STATISTICAL INFERENCE FOR FISHER MARKET EQUILIBRIUM

Anonymous authors
Paper under double-blind review

ABSTRACT

Statistical inference under market equilibrium effects has attracted increasing attention recently. In this paper we focus on the specific case of linear Fisher markets. They have been widely use in fair resource allocation of food/blood donations and budget management in large-scale Internet ad auctions. In resource allocation, it is crucial to quantify the variability of the resource received by the agents (such as blood banks and food banks) in addition to fairness and efficiency properties of the systems. For ad auction markets, it is important to establish statistical properties of the platform's revenues in addition to their expected values. To this end, we propose a statistical framework based on the concept of infinite-dimensional Fisher markets. In our framework, we observe a market formed by a finite number of items sampled from an underlying distribution (the "observed market") and aim to infer several important equilibrium quantities of the underlying long-run market. These equilibrium quantities include individual utilities, social welfare, and pacing multipliers. Through the lens of sample average approximation (SAA), we derive a collection of statistical results and show that the observed market provides useful statistical information of the long-run market. In other words, the equilibrium quantities of the observed market converge to the true ones of the long-run market with strong statistical guarantees. These include consistency, finite sample bounds, asymptotics, and confidence. As an extension we discuss revenue inference in quasilinear Fisher markets.

1 INTRODUCTION

In a Fisher market there is a set of n buyers that are interested in buying goods from a distinct seller. A market equilibrium (ME) is then a set of prices for the goods, along with a corresponding allocation, such that demand equals supply.

One important application of market equilibrium (ME) is fair allocation using the competitive equilibrium from equal incomes (CEEI) mechanism (Varian, 1974). In CEEI, each individual is given an endowment of faux currency and reports her valuations for items; then, a market equilibrium is computed, and the items are allocated accordingly. The resulting allocation has many desirable properties such as Pareto optimality, envy-freeness and proportionality. For example, Fisher market equilibrium has been used for fair work allocation, impressions allocation in certain recommender systems, course seat allocation and scarce computing resources allocation; see Appendix A for an extensive overview.

Despite numerous algorithmic results available for computing Fisher market equilibria, to the best of our knowledge, no statistical results were available for quantifying the randomness of market equilibrium. Given that CEEI is a fair and efficient mechanism, such statistical results are useful for quantifying variability in CEEI-based resource allocation. For example, for systems that assign blood donation to hospitals and blood banks (McElfresh et al., 2020), or donated food to charities in different neighborhoods (Aleksandrov et al., 2015; Sinclair et al., 2022), it is crucial to quantify the variability of the amount of resources (blood or food donation) received by the participants (hospitals or charities) of these systems as well as the variability of fairness and efficiency metrics of interest in the long run. Making statistical statements about these metrics is crucial for both evaluating and improving these systems.

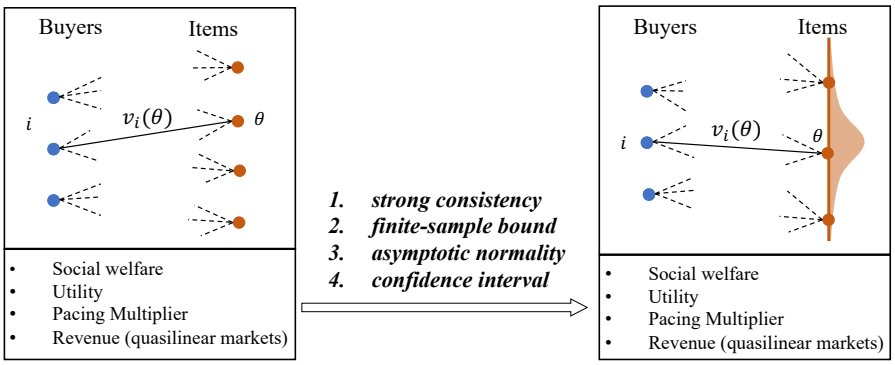

Figure 1: Our contributions. Left panel: a Fisher market with a finite number of items. Right panel: a Fisher market with a continuum of items. Middle arrow: this paper provides various forms of statistical guarantees to characterize the convergence of observed finite Fisher market (left) to the long-run market (right) when the items are drawn from a distribution corresponding to the supply function in the long-run market.

In addition to fair resource allocation, statistical results for Fisher markets can also be used in revenue inference in Internet ad auction markets. While much of the existing literature uses expected revenue as performance metrics, statistical inference on revenue is challenging due to the complex interaction among bidders under coupled supply constraints and common price signals. As shown by Conitzer et al. (2022a), in budget management through repeated first-price auctions with pacing, the optimal pacing multipliers correspond to the "prices-per-utility" of buyers in a quasilinear Fisher market at equilibrium. Given the close connection between various solution concepts in Fisher market models and first-price auctions, a statistical framework enables us to quantify the variability in long-run revenue of an advertising platform. Furthermore, a statistical framework would also help answer other statistical questions such as the study of counterfactuals and theoretical guarantees for A/B testing in Internet ad auction markets.

For a detailed survey on related work in the areas of statistical inference, applications of Fisher market models, and equilibrium computation algorithms, see Appendix A.

Our contributions are as follows.

A statistical Fisher market model. We formulate a statistical estimation problem for Fisher markets based on the continuous-item model of Gao and Kroer (2022). We show that when a finite set of goods are sampled from the continuous model, the observed ME is a good approximation of the long-run market. In particular, we develop consistency results, finite-sample bounds, central limit theorems, and asymptotically valid confidence interval for various quantities of interests, such as individual utility, Nash social welfare, pacing multipliers, and revenue (for quasilinear Fisher markets).

Technical challenges. In developing central limit theorems for pacing multipliers and utilities in Fisher markets (Theorem 5), we note that the dual objective is potentially not twice differentiable. This is a required condition, which is common in the sample average approximation or M-estimation literature. We discover three types of market where such differentiability is guaranteed. Moreover, the sample function is not differentiable, which requires us to verify a set of stochastic differentiability conditions in the proofs for central limit theorems. Finally, we achieve a fast statistical rate of the empirical pacing multiplier to the population pacing multiplier measured in the dual objective by exploiting the local strong convexity of the sample function.

Notation. For a sequence of events A_n we define the set limit by $\liminf_{n\to\infty} A_n = \bigcup_{n\geq 1}\bigcap_{j\geq n} A_j = \{A_t \text{ eventually}\}$ and $\limsup_{n\to\infty} A_n = \bigcap_{n\geq 1}\bigcup_{j\geq n} A_j = \{A_t \text{ i.o.}\}$. For vector $a, b \in \mathbb{R}^n$ we let $a{\cdot}b$ be the elementwise product and let $a_{-i} \in \mathbb{R}^{n-1}$ be the vector a with the i-th entry removed. For vector a we let $[a_{(1)}, \ldots, a_{(n)}]$ denote the sorted entries of a from greatest to least. Let $[n] = \{1, \ldots, n\}$. We use 1_t to denote the vector of ones of length t and e_j to denote the vector with one in the j-th entry and zeros in the others. For a sequence of random variables $\{X_n\}$, we say $X_n = O_p(1)$ if for any $\epsilon >$ there exists a finite M_ϵ and a finite N_ϵ such that $\mathbb{P}(|X_n| > M_\epsilon) < \epsilon$ for all $n \geq N_\epsilon$. We say $X_n = O_p(a_n)$ if $X_n/a_n = O_p(1)$. We use subscript

for indexing buyers and superscript for items. If a function f is twice continuously differentiable at a point x, we say f is C^2 at x.

2 Problem Setup

2.1 The Estimands

Following Gao and Kroer (2022), we consider a Fisher market with n buyers (individuals), each having a budget $b_i > 0$ and a (possibly continuous) set of items Θ. We let L^p (and L^p_+, resp.) denote the set of (nonnegative, resp.) L^p functions on Θ w.r.t the integrating measure μ for any $p \in [1, \infty]$ (including $p = \infty$). For example, one could take $\Theta = [0, 1]$ and $\mu =$ the Lebesgue measure on $[0, 1]$. The item *supplies* are given by a function $s \in L^\infty_+$, i.e., item $\theta \in \Theta$ has supply $s(\theta)$. The *valuation* for buyer i is a function $v_i \in L^1_+$, i.e., buyer i has valuation $v_i(\theta)$ for item $\theta \in \Theta$. For buyer i, an *allocation* of items $x_i \in L^\infty_+$ gives a utility of

$$u_i(x_i) := \langle v_i, x_i \rangle := \int_\Theta v_i(\theta) x_i(\theta) \, \mathrm{d}\mu(\theta),$$

where the angle brackets are based on the notation of applying a bounded linear functional x_i to a vector v_i in the Banach space L^1 and the integral is the usual Lebesgue integral. We will use $x \in (L^\infty_+)^n$ to denote the aggregate allocation of items to all buyers, i.e., the concatenation of all buyers' allocations. The *prices* of items are modeled as $p \in L^1_+$. The price of item $\theta \in \Theta$ is $p(\theta)$. Without loss of generality, we assume a unit total supply $\int_\Theta s \, \mathrm{d}\mu = 1$. We let $S(A) := \int_A s(\theta) \, \mathrm{d}\mu(\theta)$ be the probability measure induced by the supply s.

Imagine there is a central policymaker that sets the prices $p(\cdot)$ of items Θ. Upon receiving the price signal, the buyer i maximizes his utility $\langle v_i, x_i \rangle$ subject to the budget constraint $\langle p, x_i \rangle \le b_i$. He would demand a bundle of items coming from his demand set

$$D_i(p) := \arg\max_{x_i} \{ \langle v_i, x_i \rangle : x_i \in L^\infty_+, \ \langle p, x_i \rangle \le b_i \}.$$

Of course, due to the supply constraint, if prices are too low, there will be a shortage in supply. On the other hand, if prices are too high, a surplus occurs. Market equilibrium is the case when items are sold out exactly.

Definition 1 (The long-run market equilibrium)**.** *The market equilibrium (ME) $\mathscr{ME}(b, v, s)$ is an allocation-utility-price tuple $(x^*, u^*, p^*) \in (L^\infty_+)^n \times \mathbb{R}^n_+ \times L^1_+$ such that the following holds. (i) Supply feasibility and market clearance: $\sum_i x_i^* \le s$ and $\langle p^*, s - \sum_i x_i^* \rangle = 0$. (ii) Buyer optimality: $x_i^* \in D_i(p^*)$ and $u^* = \langle v_i, x_i \rangle$ for all i.*

Linear Fisher market equilibrium can be characterized by convex programs. We state the following result from Gao and Kroer (2022) which establishes existence and uniqueness of market equilibrium, and more importantly the convex program formulation of the equilibrium. We define the Eisenberg-Gale (EG) convex programs which as we will see are dual to each other.

$$\max_{x \in L^\infty_+(\Theta), u \ge 0} \left\{ \mathrm{NSW}(u) := \sum_{i=1}^n b_i \log(u_i) \;\middle|\; u_i \le \langle v_i, x_i \rangle \ \ \forall i \in [n], \ \ \sum_{i=1}^n x_i \le s \right\}, \tag{P-EG}$$

$$\min_{\beta > 0} \left\{ H(\beta) := \int_\Theta \Big(\max_{i \in [n]} \beta_i v_i(\theta) \Big) S(\mathrm{d}\theta) - \sum_{i=1}^n b_i \log \beta_i \right\}. \tag{P-DEG}$$

Concretely, the optimal primal variables in Eq. (P-EG) corresponds to the set of equilibrium allocations x^* and the unique equilibrium utilities u^*, and the unique optimal dual variable β^* of Eq. (P-DEG) relates to the equilibrium through

$$u_i^* = b_i / \beta_i^* \ , \quad p^*(\theta) = \max_i \beta_i^* v_i(\theta) \ , \quad x_i^*(\theta) > 0 \text{ only if } i \in \arg\max_i \beta_i^* v_i(\theta) \ .$$

We call β^* the *pacing multiplier*. Note equilibrium allocations might not be unique but equilibrium utilities and prices are unique. Given the above equivalence result, we use (x^*, u^*) to denote both the equilibrium and the optimal variables. Another feature of linear Fisher market is full budget extraction: $\int p^* \, \mathrm{d}S = \sum_{i=1}^n b_i$; we discuss quasilinear model in Appendix F.

We formally state the first-order conditions of infinite-dimensional EG programs and its relation to first-price auctions in Fact 1 in appendix. Also, we remark that there are two ways to specify the valuation component in this model: the functional form of $v_i(\cdot)$, or the distribution of values $v : \Theta \to \mathbb{R}^n_+$ when view as a random vector. More on this in Appendix E.

We are interested in estimating the following quantities of the long-run market equilibrium. (1) **Individual utilities** at equilibrium, u_i^*. (2) **Pacing multipliers** $\beta_i^* = b_i/u_i^*$. Pacing multiplier has a two-fold interpretation. Second, through the equation $p^*(\theta) = \max_i \beta_i^* v_i(\theta)$, β can also be interpreted as the *pacing policy*[1] employed by the buyers in first-price auctions. In our context, buyer i produces a bid for item θ by multiplying the value by β_i, then the item price is determined via a first-price auction. This connection is made precise in Conitzer et al. (2022a) from a game-theoretic point of view. The pacing multiplier β serves as the bridge between Fisher market equilibrium and first price pacing equilibria in auction games (Conitzer et al., 2022a) and has important usage in budget management in online auctions. (3) The (logarithm of) **Nash social welfare** (NSW) at equilibrium $\text{NSW}^* := \text{NSW}(u^*)$. NSW measures total utility of the buyers and, when used as an optimization objective, is able to promote fairness better than the social welfare, that is, the sum of buyer utilities (Caragiannis et al., 2019). Intuitively, NSW incentivizes more balancing of buyer utilities. Revenue inference for quasilinear model is discussed in Appendix F[2].

Mapping model to concrete applications. It is well-known that Fisher market is a useful mechanism for fair and efficient resource allocation. More recently, it is also shown to be intimately related to first-price auctions (Conitzer et al., 2022a). When modeling ad auction platforms, buyers' individual utilities reflect, for example, the values generated to advertisers, measured in terms of click rates, conversion rates, or other revenue metrics. A confidence interval on this quantity can be provided by the ad platform to the advertisers to better inform advertisers' decision-making. Pacing multiplier, as the ratio between the bid and the value of items, reflects the advertisers' bidding strategy. A confidence interval on this quantity could help the ad platform predict its clients bidding behavior and suggest budget management strategy to clients. Finally, Nash social welfare measures the efficiency and fairness of the whole ad platform. We can also use Fisher market to model resource allocation systems in internet companies. For example, in a job recommendation platform, we can model it as a market where we distribute viewer's attention to job post creators. In this context, social welfare measures generally the efficiency of the job recommendation system. Individual utilities track how satisfied the job post creators are with the extent to which their job posts are being recommended.

2.2 The Data

Assume we are able to observe a market formed by a finite number of items. We let $\gamma = \{\theta_1, \ldots, \theta_t\} \subset \Theta^t$ be a set of items sampled i.i.d. from the supply distribution S. We let $v_i(\gamma) = \big(v_i(\theta^1), \ldots, v_i(\theta^t)\big)$ denote the valuation for agent i of items in the set γ. For agent i, let $x_i = (x_i^1, \ldots, x_i^t) \in \mathbb{R}^t$ denote the fraction of items given to agent i. With this notation, the total utility of agent i is $\langle x_i, v_i(\gamma)\rangle$. Similar to the long-run market, we assume the observed market is at equilibrium, which we now define.

