Walrasian equilibrium price vector* p is the *Walrasian equilibrium price vector* that is coordinate-wise minimal amongst all Walrasian equilibrium price vectors.

Likewise, we call the allocation μ a *Walrasian allocation*. While there may be multiple distinct Walrasian allocations, it is known that all such allocations must maximize welfare.

In general, Walrasian equilibrium prices p are not sufficient to coordinate a corresponding allocation μ because buyers might have indifferences ($|\mathcal{D}_q(p)| > 1$). If buyers choose their bundle arbitrarily, the resulting allocation can violate supply constraints. To measure the amount of violation, we make the following two natural definitions.

Definition 2.4 (Demanders and over-demand). *The set $U(g; p)$ of demanders for a good $g \in [m]$ at price p is the set of buyers that have some demand set containing g :*

$$U(g; p) = \{q \in N : \exists D \in \mathcal{D}_q(p) \text{ where } g \in D\}.$$

Then, the over-demand $OD(g; p)$ for g at prices p is the number of demanders beyond the supply of a good:

$$OD(g; p) = \max\{|U(g; p)| - s_g, 0\}.$$

That is, the over-demand is the worst-case excess demand if bidders break ties in their demand correspondence arbitrarily.

To build intuition, we focus the first part of our paper on unit demand bidders, where

$$v_q(S) = \arg\max_{g \in S} \{v_q(g)\} \text{ for all } S \in \mathcal{G},$$

and non-empty bundles $S \in \mathcal{D}_q(p)$ are *demand goods*.

3. LOWER BOUND

To build intuition for why tie breaking at equilibrium can lead to infeasibility, we give an example of a market with n buyers in which the over-demand of a good can be $\Omega(n)$ regardless of how buyers break ties, so long as they cannot coordinate with one another after seeing the market instance.³

Lemma 3.1. *There exist unit demand valuations such that at the minimal Walrasian prices p , some good has over-demand $\Omega(n)$.*

Proof. Consider a market with n unit demand buyers $N = [n]$ and $m = n$ distinct goods. For a distinguished good $g \in [m]$, every buyer q has valuation $v_q(q) = v_q(g) = 1$ and $v_q(g') = 0$ for all $g' \notin \{q, g\}$. The minimal Walrasian equilibrium prices are $p = \mathbf{0}$, and the unique max-welfare allocation is $\mu_q = q$. At these prices, g is demanded by every buyer. Hence, $OD(g; p) = n - 1$. $\square$

Note that if buyers resolve indifferences uniformly at random, $n/2$ buyers will attempt to buy good g . While one might hope that the right choice of tie breaking rules would solve this problem, we show that it cannot. Formally, we will suppose that each buyer q has a tie breaking rule e_q^p which can depend on v_q and the price vector p . Then, we consider the over-demand of a good when all bidders use their tie breaking rules.

Definition 3.2 (Tie-breaking over-demand). *Given a set of prices p , buyers N and tie breaking rule e_q^p for buyer q , let the demanders of g be*

$$U^e(g; p) = \{q \in N : g \in \mathcal{D}_q(p) \text{ and } g = e_q^p(\mathcal{D}_q(p))\},$$

³This lack of coordination is formalized by requiring that the tie breaking rule buyer q uses to select among favorite bundles must be independent of the valuations of buyers $q' \neq q$.

and the tie breaking over-demand with respect to e^p be

$$OD^e(g; p) = \max\{|\{q \in U^e(g; p)\}| - s_g, 0\}.$$

Without loss of generality (by the min-max principle), it suffices to consider deterministic tie breaking rules when constructing a randomized lower bound instance.

Lemma 3.3. *There exists a distribution over unit demand valuations such that for any set of tie breaking rules, the expected tie breaking over-demand from n buyers is $\Omega(n)$.*

Proof. The distribution we construct will contain instances similar to the one from Lemma 3.1, with n unit demand buyers and $m = n$ distinct goods. We choose a permutation σ over the goods $[m]$ and a distinguished good $g^* \in [m]$ uniformly at random. Once σ and g^* are chosen, we define buyer valuations such that for all buyers q , $v_q(\sigma(q)) = v_q(g^*) = 1$ and $v_q(h) = 0$ for all $h \notin \{\sigma(q), g^*\}$. In this market, the minimal Walrasian equilibrium prices is $p = \mathbf{0}$ and the unique max-welfare allocation has $\mu_q = \sigma(q)$ for all $q \in N$. At equilibrium, there is a single buyer q^* for whom $\sigma(q^*) = g^*$ and $\mathcal{D}_{q^*}(p) = \{g^*\}$. For every other buyer, $\sigma(q) \neq g^*$ and q has exactly two goods in their demand correspondence: $\mathcal{D}_q(p) = \{\sigma(q), g^*\}$. Each such buyer will select a good $e_q^p(\{\sigma(q), g^*\}) \in \{\sigma(q), g^*\}$ using their tie breaking rule.

Over the randomness of the instance (σ and g^*) we have $\mathbb{P}[e_q^p(\{\sigma(q), g^*\}) = g^*] = 1/2$. We can then lower bound the expected tie breaking over-demand for the distinguished good as

$$\begin{aligned} & \mathbb{E}_{\sigma, g^*}[OD^e(g^*; p)] \\ &= \sum_{q \in N} \mathbb{P}[\sigma(q) = g^*] \\ &+ \mathbb{P}[\sigma(q) \neq g^*] \cdot \mathbb{P}[e_q^p(\{\sigma(q), g^*\}) = g^*] - \underbrace{1}_{\text{supply}} \\ &= \frac{1}{2} \cdot n(1 - 1/n) = \frac{n-1}{2} \end{aligned}$$

which completes the proof. $\square$

While this result shows that over-demand can be high without coordination, it seems rather artificial—the buyers’ valuations are extremely similar. We will soon give a simple and natural condition which rules out this example, and more generally ensures that the over-demand for any good is at most 1 at the minimal Walrasian prices regardless of tie breaking. This bound is the best possible: for any buyer valuations, minimal Walrasian prices always induce over-demand of at least 1 for every good with positive price.

Lemma 3.4. *Fix any set of buyer valuations $\{v_q : q \in [n]\}$, and let p be a minimal Walrasian equilibrium price vector. For any good g with positive price $p(g) > 0$, we have $OD(g; p) \geq 1$.*

Proof. Let μ be a Walrasian allocation for p . By the Walrasian equilibrium condition, $|q : g \in \mu_q| = s_g$ for any good with $p(g) > 0$. Suppose that $OD(g; p) = 0$, i.e., g is in some bundle in buyer q ’s demand set if and only if $g \in \mu_q$. In this case there exists $\epsilon > 0$ such that if we set $p(g) \leftarrow p(g) - \epsilon$, (p, μ) remains a Walrasian equilibrium—the allocation is unchanged, and every buyer continues to receive an allocation in their demand set. But this contradicts minimality of p . $\square$

4. UNIT DEMAND

Now that we have seen how indifference at equilibrium can lead to over-demand, we consider whether the over-demand is large for “typical” instances. To build intuition, we start with the special case

of unit demand valuations, where $v_q(S) = \arg\max_{g \in S} \{v_q(g)\}$. Such valuations can be encoded with a real number for each pair of buyer $q \in N$ and good $g \in [m]$, denoted $v_q(g)$. Thus, without loss of generality an allocation is a many-to-one matching between buyers and goods—each buyer should be matched to at most 1 good, and each good g should be matched to at most s_g buyers.

We now give conditions on unit demand valuations to ensure that when buyers buy arbitrary (singleton) demand sets from their demand correspondence at the minimal Walrasian prices, the resulting allocation has high welfare and low over-demand. Accordingly, we need to reason precisely about how the equilibrium prices depend on the valuations. Getting access to this relation is surprisingly tricky—typical characterizations of Walrasian equilibrium prices are not enough for our needs. For instance, two standard characterizations show that unit demand Walrasian prices i) are dual variables to a particular linear program (the “many-to-one matching linear program”), and ii) are computed from ascending price auction dynamics. The first observation reduces computing prices to an optimization problem, but it does not provide fine-grained information about how the prices depend on the valuations. The second observation is useful for computing prices, but the auction may proceed in a complicated manner, obscuring the relationship between the prices and valuations.

4.1 Swap Graph

Accordingly, we define a graph called the *swap graph* of a Walrasian equilibrium (p, μ) . This graph directly encodes buyer indifference induced by the equilibrium prices. Furthermore, the swap graph allows us to read off equations involving the prices and valuations. We define the swap graph as follows.

Definition 4.1 (Swap graph). *The swap graph $G = (V, E)$ defined with respect to a Walrasian equilibrium (p, μ) has a node for each good g and an additional null node $\perp$ representing the empty allocation: $V = [m] \cup \{\perp\}$. There is a directed edge $(a, b) \in E$ for $a \neq b$, $b \neq \perp$ for each buyer q that receives good a in μ but also demands b , i.e. if $\mu_q = a$ and $b \in \mathcal{D}_q(p)$ for some buyer $q \in N$. Note that while there may be parallel edges—representing the same indifference by different buyers—there are no self loops.*

Since we will phrase our arguments in terms of the swap graph in the remainder of the section, we will first recast Lemma 3.4 using our new language.

Corollary 4.2. *For buyers with unit demand valuations and a minimal Walrasian equilibrium (p, μ) , every node in the swap graph G with in-degree zero has price zero.*

Almost by definition; the over-demand of a good g is its in-degree in the swap graph.

Lemma 4.3. *Let G be the swap graph corresponding to a Walrasian equilibrium. If a node g in G has in-degree d , then $OD(g; p) \leq d$.*

Proof. By construction of the swap graph, a node with in-degree d corresponds to a good g with d buyers with g in their demand correspondence but not in their allocation. Because μ is a feasible allocation, at most s_g buyers can be allocated good g in μ , and since μ is an equilibrium allocation, g is in the demand correspondence for each of these buyers. Thus, there can be at most $s_g + d$ demanders for good g . By definition of over-demand, we have $OD(g; p) \leq d$. $\square$

So to bound the maximum over-demand of any good, it suffices to bound the in-degree for every good in the swap graph. While the

in-degree may be large in the worst case, we can introduce a simple condition on valuations that will rule out these pathological market instances.

4.2 Generic Valuations

Recall that in Section 3, we showed that over-demand can be high at the minimal Walrasian equilibrium prices. So, to provide a better bound on over-demand, we need additional assumptions; ideally, a condition that will hold “typically”. In the lower bound instance for Lemma 3.3, over-demand is large because the buyers have valuations that are too similar. Indeed, consider a market with two goods a and b where all buyers have the same *difference* in valuations between a and b . If *some* buyer is indifferent between a and b —by Lemma 3.4, this must be the case at minimal Walrasian prices—*all* buyers are indifferent. This observation motivates our genericity condition.

Definition 4.4 (Generic valuations). *A set of valuations $\{v_q(g) \in \mathbb{R} : q \in N, g \in [m]\}$ is generic if they are linearly independent over $\{-1, 0, 1\}$, i.e.*

$$\sum_{q \in N} \sum_{g \in [m]} \alpha_{q,g} v_q(g) = 0 \quad \text{for } \alpha_{q,g} \in \{-1, 0, 1\}$$

implies $\alpha_{q,g} = 0$ for all $q \in N, g \in [m]$.

Remark 4.5. *Note that this condition holds with probability 1 given any continuous perturbation of a profile of valuation functions, and so for many natural distributions, a profile of valuation functions will “generically” (i.e., with high probability) satisfy our condition. We also show how to discretely perturb a fixed set of valuations to satisfy our condition in the extended version.*

Remark 4.6. *Our definition of generic in the unit demand setting is also called independence by Ashlagi et al. [3], Braverman et al. [12], although over-demand was not the focus in these works. In fact, Braverman et al. [12] give an alternative way to bound over-demand for generic unit demand valuations (Theorem 4.10). We present the unit-demand case using the swap graph construction, in order to generalize smoothly to broader classes of valuations.*

4.3 Over-Demand

Now, we are ready to present the main technical result of this section: When buyers with generic valuations select an *arbitrary* good in their demand correspondence given minimal Walrasian equilibrium prices, over-demand is low and welfare is high. We will show that the *in-degree* for any node in our swap graph is at most 1. This will imply that no good has over-demand more than 1, regardless of its supply. We proceed via a series of properties of the swap graph. First, under genericity, the swap graph is acyclic.

Lemma 4.7. *The swap graph G , defined with respect to Walrasian equilibrium (p, μ) and generic valuations $\{v_q(g) : q \in N, g \in [m]\}$, is acyclic.*

Proof. Since the null node $\perp$ has no incoming edges by construction, it cannot be part of any cycle. Thus, suppose that there is a cycle $a_0 \rightarrow a_1 \rightarrow \dots \rightarrow a_k \rightarrow a_0$ of non-null nodes. We label the buyers so that buyer q_i is allocated a_i in the allocation μ , but also has good a_{i+1} in their demand correspondence as well (taking the subscript modulo $k+1$) for $i = 0, \dots, k$.

By construction, all buyers are distinct and $k \geq 1$. Furthermore, each edge represents an indifference relationship for some buyer. In particular,

$$v_q(a_q) - p(a_q) = v_q(a_{q+1}) - p(a_{q+1}) \quad \text{for all } q = 0, \dots, k.$$

Summing these equations and canceling prices, we have

$$\sum_{i=0}^k v_{q_i}(a_i) = \sum_{i=0}^k v_{q_i}(a_{i+1}),$$

contradicting the genericity assumption since all buyers q are distinct. Hence, the swap graph must be acyclic. $\square$

Because the swap graph is acyclic, we can choose a partial order of the nodes so that all edges go from smaller nodes to larger nodes. For the remainder of the argument, we assume nodes are labeled by such an ordering (i.e. we now have for every edge $(a_i, a_j) \in E$ implies $i < j$). Now, the price of a good can be written in terms of the valuations of smaller goods.

Lemma 4.8. *For every good g , the price $p(g)$ can be written as a linear combination of valuations $v_q(j)$ over $\{-1, 0, 1\}$ for $q \in N$ and $j < g$. Specifically, for every set of goods $g_1 < g_2 < \dots < g_k < g$ such that $g_1 \rightarrow \dots \rightarrow g_k \rightarrow g$ forms a path in the graph and g_1 has in-degree 0, there are buyers $q_1, \dots, q_k$ such that $\mu_{q_i} = g_i$ and*

$$p(g) = \sum_{i=1}^k (v_{q_i}(g_{i+1}) - v_{q_i}(g_i))$$

where $g_{k+1} = g$. If the first node $g_1 = \perp$, we define $v_{q_1}(\perp) = 0$.

Proof. We proceed by induction on the path length k . In the base case, $k = 1$ and we have path $g_1 \rightarrow g$. If $g_1 = \perp$ then buyer q_1 must be indifferent between the empty allocation and good g , so $p(g) = v_{q_1}(g)$. If $g_1 \neq \perp$, then buyer q_1 ’s indifference between g_1 and g yields

$$v_{q_1}(g_1) - p(g_1) = v_{q_1}(g) - p(g),$$

so $p(g) = v_{q_1}(g) - v_{q_1}(g_1) + p(g_1)$. Note that g_1 has in-degree zero so Corollary 4.2 implies $p(g_1) = 0$. This shows the base case.

For the inductive case, we assume that we can write prices in our desired form for any good with a path of length at most $k-1$ from a source node. Consider a good g that has path length k from a source node. Because buyer q_k is indifferent between g and g_k ,

$$v_{q_k}(g_k) - p(g_k) = v_{q_k}(g) - p(g)$$

so $p(g) = v_{q_k}(g) - v_{q_k}(g_k) + p(g_k)$. Note that $p(g_k)$ has a path length of at most $k-1$ from a source node, and so we can apply the induction hypothesis to get complete the induction:

$$p(g) = v_{q_k}(g) - v_{q_k}(g_k) + \sum_{i=1}^{k-1} (v_{q_i}(g_{i+1}) - v_{q_i}(g_i)). \quad \square$$

Finally, we can bound the in-degree of any node under genericity.

Lemma 4.9. *For generic buyer valuations, every node in the the swap graph defined with respect to a Walrasian equilibrium (p, μ) with minimal Walrasian prices has in-degree at most 1.*

Proof. Suppose otherwise and let good g be the smallest indexed good with at least two incoming edges. Then there are two sequences of goods

$$g_1 \leq \dots \leq g_k \leq g \quad \text{and} \quad g'_1 \leq \dots \leq g'_{k'} \leq g$$

such that $g_k \neq g'_{k'}$, $g_1 \rightarrow \dots \rightarrow g_k \rightarrow g$ and $g'_1 \rightarrow \dots \rightarrow g'_{k'} \rightarrow g$ form paths in the graph, and g_1 and g'_1 have in-degree 0. We will

write $g = g_{k+1} = g'_{k'+1}$. By Lemma 4.8, we can express $p(g)$ in two distinct ways:

$$p(g) = \sum_{i=1}^k (v_{q_i}(g_{i+1}) - v_{q_i}(g_i))$$

and

$$p(g) = \sum_{i=1}^{k'} (v_{q'_i}(g'_{i+1}) - v_{q'_i}(g'_i)),$$

where buyer q_i has $\mu_{q_i} = g_i$ and also has good g_{i+1} in her demand set, and buyer q'_i has $\mu_{q'_i} = g'_i$ and has good g'_{i+1} in her demand set. Taking the difference, we have

$$\sum_{i=1}^k (v_{q_i}(g_{i+1}) - v_{q_i}(g_i)) - \sum_{i=1}^{k'} (v_{q'_i}(g'_{i+1}) - v_{q'_i}(g'_i)) = 0.$$

Since we either have $q_k \neq q_{k'}$ or $g_k \neq g'_{k'}$, the above linear combination is not trivial and contradicts genericity. $\square$

Finally, by Lemma 4.3 the over-demand for any good is at most its in-degree in the swap graph, so we have bounded over-demand under generic valuations.

Theorem 4.10. *For any set of unit demand buyers with generic valuations and for p the minimal Walrasian equilibrium price vector, the over-demand for any good $g \in [m]$ is at most 1.*

As a result, when generic buyers face minimal⁴ prices p and buy a good in their demand set while resolving indifferences arbitrarily, the excess demand of any good is at most 1.

4.4 Welfare

Now that we have considered over-demand under genericity, what about welfare? If buyers break ties arbitrarily, it is not hard to see that welfare may be very bad: buyers who are indifferent between receiving a good and receiving nothing may all decide to demand nothing, giving zero welfare. However, if we simply rule out this specific kind of indifference, we can show that genericity implies near-optimal welfare.