Definition 2 (Observed Market Equilibrium)**.** *The market equilibrium $\mathscr{M}\mathscr{E}^\gamma(b, v, \mathsf{s})$ given the item set γ and the supply vector $\mathsf{s} \in \mathbb{R}^t_+$ is an allocation-utitlity-price tuple $(x^\gamma, u^\gamma, p^\gamma) \in (\mathbb{R}^t_+)^n \times \mathbb{R}^n_+ \times \mathbb{R}^t_+$ such that the following holds. (i) Supply feasibility and market clearance: $\sum_{i=1}^n x_i^\gamma \le \mathsf{s}$ and $\langle p^\gamma, 1_t - \sum_{i=1}^n x_i^\gamma\rangle = 0$. (ii) Buyer optimality: $x_i^\gamma \in D_i(p^\gamma)$ and $u_i^\gamma = \langle v_i(\gamma), x_i\rangle$ for all i, where (overloading notations)*

$$D_i(p) := \arg\max_{x_i}\{\langle v_i(\gamma), x_i\rangle : x_i \ge 0,\ \langle p, x_i\rangle \le b_i\}$$

is the demand set given the prices and the buyer's budget.

[1] In the online budget management literature, pacing means buyers produce bids for items via multiplying his value by a constant.

[2] In a linear Fisher market the budgets are extracted fully, i.e., $\int p^* \,\mathrm{d}S = \sum_i b_i$ in the long-run market and $\sum_{\tau=1}^t p^{\gamma,\tau} = \sum_i b_i$ in the observed market (see Appendix E), and therefore there is nothing to infer about revenue in this case. However, in the quasilinear utility model where buyer's utility function is $u_i(x) = \langle x - p, v_i\rangle$, buyers have the incentive to retain money and therefore one needs to study the statistical properties of revenues.

Assume we have access to $(x^\gamma, u^\gamma, p^\gamma)$ along with the budget vector b, where $(x^\gamma, u^\gamma, p^\gamma) \in \mathscr{ME}^\gamma(b, v, \frac{1}{t}1_t)$ is the market equilibrium (we explain the scaling of $1/t$ in Appendix E). Note the budget vector b and value functions $v = \{v_i(\cdot)\}_i$ are the same as those in the long-run ME. We emphasize two high-lights in this model of observation.

No convex program solving. The quantities observed are natural estimators of their counterparts in the long-run market, and so we do not need to perform iterative updates or solve optimization problems. One interpretation of this is that the actual computation is done when equilibrium is reached via the utility maximizing property of buyers; the work of computation has thus implicitly been delegated to the buyers.

For finite-dimensional Fisher market, it is well-known that the observed market equilibrium $\mathscr{ME}^\gamma(b, v, \frac{1}{t}1_t)$ can be captured by the following sample EG programs.

$$\max_{x\geq 0, u\geq 0} \left\{ \mathrm{NSW}(u) \;\middle|\; u_i \leq \langle v_i(\gamma), x_i \rangle \;\forall i, \sum_{i=1}^n x_i^\tau \leq \tfrac{1}{t}1_t \;\forall \tau \right\}, \tag{S-EG}$$

$$\min_{\beta>0} \left\{ H_t(\beta) := \frac{1}{t}\sum_{\tau=1}^t \max_{i\in[n]} \beta_i v_i(\theta^\tau) - \sum_{i=1}^n b_i \log \beta_i \right\}. \tag{S-DEG}$$

We list the KKT conditions in Appendix E. Completely parallel to the long-run market, optimal solutions to Eq. (S-EG) correspond to the equilibrium allocations and utilities, and the optimal variable β^γ to Eq. (S-DEG) relates to equilibrium prices and utilities through $u_i^\gamma = b_i/\beta_i^\gamma$, $p^{\gamma,\tau} = \max_i \beta_i^\gamma v_i(\theta^\tau)$ and $x_i^{\gamma,\tau} > 0$ only if $i \in \arg\max_i \beta_i^\gamma v_i(\theta^\tau)$. By the equivalence between market equilibrium and EG programs, we use u^γ and x^γ to denote the equilibrium and the optimal variables. Let $\mathrm{NSW}^\gamma := \mathrm{NSW}(u^\gamma) = \sum_{i=1}^n b_i \log u_i^\gamma$. All budgets in the observed market is extracted, i.e., $\sum_{\tau=1}^t p^{\gamma,\tau} = \sum_{i=1}^n b_i$.

2.3 Dual Programs: Bridging Data and the Estimands

Given the convex program characterization, a natural idea is to study the concentration behavior of observed market equilibria through these convex programs. We focus on the dual programs Eqs. (S-DEG) and (P-DEG) because they are defined in a fixed dimension, and that the constraint set is also fixed.

Define the sample function $F = f + \Psi$, where $f(\beta, \theta) = \max_i\{v_i(\theta)\beta_i\}$, and $\Psi(\beta) = -\sum_{i=1}^n b_i \log \beta_i$; the function f is the source of non-smoothness, while Ψ provides local strong convexity. Our approach is studying concentration of the convex programs in the sense that as t grows

$$\min_{\beta>0} H_t(\beta) = \frac{1}{t}\sum_{\tau=1}^t F(\beta, \theta^\tau) \quad \text{“}\Longrightarrow\text{”} \quad \min_{\beta>0} H(\beta) = \mathbb{E}[F(\beta, \theta)]\,.$$

The local strong convexity of the dual objective motivates us to do the analysis work in the neighborhood of the optimal solution β^*. In particular, the function $x \mapsto -\log x$ is not strongly convex on the positive reals, but it is on any compact subset. By working on a compact subset, we can exploit strong convexity of the dual objective and obtain better theoretical results. It is known (Fact 1) that $\underline{\beta}_i \leq \beta_i^* \leq \bar{\beta}$ where $\underline{\beta}_i = b_i / \int v_i \,\mathrm{d}S$ and $\bar{\beta} = \sum_{i=1}^n b_i / \min_i \int v_i \,\mathrm{d}S$. Define the compact set $C := \prod_{i=1}^n \left[\underline{\beta}_i/2, 2\bar{\beta}\right] \subset \mathbb{R}^n$, which must be a neighborhood of β^*. Moreover, for large-enough t we further have $\beta^\gamma \in C$ with high probability (c.f. Lemma 1).

Blanket assumptions. Recall the total supply in the long-run market is one: $\int s \,\mathrm{d}\mu = 1$. Assume the total item set produce one unit of utility in total, i.e., $\int v_i s \,\mathrm{d}\mu = 1$. Suppose budges of all buyers sum to one, i.e., $\sum_{i=1}^n b_i = 1$. Let $\underline{b} := \min_i b_i$. Note the previous budget normalization implies $\underline{b} \leq 1/n$. Finally, for easy of exposition, we assume the values are bounded $\sup_\Theta v_i(\theta) < \bar{v}$, for all i. By the normalization of values and budgets, we know $\underline{\beta}_i = b_i$ and $\bar{\beta} = 1$.

3 Consistency and Finite-Sample Bounds

In this section we introduce several natural empirical estimators based on the observed market equilibrium, and show that they satisfy both consistency and high-probability bounds. Below we state the consistency results; the formal version can be found in Appendix B.

Theorem 1 (Consistency, informal)**.** *The Nash social welfare, pacing multiplier and approximate equilibria in the observed market are strongly consistent estimators of their counterparts in the long-run market.*

High Probability Bounds Next, we refine the consistency results and provide finite sample guarantees. We start by focusing on Nash social welfare and the set of approximate market equilibria. The convergence of utilities and pacing multiplier will then be derived from the latter result.

Theorem 2. *For any failure probability $0 < \eta < 1$, let $t \geq 2\bar{v}^2 \log(4n/\eta)$. Then with probablity greater than $1 - \eta$, it holds $|\text{NSW}^\gamma - \text{NSW}^*| \leq O(1)\bar{v}\big(\sqrt{n\log((n+\bar{v})t)} + \sqrt{\log(1/\eta)}\big)t^{-1/2}$ where $O(1)$ hides only constants. Proof in Appendix I.*

Theorem 2 establishes a convergence rate $|\text{NSW}^\gamma - \text{NSW}^*| = \tilde{O}_p(\bar{v}\sqrt{n}t^{-1/2})$. The proof proceeds by first establishing a pointwise concentration inequality and then applies a discretization argument.

Theorem 3 (Concentration of Approximate Market Equilibrium)**.** *Let $\epsilon > 0$ be a tolerance parameter and $\alpha \in (0,1)$ be a failure probability. Then for any $0 \leq \delta \leq \epsilon/2$, to ensure $\mathbb{P}\big(C \cap \mathscr{B}^\gamma(\delta) \subset C \cap \mathscr{B}^*(\epsilon)\big) \geq 1 - 2\alpha$ it suffices to set*

$$t \geq O(1)(n^2 + \bar{v}^2) \min\left\{\frac{1}{\underline{b}\epsilon}, \frac{1}{\epsilon^2}\right\}\left(n \log\left(\tfrac{16(2n+\bar{v})}{\epsilon - \delta}\right) + \log\tfrac{1}{\alpha}\right), \tag{1}$$

where the set $C = \prod_{i=1}^n [\underline{\beta}_i/2, 2\bar{\beta}]$, and $O(1)$ hides only absolute constants. Proof in Appendix J.

By construction of C we know $\beta^* \in C$ holds, and so $C \cap \mathscr{B}^*(\epsilon)$ is not empty. By Lemma 1 we know that for t sufficiently large, $\beta^\gamma \in C$ with high probability, in which case the set $C \cap \mathscr{B}^\gamma(\delta)$ is not empty.

Corollary 1. *Let t satisfy Eq. (1). Then with probability $\geq 1 - 2\alpha$ it holds $H(\beta^\gamma) \leq H(\beta^*) + \epsilon$.*

By simply taking $\delta = 0$ in Theorem 3 we obtain the above corollary. More importantly, it establishes the fast statistical rate $H(\beta^\gamma) - H(\beta^*) = \tilde{O}_p(t^{-1})$ for t sufficiently large, where we use $\tilde{O}_p$ to ignore logarithmic factors. In words, when measured in the population dual objective where we take expectation w.r.t. the item supply, β^γ converges to β^* with the fast rate $1/t$. This is in contrast to the usual $1/\sqrt{t}$ rate obtained in Theorem 2, where β^γ is measured in the sample dual objective. There the $1/\sqrt{t}$ rate is the best obtainable.

By the strong-convexity of dual objective, the containment result can be translated to high-probability convergence of the pacing multipliers and the utility vector.

Corollary 2. *Let t satisfy Eq. (1). Then with probability $\geq 1 - 2\alpha$ it holds $\|\beta^\gamma - \beta^*\|_2 \leq \sqrt{\frac{8\epsilon}{\underline{b}}}$ and $\|u^\gamma - u^*\|_2 \leq \frac{4}{\underline{b}}\sqrt{8\epsilon/\underline{b}}$.*

We compare the above corollary with Theorem 9 from Gao and Kroer (2022) which establishes the convergence rate of the stochastic approximation estimator based on dual averaging algorithm (Xiao, 2010). In particular, they show that the average of the iterates, denoted β_{DA}, enjoys a convergence rate of $\|\beta_{\text{DA}} - \beta^*\|_2^2 = \tilde{O}_p\big(\frac{\bar{v}^2}{\underline{b}^2}\frac{1}{t}\big)$, where t is the number of sampled items. The rate achieved in Corollary 2 is $\|\beta^\gamma - \beta^*\|_2^2 = \tilde{O}_p\big(\frac{n(n^2+\bar{v}^2)}{\underline{b}^2}\frac{1}{t}\big)$ for t sufficiently large. Noting $n \leq \underline{b}^{-1}$ due to the normalization $\sum_{i=1}^n b_i = 1$, we see that our rate is worse off by a factor of $n(1 + \frac{n^2}{\bar{v}^2})$. And yet our estimates is produced by the strategic behavior of the agents without any extra computation at all. Moreover, in the computation of the dual averaging estimator the knowledge of values $v_i(\theta)$ is required, while again β^γ can be just observed naturally.

4 Asymptotics and Inference

4.1 Asymptotics and Analytical Properties of the Dual Objective

In this section we derive asymptotic normality results for Nash social welfare, utilities and pacing multipliers. As we will see, a central limit theorem (CLT) for Nash social welfare holds under

basically no additional assumptions. However, the CLTs of pacing multipliers and utilities will require twice continuous differentiability of the population dual objective H, with a nonsingular Hessian matrix. We present CLT results under such a premise, and then provide three sufficient conditions under which H is C^2 at the optimum.

Theorem 4 (Asymptotic Normality of Nash Social Welfare)**.** *It holds that*

$$\sqrt{t}(\mathrm{NSW}^\gamma - \mathrm{NSW}^*) \xrightarrow{\mathrm{d}} \mathscr{N}(0, \sigma^2_{\mathrm{NSW}}) \ , \tag{2}$$

where $\sigma^2_{\mathrm{NSW}} = \int_\Theta (p^*)^2 \,\mathrm{d}S(\theta) - \big(\int_\Theta p^* \,\mathrm{d}S(\theta)\big)^2 = \int_\Theta (p^*)^2 \,\mathrm{d}S(\theta) - 1$. *Proof in Appendix K.*

To present asymptotics for β and u we need a bit more notation. Let $\Theta_i(\beta) := \{\theta \in \Theta : v_i(\theta)\beta_i \geq v_k(\theta)\beta_k, \forall k \neq i\}$, i.e., the *potential* winning set of buyer i when the pacing multiplier are β. Let $\Theta_i^* := \Theta_i(\beta^*)$. We will see later that if the dual objective is sufficiently smooth at β^*, then the winning sets, $\Theta_i^*, i \in [n]$, will be disjoint (up to a measure-zero set). Define the *variance of winning values* for buyer i as follows

$$\Omega_i^2 = \int_{\Theta_i^*} v_i^2(\theta) \,\mathrm{d}S(\theta) - \left(\int_{\Theta_i^*} v_i(\theta) \,\mathrm{d}S(\theta)\right)^2 .$$

Theorem 5 (Asymptotic Normality of Individual Behavior)**.** *Assume H is C^2 at β^* with non-singular Hessian matrix $\mathscr{H} = \nabla^2 H(\beta^*)$. Then $\sqrt{t}(\beta^\gamma - \beta^*) \xrightarrow{\mathrm{d}} \mathscr{N}\big(0, \Sigma_\beta\big)$ and $\sqrt{t}(u^\gamma - u^*) \xrightarrow{\mathrm{d}} \mathscr{N}\big(0, \Sigma_u\big)$, where $\Sigma_\beta = \mathscr{H}^{-1}\mathrm{Diag}(\{\Omega_i^2\}_{i=1}^n)\mathscr{H}^{-1}$ and $\Sigma_u = \mathrm{Diag}(\{-b_i/(\beta_i^*)^2\})\mathscr{H}^{-1}$ $\mathrm{Diag}(\{\Omega_i^2\}_{i=1}^n)\mathscr{H}^{-1}\mathrm{Diag}(\{-b_i/(\beta_i^*)^2\})$. Proof in Appendix K.*

In Theorem 5 we require a strong regularity condition: twice differentiability of H, which seems hard to interpret at first sight. Next we derive a set of simpler sufficient conditions for the twice differentiability of the dual objective. Intuitively, the expectation operator will smooth out the kinks in the piecewise linear function $f(\cdot, \theta)$; even if f is non-smooth, it is reasonable to hope the expectation counterpart $\bar{f}$ is smooth, facilitating statistical analysis.

First we introduce notation for characterizing smoothness of $\bar{f}$. Define the gap between the highest and the second-highest bid under pacing multiplier β by

$$\epsilon(\beta, \theta) := \big(v(\theta)\cdot\beta\big)_{(1)} - \big(v(\theta)\cdot\beta\big)_{(2)} \ , \tag{3}$$

here $v(\theta)\cdot\beta$ is the elementwise product of $v(\theta)$ and β, and $(v(\theta)\cdot\beta)_{(1)}$ and $(v(\theta)\cdot\beta)_{(2)}$ are the greatest and second-greatest entries of $v(\theta)\cdot\beta$, respectively. When there is a tie for an item θ, we have $\epsilon(\beta, \theta) = 0$. When there is no tie for an item θ, the gap $\epsilon(\beta, \theta)$ is strictly positive. Let $G(\beta, \theta) \in \partial f(\beta, \theta)$ be an element in the subgradient set. The gap function characterizes smoothness of f: $f(\cdot, \theta)$ is differentiable at $\beta \Leftrightarrow \epsilon(\beta, \theta)$ is strictly positive, in which case $G(\beta, \theta) = \nabla_\beta f(\beta, \theta) = e_{i(\beta,\theta)} v_{i(\beta,\theta)}$ with e_i being the i-th unit vector and $i(\beta, \theta) = \arg\max_i \beta_i v_i(\theta)$. When $f(\cdot, \theta)$ is differentiable at β a.s., the potential winning sets $\{\Theta_i(\beta)\}_i$ are disjoint (up to a measure-zero set).

Theorem 6 (First-order differentiability)**.** *The dual objective H is differentiable at a point β if and only if*

$$\frac{1}{\epsilon(\beta, \theta)} < \infty, \quad \textit{for } S\textit{-almost every } \theta \ . \tag{NO-TIE}$$

When Eq. (NO-TIE) holds, $\nabla \bar{f}(\beta) = \mathbb{E}[G(\beta, \theta)]$. Proof and further technical remarks in Appendix L.

Given the neat characterization of differentiability of dual objective via the gap function $\epsilon(\beta, \theta)$, it is then natural to explore higher-order smoothness, which was needed for some asymptotic normality results. We provide three classes of markets whose dual objective H enjoys twice differentiability.

Theorem 7 (Second-order differentiability, Informal)**.** *If any one of the following holds, then H is C^2 at β^*. (i) A stronger form of Eq. (NO-TIE) holds, e.g., $\mathbb{E}[\epsilon(\beta, \epsilon)^{-1}]$ or $\operatorname{ess\,sup}_\theta\{\epsilon(\beta, \theta)^{-1}\}$ is finite in a neighborhood of β^*. (ii) The distribution of $v = (v_1, \ldots, v_n) : \Theta \to \mathbb{R}_+^n$ is smooth enough. (iii) $\Theta = [0, 1]$ and the valuations $v_i(\cdot)$'s are linear functions.*

4.2 INFERENCE

In this section we discuss constructing confidence intervals for Nash social welfare, the pacing multipliers, and the utilities. We remark that the observed NSW, NSW^γ, is a negatively-biased estimate of the NSW, NSW^*, of the long-run ME, i.e., $\mathbb{E}[\text{NSW}^\gamma] - \text{NSW}^* \leq 0$.[3] Moreover, it can be shown that, when the items are i.i.d. $\mathbb{E}[\min H_t] \leq \mathbb{E}[\min H_{t+1}]$ using Proposition 16 from Shapiro (2003). Monotonicity tells us that increasing the size of market produces on average less biased estimates of the long-run NSW.

To construct a confidence interval for Nash social welfare one needs to estimate the asymptotic variance. We let $\hat{\sigma}^2_{\text{NSW}} := \frac{1}{t}\sum_{\tau=1}^t \big(F(\beta^\gamma, \theta^\tau) - H_t(\beta^\gamma)\big)^2 = \big(\frac{1}{t}\sum_{\tau=1}^t (p^{\gamma,\tau})^2\big) - 1$. where $p^{\gamma,\tau}$ is the price of item θ^τ in the observed market. We emphasize that in the computation of the variance estimator $\hat{\sigma}^2_{\text{NSW}}$ one does not need knowledge of values $\{v_i(\theta^\tau)\}_{i,\tau}$. All that is needed is the equilibrium prices $p^\gamma = (p^{\gamma,1}, \ldots, p^{\gamma,t})$ of the items. Given the variance estimator, we construct the confidence interval $[\text{NSW}^\gamma \pm z_{\alpha/2}\frac{\hat{\sigma}_{\text{NSW}}}{\sqrt{t}}]$, where z_α is the α-th quantile of a standard normal. The next theorem establishes validity of the variance estimator.

Theorem 8. *It holds that* $\hat{\sigma}_{\text{NSW}} \xrightarrow{\text{P}} \sigma^2_{\text{NSW}}$. *Given* $0 < \alpha < 1$, *it holds that* $\lim_{t\to\infty} \mathbb{P}\big(\text{NSW}^* \in [\text{NSW}^\gamma \pm z_{\alpha/2}\hat{\sigma}_{\text{NSW}}/\sqrt{t}\,]\big) = 1 - \alpha$. *Proof in Appendix M.*

Estimation of the variance matrices for β and u is more complicated. The main difficulty lies in estimating the inverse Hessian matrix. Due to the non-smoothness of the sample function, we cannot exchange the twice differential operator and expectation, and thus the plug-in estimator, i.e., the sample average Hessian, is a biased estimator for the Hessian of the population function in general. In Appendix D we provide a brief discussion of variance estimation under the following two simplified scenarios. First, in the case where $\mathbb{E}[\epsilon(\beta, \theta)^{-1}] < \infty$ holds in a neighborhood of β^*, which we recall is a stronger form Eq. (NO-TIE), we prove that a plug-in type variance estimator is valid. Second, if we have knowledge of $\{v_i(\theta^\tau)\}_{i,\tau}$, then we give a numerical difference estimator for the Hessian which is consistent.

5 GUIDE TO PRACTITIONERS

We close our theoretical development with practical instructions on how to construct confidence interval with results presented so far. Revenue inference is only well-defined in the context of quasilinear market, which we explore in Appendix F.

Inference on NSW. Recall b_i is the budget of agent i and let u_i^γ be the observed utility. Let $\text{NSW}^\gamma = \sum_{i=1}^n b_i \log u_i^\gamma$ be the NSW in the observed market. This will be a good estiamte of the limit NSW and as the number of items grows, it converges quickly at a rate of $1/\sqrt{t}$. To construct a CI we need to estimate the asymptotic variance. Let $\hat{\sigma}^2_{\text{NSW}} = \frac{1}{t}\sum_{\tau=1}^t (p^{\gamma,\tau} - \bar{p}^\gamma)^2$ where $p^{\gamma,\tau}$ is the price of item θ^τ and $\bar{p}^\gamma$ is the average price. In words, $\hat{\sigma}^2_{\text{NSW}}$ is the variance of prices. Then a $(1-\alpha)$ CI for the NSW in the long-run market will be $[\text{NSW}^\gamma \pm z_{\alpha/2}\hat{\sigma}_{\text{NSW}}/\sqrt{t}]$, where z_α is the $(1-\alpha)$-th quantile of a standard normal.

Inference on β^* and u^*: simplified inference with a bid-gap condition. Suppose in the ad auction platform one observes that for each item the winning bid is always higher than the swecond bid by some amount, then we could use the following inference procedure. For a precise statement of the condition see Theorem 10. When such a condition holds, the variance expression in Theorem 5 simplifies. Let $\hat{\Omega}_i$ be the variance of item utility that are allocated to buyer i, i.e., $\hat{\Omega}_i^2 := \frac{1}{t}\sum_{\tau=1}^t (tu_i^{\gamma,\tau} - u_i^\gamma)^2$, where $u_i^{\gamma,\tau} = x_i^{\gamma,\tau} v_i(\theta^\tau)$ is the utility buyer i obtains from item θ^τ. Define the asymptotic variance estimators $\hat{\Sigma}_\beta = \text{Diag}(\{\hat{\Omega}_i^2 (\beta_i^\gamma)^4 / b_i^2\})$ and $\hat{\Sigma}_u = \text{Diag}(\{\hat{\Omega}_i^2\})$. The confidence regions for β^* and u^* are ellipsoids that center around β^γ and u^γ. Concretely, $\text{CR}_\beta = \beta^\gamma + (\chi_{n,\alpha}/\sqrt{t})\hat{\Sigma}_\beta^{1/2}\mathbb{B}$ and the confidence region for u^* is $\text{CR}_u = u^\gamma + (\chi_{n,\alpha}/\sqrt{t})\hat{\Sigma}_u^{1/2}\mathbb{B}$, where $\chi_{k,\alpha}$ is the $(1-\alpha)$-th quantile of a chi-square with degree k and $\mathbb{B}$ is the unit ball in $\mathbb{R}^n$.

Inference on β^* and u^*: a robust approach. Under fairly general setting, which is justified in Section 4.1, the following inference procedure is valid. Choose a smoothing level η_t, to be

[3] Note $\mathbb{E}[\text{NSW}^\gamma] - \text{NSW}^* = \mathbb{E}[\min_\beta H_t(\beta)] - H(\beta^*) \leq \min_\beta \mathbb{E}[H_t(\beta)] - H(\beta^*) = 0$.

chosen roughly of order $\eta_t = \omega(1/\sqrt{t})$, we construct the following matrix $\hat{\mathscr{H}}$ whose (i,j)-th entry is $(\hat{\mathscr{H}})_{ij} := \frac{1}{4\eta_t^2}(H_t(\beta^\gamma + \eta_t(e_i + e_j)) - H_t(\beta^\gamma + \eta_t(-e_i + e_j)) - H_t(\beta^\gamma + \eta_t(e_i - e_j)) + H_t(\beta^\gamma + \eta_t(-e_i - e_j)))$. This will be an estimate of the Hessian matrix of the dual objective. Here H_t defined in Eq. (S-EG) needs only values of the realized items but not the value functions $v_i(\cdot)$'s. Let $\hat{\Sigma}_\beta = \hat{\mathscr{H}}^{-1}\mathrm{Diag}(\{\hat{\Omega}_i^2\}_{i=1}^n)\hat{\mathscr{H}}^{-1}$ and $\hat{\Sigma}_u = \mathrm{Diag}(\{b_i/(\beta_i^\gamma)^2\})\hat{\mathscr{H}}^{-1}\mathrm{Diag}(\{\hat{\Omega}_i^2\}_{i=1}^n)$ $\hat{\mathscr{H}}^{-1}\mathrm{Diag}(\{b_i/(\beta_i^\gamma)^2\})$ be the estimated covariance matrices. We can then construct confidence regions CR_β and CR_u as above.

Bootstrap. Though not studied in this paper, bootstrap inference is a valuable alternative. In a bootstrap procedure, one samples the items with replacement and solve the EG programs (S-EG) or (S-DEG). By repeating this we obtain a sequence of estimates say $\beta^1, \ldots, \beta^B$, where B is the number of bootstrap batches. Then a confidence interval is constructed by the sample quantiles of the sequence $\{\beta^k\}_{k\in[B]}$.

6 Experiments

We conduct experiments to validate the theoretical findings, namely, the convergence of NSW^γ to NSW^* (Theorem 9) and CLT (Eq. (2)). All figures can be found in Appendix O.

Verify convergence of NSW to its infinite-dimensional counterpart in a linear Fisher market. First, we generate an infinite-dimensional market $\mathscr{M}_1$ of $n = 50$ buyers each having a linear valuation $v_i(\theta) = a_i\theta + c_i$ on $\Theta = [0,1]$, with randomly generated a_i and c_i such that $v_i(\theta) \geq 0$ on $[0,1]$. Their budgets b_i are also randomly generated. We solve for NSW^* using the tractable convex conic formulation described in Gao and Kroer (2022, Section 4). Then, following Section 2.2, for the j-th ($j \in [k]$) sampled market of size t, we randomly sample $\{\theta_j^{t,\tau}\}_{\tau\in[t]}$ uniformly and independently from $[0,1]$ and obtain markets with n buyers and t items, with individual valuations $v_i(\theta_j^{t,\tau}) = a_i\theta_j^{t,\tau} + c_i$, $j \in [t]$. We take $t = 100, 200, \ldots, 5000$ and $k = 10$. We compute their equilibrium Nash social welfare, i.e., NSW^γ, and their means and standard errors over k repeats across all t. As can be seen from Fig. 2, NSW^γ values quickly approach NSW^*, which align with the a.s. convergence of NSW^γ in Theorem 9. Moreover, NSW^γ values increase as t increase, which align with the monotonicity observation in the beginning of Section 4.2.

Verify asymptotic normality of NSW in a linear Fisher market. Next, for the same infinite-dimensional market $\mathscr{M}_1$, we set $t = 5000$, sample $k = 50$ markets of t items analogously, and compute their respective NSW^γ values. We plot the enpirical distribution of $\sqrt{t}(\mathrm{NSW}^\gamma - \mathrm{NSW}^*)$ and the probability density of $\mathscr{N}(0, \sigma^2_{\mathrm{NSW}})$, where σ^2_{NSW} is defined in Theorem 4.[4] Theorem 4 shows that $\sqrt{t}(\mathrm{NSW}^\gamma - \mathrm{NSW}^*) \xrightarrow{\mathrm{d}} \mathscr{N}(0, \sigma^2_{\mathrm{NSW}})$. As can be seen in Fig. 5, the empirical distribution is close to the limiting normal distribution. A simple Kolmogorov-Smirnov test shows that the empirical distribution appears normal, that is, the alternative hypothesis of it not being a normal distribution is not statistically significant. This is further corroborated by the Q-Q plot in Fig. 4, as the plots of the quantiles of $\sqrt{t}(\mathrm{NSW}^\gamma - \mathrm{NSW}^*)$ values against theoretical quantiles of $\mathscr{N}(0, \sigma^2_{\mathrm{NSW}})$ appear to be a straight line.

Verify NSW convergence in a multidimensional linear Fisher market. Finally, we consider an infinite-dimensional market $\mathscr{M}_2$ with multidimensional linear valuations $v_i(\theta) = a_i^\top\theta + c_i$, $a_i \in \mathbb{R}^{10}$. We similarly sample markets of sizes $t = 100, 200, ..., 5000$ from $\mathscr{M}_2$, where the items θ_j^t, $j \in [k]$ are sampled uniformly and independently from $[0,1]^{10}$. As can be seen from Fig. 2, NSW^γ values increase and converge to a fixed value around -1.995. In this case, the underlying true value NSW^* (which should be around -1.995) cannot be captured by a tractable optimization formulation.

[4] To compute σ^2_{NSW}, we use the fact that $p^* = \max_i \beta_i^* v_i(\theta)$ is a piecewise linear function, since v_i are linear. Following (Gao and Kroer, 2022, Section 4), we can find the breakpoints of the pure equilibrium allocation $0 = a_0 < a_1 < \cdots < a_{50} = 1$, and the corresponding interval of each buyer i. Then, $\int_0^1 (p^*(\theta))^2 dS(\theta)$ amounts to integrals of quadratic functions on intervals.

REFERENCES

Martin Aleksandrov, Haris Aziz, Serge Gaspers, and Toby Walsh. Online fair division: Analysing a food bank problem. *arXiv preprint arXiv:1502.07571*, 2015.

Amine Allouah, Christian Kroer, Xuan Zhang, Vashist Avadhanula, Anil Dania, Caner Gocmen, Sergey Pupyrev, Parikshit Shah, and Nicolas Stier. Robust and fair work allocation, 2022. URL https://arxiv.org/abs/2202.05194.