In our calculations of welfare, $\text{Welfare}_N(B)$, for some (possibly infeasible) allocation B , we assume over-demand is resolved in a worst-case way—i.e. we assume goods are allocated to buyers that demand them in the way that minimizes welfare, while obeying the supply constraints. For the welfare argument only, we assume that a buyer chooses a *nonempty* demand bundle whenever possible. Subject to this restriction, buyers can break ties however they like. Then, we can lower bound the welfare of these buyers N in terms of the optimal welfare Opt-Welfare_N .

Theorem 4.11. *Consider any set of buyers N with generic unit demand valuations bounded in $[0, H]$ and minimal Walrasian equilibrium prices p . For each buyer q , let $b_q \in \mathcal{D}_q(p)$ be some arbitrary set in their demand correspondence, assuming only that $b_q \neq \emptyset$ whenever $|\mathcal{D}_q(p)| > 1$. Then the welfare obtained by the resulting allocation is nearly optimal:*

$$\text{Welfare}_N(b_1, b_2, \dots, b_n) \geq \text{Opt-Welfare}_N - 2 \cdot m \cdot H.$$

⁴We show in the full version that even with generic valuations, *non-minimal* Walrasian prices can still induce high over-demand, further justifying our focus on minimal Walrasian prices.

5. TOWARDS GROSS SUBSTITUTES

Now that we have seen our techniques in the unit demand setting, we will generalize our results to more general valuations. Ideally, we would like to extend our results to buyers who have *gross substitutes* valuations.

Definition 5.1 (Gross substitutes). *A valuation function v_q satisfies gross substitutes (GS) if for every price vectors $p' \geq p$ and $S \in \mathcal{D}_q(p)$, there is a bundle $S' \in \mathcal{D}_q(p')$ with*

$$S \cap \{g \in [m] : p(g) = p'(g)\} \subseteq S'.$$

Note that unit demand valuations also satisfy GS.

If all buyers have GS valuations, there always exist Walrasian equilibria, and the minimal equilibrium can be found by following a tâtonnement procedure [29]. Gul and Stacchetti [25] show that the class GS is in some sense the most general class of valuations that are guaranteed to have WE.

While GS is an attractive class to target, its definition is axiomatic rather than constructive; Definition 5.1 shows us what properties a GS valuation must satisfy, rather than how to concretely construct GS valuations. This poses a problem for defining genericity: it is not obvious which, if any, valuations in GS satisfy a candidate definition of genericity, so it may be that generic valuations do not even have Walrasian equilibria!

Ultimately, we will prove our results for buyers with *matroid based valuations* (MBV), a subclass of gross substitutes valuations that is conjectured to be equal to all of gross substitutes [38]. Such valuations can be explicitly constructed from a set of numeric weights and a matroid, giving us a path to define generic valuations and certify that they are contained in GS.

However, there are still several obstacles to overcome, and the arguments are significantly more involved than for unit demand. Roughly, the central difficulty is establishing a connection between valuations on bundles and valuations on the items in the bundle, since we will require genericity on values for single goods. While this connection is immediate in the case of unit demand—favorite bundles are simply single goods—the situation for MBV is more delicate. For instance, some bundles may contain “irrelevant” goods, which don’t contribute at all to the valuation. Hence, we propose a more complex swap graph to capture the richer structure of the valuations.

Despite the complications, our high-level argument parallels the unit demand case. We will first define a swap graph and connect it to over-demand. Then, we define a generic version of MBV (GMBV) and prove properties about the swap graph for valuations in GMBV. We only restrict to MBV when defining genericity; our arguments in the first step apply to GS valuations.

5.1 Swap Graph with GS Valuations

To define the swap graph for gross substitute valuations, the core idea is to have an edge (a, b) represent a single *swap* of good a for good b in some larger bundle, naturally generalizing our construction for unit demand. The main challenge is ensuring that we faithfully model buyer indifferences—which are between *bundles* of goods—via indifferences of single swaps. More precisely, in order to bound over-demand with arbitrary tie-breaking, we must ensure that if a buyer is indifferent between her Walrasian allocation and some other bundle B , then there must be at least one incoming edge to every good in B . If the swap graph satisfies this property, we can describe potential demand for goods in terms of in-degree of nodes like we did for unit demand.

At first, we might hope that if a buyer is indifferent between bundles B_1 and B_2 , every good in B_1 can be exchanged for a good

in B_2 while preserving utility; this would ensure that each good in B_2 has an incoming edge. While this property turns out to not be true for general bundles, it is true for the “smallest” bundles in a buyer’s demand correspondence.

Definition 5.2 (Minimal demand correspondence). *For price vector p and buyer $q \in N$, the minimum demand correspondence is*

$$\mathcal{D}_q^*(p) = \{S \in \mathcal{D}_q(p) : T \notin \mathcal{D}_q(p) \text{ for all } T \subsetneq S\}.$$

We call bundles $B \in \mathcal{D}_q^*(p)$ minimum demand bundles, or minimum bundles for short.

Crucially, the minimum demand correspondence for any buyer with a GS valuation forms the bases of a matroid [9]. (We give a self-contained proof in the extended version.)

Lemma 5.3 (Ben-Zwi et al. [9]). *Let $v_q : \mathcal{G} \rightarrow [0, H]$ be a GS valuation. For any price vector p , the minimum demand correspondence $\mathcal{D}_q^*(p)$ forms the set of bases of some matroid.*

By standard facts from matroid theory [39], all elements of $\mathcal{D}_q^*(p)$ are the same size. More importantly, the bases satisfy the exchange property: for B_1, B_2 in the basis set $\mathcal{B}$ of a matroid and for every $b \in B_1 \setminus B_2$, there exists $b' \in B_2 \setminus B_1$ such that

$$B_1 \cup b' \setminus b \in \mathcal{B} \quad \text{and} \quad B_2 \cup b \setminus b' \in \mathcal{B}.$$

Thus if all bundles in a buyer’s demand correspondence have the same size, then we can take $\mathcal{B}$ to be $\mathcal{D}_q^*(p)$ and we have exactly what we need: for any two bundles in the correspondence, every good can be swapped for some other good while remaining in the correspondence. Unfortunately, buyers may prefer bundles of different sizes. We will now introduce the swap graph, and then explain how it models buyers who prefer bundles of different sizes.

Definition 5.4 (Swap graph). *Let buyers have GS valuations, (p, μ) be a WE, and for each buyer $q \in N$ fix a minimum demand bundle $M_q \in \mathcal{D}_q^*(p)$ where $M_q \subseteq \mu_q$. Define the swap graph $G(p, \mu, (M_q)_{q \in N})$ to have a node for every good $g \in [m]$ and an additional node which we refer to as the null node $\perp$. There is a directed edge from (a, b) for every buyer $q \in N$ such that $a \in M_q$, $b \notin \mu_q$, and there exists $B \in \mathcal{D}_q^*(p)$ with $b \in B$ where*

$$M_q \cup b \setminus a \in \mathcal{D}_q^*(p).$$

Further, the graph contains an edge from $\perp$ to good b for each buyer $q \in N$ with $b \notin \mu_q$ such that (i) $b \in B$ with $B \in \mathcal{D}_q(p)$ and $B \setminus b \in \mathcal{D}_q^*(p)$, and (ii) b has positive price.

There are two main conditions when specifying edges from the null good: (i) involves demand bundles, while (ii) requires positive price. These two conditions address two distinct problems.

The first condition models cases where the demand correspondence has bundles of different sizes. Suppose there is some bundle B strictly larger than the minimum cardinality demand bundle. We need all goods in B to have an incoming edge, to reflect over-demand if a buyer selects B . We cannot ensure these edges via matroid properties, since B is not a minimum cardinality demand bundle and so it is not a basis of the matroid. However, there is always a minimum bundle $B^{\min}$ contained in B , which is a matroid basis, so we have incoming edges to $B^{\min}$.

The following interpolation lemma shows that for every good $g \in B \setminus B^{\min}$ the bundle $B^{\min} \cup g$ is also a demand bundle, so the swap graph will also have edges from $\perp$ to g , covering all goods in B as desired.

Lemma 5.5. *Consider a buyer with a GS valuation. If we have two bundles B, B' such that $B \subseteq B'$, and B, B' are both bundles in the buyer’s demand correspondence at price p , then*

$$u(B; p) = u(B''; p) = u(B'; p)$$

for any B'' with $B \subseteq B'' \subseteq B'$.

Proof. Let B'' satisfy the conditions in the lemma statement and define $\{b_1, \dots, b_k\} = B'' \setminus B$; we proceed by induction on k . Since GS valuations are submodular,

$$u(B'; p) - u(B' \setminus b_1; p) \leq u(B \cup b_1; p) - u(B; p),$$

so

$$u(B'; p) \leq u(B \cup b_1; p) - u(B; p) + u(B' \setminus b_1; p).$$

Since B is most demanded, $u(B \cup b_1; p) - u(B; p) \leq 0$. If $B \cup b_1$ is strictly worse, then $B' \setminus b_1$ must have strictly higher utility than B' , which contradicts B' being a most demanded bundle. So, we must have $u(B \cup b_1; p) = u(B; p)$.

By induction we assume that $B \cup \{b_1, \dots, b_{t-1}\}$ is a most demanded bundle. Using the same argument as in the base case, we can use submodularity to show that

$$u(B \cup \{b_1, \dots, b_{t-1}\}; p) = u(B \cup \{b_1, \dots, b_t\}; p).$$

Hence, B'' is a most demanded bundle at price p . $\square$

The second condition for edges from the null good—giving edges only to goods with positive price—handles a subtle case. Up to this point, we have argued informally that if B is a bundle in $\mathcal{D}_q^*(p)$, then all goods in $B \setminus M_q$ will have an incoming edge. We plan to bound the over-demand by the in-degree, but there is an important corner case: goods with price 0. Such goods can be freely added to any buyer’s bundle, ruining the bound on over-demand. However, all is not lost: the problem stems from buyers demanding goods with zero marginal valuation. We call bundles with such goods *degenerate*, and assume (or require) that buyers do not select them.

Definition 5.6 (Non-degenerate). *A bundle S is non-degenerate with respect to S' if goods in S' have non-zero marginal value:*

$$v(S \setminus g) < v(S) \quad \text{for each } g \in S \cap S'.$$

When $S' = [m]$, we will say S is non-degenerate. The non-degenerate correspondence for buyer $q \in N$ at price p is the set of demand bundles defined by

$$\mathcal{D}_q^\bullet(p) = \{S \in \mathcal{D}_q(p) : v_q(S \setminus g) < v_q(S) \quad \forall g \in S\}.$$

Minimum bundles are an example of non-degenerate bundles.

Lemma 5.7. *Let v be a GS valuation function. If S is a minimum demand bundle with respect to this valuation and some set of prices, then S is non-degenerate.*

Proof. Fix a buyer and a price vector p over goods and some minimum demand bundle S . By hypothesis, the buyer’s utility $u(S; p)$ must strictly decrease if we eliminate a good from S , i.e. $\forall g \in S$ we have $u(S \setminus g; p) < u(S; p)$. We need to show that $v(S \setminus g) < v(S)$ for every $g \in S$ to prove that S is non-degenerate. Because S is a minimum demand bundle,

$$\begin{aligned} u(S; p) &= v(S) - p(S) > v(S \setminus g) - p(S \setminus g) \\ &> v(S \setminus g) + p(g) \geq v(S \setminus g) \end{aligned}$$

using non-negativity of the prices. $\square$

If we only consider buyers who purchase bundles in $\mathcal{D}_q^\bullet(p)$, there is no problem with zero-price goods g : for buyers with non-degenerate bundles that contain g , that good must automatically be in any minimum bundle for those buyers. Hence, buyers cannot be indifferent between non-degenerate bundle B that contains g and $B \setminus g$. So, we only need edges in the swap graph from $\perp$ to goods with positive price.

We are almost ready to formally connect the swap graph with over-demand, but there is one last wrinkle. Since a buyer is assigned a bundle rather than a single good, they may have two different swaps to the same good: perhaps $M_q \setminus a_1 \cup b$ and $M_q \setminus a_2 \cup b$ are both demanded. We want to count this buyer as causing demand 1 rather than 2, since they will select a bundle of distinct goods. This consideration motivates our definition of *buyer in-degree*.

Definition 5.8 (Buyer in-degree). *Consider buyers with GS valuations $\{v_q\}_{q \in N}$, the minimal Walrasian prices p and allocation μ , and the corresponding swap graph G . The buyer in-degree of node b is the number of distinct buyers with an edge directed to b .*

To prevent buyers from piling on zero-priced goods with zero marginal value, we will require buyers to take non-degenerate bundles. Then, the following definition of over-demand is natural.

Definition 5.9 (Non-degenerate over-demand). *The non-degenerate over-demand $OD^\bullet(g; p)$ for a good $g \in [m]$ at Walrasian prices p is*

$$OD^\bullet(g; p) = \max\{|U^\bullet(g; p)| - s_g, 0\},$$

where

$$U^\bullet(g; p) = \{q \in N : \exists B \in \mathcal{D}_q^\bullet(p) \text{ and } g \in B\}.$$

Like the unit demand case, this is simply the worst-case over-demand assuming that buyers choose an arbitrary non-degenerate bundle from their demand correspondence, without any assumption on how they break ties. Finally, we can relate the swap graph to over-demand.

Lemma 5.10. *Let buyers have GS valuations and let (p, μ) be a Walrasian equilibrium with minimal prices. If a node $g \in [m]$ in the swap graph G has buyer in-degree at most d , then $OD^\bullet(g; p) \leq d$.*

Proof. Consider any good g . We wish to bound the size of $U^\bullet(g; p)$, the set of buyers who have g in a non-degenerate bundle. These buyers fall into two classes: buyers with $g \in M_q$, and buyers with $g \notin M_q$. The number of buyers of the first kind is at most the supply s_g , since M allocates only fewer copies of goods than the Walrasian allocation μ .

To bound the number of buyers of the second kind, we will show that there is an edge to g in the swap graph for each such buyer. Consider any buyer q with $g \notin M_q$ and some non-degenerate bundle $B_q^\bullet \in \mathcal{D}_q^\bullet(p)$ where $g \in B_q^\bullet$. Consider any minimum demand bundle $B_q^* \subseteq B_q^\bullet$. If $g \in B_q^*$, then there is an edge corresponding to q from M_q to g , since both M_q and B_q^* lie in $\mathcal{D}_q^\bullet(p)$ (which forms the set of bases for some matroid by Lemma 5.3).

Otherwise, if $g \notin B_q^*$, Lemma 5.5 shows that $B_q^*(p) \setminus g$ is also a demand bundle:

$$v_q(B_q^* \setminus g) - p(B_q^* \setminus g) = v_q(B_q^*) - p(B_q^*)$$

Since B_q^* is non-degenerate,

$$v_q(B_q^* \setminus g) < v_q(B_q^*)$$

and so $p(g) > 0$. At the same time, Lemma 5.5 also shows that $B_q^* \cup g$ is a demand bundle. Combining these last two facts, there

must be an edge corresponding to q from $\perp$ to g in the swap graph. So, the number of demanders of the second kind ($g \notin M_q$) is at most the buyer in-degree of g , which is at most d by assumption.

Now, we count up the demanders of both kinds: at most $s_g + d$ buyers demand g in a non-degenerate bundle. Hence, the non-degenerate over-demand is at most d . $\square$

5.2 Matroid Based Valuations

Now that we have seen the swap graph, we move on to describe our genericity condition on valuations. As discussed above, it is not clear how to define genericity for general GS valuations due to the axiomatic nature of the definition. We will define genericity for a subclass of GS valuations studied by Ostrovsky and Paes Leme [38] called *matroid based valuations* (MBV). Such valuations are defined in terms of a simpler class of valuations, known as VIWM.

Definition 5.11 (VIWM). *Let X be some ground set and let w be a function $w : X \rightarrow \mathbb{R}_{\geq 0}$. Consider a matroid $\mathcal{M} = (\mathcal{I}, X)$ and define $\mathcal{M}_S = (\mathcal{I}_S, S)$ as the restricted matroid to set $S \subseteq X$ where $\mathcal{I}_S = \{T \in \mathcal{I} : T \subseteq S\}$. The valuation induced by weighted matroid $(\mathcal{M}, w)$ is*

$$v(S) = \max_{T \in \mathcal{I}_S} \sum_{j \in T} w(j).$$

Some examples of valuations induced by a weighted matroid include unit demand valuations (the matroid has independent sets that are singletons and the empty set) and additive valuations (the matroid has the power set 2^X as its independent sets). For any matroid $\mathcal{M} = (\mathcal{I}, X)$, the valuations induced by $\mathcal{M}$ lie in GS [23, 34, 36]. However, Ostrovsky and Paes Leme [38] show that VIWM is a strict subset of GS by showing that two operations on valuation functions preserve GS, but do not preserve VIWM.

Definition 5.12 (Merging (convolution)). *For any $S \subseteq X$ and valuation functions $v', v'' : 2^X \rightarrow \mathbb{R}_{\geq 0}$, the merging of v', v'' is the valuation defined by*

$$v^{merg}(S) = \max_{(S', S'')=S} \{v'(S') + v''(S'')\}.$$

Definition 5.13 (Endowment (marginal valuation)). *Suppose $S \subseteq X$ and valuation $v' : 2^{X \cup Y} \rightarrow \mathbb{R}_{\geq 0}$, where $X \cap Y = \emptyset$. Then, the endowment of v' by Y is the valuation defined by*

$$v^{end}(S) = v'(S \cup Y) - v'(Y).$$

Accordingly, Ostrovsky and Paes Leme [38] propose a new class of valuations closing VIWM under the two operations, called MBV, and conjecture that MBV is equal to all of GS.