Peter M Aronow and Cyrus Samii. Estimating average causal effects under general interference, with application to a social network experiment. *The Annals of Applied Statistics*, 11(4):1912–1947, 2017.

Susan Athey, Dean Eckles, and Guido W Imbens. Exact p-values for network interference. *Journal of the American Statistical Association*, 113(521):230–240, 2018.

Yossi Azar, Niv Buchbinder, and Kamal Jain. How to allocate goods in an online market? *Algorithmica*, 74(2):589–601, 2016.

Siddhartha Banerjee, Vasilis Gkatzelis, Artur Gorokh, and Billy Jin. Online nash social welfare maximization with predictions. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1–19. SIAM, 2022.

Heinz H Bauschke, Patrick L Combettes, et al. *Convex analysis and monotone operator theory in Hilbert spaces*, volume 408. Springer, 2011.

Amir Beck. *First-Order Methods in Optimization*, volume 25. SIAM, 2017.

Xiaohui Bei, Jugal Garg, and Martin Hoefer. Ascending-price algorithms for unknown markets. *ACM Transactions on Algorithms (TALG)*, 15(3):1–33, 2019a.

Xiaohui Bei, Jugal Garg, Martin Hoefer, and Kurt Mehlhorn. Earning and utility limits in fisher markets. *ACM Transactions on Economics and Computation (TEAC)*, 7(2):1–35, 2019b.

Dimitri P Bertsekas. Stochastic optimization problems with nondifferentiable cost functionals. *Journal of Optimization Theory and Applications*, 12(2):218–231, 1973.

Benjamin Birnbaum, Nikhil R Devanur, and Lin Xiao. Distributed algorithms via gradient descent for fisher markets. In *Proceedings of the 12th ACM Conference on Electronic Commerce*, pages 127–136. ACM, 2011.

Christian Borgs, Jennifer Chayes, Nicole Immorlica, Kamal Jain, Omid Etesami, and Mohammad Mahdian. Dynamics of bid optimization in online advertisement auctions. In *Proceedings of the 16th international conference on World Wide Web*, pages 531–540, 2007.

Eric Budish. The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes. *Journal of Political Economy*, 119(6):1061–1103, 2011.

Eric Budish, Gérard P Cachon, Judd B Kessler, and Abraham Othman. Course match: A large-scale implementation of approximate competitive equilibrium from equal incomes for combinatorial allocation. *Operations Research*, 65(2):314–336, 2016.

Ioannis Caragiannis, David Kurokawa, Hervé Moulin, Ariel D Procaccia, Nisarg Shah, and Junxing Wang. The unreasonable fairness of maximum Nash welfare. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pages 305–322. ACM, 2016.

Ioannis Caragiannis, David Kurokawa, Hervé Moulin, Ariel D Procaccia, Nisarg Shah, and Junxing Wang. The unreasonable fairness of maximum nash welfare. *ACM Transactions on Economics and Computation (TEAC)*, 7(3):1–32, 2019.

Sarah H Cen and Devavrat Shah. Regret, stability & fairness in matching markets with bandit learners. In *International Conference on Artificial Intelligence and Statistics*, pages 8938–8968. PMLR, 2022.

Lihua Chen, Yinyu Ye, and Jiawei Zhang. A note on equilibrium pricing as convex optimization. In *International Workshop on Web and Internet Economics*, pages 7–16. Springer, 2007.

Yun Kuen Cheung, Richard Cole, and Nikhil R Devanur. Tatonnement beyond gross substitutes? Gradient descent to the rescue. *Games and Economic Behavior*, 123:295–326, 2020.

Richard Cole and Lisa Fleischer. Fast-converging tatonnement algorithms for one-time and ongoing market problems. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 315–324, 2008.

Richard Cole and Vasilis Gkatzelis. Approximating the Nash social welfare with indivisible items. *SIAM Journal on Computing*, 47(3):1211–1236, 2018.

Richard Cole, Nikhil R Devanur, Vasilis Gkatzelis, Kamal Jain, Tung Mai, Vijay V Vazirani, and Sadra Yazdanbod. Convex program duality, fisher markets, and Nash social welfare. In *18th ACM Conference on Economics and Computation, EC 2017*. Association for Computing Machinery, Inc, 2017.

Vincent Conitzer, Christian Kroer, Debmalya Panigrahi, Okke Schrijvers, Nicolas E Stier-Moses, Eric Sodomka, and Christopher A Wilkens. Pacing equilibrium in first price auction markets. *Management Science*, 2022a.

Vincent Conitzer, Christian Kroer, Eric Sodomka, and Nicolas E Stier-Moses. Multiplicative pacing equilibria in auction markets. *Operations Research*, 70(2):963–989, 2022b.

Xiaowu Dai and Michael Jordan. Learning in multi-stage decentralized matching markets. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 12798–12809. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper/2021/file/6a571fe98a2ba453e84923b447d79cff-Paper.pdf.

Xiaotie Deng, Christos Papadimitriou, and Shmuel Safra. On the complexity of price equilibria. *Journal of Computer and System Sciences*, 67(2):311–324, 2003.

Nikhil R Devanur, Christos H Papadimitriou, Amin Saberi, and Vijay V Vazirani. Competitive equilibrium via a primal-dual algorithm for a convex program. *Journal of the ACM (JACM)*, 55 (5):1–18, 2008.

Nikhil R Devanur, Jugal Garg, Ruta Mehta, Vijay V Vaziranb, and Sadra Yazdanbod. A new class of combinatorial markets with covering constraints: Algorithms and applications. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2311–2325. SIAM, 2018.

Edmund Eisenberg. Aggregation of utility functions. *Management Science*, 7(4):337–350, 1961.

Edmund Eisenberg and David Gale. Consensus of subjective probabilities: The pari-mutuel method. *The Annals of Mathematical Statistics*, 30(1):165–168, 1959.

Yuan Gao and Christian Kroer. First-order methods for large-scale market equilibrium computation. In *Neural Information Processing Systems 2020, NeurIPS 2020*, 2020.

Yuan Gao and Christian Kroer. Infinite-dimensional fisher markets and tractable fair division. *Operation Research, Forthcoming*, 2022.

Yuan Gao, Christian Kroer, and Alex Peysakhovich. Online market equilibrium with application to fair division. *arXiv preprint arXiv:2103.12936*, 2021.

Ali Ghodsi, Matei Zaharia, Benjamin Hindman, Andy Konwinski, Scott Shenker, and Ion Stoica. Dominant resource fairness: Fair allocation of multiple resource types. In *Nsdi*, volume 11, pages 24–24, 2011.

Artur Gorokh, Siddhartha Banerjee, and Krishnamurthy Iyer. The remarkable robustness of the repeated fisher market. *Available at SSRN 3411444*, 2019.

Wenshuo Guo, Kirthevasan Kandasamy, Joseph E Gonzalez, Michael I Jordan, and Ion Stoica. Online learning of competitive equilibria in exchange economies. *arXiv preprint arXiv:2106.06616*, 2021.

Nils Lid Hjort and David Pollard. Asymptotics for minimisers of convex processes. *arXiv preprint arXiv:1107.3806*, 2011.

Yuchen Hu, Shuangning Li, and Stefan Wager. Average direct and indirect causal effects under interference. *Biometrika*, 2022.

Michael G Hudgens and M Elizabeth Halloran. Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482):832–842, 2008.

Sungjin Im, Janardhan Kulkarni, and Kamesh Munagala. Competitive algorithms from competitive equilibria: Non-clairvoyant scheduling under polyhedral constraints. *Journal of the ACM (JACM)*, 65(1):1–33, 2017.

Meena Jagadeesan, Alexander Wei, Yixin Wang, Michael Jordan, and Jacob Steinhardt. Learning equilibria in matching markets from bandit feedback. *Advances in Neural Information Processing Systems*, 34:3323–3335, 2021.

Kamal Jain. A polynomial time algorithm for computing an arrow–debreu market equilibrium for linear utilities. *SIAM Journal on Computing*, 37(1):303–318, 2007.

Ian Kash, Ariel D Procaccia, and Nisarg Shah. No agent left behind: Dynamic fair division of multiple resources. *Journal of Artificial Intelligence Research*, 51:579–603, 2014.

Sujin Kim, Raghu Pasupathy, and Shane G. Henderson. *A Guide to Sample Average Approximation*, pages 207–243. Springer New York, New York, NY, 2015. doi: 10.1007/978-1-4939-1384-8_8.

Christian Kroer and Alexander Peysakhovich. Scalable fair division for 'at most one' preferences. *arXiv preprint arXiv:1909.10925*, 2019.

Christian Kroer and Nicolas E Stier-Moses. Market equilibrium models in large-scale internet markets. *Innovative Technology at the Interface of Finance and Operations. Springer Series in Supply Chain Management. Springer Natures, Forthcoming*, 2022.

Christian Kroer, Alexander Peysakhovich, Eric Sodomka, and Nicolas E Stier-Moses. Computing large competitive equilibria using abstractions. *Operations Research*, 2021.

Michael P Leung. Treatment and spillover effects under network interference. *Review of Economics and Statistics*, 102(2):368–380, 2020.

Shuangning Li and Stefan Wager. Random graph asymptotics for treatment effect estimation under network interference. *The Annals of Statistics*, 50(4):2334–2358, 2022.

Luofeng Liao, Yuan Gao, and Christian Kroer. Nonstationary dual averaging and online fair allocation. *arXiv preprint arXiv:2202.11614v1*, 2022.

Lydia T Liu, Feng Ruan, Horia Mania, and Michael I Jordan. Bandit learning in decentralized matching markets. *J. Mach. Learn. Res.*, 22:211–1, 2021.

Zhihan Liu, Miao Lu, Zhaoran Wang, Michael Jordan, and Zhuoran Yang. Welfare maximization in competitive equilibrium: Reinforcement learning for markov exchange economy. In *International Conference on Machine Learning*, pages 13870–13911. PMLR, 2022.

Duncan C McElfresh, Christian Kroer, Sergey Pupyrev, Eric Sodomka, Karthik Abinav Sankararaman, Zack Chauvin, Neil Dexter, and John P Dickerson. Matching algorithms for blood donation. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 463–464, 2020.

Yifei Min, Tianhao Wang, Ruitu Xu, Zhaoran Wang, Michael I Jordan, and Zhuoran Yang. Learn to match with no regret: Reinforcement learning in markov matching markets. *arXiv preprint arXiv:2203.03684*, 2022.

Evan Munro, Stefan Wager, and Kuang Xu. Treatment effects in market equilibrium. *arXiv preprint arXiv:2109.11647*, 2021.

Riley Murray, Christian Kroer, Alex Peysakhovich, and Parikshit Shah. Robust market equilibria with uncertain preferences. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 2192–2199, 2020a.

Riley Murray, Christian Kroer, Alex Peysakhovich, and Parikshit Shah. https://research.fb.com/blog/2020/09/robust-market-equilibria-how-to-model-uncertain-buyer-preferences/, Sep 2020b.

Yurii Nesterov and Vladimir Shikhman. Computation of fisher–gale equilibrium by auction. *Journal of the Operations Research Society of China*, 6(3):349–389, 2018.

Whitney K Newey and Daniel McFadden. Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245, 1994.

Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V Vazirani. *Algorithmic game theory*. Cambridge University Press, 2007.

Abraham Othman, Tuomas Sandholm, and Eric Budish. Finding approximate competitive equilibria: efficient and fair course allocation. In *AAMAS*, volume 10, pages 873–880, 2010.

Abraham Othman, Christos Papadimitriou, and Aviad Rubinstein. The complexity of fairness through equilibrium. *ACM Transactions on Economics and Computation (TEAC)*, 4(4):1–19, 2016.

David C Parkes, Ariel D Procaccia, and Nisarg Shah. Beyond dominant resource fairness: Extensions, limitations, and indivisibilities. *ACM Transactions on Economics and Computation (TEAC)*, 3(1):1–22, 2015.

Alexander Peysakhovich and Christian Kroer. Fair division without disparate impact. *Mechanism Design for Social Good*, 2019.

R Tyrrell Rockafellar. *Convex analysis*, volume 18. Princeton university press, 1970.

R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

Roshni Sahoo and Stefan Wager. Policy learning with competing agents. *arXiv preprint arXiv:2204.01884*, 2022.

Alexander Shapiro. Asymptotic properties of statistical estimators in stochastic programming. *The Annals of Statistics*, 17(2):841–858, 1989.

Alexander Shapiro. Monte carlo sampling methods. *Handbooks in operations research and management science*, 10:353–425, 2003.

Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczynski. *Lectures on stochastic programming: modeling and theory*. SIAM, 2021.

Vadim I Shmyrev. An algorithm for finding equilibrium in the linear exchange model with fixed budgets. *Journal of Applied and Industrial Mathematics*, 3(4):505, 2009.

Sean R Sinclair, Siddhartha Banerjee, and Christina Lee Yu. Sequential fair allocation: Achieving the optimal envy-efficiency tradeoff curve. In *Abstracts of the 2022 SIGMETRICS/Performance Joint International Conference on Measurement and Modeling of Computer Systems*, 2022.

Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

Hal R Varian. Equity, envy, and efficiency. *Journal of Economic Theory*, 9(1):63–91, 1974.

Vijay V. Vazirani. *Combinatorial Algorithms for Market Equilibria*, pages 103–134. Cambridge University Press, 2007. doi: 10.1017/CBO9780511800481.007.

Stefan Wager and Kuang Xu. Experimenting in equilibrium. *Management Science*, 67(11):6694–6715, November 2021. doi: 10.1287/mnsc.2020.3844. URL https://doi.org/10.1287/mnsc.2020.3844.

Jinde Wang. Distribution sensitivity analysis for stochastic programs with complete recourse. *Mathematical Programming*, 31(3):286–297, 1985.

Fang Wu and Li Zhang. Proportional response dynamics leads to market equilibrium. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 354–363, 2007.

Lin Xiao. Dual averaging methods for regularized stochastic learning and online optimization. *Journal of Machine Learning Research*, 11:2543–2596, 2010.

Yinyu Ye. A path to the arrow–debreu competitive market equilibrium. *Mathematical Programming*, 111(1):315–348, 2008.

Li Zhang. Proportional response dynamics in the fisher farket. *Theoretical Computer Science*, 412 (24):2691–2698, 2011.

A Related Work

Our paper is related to following lines of research.

Statistical inference for SAA. Asymptotics of sample average function minimizers are well-studied by the stochastic programming community (see, e.g., Shapiro et al. (2021, Chapter 5), Shapiro (2003) and Kim et al. (2015)) and the statistics community (see, e.g., Van der Vaart (2000) and Newey and McFadden (1994)). Despite the powerful tools developed by researchers, the problem of developing asymptotics for the convex EG program exhibits special challenges. The sample function consists of two parts: a non-smooth part coming from a piecewise linear function, and a locally-strongly convex part, which as we will see comes from the function $x \mapsto -\log x$. The non-smoothness of sample function requires us to investigate sufficient conditions for the second-order differentiability of the expected function and verify several technical regularity conditions, both of which are key hypotheses for most asymptotic normality of minimizers of non-smooth sample function. Moreover, the strong convexity of the sample function requires us to develop sharp finite-sample guarantees that exploit the strong convexity structure.