Definition 5.14 (Matroid based valuation). *Matroid based valuations (MBV) is the smallest class of valuation functions containing VIWM closed under finite merges and endowments.*

Conjecture 5.15 (Ostrovsky and Paes Leme [38]). *The class of MBV is precisely the class of GS valuations: $GS = MBV$.*

5.3 Generic MBV

Now that we have defined MBV, defining genericity is straightforward: we simply require genericity on the set of weights. Each valuation in MBV can be decomposed as a set of weights over the goods in $[m]$ as well as the goods that have been *endowed*. We will use the notation $W_q(A_q)$ to denote the set of weights belonging to buyer q that are evaluated on elements in the set A_q and $W(A)$ to denote the set of weights for all buyers, i.e.

$$W_q(A_q) = \{w_q^{(t)}(a) : a \in A, t \in [\tau_q]\},$$

and

$$W(A) = \bigcup_{q \in N} W_q(A_q).$$

Definition 5.16 (Generic matroid based valuations). *We say that a collection of valuations $\{v_q : \mathcal{G} \rightarrow [0, H] : q \in N\}$ are generic matroid based valuations (GMBV) if each v_q is a matroid based valuation and the weights across buyers over all goods $[m]$ and endowments T_q introduced in endowment operations are linearly independent over the integers. We will write the set $A_q = [m] \cup T_q$. Formally, the weights satisfy*

$$\sum_{w \in W(A)} \gamma_w w = 0$$

for $\gamma_w \in \mathbb{Z}$ if and only if the coefficients are all zero.

5.4 Bounding the Buyer In-Degree

To conclude, we show several properties of the swap graph. The arguments are rather involved (see the extended version for details), but the overall plan follows the unit demand case. Finally, we can bound the buyer in-degree and hence over-demand.

Theorem 5.17. *For GMBV buyers and Walrasian equilibrium (p, μ) with minimal Walrasian prices, each node in the swap graph G has buyer in-degree at most 1.*

6. WELFARE AND OVER-DEMAND GENERALIZATION

In this section, we show that Walrasian prices *generalize*: the equilibrium prices for a market N of buyers induces similar behavior when presented to a new sample N' of buyers, both in terms of the demand for each good and in terms of welfare. More precisely, a set of prices which minimizes over-demand of each good *and* maximizes welfare when each buyer purchases her most-preferred bundle retains these properties (approximately) on a *new* market N' when each buyer purchases her most-preferred bundle, if buyers in N and N' are drawn independently from the same distribution.

6.1 Results for Arbitrary Valuations

To formalize how over-demand and welfare induced by prices p vary between two markets, we first fix a tie-breaking rule e that buyers use to select amongst demanded bundles for a valuation from set $\mathcal{V}$. We define classes of *functions* parameterized by pricings mapping valuations to (a) bundles purchased, (b) whether or not a particular good g is purchased, and (c) the value a buyer gets for her purchased bundle.

Let $e : 2^{\mathcal{G}} \times \mathcal{V} \rightarrow \mathcal{G}$ denote some tie-breaking function, which satisfies some mild technical conditions defined in the extended version. We call the bundle that e selects from $\mathcal{D}_q(p)$ the *canonical* bundle for q at p , denoted by $B_q^*(p) = e(\mathcal{D}_q(p), v_q)$. Then, for each good g , let $h_{g,p}(v_q) = \mathbf{1}[g \in B_q^*(p)]$ indicate whether or not g is in q 's canonical bundle at prices p . Let $C^e(g; p; N) = \sum_{q \in N} h_{g,p}(v_q)$, i.e. the number of buyers in N whose canonical bundle at p contains g . For a sample of n buyers $\{v_q\} \stackrel{i.i.d.}{\sim} \Pi$, let $C^e(g; p; \Pi)$ represent the *expected* number of copies of g demanded at prices p if buyers demand canonical bundles. Similarly, let the *welfare* of a pricing p on a market N be $\text{Welfare}_N(p) = \sum_{q \in N} v_q(\widehat{B}_q^*(p))$ for $\widehat{B}_q^*(p)$ a worst-case resolution of the over-demand for each good.

Before presenting the technical details, we show how the behavior induced by prices generalizes for buyers with arbitrary valuations. Our first theorem bounds the over-demand when the Walrasian prices computed for a market are applied to a new market.

Theorem 6.1. *Fix a pricing p and two sampled markets N, N' of buyers with arbitrary valuations, with $|N| = |N'| = n$. For good g , suppose $C^e(g; p; N) \leq s_g + 1$. Then, for any $\alpha \in (0, \frac{4}{5})$, if $s_g = \Omega(m \ln \frac{1}{\delta} / \alpha^2)$, with probability $1 - \delta$,*

$$C^e(g; p; N') \leq (1 + \alpha)s_g.$$

Our second theorem is an analogous generalization result, for welfare instead of over-demand.

Theorem 6.2. *Fix two markets $N, N' \sim \Pi$ for which $|N| = |N'| = n$. Suppose for each good g , the prices p satisfy $C^e(g; p; N) \leq s_g + 1$. With probability $1 - \delta$, for any $\alpha \in (0, 4/5)$, if p are welfare-optimal for N and*

$$\text{Opt-Welfare}_N = \Omega\left(\frac{H^3 n^{0.5} m^4 \ln^2(m) \ln^2 \frac{1}{\delta}}{\alpha^2}\right)$$

then

$$\text{Welfare}_{N'}(p) \geq (1 - \alpha) \text{Opt-Welfare}_{N'}.$$

These proofs can be found in the full version. The first result (Theorem 6.1), for over-demand, relies on bounding the VC dimension of the class of good g 's demand indicator functions (the class contains a function for each pricing, labeling a valuation v positive if and only if g is in the canonical bundle for v at those prices). To bound the VC dimension, we first prove that the class of *bundle* predictors (the class contains one function for each pricing p , which maps v to v 's canonical bundles at p) is *linearly separable* in a space of $m + 1$ dimensions. Using a recent result by Daniely and Shalev-Shwartz [15], such a class admits a compression scheme of size $m + 1$ and can be (ϵ, δ) -PAC learned with $O(\frac{m}{\epsilon} \ln \frac{1}{\delta})$ samples. Since item predictor functions are simply projections of bundle prediction functions, their PAC-complexity of the item predictors is an upper bound on the PAC-complexity of the bundle predictors. Since item predictors are binary valued, we can apply the classical result of [21] which shows the equivalence of learning and uniform convergence for binary prediction problems, so we can upper bound the class's VC dimension as a function of its PAC sample complexity.

The second result (Theorem 6.2), for welfare, proceeds by bounding the pseudo-dimension of the class of welfare predictor functions (containing, for each pricing a function which map valuations to the *value* of a buyer purchasing her canonical bundle at these prices). The argument uses the existence of a compression scheme to upper bound the number of possible distinct labelings of valuations by bundles. Fixing a bundle labeling of a valuation also fixes the welfare of that valuation; thus, one can upper-bound the size of a “shatterable” set and the pseudo-dimension of the class of welfare predictors.

6.2 Learning Theory Preliminaries

This section reviews several well-known definitions from learning theory. We will show that:

- the demand of a good g from a market N at prices p should be close to the demand for g on a new market N' ; and
- the welfare-optimal prices for N should be approximately welfare-optimal for N'

under appropriate conditions.

Viewed from a learning-theoretic perspective, both of these statements are about the *generalization guarantees* of a class of mechanisms. Learning theory provides many tools to formalize how well properties of prices generalize from one market to the next. In what

follows, we will first describe standard tools for measuring the generalizability of binary-valued functions, followed by the analogous tools for understanding the generalizability of real-valued functions. The former will be useful in measuring the concentration of demand for a good g , while the latter will be useful for measuring the concentration of welfare.

Suppose there is a domain $\mathcal{V}$ (for our purposes, the valuation functions), an unknown target function $c : \mathcal{V} \rightarrow \{0, 1\}$, and an unknown distribution Π over $\mathcal{V}$. We wish to understand how many labeled samples $(v, c(v))$, with $v \sim \Pi$, are necessary and sufficient to be able to compute a $\hat{c}$ which agrees with c almost everywhere with respect to Π . The distribution-independent sample complexity of learning c depends fundamentally on the “complexity” of the set of binary functions $\mathcal{F}$ from which we are choosing $\hat{c}$. We review two standard complexity measures next.

Let N be a set of n samples from $\mathcal{V}$. The set N is said to be *shattered* by $\mathcal{F}$ if, for every subset $T \subseteq N$, there is some $c_T \in \mathcal{F}$ such that $c_T(v) = 1$ if $v \in T$ and $c_T(v') = 0$ if $v' \notin T$. That is, ranging over all $c \in \mathcal{F}$ induces all $2^{|N|}$ possible projections onto N . The *VC dimension* of $\mathcal{F}$, denoted $\text{VC}(\mathcal{F})$, is the size of the largest set S that can be shattered by $\mathcal{F}$.

Let $\text{err}_N(\hat{c}) = (\sum_{v \in N} |c(v) - \hat{c}(v)|) / |N|$ denote the empirical error of $\hat{c}$ on N , and let $\text{err}(\hat{c}) = \mathbb{E}_{v \sim \Pi} [|c(v) - \hat{c}(v)|]$ denote the *true* expected error of $\hat{c}$ with respect to Π . We say that $\mathcal{F}$ is (ϵ, δ) -PAC learnable with sample complexity n if there exists an algorithm $\mathcal{A}$ such that, for all distributions Π and all target functions $c \in \mathcal{F}$, when $\mathcal{A}$ is given a sample S of size n it produces some $\hat{c} \in \mathcal{F}$ such that $\text{err}(\hat{c}) < \epsilon$ with probability $1 - \delta$ over the choice of the sample. The PAC sample complexity of a class $\mathcal{F}$ can be bounded as a polynomial function of $\text{VC}(\mathcal{F})$, ϵ , and $\ln \frac{1}{\delta}$ [43]; furthermore, any algorithm which (ϵ, δ) -PAC learns $\mathcal{F}$ over all distributions Π *must* use nearly as many samples to do so. The following theorem states this well-known result formally.