Applications of Fisher Market Equilibrium Fisher market equilibrium is related to a game-theoretic solution concept called pacing equilibrium which is a useful model for online ad auction platforms (Borgs et al., 2007; Conitzer et al., 2022b;a). In addition to ad auction markets, Fisher market equilibrium model has other usages in the tech industry, such as the allocation of impressions to content in certain recommender systems (Murray et al., 2020b), robust and fair work allocation in content review (Allouah et al., 2022). We refer readers to Kroer and Stier-Moses (2022) for a comprehensive review. Outside the tech industry, Fisher market equilibria also have applications to scheduling problems (Im et al., 2017), fair course seat allocation (Othman et al., 2010; Budish et al., 2016), allocating donations to food banks (Aleksandrov et al., 2015), sharing scarce compute resources (Ghodsi et al., 2011; Parkes et al., 2015; Kash et al., 2014; Devanur et al., 2018), and allocating blood donations to blood banks (McElfresh et al., 2020).

The statistical framework developed in this paper provides a guideline to quantify the uncertainty in equilibrium-based allocations in the above-mentioned applications.

Algorithmic Results for Fisher Market The problem of equilibrium computation has been of interest in economics for a long time (see, e.g., Nisan et al. (2007)). There is a large literature focusing on computation of equilibrium in Fisher markets through combinatorial algorithms (Vazirani (2007); Devanur et al. (2008); Jain (2007); Ye (2008); Deng et al. (2003)), convex optimization formulations (Eisenberg and Gale, 1959; Eisenberg, 1961; Shmyrev, 2009; Cole et al., 2017), gradient-based methods (Wu and Zhang, 2007; Zhang, 2011; Aleksandrov et al., 2015; Birnbaum et al., 2011; Nesterov and Shikhman, 2018; Gao and Kroer, 2020), tâtonnement process-based methods (Borgs et al., 2007; Bei et al., 2019a; Cole and Fleischer, 2008; Cheung et al., 2020), and abstraction methods (Kroer et al., 2021). Extensions to settings such as quasilinear utilities (Chen et al., 2007; Cole et al., 2017), limited utilities (Bei et al., 2019b), indivisible items (Cole and Gkatzelis, 2018), or

imperfectly specified utility functions (Caragiannis et al., 2016; Murray et al., 2020a; Kroer and Peysakhovich, 2019; Peysakhovich and Kroer, 2019) are also available. Several works study fair online allocation of divisible goods (Azar et al., 2016; Sinclair et al., 2022; Banerjee et al., 2022; Liao et al., 2022) and indivisible goods (Budish, 2011; Othman et al., 2016; Gorokh et al., 2019) by Fisher market equilibrium-based methods.

Most related to our work is Gao and Kroer (2022), where the authors extend the classical Fisher market model to a measurable (possibly continuous) item space and shows that infinite-dimensional EG-type convex programs capture ME under this setting. This paper proposes a statistical model based on their infinite-dimensional Fisher market and investigate the statistical inference problem.

Statistical Learning and Inference in Equilibrium Models Wager and Xu (2021); Munro et al. (2021); Sahoo and Wager (2022) take a mean-field game modeling approach and perform policy learning with a gradient descent method. In particular, Munro et al. (2021) study the causal effects of binary intervention on the supply-demand market equilibrium. Wager and Xu (2021) study the effect of supply-side payments on the market equilibrium. Sahoo and Wager (2022) study the learning of capacity-constrained treatment assignment while accounting for strategic behavior of agents. Different from the above mean-field modeling papers, in the linear or quaislinear Fisher market models, equilibrium concept is defined for a finite number of agents, allowing us to avoid a mean-field modeling approach. Moreover, the Fisher markets equilibria we study are captured by convex programs, so we can leverage well-established tools from the stochastic programming and the empirical processes literature. Finally, in Fisher market equilibria, a concrete parametric model of demand is imposed as opposed to previous works that take a more or less nonparametric approach, and therefore we could obtain results that characterize each agent's behavior, e.g., a central limit result for as market size grows (c.f. Theorem 5).

By a different group of researchers, the question of statistical learning and inference has been investigated for other equilibrium models, such as general exchange economy (Guo et al., 2021; Liu et al., 2022) and matching markets (Cen and Shah, 2022; Dai and Jordan, 2021; Liu et al., 2021; Jagadeesan et al., 2021; Min et al., 2022). Our paper focuses on a specific type of exchange economy called infinite-dimensional Fisher market, which is a model for the long-run market behavior.

Our work is also related to the rich literature of inference under interference (Hudgens and Halloran, 2008; Aronow and Samii, 2017; Athey et al., 2018; Leung, 2020; Hu et al., 2022; Li and Wager, 2022). In the Fisher market model, the interference among agents is caused by the supply constraint and the utility-maximizing behavior of agents given the price signal. In other words, in Fisher markets we put a parametric model on the interference structure which allows us to derive a rich collection of results.

B Consistency

We define approximate market equilibria. For any utility vector u achieved by a feasible allocation, we define $\beta_u = [\frac{b_1}{u_1}, \dots, \frac{b_n}{u_n}]$. We say that a utility vector u is an ϵ-approximate equilibrium utility vector if $H_t(\beta_u) \leq \inf_\beta H_t(\beta) + \epsilon$. It can be shown that for any feasible utilities u, we have $H_t(\beta_u) \geq H_t(\beta^\gamma)$, and u is the equilibrium utility vector if and only if $H_t(\beta_u) = H_t(\beta^\gamma)$. To that end, let

$$\mathscr{B}^\gamma(\epsilon) := \{\beta > 0 : H_t(\beta) \leq \inf_\beta H_t(\beta) + \epsilon\},\ \mathscr{B}^*(\epsilon) := \{\beta > 0 : H(\beta) \leq \inf_\beta H(\beta) + \epsilon\}\ . \quad (4)$$

be the sets of ϵ-approximate solutions to Eqs. (S-DEG) and (P-DEG), respectively.

Thanks to the convexity of the dual objectives H and H_t, we can provide a set of consistency results based on the theory of epi-convergence (Rockafellar and Wets, 2009).

Theorem 9 (Consistency)**.** *It holds that*

- *9.1 Empirical NSW and empirical individual utilities converge almost surely to their long-run market counterparts, i.e.,* $\sum_{i=1}^n b_i \log(u_i^\gamma) \xrightarrow{\text{a.s.}} \sum_{i=1}^n b_i \log(u_i^*)$ *and* $u_i^\gamma \xrightarrow{\text{a.s.}} u_i^*$.
- *9.2 The empirical pacing multiplier converges almost surely, i.e.,* $\beta_i^\gamma \xrightarrow{\text{a.s.}} \beta_i^*$.
- *9.3 Convergence of approximate market equilibrium:* $\limsup_t \mathscr{B}^\gamma(\epsilon) \subset \mathscr{B}^*(\epsilon)$ *for all* $\epsilon \geq 0$ *and* $\limsup_t \mathscr{B}^\gamma(\epsilon_t) \subset \mathscr{B}^*(0) = \{\beta^*\}$ *for all* $\epsilon_t \downarrow 0$. *Recall the approximate solutions set,* $\mathscr{B}^\gamma$ *and* $\mathscr{B}^*$*, are defined in Eq.* (4).

Proof in Appendix H.

We briefly comment on Part 9.3. The set limit result can be interpreted from a set distance point of view. We define the inclusion distance from a set A to a set B by $d_{\subset}(A, B) := \inf_{\epsilon}\{\epsilon \geq 0 : A \subset \{y : \mathrm{dist}(y, B) \leq \epsilon\}\}$ where $\mathrm{dist}(y, B) := \inf\{\|y - b\| : b \in B\}$. Intuitively, $d_{\subset}(A, B)$ measures how much one should enlarge B such that it covers A. Then for any sequence $\epsilon_n \downarrow 0$, by the second claim in Part 9.3, we know $d_{\subset}(\mathscr{B}^{\gamma}(\epsilon_t), \{\beta^*\}) \to 0$. This shows that the set of approximate solutions of H_t with increasing accuracy centers around β^* as market size grows.

C Extended Analytical Properties of the Dual Objective

We briefly comment on the three candidate sufficient conditions in Theorem 7; for a rigorous statement we refer readers to Appendix C. Based on the differentiability characterization, it is natural to search for a stronger form of Eq. (NO-TIE) and hope that such a refinement could lead to second-order differentiability. Condition (i) gives two such refinements. Condition (ii) is motivated by the idea that expectation operator tends to produce smooth functions. The exact smoothness requirement is presented in the appendix, which we show is easy to verify for several common distributions. Finally, Condition (iii) considers the linear-valuations setting of Gao and Kroer (2022), where the authors provide tractable convex programs for computing the infinite-dimensional equilibrium. Here we give another interesting properties of this setup by showing that the dual objective is C^2. Piecewise linear value functions are discussed in the appendix.

Let $I(\beta, \theta) = \arg\max_i \beta_i v_i(\theta)$ be the set of maximizing indices, which could be non-unique. We say there is *no tie for item θ at β* if $I(\beta, \theta)$ is single-valued, in which case we use $i(\beta, \theta)$ to denote the unique maximizing index. Moreover, by Theorem 3.50 from Beck (2017), the subgradient $\partial_\beta f(\beta, \theta)$ is the convex hull of the set $\{v_i e_i, i \in I(\beta, \theta)\}$. When $I(\beta, \theta)$ is single-valued, the subgradient set is a singleton, and thus f is differentiable.

Now we have different equivalent ways to describe when f is differentiable: $f(\cdot, \theta)$ is differentiable at $\beta \Leftrightarrow I(\beta, \theta)$ is single-valued $\Leftrightarrow \epsilon(\beta, \theta)$ is strictly positive $\Leftrightarrow$ the sets $\{\Theta_i(\beta) = \{\theta : \beta_i v_i(\theta) \geq \beta_k v_k(\theta), \forall k \neq i\}\}$ are disjoint. When $f(\cdot, \theta)$ is differentiable, we have $G(\beta, \theta) = \nabla_\beta f(\beta, \theta) = e_{i(\beta,\theta)} v_{i(\beta,\theta)}$.

Markets with stability

A natural idea is to search for a stronger form of Eq. (NO-TIE) and hope that such a refinement could lead to second-order differentiability. In particular, this section is concerned with statement (i) of Theorem 7. First we show the condition based on the expectation.

Theorem 10. *If the following integrability condition holds in a neighborhood of β^**

$$\mathbb{E}\left[\frac{1}{\epsilon(\beta, \theta)}\right] = \int_{\Theta} \frac{1}{\epsilon(\beta, \theta)} \, \mathrm{d}S(\theta) < \infty \,, \tag{INT}$$

then H is twice continuously differentiable at β^*. Furthermore, it holds $\nabla^2 \bar{f}(\beta^*) = 0$.

Proof in Appendix L.

By the above theorem, if Eq. (INT) holds, then the variance matrices in Theorem 5 can be simplified as

$$\Sigma_\beta = \mathrm{Diag}(\{\Omega_i^2 (\beta_i^*)^4 / (b_i)^2\}_{i=1}^n) \,, \; \Sigma_u = \mathrm{Diag}(\{\Omega_i^2\}_i) \,. \tag{5}$$

In this case components of β^γ are asymptotically independent.

We compare the integrability condition in the above theorem with Eq. (NO-TIE). Both Eq. (INT) and Eq. (NO-TIE) can be interpreted as a form of robustness of the market equilibrium. The quantity $\epsilon(\beta, \theta)$ measures the advantage the winner of item θ has over other losing bidders. The larger $\epsilon(\beta, \theta)$ is, the more slack there is in terms of perturbing the pacing multiplier before affecting the allocation at θ. In contrast to Eq. (NO-TIE) which only imposes an item-wise requirement on the winning margin, the above assumption requires the margin exists in a stronger sense. Concretely, such a moment condition on the margin function ϵ represents a balance between how small the margin could be and the size of item sets for which there is a small winning margin.

Second we consider the condition based on the essential supremum. For any buyer i and his winning set Θ_i^*, there exists a positive constant $\epsilon_i > 0$ such that

$$\beta_i^* v_i(\theta) \geq \max_{k \neq i} \beta_k^* v_k(\theta) + \epsilon_i \,, \forall \theta \in \Theta_i^* \quad \Leftrightarrow \quad \operatorname*{ess\,sup}_{\theta \in \Theta} 1/\epsilon(\beta, \theta) < K < \infty \tag{GAP}$$

It requires that the buyer wins the items without tying bids uniformly over the winning item set. The existence of a constant $K < \infty$ such that $1/\epsilon(\beta, \theta) < K$ for almost all items makes a stronger requirement than Eq. (INT). From a practical perspective, it is also evidently a very strong assumption: for example, it won't occur with many natural continuous valuation functions. Instead, the condition requires the valuation functions to be discontinuous at the points in Θ where the allocation changes. Empirically, since β^γ is a good approximation of β^* for a market of sufficiently large size, Eq. (GAP) can be approximately verified by replacing β^* with β^γ. As a trade-off, Eq. (INT) is a weaker condition than Eq. (GAP) but is harder to verify in practical application.

Below we present two examples where Eq. (INT) holds.

Example 1 (Discrete Values)**.** *Suppose the values are supported on a discrete set, i.e.,* $[v_1, \ldots, v_n] \in \{V_1, \ldots, V_K\} \subset \mathbb{R}^n$ *a.s. Suppose there is no tie for each item at* β^**. Then Eq.* (GAP) *and thus Eq.* (INT) *hold.* ■

Example 2 (Continuous Values)**.** *Here we give a numeric example of market with two buyers where Eq.* (INT) *holds. Suppose the values are uniformly distributed over the sets* $A_1 = \{v \in \mathbb{R}_+^2 : v_2 \leq 1, v_2 \geq 2v_1\}$ *and* $A_2 = \{v \in \mathbb{R}_+^2 : v_1 \leq 1, v_2 \leq \frac{1}{2}v_1\}$*. One can verify on* $B = \{\beta \in \mathbb{R}^2 : \frac{1}{2}\beta_1 < \beta_2 < 2\beta_1\}$ *Eq.* (INT) *holds. To further verify this, by calculus, we can show the map* $\bar{f}(\beta) = \mathbb{E}[\max\{v_1\beta_1, v_2\beta_2\}]$ *is*

$$\bar{f}(\beta) = \begin{cases} \left(\frac{5}{12} - \frac{1}{3}\frac{\beta_1}{\beta_2}\right)\beta_1 + \frac{2\beta_2}{3\beta_1}\beta_2 & \text{if } \beta_2 \geq 2\beta_1 \\ \frac{1}{3}(\beta_1 + \beta_2) & \text{if } \beta \in B, \text{ i.e., } \frac{1}{2}\beta_1 < \beta_2 < 2\beta_1 \\ \left(\frac{5}{12} - \frac{1}{3}\frac{\beta_2}{\beta_1}\right)\beta_2 + \frac{2\beta_1}{3\beta_2}\beta_1 & \text{if } \beta_2 \leq \frac{1}{2}\beta_1 \end{cases} .$$

We see that $\nabla^2 \bar{f} = 0$ *on* B *which agrees with Theorem 10.*

However, Eq. (INT) *fails to capture the fact that* $\bar{f}$ *is* C^2 *in other regions as well. To see this, note that in the region* $\{\beta \in \mathbb{R}_{++}^2 : \beta_2 > 2\beta_1\}$*, the Hessian is*

$$\nabla^2 \bar{f}(\beta) = \begin{bmatrix} \frac{2}{3}\frac{{\beta_2}^2}{{\beta_1}^3} & -\frac{2}{3}\frac{\beta_2}{{\beta_1}^2} \\ -\frac{2}{3}\frac{\beta_2}{{\beta_1}^2} & \frac{2}{3\beta_1} \end{bmatrix} = \frac{2}{3}\begin{bmatrix} \frac{\beta_2}{\sqrt{\beta_1^3}} \\ -\frac{1}{\sqrt{\beta_1}} \end{bmatrix}^\top \begin{bmatrix} \frac{\beta_2}{\sqrt{\beta_1^3}} \\ -\frac{1}{\sqrt{\beta_1}} \end{bmatrix} .$$

The Hessian on the region $\{\beta_2 < \frac{1}{2}\beta_1\}$ *has a completely symmetric expression by switching* β_1 *and* β_2*. From here we can see the function* $\bar{f}$ *is* C^2 *except on the lines* $\beta_2 = 2\beta_1$ *and* $\beta_2 = \beta_1/2$*. Thus, the condition in Eq.* (INT) *does not provide the full picture of when twice differentiability holds.* ■

Markets with linear values

Now we consider the condition (iii) of Theorem 7: linear valuations. To study linear valuations, we adopt the setup in Section 4 from Gao and Kroer (2022). Suppose the item space is $\Theta = [0, 1]$ with supply $s(\theta) = 1$. The valuation of each buyer i is linear and nonnegative: $v_i(\theta) = c_i\theta + d_i \geq 0$. Moreover, assume the valuations are normalized so that $\int_{[0,1]} v_i \, \mathrm{d}\theta = 1 \Leftrightarrow c_i/2 + d_i = 1$. Assume the intercepts of v_i are ordered such that $2 \geq d_1 > \cdots > d_n \geq 0$.

We briefly review the structure of equilibrium allocation in this setting. By Lemma 5 from Gao and Kroer (2022), there is a unique partition $0 = a_0^* < a_1^* < \cdots < a_n^* = 1$ such that buyer i receives $\Theta_i = \left[a_{i-1}^*, a_i^*\right]$. In words, the item set $[0, 1]$ will be partitioned into n segments and assigned to buyers 1 to n one by one starting from the leftmost segments. Intuitively, buyer 1 values items on the left of the interval more than those on the right, which explains the allocation structure. Moreover, the equilibrium prices $p^*(\cdot)$ are convex piecewise linear with exactly n linear pieces, corresponding to intervals that are the pure equilibrium allocations to the buyers.

Theorem 11. *In the market set up as above, the dual objective* H *is* C^2 *at* β^**.*

Proof in Appendix L.

The above result also extends to most cases of piecewise linear (PWL) valuations discussed in Section 4.3 of Gao and Kroer (2022)). In the PWL setup there is a partition of $[0, 1]$, $A_0 = 0 \leq A_1 \leq \cdots \leq A_{K-1} \leq A_K = 1$, such that all $v_i(\theta)$'s are linear on $[A_{k-1}, A_k]$. At the equilibrium of a market with PWL valuations, we call an item θ an *allocation breakpoint* if there is a tie, i.e., $I(\beta^*, \theta)$ is multivalued. Now suppose the following two conditions hold: (i) none of the allocation breakpoints coincide with any of the valuation breakpoints $\{A_k\}$, and (ii) at any allocation breakpoint there are exactly two buyers in a tie. Under these two conditions, one can show that in a small enough neighborhood of the optimal pacing multiplier β^*, the allocation breakpoints are differentiable functions of the pacing multiplier. This in turn implies twice differentiability of the dual objective by repeating the argument in the proof of Theorem 11. However, if either condition (i) or (ii) mentioned above breaks, the dual objective is not twice differentiable.

Markets with smoothly distributed values

Now we consider condition (ii) of Theorem 7: smoothing via the expectation operator. Given that the dual objective H is the expectation of the non-smooth function f (plus a smooth term Ψ), we expect that under certain conditions on the expectation operator H will be twice differentiable. In this section, we make this precise. First we introduce some extra notations. For each $i \in [n]$, define the map $\sigma_i : \mathbb{R}^n_+ \to \mathbb{R}^n_+$,

$$\sigma_i(v) = [v_1 v_i,\, \ldots,\, v_{i-1} v_i,\, v_i,\, v_{i+1} v_i,\, \ldots,\, v_n v_i]^\top$$

for $i \in [n]$, which multiplies all except the i-th entry of v by v_i.

Definition 3 (Regularity)**.** *Let $f : \mathbb{R}^n_+ \to \mathbb{R}_+$ be the probability density function (w.r.t. the Lebesgue measure) of a positive-valued random vector with finite first moment. We say the density f is regular if for all $h_i(v_{-i}) := \int_0^\infty f\big(\sigma_i(v)\big) v_i^n \,\mathrm{d}v_i$, $i \in [n]$, it holds (i) h_i is continuous on $\mathbb{R}^{n-1}_{++}$, and (ii) all lower dimensional density functions of h_i are continuous (treating h_i as a scaled probability density function).*

Theorem 12. *Let H be differentiable in a neighborhood of β^*. Assume the random vector $[v_1, \ldots, v_n] : \Theta \to \mathbb{R}^n_+$ has a distribution absolutely continuous w.r.t. the Lebesgue measure on $\mathbb{R}^n$ with density function f_v. If f_v is regular, then H is twice continuously differentiable on $\mathbb{R}^n_{++}$.*

Proof in Appendix L.

The above regularity conditions are easy to verify when the values are i.i.d. draws from a distribution. In that case, many smooth distributions supported on the positive reals fall under the umbrella of the described regularity. Below we examine three cases: the truncated Gaussian distribution, the exponential distribution and the uniform distribution.

When values are i.i.d. truncated standard Gaussians, the joint density $f(v) = c_1 \prod_{i=1}^n \exp(-v_i^2/2)$ and $h_i(v_{-i}) = c_1 \int_{\mathbb{R}_+} v_i^n \exp(-\frac{1}{2} v_i^2 (1 + \sum_{k \neq i} v_k^2))\,\mathrm{d}v_i = c_2 (\sum_{k \neq i} v_k^2)^{-n/2}$, which are regular. Here c_i, $i = 1, 2$, are appropriate constants. Similarly, for the i.i.d. exponential case with the rate parameter equal to one, the density $f(v) = \prod_{i=1}^n \exp(-v_i)$ and $h_i(v_{-i}) = (\sum_{k \neq i} v_k)^{-n}$ satisfy the required continuity conditions. Finally, suppose the values are i.i.d. uniforms on $[0, 1]$. The joint density is $f(v) = \prod_{i=1}^n \mathbb{1}\{0 < v_i < 1\}$ and for example, if $i = 1$, $h_1(v_{-1}) = (\min\{1, v_2^{-1}, \ldots, v_n^{-1}\})^{n+1}/(n+1)$, which also satisfies the required continuity conditions.

D Variance Estimation for β and u

First, we discuss the restrictive case where a stronger form of Eq. (NO-TIE) holds: $\mathbb{E}[\epsilon(\beta, \theta)^{-1}] < \infty$ in a neighborhood of β^*, which we recall is a sufficient condition for twice differentiability (Theorem 7). Note that in the observed market the equilibrium allocation x^γ might not be unique, and for our purpose we let x^γ be any equilibrium allocation. We construct the following estimator for Ω_i. Let $u_i^{\gamma,\tau} := x_i^\tau v_i(\theta^\tau)$ be the utility of buyer i obtained from item θ^τ. Then $u_i^\gamma = \sum_{\tau=1}^t u_i^{\gamma,\tau}$. Under the assumption that H is differentiable at β^* (c.f. Theorem 6), the equilibrium allocation is unique and pure, i.e., $x_i^* = \mathbb{1}\{\Theta_i^*\}$. By rewriting $\Omega_i^2 = \int \big(v_i x_i^* - (\int v_i x_i^* \,\mathrm{d}S)\big)^2 \,\mathrm{d}S$, it is natural to consider the estimator $\hat{\Omega}_i^2 := \frac{1}{t} \sum_{\tau=1}^t (t u_i^{\gamma,\tau} - u_i^\gamma)^2$.

Theorem 13. *If $\epsilon(\beta, \theta)^{-1}$ is finite a.s. in a neighborhood of β^*, then $\hat{\Omega}_i^2 \xrightarrow{\mathrm{P}} \Omega_i^2$. Proof in Appendix M.*

Having derived a consistent estimator for Ω_i^2, we can construct confidence interval for β^γ and u^γ. By Theorem 10 and Eq. (5), the variance matrices are

$$\Sigma_\beta = \mathrm{Diag}(\{\Omega_i^2(\beta_i^*)^4/(b_i)^2\}_{i=1}^n)\ ,\ \Sigma_u = \mathrm{Diag}(\{\Omega_i^2\}_i)\ ,$$

and the plug-in type estimators for Σ_β and Σ_u take the form $\hat{\Sigma}_\beta = \mathrm{Diag}(\{\hat{\Omega}_i^2(\beta_i^\gamma)^4/b_i^2\})$ and $\hat{\Sigma}_u = \mathrm{Diag}(\{\hat{\Omega}_i^2\})$.

Second, we discuss the case where we have the knowledge of $\{v_i(\theta^\tau)\}_{i,\tau}$ under the general assumption that H is C^2 at β^*. Following the discussion in Section 7.3 of Newey and McFadden (1994), we estimate the Hessian matrix by computing numerical difference. We choose a smoothing level η_t and define

$$\begin{aligned}(\hat{\mathscr{H}})_{ij} := \frac{1}{4\eta_t^2}\Big(&H_t(\beta^\gamma + \eta_t(e_i + e_j)) - H_t(\beta^\gamma + \eta_t(-e_i + e_j)) \\ &- H_t(\beta^\gamma + \eta_t(e_i - e_j)) + H_t(\beta^\gamma + \eta_t(-e_i - e_j))\Big)\ ,\end{aligned}$$

which serves as an estimator of the (i, j)-th entry of the Hessian $\mathscr{H} = \nabla^2 H(\beta^*)$. By Theorem 7.4 from Newey and McFadden (1994), one can show that if $\eta_t \to 0$ and $\sqrt{t}\eta_t \to \infty$, then the estimator is consistent, i.e, $\hat{\mathscr{H}} \xrightarrow{\mathrm{P}} \mathscr{H}$. Note in order to compute the value of H_t at a perturbed β^γ we need access to the values of the buyers. Since those values may not always be available in practice, this estimator is not as practical as our other estimators which rely purely on equilibrium quantities. Moreover, the estimator requires tuning of the smoothing parameter η_t.

E Further Properties of EG programs

Fact 1. *Both optima in Eqs.* (P-EG) *and* (P-DEG) *are attained. Let* $(x_{\mathrm{EG}}^*, u_{\mathrm{EG}}^*)$ *and* β^* *attain the optima the EG programs Eqs.* (P-EG) *and* (P-DEG)*, respectively.*

- *First-order conditions. Given* $(x_{\mathrm{EG}}, u_{\mathrm{EG}})$ *feasible to* (P-EG) *and* β *feasible to* (P-DEG)*, they are both optimal if and only if the following KKT conditions hold: (i)* $\langle p_{\mathrm{EG}}, s - \sum_i x_{\mathrm{EG},i}\rangle = 0$ *where* $p_{\mathrm{EG}} = \max_i \beta_i v_i$*, (ii)* $\langle p_{\mathrm{EG}} - \beta_i v_i, x_{\mathrm{EG},i}\rangle = 0$*, and (iii)* $\langle v_i, x_{\mathrm{EG},i}\rangle = u_{\mathrm{EG},i} = b_i/\beta_i$*.*
- *Uniqueness. The equilibrium utility and prices are unique. The optimal solution* β^* *to Eq.* (P-DEG) *is unique.*
- *Strong duality.* $\sum_{i=1}^n b_i \log u_{\mathrm{EG},i}^* = H(\beta^*) + \sum_{i=1}^n b_i(\log b_i - 1)$*.*
- *Equilibrium. Given any optimal solutions* $(x_{\mathrm{EG}}^*, u_{\mathrm{EG}}^*, \beta^*)$ *to Eqs.* (P-EG) *and* (P-DEG)*, let* $p_{\mathrm{EG}}^*(\cdot) = \max_i \beta_i^* v_i(\cdot)$*. Then* $(x_{\mathrm{EG}}^*, u_{\mathrm{EG}}^*, p_{\mathrm{EG}}^*)$ *is a ME. Conversely, for a ME* (x^*, u^*, p^*)*, it holds that (i)* (x^*, u^*) *is an optimal solution of* (P-EG) *and (ii)* $\beta_{\mathrm{ME}}^* := b_i/\langle v_i, x_i^*\rangle$ *is the optimal solution of* (P-DEG)*.*
- *Bounds on* β^**. Define* $\underline{\beta}_i := b_i/\int v_i \,\mathrm{d}S$ *and* $\bar{\beta} := \sum_{i=1}^n b_i/\min_i\{\int v_i \,\mathrm{d}S\}$*. Then* $\underline{\beta}_i \le \beta_i^* \le \bar{\beta}$*.*

Based on the set of KKT conditions we comment on the structure of market equilibrium. Condition (ii) describes how the pacing multiplier relates to equilibrium allocation; buyer i only receives items within its 'winning set' $\{\theta : p^*(\theta) = \beta_i^* v_i(\theta)\}$. This also hints at a connection between Fisher market and first-price auction: the equilibrium allocation can be thought of as the result of a first-price auction where each buyer bids $\beta_i^* v_i(\theta)$ and then item goes to the highest bidder (with appropriate tie-break). Condition (iii) shows pacing multipliers β^* can be interpreted as price-per-utility. Finally, all budgets are extracted, i.e., $\langle p^*, s\rangle = \sum_{i=1}^n b_i$. To see this, we apply all three KKT conditions and obtain $\langle p^*, s\rangle = \langle p^*, \sum_{i=1}^n x_i^*\rangle = \sum_{i=1}^n \beta_i^* \langle v_i, x_i^*\rangle = \sum_{i=1}^n \beta_i^*(b_i/\beta_i^*) = \sum_{i=1}^n b_i$. Intuitively, this is due to the fact that buyers only receives utilities from obtaining goods but not retaining money. In Appendix F we study an extension called *quasilinear market* where buyers have the incentive to retain money.

Parallel to the population EG programs, we state the optimality conditions for the sample EG programs.