Theorem 6.3 (Ehrenfeucht et al. [21], Hanneke [26]). *Suppose $\mathcal{F}$ is a class of binary functions. Then, $\mathcal{F}$ can be (ϵ, δ) -PAC learned with a sample of size*

$$n = O\left(\frac{\text{VC}(\mathcal{F}) + \ln \frac{1}{\delta}}{\epsilon}\right)$$

whenever $c \in \mathcal{F}$. Furthermore, any (ϵ, δ) -PAC learning algorithm for $\mathcal{F}$ must have sample complexity

$$n = \Omega\left(\frac{\text{VC}(\mathcal{F}) + \ln \frac{1}{\delta}}{\epsilon}\right).$$

There is a stronger sense in which a class $\mathcal{F}$ can be learned, called *uniform learnability*. This property implies that with a sufficiently large sample, the error of *every* $c \in \mathcal{F}$ on the sample is close to the true error of c . We say $\mathcal{F}$ is (ϵ, δ) -uniformly learnable with sample complexity n if for every distributions Π and a sample N of size n , $|\text{err}_N(c) - \text{err}(c)| < \epsilon$ for every $c \in \mathcal{F}$, with probability $1 - \delta$. Notice that if $\mathcal{F}$ is (ϵ, δ) -uniformly learnable with n samples, then it is also (ϵ, δ) -PAC learnable with n samples. We will use a well-known upper bound on the uniform sample complexity of a class as a function of its VC dimension.

Theorem 6.4 (E.g. Vapnik and Chervonenkis [43]). *Suppose $\mathcal{F}$ is a class of binary functions. Then, $\mathcal{F}$ can be (ϵ, δ) -uniformly learned with a sample of size*

$$n = O\left(\frac{\text{VC}(\mathcal{F}) \ln \frac{1}{\epsilon} + \ln \frac{1}{\delta}}{\epsilon^2}\right).$$

Both PAC learnability and uniform learnability of binary-valued functions are characterized by the class’s VC dimension. When

learning real-valued functions, for example, to guarantee convergence of the welfare of pricings), we use a real-valued analogue to VC dimension which will give a sufficient condition for uniform convergence, called the *pseudo-dimension* [40]. Formally, let $c : \mathcal{V} \rightarrow [0, H]$ be a real-valued function over $\mathcal{V}$, and $\mathcal{F}$ be the class we are learning over. Let S be a sample drawn from Π , $|N| = n$, labeled according to c . Both the empirical and true errors of a hypothesis $\hat{c}$ are defined as before, though $|\hat{c}(v) - c(v)|$ is now in the interval $[0, H]$ rather than in $\{0, 1\}$. Let $(r_1, \dots, r_n) \in [0, H]^n$ be a set of *targets* for N . We say $(r_1, \dots, r_n)$ *witnesses* the shattering of N by $\mathcal{F}$ if, for each $T \subseteq N$, there exists some $c_T \in \mathcal{F}$ such that $c_T(v_q) \geq r_q$ for all $v_q \in T$ and $c_T(v_q) < r_q$ for all $v_q \notin T$. If there exist such witnesses, we say N is *shatterable* by $\mathcal{F}$. The *pseudo-dimension* of $\mathcal{F}$, denoted $\text{PD}(\mathcal{F})$, is the size of the largest set S which is shatterable by $\mathcal{F}$. We will derive sample complexity upper bounds from the following theorem, which connects the sample complexity of uniform learning over a class of real-valued functions to the pseudo-dimension of the class.

Theorem 6.5 (E.g. Anthony and Bartlett [1]). *Let $\mathcal{F}$ be a class of real-valued functions with range in $[0, H]$ and pseudo-dimension $\text{PD}(\mathcal{F})$. For every $\epsilon > 0, \delta \in [0, 1]$, the sample complexity of (ϵ, δ) -uniformly learning the class $\mathcal{F}$ is*

$$n = O\left(\left(\frac{H}{\epsilon}\right)^2 \left(\text{PD}(\mathcal{F}) \ln \frac{H}{\epsilon} + \ln \frac{1}{\delta}\right)\right).$$

Moreover, a conceptually simple algorithm achieves the guarantee in Theorem 6.5: simply output the function $c \in \mathcal{F}$ with the smallest empirical error on the sample. These algorithms are called *empirical risk minimizers*.

6.3 Learning from Pricings and Learning with Compression

We now introduce several other powerful tools from learning theory, recast in the language of mechanism design. Let $(v_1, B_1), \dots, (v_n, B_n), v_q \in \mathcal{V}, B_q \in \mathcal{G}$ represent samples drawn from some distribution Π over $\mathcal{V}$, labeled by bundles $B_q \in \mathcal{G}$. For each set of prices p , the functions h_p, f_p will map valuations to utility-maximizing bundles and the value the buyer of utility-maximizing bundles, respectively. We prove bounds on the sample complexity of uniform convergence over three classes of functions:

- 1) the class of functions which map valuations v_q to $v_q(B_q)$ (where B_q is utility-maximizing):

$$\mathcal{H}_V = \{f_p : \mathcal{V} \rightarrow \mathbb{R}, f_p(v_q) = v_q(B^*), B^* = e(\mathcal{D}_q(p))\};$$

- 2) the class of bundle predictors:

$$\mathcal{H} = \{h_p : \mathcal{V} \rightarrow \mathcal{G}, h_p(v_q) = e(\mathcal{D}_q(p))\};$$

- 3) the projection of $\mathcal{H}$ to its component *good* predictors:

$$\mathcal{H}_g = \{h_{g,p} : \mathcal{V} \rightarrow \{0, 1\}, h_p(v) = (h_{g,p}(v))_g\}.$$

We will show that we can learn the class $\mathcal{H}$ using a *compression scheme*, a tool for proving sample complexity bounds in the multi-label setting.

Definition 6.6. A compression scheme for $\mathcal{F} : \mathcal{V} \rightarrow X$, of size d consists of

- a compression function

$$\text{compress} : (\mathcal{V} \times X)^n \rightarrow (\mathcal{V} \times X)^d,$$

where $\text{compress}(N) \subseteq N$ and $d \leq n$; and

- a decompression function

$$\text{decompress} : (\mathcal{V} \times \mathcal{X})^d \rightarrow \mathcal{F}.$$

For any $f \in \mathcal{F}$ and any sample $(v_1, f(v_1)), \dots, (v_n, f(v_n))$, the functions satisfy

$$\text{decompress} \circ \text{compress}((v_1, f(v_1)), \dots, (v_n, f(v_n))) = f'$$

where $f'(v_q) = f(v_q)$ for each $q \in [n]$.

Intuitively, a compression function selects a subset of d most relevant points from a sample, and based on these points, the decompression scheme selects a hypothesis. When such a scheme exists, the learning algorithm **decompress** $\circ$ **compress** is an empirical risk minimizer. Furthermore, this compression-based learning algorithm has sample complexity bounded by a function of d , which plays a role analogous to VC dimension in the sample complexity guarantees.

Theorem 6.7 (Littlestone and Warmuth [30]). *Suppose $\mathcal{F}$ has a compression scheme of size d . Then, the PAC complexity of $\mathcal{F}$ is at most $n = O\left(\frac{d \ln \frac{1}{\epsilon} + \ln \frac{1}{\delta}}{\epsilon}\right)$.*

While compression schemes imply useful sample complexity bounds, it can be hard to show that a particular hypothesis class admits a compression scheme. One general technique is to show that the class is linearly separable in a higher-dimensional space.

Definition 6.8. *A class $\mathcal{F}$ is d -dimensionally linearly separable if there exists a function $\psi : \mathcal{V} \times \mathcal{G} \rightarrow \mathbb{R}^d$ such that for any $f \in \mathcal{F}$, there exists some $w_f \in \mathbb{R}^d$ with $f(v) \in \arg\max_y \langle w_f, \psi(v, y) \rangle$ and $|\arg\max_y \langle w_f, \psi(v, y) \rangle| = 1$.*

It is known that a d -dimensional linearly separable class admits a compression scheme of size d .

Theorem 6.9 (Daniely and Shalev-Shwartz [15]). *Suppose $\mathcal{F}$ has a d -dimensional realizable linear separator ψ . Then, there exists a compression scheme for $\mathcal{F}$ of size d .*

6.4 Over-Demand Concentration

In this section, we outline the proof of over-demand concentration (Theorem 6.1). The proof proceeds in two steps. First, we bound the VC dimension of the class of good predictors $\mathcal{H}_g$ by $O(m)$ (Theorem 6.10). For arbitrary valuations, the argument first shows that the class of bundle predictors $\mathcal{H}$ is $(m+1)$ -linearly separable and is thus PAC learnable with a sample size proportional to m . Since $h_{g,p} \in \mathcal{H}_g$ is a projection of $h \in \mathcal{H}$, $\mathcal{H}_g$ must also be PAC learnable with the same number of samples. Since functions in $\mathcal{H}_g$ are binary classifiers and PAC learning is completely characterized by VC dimension in the binary setting, we can bound the VC dimension of $\mathcal{H}_g$. Then, since $\mathcal{H}_g$ has small VC dimension, it is possible to bound the maximum difference between the sampled demand on N and the new demand on N' for any $h_{g,p} \in \mathcal{H}_g$ by $\tilde{O}(\text{VC}(\mathcal{H}_g) \cdot \sqrt{s_g})$ (Theorem 6.12, whose proof follows standard arguments for bounding PAC sample complexity, using Bernstein's inequality in place of Hoeffding's inequality to achieve an upper bound on error which scales as $\sqrt{s_g}$ rather than $\sqrt{n}$).