Fact 2. *Consider the sample EG programs Eqs.* (S-EG) *and* (S-DEG)*. It holds*

- *First-order conditions. Given* $(x^\gamma_{\text{EG}}, u^\gamma_{\text{EG}})$ *feasible to* (S-EG) *and* β^γ *feasible to* (S-DEG)*, they are both optimal if and only if the following KKT conditions hold: (i)* $\langle p^\gamma_{\text{EG}}, \mathsf{s} - \sum_i x^\gamma_{\text{EG},i}\rangle = 0$ *where* $p^\gamma_{\text{EG}} = \max_i \beta^\gamma_i v_i$*, (ii)* $\langle p^\gamma_{\text{EG}} - \beta^\gamma_i v_i, x^\gamma_{\text{EG},i}\rangle = 0$*, (iii) and* $\langle v_i, x^\gamma_{\text{EG},i}\rangle = u^\gamma_{\text{EG},i} = b_i/\beta^\gamma_i$.
- *Strong duality.* $\sum_{i=1}^n b_i \log u^\gamma_{\text{EG},i} = H_t(\beta^\gamma) + \sum_{i=1}^n b_i(\log b_i - 1)$.
- *Uniqueness. The equilibrium utility and prices are unique. The optimal solution* β^γ *to Eq.* (S-DEG) *is unique.*
- *Equilibrium. Any optimal solutions* $(x^\gamma_{\text{EG}}, u^\gamma_{\text{EG}}, p^\gamma_{\text{EG}})$ *to Eqs.* (S-EG) *and* (S-DEG) *is a ME. Conversely, for a ME* $(x^\gamma, u^\gamma, p^\gamma)$*, it holds that (i)* (x^γ, u^γ) *is an optimal solution of* (S-EG) *and (ii)* $\beta^\gamma_{\text{ME}} := b_i/\langle v_i, x^\gamma_i\rangle$ *is the optimal solution of* (S-DEG).
- *Bounds on* β^γ *and* u^γ*. It holds* $\frac{b_i}{\sum_{i=1}^n b_i}\sum_{\tau=1}^t \mathsf{s}^\tau v_i(\theta^\tau) \le u^\gamma_i \le \sum_{\tau=1}^t \mathsf{s}^\tau v_i(\theta^\tau)$ *and* $\frac{b_i}{\sum_{\tau=1}^t \mathsf{s}^\tau v_i(\theta^\tau)} \le \beta^\gamma_i \le \frac{\sum_i b_i}{\sum_{\tau=1}^t \mathsf{s}^\tau v_i(\theta^\tau)}$

We comment on the scaling $1/t$. Recall that in the long-run market the total supply of items is one, while each buyer i has budget b_i. To match budget sizes and markets sizes, we require that in the observed market, the ratio between total supply and buyer i's budget is also $1 : b_i$. Note that in a linear Fisher market, we can scale all budgets by any positive constant and the equilibrium does not change, except that prices are scaled by the same amount. Formally, if $(x, u, p) \in \mathscr{ME}^\gamma(b, v, \mathsf{s})$, then $(x, (\alpha_1\beta u_1, \ldots, \alpha_n\beta u_n), \delta p) \in \mathscr{ME}^\gamma(\delta b, (\alpha_1 v_1, \ldots, \alpha_n v_n), \beta\mathsf{s})$ for positive scalars $\delta, \beta, \{\alpha_i\}$. Thus, our particular choice of supply normalization is not crucial. For example, we could equivalently work with the market $\mathscr{ME}^\gamma(tb, v, 1_t)$, with trivial scaling adjustments in derivation of the results.

We remark that there are two ways to specify the valuation component of infinite-dimensional Fisher market. The first one is simply imposing functional form assumptions on $v_i(\cdot)$. For example, one could take $\Theta = [0, 1]$ and let $v_i(\cdot)$ be a linear function or a piecewise linear function (Gao and Kroer, 2022, Section 4). If we view $v = (v_1, \ldots, v_n) : \Theta \to \mathbb{R}^n_+$ as a random vector, then an alternative way is to simply specify the distribution of v. Formally, let v and v' be identically distributed random vectors representing the values of buyers. By the form of EG programs Eqs. (P-EG) and (P-DEG), if (u^*, β^*) are equilibrium utilities and pacing multipliers of the market $\mathscr{ME}(b, v, s)$ and $(u^{*'}, \beta^{*'})$ are those of $\mathscr{ME}(b, v', s)$, then $(u^*, \beta^*) = (u^{*'}, \beta^{*'})$. Even though the equilibrium allocations and prices are different in the two markets, the quantities we care about, e.g., individual utilities and Nash social welfare are the same. Moreover, applying the same reasoning to the case of quasilinear market (see Appendix F), it will be clear that identical value distributions implies identical revenues in market equilibrium. When the distribution of v is absolutely continuous w.r.t. the Lebesgue measure on $\mathbb{R}^n$ we use f_v to denote the density function.

F Extension: Revenue Inference in Quasilinear Fisher Market

As we mentioned previously, in a linear Fisher market all buyer budgets are extracted, i.e., $\sum_{\tau=1}^t p^{\gamma,\tau}$ equals $\sum_{i=1}^n b_i$ in the observed market (and similarly for the underlying market), and there is thus nothing to infer about revenue if we know the budgets of each buyer. A quasilinear (QL) utility is one such that the cost of purchasing goods is deducted from the utility, i.e., $u_i(x) = \langle x - p, v_i\rangle$. This may give buyers an incentive to leave some budget unspent. In the finite-dimensional case, Chen et al. (2007) and Cole et al. (2017) show that there is an variant of EG program that captures the market equilibrium with QL utility. Furthermore, Conitzer et al. (2022a) showed that budget management in ad auctions with first-price auctions can be computed by Fisher markets with QL utilities. A QL variant of infinite-dimensional markets and an EG program are given by Gao and Kroer (2022).

Quaislinear market equilibria (QME) are defined analogously to the linear variant via market clearance conditions and buyer optimality; we present the formal finite and infinite-dimensional definitions in Appendix N. The demand sets are $\arg\max\{\langle v_i - p, x_i\rangle : x_i \in L^\infty_+, \langle p, x_i\rangle \le b_i\}$ in the long-run QME and $\arg\max\{\langle v_i(\gamma) - p, x_i\rangle : x_i \ge 0, \langle p, x_i\rangle \le b_i\}$ in the observed QME. QME has several distinctions from the linear ME. First, in QME we cannot normalize both valuations and budgets, since buyers' budgets have value outside the current market. Second, budgets are not fully

extracted in QME, which motivates the need for statistical analysis. Third, the pacing multipliers are restricted to $\beta \leq 1$, and may lie on the resulting boundary.

Define the revenues from the observed and the long-run market as follows: $\mathrm{REV}^\gamma := \frac{1}{t}\sum_{\tau=1}^t p^{\gamma,\tau}, \mathrm{REV}^* := \int_\Theta p^* \,\mathrm{d}S(\theta)$. Assume $\sum_{i=1}^n b_i = 1$ and unit supply $\int s \,\mathrm{d}\mu = 1$. Let $\nu_i := \int v_i \,\mathrm{d}S$ be the average value of buyer i. Let $\bar{\nu} = \max_i \nu_i$. Assume we observe the market $\mathscr{QME}^\gamma(b, v, \frac{1}{t}1_t) = (x^\gamma, u^\gamma, p^\gamma)$. Then we show that consistency and high-probability bounds hold for the revenue estimator.

Theorem 14 (Revenue Convergence)**.** *It holds that* $\mathrm{REV}^\gamma \xrightarrow{\text{a.s.}} \mathrm{REV}^*$ *and* $|\mathrm{REV}^\gamma - \mathrm{REV}^*| = \tilde{O}_p\left(\frac{\bar{v}\sqrt{n}(\bar{v}+2\bar{\nu}n+1)}{\underline{b}}\frac{1}{\sqrt{t}}\right)$ *for t sufficiently large. Proofs are in Appendix N.*

We leave CLT results for revenue estimates in quasilinear markets as an open problem. The main challenge compared to the linear case is that the optimal pacing multipliers can lie on the boundary of the constraint set. More precisely, if the equilibrium pacing multiplier of a buyer is in the interior, then his budget is fully extracted. On the other hand, if it is on the boundary, the buyer retains a portion of his budget at equilibrium. When the optimum of the expectation function lies on the boundary of the constraint set, the asymptotic variance of the sample average optimum takes on a complicated expression (Shapiro, 1989, Theorem 3.3), which makes variance estimation difficult.

G Technical Lemmas

Lemma 1. *Define the event* $A_t = \{\beta^\gamma \in C\}$ *. (i) If* $t \geq 2\bar{v}^2\log(2n/\eta)$*, then* $\mathbb{P}(A_t) \geq \mathbb{P}(\frac{1}{2} \leq \frac{1}{t}\sum_{\tau=1}^t v_i(\theta^\tau) \leq 2, \forall i) \geq 1-\eta$*. (ii) It holds* $\mathbb{P}(A_t \text{ eventually}) = 1$*. Proof in Appendix G.*

Proof of Lemma 1. Recall the event $A_t = \{\beta^\gamma \in C\}$. Define $\bar{v}_i^t = \frac{1}{t}\sum_{\tau=1}^t v_i(\theta^\tau)$.

First we notice concentration of values implies membership of β^γ to C, i.e., $\{1/2 \leq \bar{v}_i^t \leq 2, \forall i\} \subset \{\beta^\gamma \in C\}$ due to Fact 2. Concretely, $u_i^\gamma \leq \frac{1}{t}\sum_{\tau=1}^t v_i(\theta^\tau)$ and $u_i^\gamma \geq \frac{1}{t}\frac{b_i}{\sum_{i=1}^n b_i}\sum_{\tau=1}^t v_i(\theta^\tau)$, and through the equation $\beta_i^\gamma = b_i/u_i^\gamma$ the inclusion follows. Note $0 \leq v_i(\theta^\tau) \leq \bar{v}$ is a bounded random variable with mean $\mathbb{E}[v_i(\theta^\tau)] = 1$. By Hoeffding's inequality we have $\mathbb{P}(|\bar{v}_i^t - 1| \geq \delta) \leq 2\exp(-\frac{2\delta^2 t}{\bar{v}^2})$. Next we use a union bound and obtain

$$\mathbb{P}(\beta^\gamma \notin C) \leq \mathbb{P}\left(\bigcup_{i=1}^n \{|\bar{v}_i^t - 1| \geq \delta\}\right) \leq 2n\exp\left(-\frac{2\delta^2 t}{\bar{v}^2}\right). \tag{6}$$

By setting $2n\exp(-\frac{2\delta^2 t}{\bar{v}^2}) = \eta$ and $\delta = 1/2$ and solving for t we obtain item (i) in claim.

To show item (ii), we use the Borel-Cantelli lemma. By choosing $\delta = 1/2$ in the Eq. (6) we know $\mathbb{P}(A_t^c) \leq \mathbb{P}(\{1/2 \leq \bar{v}_i^t \leq 2, \forall i\}^c) \leq 2n\exp(-t/(2\bar{v}^2))$. Then we have

$$\sum_{t=1}^\infty \mathbb{P}(A_t^c) < \infty\,.$$

By the Borel-Cantelli lemma it follows that $\mathbb{P}(\{A_t^c \text{ infinitely often}\}) = 0$, or equivalently $\mathbb{P}(A_t \text{ eventually}) = 1$. □

Lemma 2 (Smoothness and Curvature)**.** *It holds that both H and H_t are L-Lipschitz and λ-strongly convex w.r.t the ℓ_∞-norm on C with $L = 2n + \bar{v}$ and $\lambda = \underline{b}/4$. Moreover, H_t and H are $(\bar{v} + 2\sqrt{n})$-Lipschitz w.r.t. ℓ_2-norm.*

Proof of Lemma 2. Now we verify that H_t and H are $(\bar{v}+2n)$-Lipschitz on the compact set C w.r.t. the ℓ_∞-norm. For $\beta, \beta' \in C$,

$$\begin{aligned}
&|H_t(\beta) - H_t(\beta')| \\
&\leq \frac{1}{t}\sum_{\tau=1}^{t}\big|\max_i\{v_i(\theta^\tau)\beta_i\} - \max_i\{v_i(\theta^\tau)\beta_i'\}\big| + \sum_{i=1}^{n} b_i\big|\log\beta_i - \log\beta_i'\big| \\
&\leq \bar{v}\|\beta-\beta'\|_\infty + \sum_{i=1}^{n} b_i \cdot \frac{1}{\underline{\beta}_i/2}|\beta_i - \beta_i'| \\
&= (\bar{v}+2n)\|\beta-\beta'\|_\infty \ .
\end{aligned}$$

This concludes the $(\bar{v}+2n)$-Lipschitzness of H_t on C. Similar argument goes through for H. From the above reasoning we can also conclude $|H_t(\beta) - H_t(\beta')| \leq \bar{v}\|\beta-\beta'\|_2 + 2\|\beta-\beta'\|_1 \leq (\bar{v}+2\sqrt{n})\|\beta-\beta'\|_2$. This concludes $(\bar{v}+2\sqrt{n})$-Lipschitzness of H_t w.r.t. ℓ_2-norm.

Recall $H = \bar{f} + \Psi$ where $\bar{f}(\beta) = \mathbb{E}[\max_i\{v_i(\theta)\beta_i\}]$ and $\Psi(\beta) = -\sum_{i=1}^{n} b_i \log\beta_i$. The function Ψ is smooth with the first two derivatives

$$\nabla\Psi(\beta) = -[b_1/\beta_1, \ldots, b_n/\beta_n]^\top, \quad \nabla^2\Psi(\beta) = \mathrm{Diag}(\{b_i/(\beta_i)^2\}) \ .$$

It is clear that for all $\beta \in C$ it holds $\beta_i \leq 2$. So $\nabla^2\Psi(\beta) \succ \min_i\{b_i/4\}I = \lambda I$. To verify the storng-convexity w.r.t $\|\cdot\|_\infty$ norm, we note for all $\beta', \beta \in C$.

$$H(\beta') - H(\beta) - \langle z + \nabla\Psi(\beta), \beta' - \beta\rangle \geq (\lambda/2)\|\beta'-\beta\|_2^2 \geq (\lambda/2)\|\beta'-\beta\|_\infty^2 \ ,$$

where $z \in \partial\bar{f}(\beta)$ and $z + \nabla\Psi(\beta) \in \partial H(\beta)$. This completes the proof. □

Definition 4 (Definition 7.29 in Shapiro et al. (2021))**.** *A sequence $f_k : \mathbb{R}^n \to \bar{\mathbb{R}}$, $k = 1, \ldots,$ of extended real valued functions epi-converge to a function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, if for any point $x \in \mathbb{R}^n$ the following conditions hold*

(1) For any sequence $x_k \to x$, it holds $\liminf_{k\to\infty} f_k(x_k) \geq f(x)$,

(2) There exists a sequence $x_k \to x$ such that $\limsup_{k\to\infty} f_k(x_k) \leq f(x)$.

Definition 5 (Definition 3.25, Rockafellar and Wets (2009), see also Definition 11.11 and Proposition 14.16 from Bauschke et al. (2011))**.** *A function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is level-coercive if $\liminf_{\|x\|\to\infty} f(x)/\|x\| > 0$. It is equivalent to $\lim_{\|x\|\to+\infty} f(x) = +\infty$.*

Lemma 3 (Corollary 11.13, Rockafellar and Wets (2009))**.** *For any proper, lsc function f on $\mathbb{R}^n$, level coercivity implies level boundedness. When f is convex the two properties are equivalent.*

Lemma 4 (Theorem 7.17, Rockafellar and Wets (2009))**.** *Let $h_n : \mathbb{R}^d \to \bar{\mathbb{R}}$, $h : \mathbb{R}^d \to \bar{\mathbb{R}}$ be closed convex and proper. Then $h_n \xrightarrow{\text{epi}} h$ is equivalent to either of the following conditions.*

(1) There exists a dense set $A \subset \mathbb{R}^d$ such that $h_n(v) \to h(v)$ for all $v \in A$.

(2) For all compact $C \subset \mathrm{Dom}\, h$ not containing a boundary point of $\mathrm{Dom}\, h$, it holds

$$\lim_{n\to\infty}\sup_{v\in C}|h_n(v) - h(v)| = 0 \ .$$

Lemma 5 (Proposition 7.33, Rockafellar and Wets (2009))**.** *Let $h_n : \mathbb{R}^d \to \bar{\mathbb{R}}$, $h : \mathbb{R}^d \to \bar{\mathbb{R}}$ be closed and proper. If h_n has bounded sublevel sets and $h_n \xrightarrow{\text{epi}} h$, then $\inf_v h_n(v) \to \inf_v h(v)$.*

Lemma 6 (Theorem 7.31, Rockafellar and Wets (2009))**.** *Let $h_n : \mathbb{R}^d \to \bar{\mathbb{R}}$, $h : \mathbb{R}^d \to \bar{\mathbb{R}}$ satisfy $h_n \xrightarrow{\text{epi}}$ and $-\infty < \inf h < \infty$. Let $S_n(\varepsilon) = \{\theta \mid h_n(\theta) \leq \inf h_n + \varepsilon\}$ and $S(\varepsilon) = \{\theta \mid h(\theta) \leq \inf h + \varepsilon\}$. Then $\limsup_n S_n(\varepsilon) \subset S(\varepsilon)$ for all $\varepsilon \geq 0$, and $\limsup_n S_n(\varepsilon_n) \subset S(0)$ whenever $\varepsilon_n \downarrow 0$.*

Lemma 7 (Theorem 5.7, Shapiro et al. (2021), Asymptotics of SAA Optimal Value)**.** *Consider the problem*

$$\min_{x\in X} f(x) = \mathbb{E}[F(x,\xi)]$$

where X is a nonempty closed subset of $\mathbb{R}^n$, ξ is a random vector with probability distribution P on a set Ξ and $F : X \times \Xi \to \mathbb{R}$. Assume the expectation is well-defined, i.e., $f(x) < \infty$ for all $x \in X$. Define the sample average approxiamtion (SAA) problem

$$\min_{x \in X} f_N(x) = \frac{1}{N} \sum_{i=1}^{N} F(x, \xi_i)$$

where ξ_i are i.i.d. copies of the random vector ξ. Let v_N (resp., v^*) be the optimal value of the SAA problem (resp., the original problem). Assume the following.