Theorem 6.10. *The VC dimension of $\mathcal{H}_g$ is at most $O(m)$.*

The proof of Theorem 6.10 relies on the following lemma about the linear separability of the class.

Lemma 6.11. *$\mathcal{H}$ is $(m+1)$ -linearly separable.*

Proof of Theorem 6.10. Lemma 6.11 states that $\mathcal{H}$, the class which predicts bundles, is $(m+1)$ -linearly separable. Then, Theorem 6.9 implies that there exists a compression scheme of size $m+1$ for $\mathcal{H}$.

Thus, one can PAC-learn $\mathcal{H}$ with at most $n = O\left(\frac{m \ln \frac{1}{\epsilon} + \ln \frac{1}{\delta}}{\epsilon}\right)$ samples, by Theorem 6.7. Since $\mathcal{H}_g$ is a projection of $\mathcal{H}$ onto its the g th coordinate, PAC-learning $\mathcal{H}_g$ cannot require more samples than PAC-learning $\mathcal{H}$, so this sample complexity bound on n also applies to $\mathcal{H}_g$. Then, Theorem 6.3 implies that the VC dimension of the binary class $\mathcal{H}_g$ satisfies

$$c \cdot \left(\frac{m + \ln \frac{1}{\delta}}{\epsilon}\right) \geq \frac{\text{VC}(\mathcal{H}_g) + \ln \frac{1}{\delta}}{\epsilon}$$

for some constant c . The claim follows by basic algebra. $\square$

We briefly sketch the proof of Lemma 6.11 (the full proof can be found in the extended version). In order to show $\mathcal{H}$ is $(m+1)$ -linearly separable, we must define two things. First, we define a mapping $\Psi : \mathcal{V} \times \mathcal{G} \rightarrow \mathbb{R}^{m+1}$ where $\Psi(v, B)_g$ encodes whether or not $g \in B$ for each good $g \in [m]$, and $\Psi(v, B)_{m+1} = v(B)$ encodes the buyer's value for a bundle B . Second, we define a weight vector $w^p \in \mathbb{R}^{m+1}$ for each price vector p where $w_g^p = -p_g$ encodes the *cost* of good $g \in [m]$ at these prices, while $w_{m+1}^p = 1$. Accordingly, the dot product of $\Psi(v, B)$ and w^p encodes the utility of a buyer v buying a bundle B at prices p : $\Psi(v, B) \cdot w^p = v(B) - p(B)$. Thus, a utility-maximizing bundle B^* maximizes the dot product $B^* \in \arg\max_B \Psi(v, B) \cdot w^p$.

Unfortunately, there are two obstacles with this plan of attack. First, the statement of Theorem 6.9 assumes that the maximum is *unique*—if not, prediction is not even well-defined. Second, in order to obtain welfare guarantees, we need some assumptions on how ties are broken, namely, that a buyer buys her canonical bundle. To solve both problems simultaneously, we describe how to perturb Ψ and w^p to ensure that the argmax is unique, and that $B^* = \arg\max_B (\Psi(v, B) \cdot w^p) = e(\arg\max_B v(B) - p(B))$; namely, that ties are broken appropriately by the linear mapping Ψ .

The second ingredient in proving the demand for good g on N is close to N' is the following theorem, which states that the empirical average of $\ell(f, \cdot)$ on a sample N should be close to the expectation of ℓ on the distribution Π for all functions $f \in \mathcal{F}$, for any function $\ell : \mathcal{F} \times (\mathcal{V}, \mathcal{G}) \rightarrow \{0, 1\}$.

Theorem 6.12. *Consider a hypothesis class $\mathcal{F}$, and $\ell(f, x) \in \{0, 1\}$. Let $N \sim \Pi$ be a sample such that $|N| = n$. Then, with probability $1 - \delta$, for all $f \in \mathcal{F}$,*

$$\ell_\Pi(f) - \ell_N(f) \leq \frac{3\text{VC}(\mathcal{F}) \ln \frac{1}{\delta} \sqrt{\ell_N(f)}}{\sqrt{n}}.$$

The bound on the distance depends on the size of the sample N as well as the VC dimension of the class $\mathcal{F}$. The proof of this theorem and the formal proof of Theorem 6.1 can be found in the appendix; the argument follows straightforwardly from Shalev-Shwartz and Ben-David [42] (specifically, Theorem 6.11 replacing the bound on the difference in losses for a fixed hypothesis by the bound in Lemma B.10).

6.5 Welfare Concentration

Somewhat unusually, our proof of Theorem 6.1 did not go through a combinatorial shattering argument to prove a bound on VC dimension. In contrast, our proof that the *welfare* of Walrasian prices generalizes (Theorem 6.2) relies on an explicit shattering argument. First, we bound the pseudo-dimension of the class of welfare predictors $\mathcal{H}_V$ (Theorem 6.13). This follows from a combinatorial shattering argument using the linear separator for $\mathcal{H}$. Then,

we show that classes with small pseudo-dimension yield strong concentration. Similar to the results for over-demand, the latter statement follows from a standard sample complexity argument using Bernstein’s inequality in place of Hoeffding’s inequality to get a multiplicative bound. We then show that optimal welfare is well-concentrated (Lemma 6.15). Thus, if the welfare on N' for prices p is close to the optimal welfare for N , then the welfare of N' at p must be nearly optimal for N' .

We begin by proving a bound on the pseudo-dimension of the welfare predictors $\mathcal{H}_V$.

Theorem 6.13. *The pseudo-dimension of $\mathcal{H}_V$ is $O(m^2 \ln m)$ for general valuations $\mathcal{V}$ and $O(m \ln^3 m)$ for unit demand valuations.*

The proof of Theorem 6.13 (see the extended version) relies on the following combinatorial lemma.

Lemma 6.14. *$\mathcal{H}$ can induce at most $\binom{n}{m+1} \cdot 2^{m(m+1)}$ bundle labelings of n sampled valuations, and at most $\binom{n}{m+1} \cdot m^{m+1}$ for unit demand buyers.*

Proof. Let $h_p(N) = (h_p(v_1), \dots, h_p(v_n))$ (by abuse of notation) represent the labeling of a sample N by h_p . Lemma 6.11 shows that $\mathcal{H}$ is $(m+1)$ -linearly separable. By Theorem 6.9, there exists a $(m+1)$ -sized compression scheme Ψ for learning the class $\mathcal{H}$. Thus, there exists a pair of functions (**decompress**, **compress**) such that for any price vector p , there exists $N_p \subseteq N$ with $|N_p| = m+1$ such that (i) **compress**($h_p(N)$) = $h_p(N_p)$, (ii) **decompress**(**compress**($h_p(N)$)) = **decompress**($h_p(N_p)$) = f , and $f(v_q) = B_q = h_p(v_q)$ for every $v_q \in N$.

We will bound the number of distinct labelings $\mathcal{H}$ can induce on a sample $N = (v_1, \dots, v_n)$ of size n . To do so, we the *decompression* function (which depends only on the separator Ψ) to upper-bound the total number of labelings $\mathcal{H}$ can induce on S .

Our approach will be to upper bound the size of the *range* of **decompress** $\circ$ **compress** when restricted to subsamples of a particular population N by some quantity D . By definition of a compression scheme, for every labeling of buyers in N by bundles that can be induced by some price vector, there is an element of the range of **decompress** $\circ$ **compress** that recovers that labeling on N . So, we can upper bound the total number of distinct labelings that can be induced on N by price vectors by upper bounding the range of **decompress** $\circ$ **compress** on N . The idea is that each *compression* is defined by a subset N' of at most $m+1$ of the buyers in N , and a labeling of the buyers with bundles. So, bounding the number of such labeled subsets suffices to bound the range of **decompress** $\circ$ **compress**.

We prove such a bound for $D = \binom{n}{m+1} \cdot (2^m)^{m+1}$. Fix any subset of unlabeled examples $N' \subseteq N$, where $|N'| = m+1$. For any $x \in N'$, there are at most 2^m labelings of x from $\mathcal{G}$. Thus, there are at most $(2^m)^{m+1} = 2^{m(m+1)}$ labelings of all of N' from $\mathcal{G}$.⁵ Thus, **decompress** $\circ$ **compress** can output at most $\binom{n}{m+1} \cdot 2^{m(m+1)}$ distinct functions based on labeled subsets of N . $\square$

Theorem 6.13, along with known results for uniform convergence over classes with small pseudo-dimension, implies that the welfare of any set of prices p on a sufficiently large sample is very close to the welfare p induces on a fresh sample of the same size. The next lemma shows that the *optimal* welfare allocation on the training and test samples have very similar welfare, since the optimal welfare is a well-concentrated quantity.

⁵If buyers are unit demand, we have a sharper bound: there are at most m labelings per $x \in N'$, or m^{m+1} labelings in total.

Lemma 6.15. *Recall that Opt-Welfare_N denotes the optimal welfare achievable by an allocation of a fixed set of goods in a market to buyers with valuations $N = (v_1, \dots, v_n)$ bounded in $[0, H]$. Then, the optimal welfare on two samples N, N' of size n drawn i.i.d. from distribution Π satisfies*

$$\Pr[|\text{Opt-Welfare}_N - \text{Opt-Welfare}_{N'}| \geq 4\epsilon] \leq \exp\left(-\frac{2\epsilon^2}{nH^2}\right)$$

for all $\epsilon \in (0, 1)$. In particular, $|\text{Opt-Welfare}_N - \text{Opt-Welfare}_{N'}| \leq H\sqrt{n \ln \frac{1}{\delta}}$ with probability at least $1 - \delta$.

The proof of Lemma 6.15 uses McDiarmid’s inequality, and can be found in the extended version. We now prove the main result of this section, which guarantees first that for any p , the welfare induced on N is similar to the welfare induced on N' . In particular, if p is the welfare-optimal pricing for N , then it is also nearly welfare-optimal for N' .

Finally, we sketch the proof of Theorem 6.2. The optimal welfare for the two markets must be close, by Lemma 6.15. The welfare of the pricing p output for the first market is nearly optimal even when over-demand is resolved adversarially, since there is over-demand of at most 1 for each good by assumption. By Theorem 6.13, the pseudo-dimension of the class is at most $m^2 \ln^2 m$ for general valuations and at most $m \ln^3 m$ for unit-demand valuations. A pseudo-dimension analogue of Theorem 6.12 implies that the welfare of applying p to the second market will be close to the welfare of applying p to the initial market.

References

- [1] Martin Anthony and Peter L Bartlett. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, 1999.
- [2] Kenneth J Arrow and Gerard Debreu. Existence of an equilibrium for a competitive economy. *Econometrica*, pages 265–290, 1954.
- [3] Itai Ashlagi, Mark Braverman, and Avinatan Hassidim. Ascending unit demand auctions with budget limits. Technical report, 2009.
- [4] Moshe Babaioff, Liad Blumrosen, Shaddin Dughmi, and Yaron Singer. Posting prices with unknown distributions. In *Symposium on Innovations in Computer Science, Beijing, China*. Tsinghua University Press, January 2011.
- [5] Maria-Florina Balcan, Avrim Blum, Jason D Hartline, and Yishay Mansour. Mechanism design via machine learning. In *IEEE Symposium on Foundations of Computer Science (FOCS), Pittsburgh, Pennsylvania*, pages 605–614, 2005.
- [6] Maria-Florina Balcan, Nikhil Devanur, Jason D Hartline, and Kunal Talwar. Random sampling auctions for limited supply. Technical report, Carnegie Mellon University, September 2007.
- [7] Maria-Florina Balcan, Avrim Blum, Jason D Hartline, and Yishay Mansour. Reducing mechanism design to algorithm design via machine learning. *Journal of Computer and System Sciences*, 74(8):1245–1270, 2008.
- [8] Maria-Florina Balcan, Amit Daniely, Ruta Mehta, Ruth Uner, and Vijay V Vazirani. Learning economic parameters from revealed preferences. In *Conference on Web and Internet Economics (WINE), Beijing, China*, pages 338–353, 2014.

- [9] Oren Ben-Zwi, Ron Lavi, and Ilan Newman. Ascending auctions and walrasian equilibrium. 2013.
- [10] Avrim Blum and Jason D Hartline. Near-optimal online auctions. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, Vancouver, British Colombia, pages 1156–1163, 2005.
- [11] Avrim Blum, Vijay Kumar, Atri Rudra, and Felix Wu. Online learning in online auctions. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, Baltimore, Maryland, 2003.
- [12] Mark Braverman, Jing Chen, and Sampath Kannan. Optimal provision-after-wait in healthcare. In *ACM SIGACT Innovations in Theoretical Computer Science (ITCS)*, Princeton, New Jersey, pages 541–542, 2014.
- [13] Nicolò Cesa-Bianchi, Claudio Gentile, and Yishay Mansour. Regret minimization for reserve prices in second-price auctions. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, New York, New York, pages 1190–1204, 2013.
- [14] Richard Cole and Tim Roughgarden. The sample complexity of revenue maximization. In *ACM SIGACT Symposium on Theory of Computing (STOC)*, New York, New York, pages 243–252, 2014.
- [15] Amit Daniely and Shai Shalev-Shwartz. Optimal learners for multiclass problems. In *Conference on Computational Learning Theory (CoLT)*, Barcelona, Spain, pages 287–316, 2014.
- [16] Xiaotie Deng, Christos Papadimitriou, and Shmuel Safra. On the complexity of equilibria. In *ACM SIGACT Symposium on Theory of Computing (STOC)*, Montréal, Québec, pages 67–71, 2002.
- [17] Nikhil Devanur, Jason Hartline, Anna Karlin, and Thach Nguyen. Prior-independent multi-parameter mechanism design. In *Conference on Web and Internet Economics (WINE)*, Singapore, pages 122–133, 2011.
- [18] Nikhil R Devanur and Thomas P Hayes. The Adwords problem: Online keyword matching with budgeted bidders under random permutations. In *ACM SIGecom Conference on Economics and Computation (EC)*, Stanford, California, pages 71–78, 2009.
- [19] Peerapong Dhangwatnotai, Tim Roughgarden, and Qiqi Yan. Revenue maximization with a single sample. *Games and Economic Behavior*, 2014.
- [20] Shaddin Dughmi, Li Han, and Noam Nisan. Sampling and representation complexity of revenue maximization. In *Conference on Web and Internet Economics (WINE)*, Beijing, China, pages 277–291, 2014.
- [21] Andrzej Ehrenfeucht, David Haussler, Michael Kearns, and Leslie Valiant. A general lower bound on the number of examples needed for learning. *Information and Computation*, 82(3):247–261, 1989.
- [22] Edith Elkind. Designing and learning optimal finite support auctions. In *ACM SIGecom Conference on Economics and Computation (EC)*, San Diego, California, pages 736–745, 2007.
- [23] Satoru Fujishige and Zaifu Yang. A note on Kelso and Crawford’s Gross Substitutes condition. *Mathematics of Operations Research*, 28(3):463–469, 2003.
- [24] Andrew V Goldberg, Jason D Hartline, Anna R Karlin, Michael Saks, and Andrew Wright. Competitive auctions. *Games and Economic Behavior*, 55(2):242–269, 2006.
- [25] Faruk Gul and Ennio Stacchetti. Walrasian equilibrium with gross substitutes. *Journal of Economic Theory*, 87(1):95–124, 1999.
- [26] Steve Hanneke. The optimal sample complexity of PAC learning. 2015.
- [27] Jason D Hartline and Tim Roughgarden. Simple versus optimal mechanisms. In *ACM SIGecom Conference on Economics and Computation (EC)*, Stanford, California, 2009.
- [28] Zhiyi Huang, Yishay Mansour, and Tim Roughgarden. Making the most of your samples. In *ACM SIGecom Conference on Economics and Computation (EC)*, Portland, Oregon, pages 45–60, 2015.
- [29] Alexander Kelso and Vincent Crawford. Job matching, coalition formation, and gross substitutes. *Econometrica*, 50(6):1483–1504, 1982.
- [30] Nick Littlestone and Manfred Warmuth. Relating data compression and learnability. Technical report, University of California, Santa Cruz, 1986.
- [31] Andres Munoz Medina and Mehryar Mohri. Learning theory and algorithms for revenue optimization in second price auctions with reserve. In *International Conference on Machine Learning (ICML)*, Beijing, China, pages 262–270, 2014.
- [32] Jamie Morgenstern and Tim Roughgarden. The pseudo-dimension of nearly-optimal auctions. In *Conference on Neural Information Processing Systems (NIPS)*, Montréal, Québec, pages 136–144, 2015.
- [33] Kenneth Mount and Stanley Reiter. The informational size of message spaces. *Journal of Economic Theory*, 8(2):161–192, 1974.
- [34] Kazuo Murota. Convexity and Steinitz’s exchange property. In William H Cunningham, S Thomas McCormick, and Maurice Queyranne, editors, *Conference on Integer Programming and Combinatorial Optimization (IPCO)*, Vancouver, British Columbia, volume 1084 of *Lecture Notes in Computer Science*, pages 260–274. Springer Berlin Heidelberg, 1996.
- [35] Kazuo Murota. Valuated matroid intersection I: Optimality criteria. *SIAM Journal on Discrete Mathematics*, 9(4):545–561, 1996.
- [36] Kazuo Murota and Akiyoshi Shioura. M-Convex function on generalized polymatroid. *Mathematics of Operations Research*, 24(1):95–105, 1999.
- [37] Noam Nisan and Ilya Segal. The communication requirements of efficient allocations and supporting prices. *Journal of Economic Theory*, 129(1):192–224, 2006.
- [38] Michael Ostrovsky and Renato Paes Leme. Gross substitutes and endowed assignment valuations. *Theoretical Economics*, 10(3):853–865, September 2015.
- [39] James G Oxley. *Matroid Theory*. Oxford Graduate Texts in Mathematics. Oxford University Press, 1997.

- [40] David Pollard. *Convergence of stochastic processes*. David Pollard, New Haven, Connecticut, 1984.
- [41] Tim Roughgarden and Okke Schrijvers. Ironing in the dark. Under submission, 2015.
- [42] Shai Shalev-Shwartz and Shai Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014.
- [43] Vladimir N Vapnik and A Ya Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability & Its Applications*, 16(2):264–280, 1971.