- *7.a The set X is compact.*
- *7.b For some point $x \in X$ the expectation $\mathbb{E}[F(x, \xi)^2]$ is finite.*
- *7.c There is a measurable function $C : \Xi \to \mathbb{R}_+$ such that $\mathbb{E}[C(\xi)^2] < \infty$ and $|F(x, \xi) - F(x', \xi)| \leq C(\xi)||x - x'||$ for all $x, x' \in X$ and almost every $\xi \in \Xi$.*
- *7.d The function f has a unique minimizer x^* on X.*

Then

$$v_N = f_N(x^*) + o_p(N^{-1/2}) \ , \quad \sqrt{N}(v_N - v^*) \xrightarrow{\mathrm{d}} \mathscr{N}\big(0, \mathrm{var}\big(F(x^*, \xi)\big)\big) \ .$$

H Proof of Theorem 9

Proof of Theorem 9. We show epi-convergence (see Definition 4) of H_t to H. Epi-convergence is closely related to the question of whether we have convergence of the set of minimizers. In particular, epi-convergence is a suitable notion of convergence under which one can guarantee that the set of minimizers of the sequence of approximate optimization problems converges to the minimizers of the original problem.

To work under the framework of epi-convergence, we extend the definition of H_t and H to the entire Euclidean space as follows. We extend $\log$ to the entire real by defining $\log(x) = -\infty$ if $x < 0$. Let

$$\tilde{F}(\beta, \theta) = \begin{cases} F(\beta, \theta) = \max_i v_i(\theta)\beta_i - \sum_{i=1}^n b_i \log \beta_i & \text{if } \beta \in \mathbb{R}^n_{++} \\ +\infty & \text{else} \end{cases} ,$$

and

$$\tilde{H}(\beta) : \mathbb{R}^n \to \bar{\mathbb{R}}, \beta \mapsto \begin{cases} H(\beta) & \text{if } \beta \in \mathbb{R}^n_{++} \\ +\infty & \text{else} \end{cases} , \quad \tilde{H}_t(\beta) : \mathbb{R}^n \to \bar{\mathbb{R}}, \beta \mapsto \begin{cases} H_t(\beta) & \text{if } \beta \in \mathbb{R}^n_{++} \\ +\infty & \text{else} \end{cases} .$$

It is clear that for $\beta \in \mathbb{R}^n$ it holds $\tilde{H}(\beta) = \mathbb{E}[\tilde{F}(\beta, \theta)]$ and $\tilde{H}_t(\beta) = \frac{1}{t}\sum_{\tau=1}^t \tilde{F}(\beta, \theta^\tau)$. In order to prove the result, we will invoke Lemmas 4, 5, and 6. To invoke those lemmas, we will need the following four properties that we each prove immediately after stating them.

1. Check that $\tilde{H}$ is closed, proper and convex, and $\tilde{H}_t$ is closed, proper and convex almost surely. Convexity and properness of the functions $\tilde{H}_t$ and $\tilde{H}$ is obvious. Recall for a proper convex function, closedness is equivalent to lower semicontinuity (Rockafellar, 1970, Page 52). It is obvious that $\tilde{H}_t$ is continuous and thus closed almost surely.

   It remains to verify lower semicontinuity of $\tilde{H}$, i.e., for all $\beta \in \mathbb{R}^n$, $\liminf_{\beta' \to \beta} \tilde{H}(\beta') \geq \tilde{H}(\beta)$. For any $\beta \in \mathbb{R}^n$, we have that $\tilde{f}(\beta, \theta) := \max_i v_i(\theta)\beta_i + \delta_{\mathbb{R}^n_+}(\beta) \geq 0$, where $\delta_A(\beta) = \infty$ if $\beta \notin A$ and 0 if $\beta \in A$. With this definition of $\tilde{f}$ we have $\tilde{F}(\beta, \theta) = \tilde{f}(\beta, \theta) - \sum_{i=1}^n b_i \log \beta_i$. Applying Fatou's lemma (for extended real-valued random variables), we get $\liminf_{\beta' \to \beta} \mathbb{E}[\tilde{f}(\beta', \theta)] \geq \mathbb{E}[\liminf_{\beta' \to \beta} \tilde{f}(\beta', \theta)] \geq \mathbb{E}[\tilde{f}(\beta, \theta)]$ where in the last step

we used lower semicontinuity of $\beta \mapsto \tilde{f}(\beta, \theta)$. And thus

$$\begin{aligned}
&\liminf_{\beta' \to \beta} \tilde{H}(\beta') \\
&= \liminf_{\beta' \to \beta} \mathbb{E}\left[\tilde{f}(\beta', \theta) - \sum_{i=1}^{n} b_i \log \beta'_i\right] \\
&\geq \liminf_{\beta' \to \beta} \mathbb{E}[\tilde{f}(\beta', \theta)] - \sum_{i=1}^{n} b_i \log \beta_i \\
&\geq \mathbb{E}\left[\tilde{f}(\beta, \theta) - \sum_{i=1}^{n} b_i \log \beta_i\right] \\
&= \tilde{H}(\beta)
\end{aligned}$$

This shows $\tilde{H}$ is lower semicontinuous.

2. Check $\tilde{H}_t$ pointwise converges to $\tilde{H}$ on $\mathbb{Q}^n$. Let $\mathbb{Q}^n$ be the set of n-dimensional vectors with rational entries. For a fixed $\beta \in \mathbb{R}^n$, define the event $E_\beta := \{\lim_{t\to\infty} \tilde{H}_t(\beta) = \tilde{H}(\beta)\}$. Since $v_i(\theta^\tau) \leq \bar{v}$ almost surely by assumption, the strong law of large numbers implies that $\mathbb{P}(E_\beta) = 1$. Define

$$E := \left\{ \lim_{t\to\infty} \tilde{H}_t(\beta) = \tilde{H}(\beta), \text{ for all } \beta \in \mathbb{Q}^n \right\} = \bigcap_{\beta \in \mathbb{Q}^n} E_\beta .$$

Then by a union bound we obtain $\mathbb{P}(E^c) = \mathbb{P}(\bigcup_{\beta\in\mathbb{Q}^n} E_\beta^c) \leq \sum_{\beta\in\mathbb{Q}^n} \mathbb{P}(E_\beta^c) = 0$, implying E has measure one.

3. Check $-\infty < \inf_\beta \tilde{H} < \infty$. This is obviously true since valuations are bounded.

4. Check that for almost every sample path ω, $\tilde{H}_t$ has bounded sublevel sets (eventually). By Lemma 3, this property is equivalent to eventual coerciveness of $\tilde{H}_t$, i.e., there is a (random) N such that for all $t \geq N$, it holds $\lim_{\|\beta\|\to\infty} \tilde{H}_t(\beta) = +\infty$. By Lemma 1, we know for almost every ω, there is a finite constant N_ω such that for all $t \geq N_\omega$ it holds $\bar{v}_i^t \geq 1/2$. Then it holds for this ω, all $t \geq N_\omega$, and all $\beta \in \mathbb{R}^n$,

$$\begin{aligned}
\tilde{H}_t(\beta) &= \frac{1}{t}\sum_{\tau=1}^{t} \max_i v_i(\theta^\tau)\beta_i - \sum_{i=1}^{n} b_i \log \beta_i \\
&\geq \max_i (\bar{v}_i^t \beta_i) - \sum_{i=1}^{n} b_i \log \beta_i \\
&\geq \frac{1}{2}\|\beta\|_\infty - \sum_{i=1}^{n} b_i \log \beta_i \to +\infty \quad \text{as } \|\beta\| \to \infty .
\end{aligned}$$

This implies $\tilde{H}_t$ has bounded sublevel sets.

With the above Item 1 and Item 2 we invoke Lemma 4 and obtain that

$$\mathbb{P}\big(\tilde{H}_t(\beta) \xrightarrow{\text{epi}} \tilde{H}(\beta)\big) = 1 , \tag{7}$$

and that the convergence is uniform on any compact set.

The epi-convergence result Eq. (7) along with Item 4 allows us to invoke Lemma 5 and obtain

$$\inf_{\beta\in\mathbb{R}^n} \tilde{H}_t(\beta) \to \inf_{\beta\in\mathbb{R}^n} \tilde{H}(\beta) \text{ a.s.} \tag{8}$$

which also implies $\inf_{\mathbb{R}^n_{++}} H_t \to \inf_{\mathbb{R}^n_{++}} H$ a.s.

With the epi-convergence result Eq. (7) along with Item 3 we invoke Lemma 6 and obtain

$$\begin{aligned}
\limsup_t \mathscr{B}^\gamma(\epsilon) &\subset \mathscr{B}^*(\epsilon) \text{ for all } \epsilon \geq 0 , \\
\limsup_t \mathscr{B}^\gamma(\epsilon_t) &\subset \mathscr{B}^*(0) \text{ for all } \epsilon_t \downarrow 0 .
\end{aligned} \tag{9}$$

Putting together. At this stage all statements in the theorem are direct implications of the above results.

Proof of Part 9.1

Convergence of Nash social welfare follows from Eq. (8) and strong duality, i.e., $\text{NSW}^\gamma = \inf_{\beta\in\mathbb{R}^n_{++}} H_t(\beta) + \sum_{i=1}^n (b_i \log b_i - b_i)$ and $\text{NSW}^* = \inf_{\beta\in\mathbb{R}^n_{++}} H(\beta) + \sum_{i=1}^n (b_i \log b_i - b_i)$.

Proof of Part 9.2

Now we show consistency of the pacing multiplier via Lemma 4 and Lemma 5. Recall the compact set $C = \prod_{i=1}^n [\underline{\beta}_i/2, 2\bar{\beta}] = \prod_{i=1}^n [b_i/2, 2] \subset \mathbb{R}^n$. By construction, $\beta^* \in C$. First note that for almost every sample path ω, $1/2 \le \bar{v}_i^t \le 2$ eventually, and thus $\beta_i^\gamma = b_i/u_i^\gamma \le b_i/(b_i\bar{v}_i^t) \le 2$ and $\beta_i^\gamma \ge b_i/2$ eventually. So $\beta^\gamma \in C$ eventually. Now we can invoke Lemma 4 Item (2) to get

$$\lim_{t\to\infty} \sup_{\beta\in C} |H_t(\beta) - H(\beta)| \to 1 \quad \text{a.s.} \tag{10}$$

Now we can show that the value of H on the sequence β^γ converges to the value at β^*:

$$0 \le \lim_{t\to\infty} H(\beta^\gamma) - H(\beta^*) = \lim_{t\to\infty} [H(\beta^\gamma) - H_t(\beta^\gamma)] + \lim_{t\to\infty} [H_t(\beta^\gamma) - H(\beta^*)] = 0\ .$$

Here the first term tends to zero due to (10), and the second term by Eq. (8). For any limit point of the sequence $\{\beta^\gamma\}_t$, β^∞, by lower semicontinuity of H,

$$0 \le H(\beta^\infty) - H(\beta^*) \le \liminf_{t\to\infty} H(\beta^\gamma) - H(\beta^*) = 0\ .$$

So it holds that $H(\beta^\infty) = H(\beta^*)$ for all limit points β^∞. By uniqueness of the optimal solution β^* (see Fact 1), we have $\beta^\gamma \to \beta^*$ a.s.

Proof of Part 9.3

Convergence of approximate equilibrium follows from Eq. (9). □

I Proof of Theorem 2

Proof of Theorem 2. Recall the set $C = \prod_{i=1}^n [\underline{\beta}_i/2, 2\bar{\beta}] = \prod_{i=1}^n [b_i/2, 2] \subset \mathbb{R}^n$ and the event $A_t = \{\beta^\gamma \in C\}$. By Lemma 1 we know that if $t \ge 2\bar{v}^2 \log(4n/\eta)$ then event A_t happens with probability $\ge 1 - \eta/2$. Now the proof proceeds in two steps.

Step 1. A covering number argument. Let $\mathscr{B}^o$ be an ϵ-covering of the compact set C, i.e, for all $\beta \in C$ there is a $\beta^o(\beta) \in \mathscr{B}^o$ such that $\|\beta - \beta^o(\beta)\|_\infty \le \epsilon$. It is easy to see that such a set can be chosen with cardinality bounded by $|\mathscr{B}^o| \le (2/\epsilon)^n$.

Recall H_t and H are L-Lipschitz w.r.t. ℓ_∞-norm on C. Using this fact we get the following uniform concentration bound over the compact set C.

$$\begin{aligned}
&\sup_{\beta\in C} |H_t(\beta) - H(\beta)| \\
&\le \sup_{\beta\in C} \left\{ |H_t(\beta) - H_t(\beta^o(\beta))| + |H(\beta) - H(\beta^o(\beta))| + |H_t(\beta^o(\beta)) - H(\beta^o(\beta))| \right\} \\
&\le 2(\bar{v} + 2n)\epsilon + \sup_{\beta^o\in\mathscr{B}^o} |H_t(\beta^o) - H(\beta^o)|\ .
\end{aligned}$$

Next we bound the second term in the last expression. For some fixed $\beta \in C$, let $X^\tau := \max_i v_i(\theta^\tau)\beta_i$ and let its mean be μ. Note $0 \le X^\tau \le \bar{v}\|\beta\|_\infty \le 2\bar{v}$ due to $\beta \in C$. So X^τ's are bounded random variables. By Hoeffding's inequality we have

$$\mathbb{P}\big(|H_t(\beta) - H(\beta)| \ge \delta\big) = \mathbb{P}\left(\Big|\frac{1}{t}\sum_{\tau=1}^t X^\tau - \mu\Big| \ge \delta\right) \le 2\exp\left(-\frac{\delta^2 t}{2\bar{v}^2}\right).$$

By a union bound we get

$$\mathbb{P}\left(\sup_{\beta^o\in\mathscr{B}^o} |H_t(\beta^o) - H(\beta^o)| \ge \delta\right) \le 2|\mathscr{B}^o| \exp\left(-\frac{\delta^2 t}{2\bar{v}^2}\right) \le 2\exp\left(-\frac{\delta^2 t}{2\bar{v}^2} + n\log(2/\epsilon)\right).$$

Define the event

$$E_t := \left\{ \sup_{\beta^o \in \mathscr{B}^o} |H_t(\beta^o) - H(\beta^o)| \le \frac{2\bar{v}}{\sqrt{t}} \sqrt{\log(4/\eta) + n\log(2/\epsilon)} =: \iota \right\} . \tag{11}$$

By setting $2\exp(-\delta^2 t/(2\bar{v}^2) + n\log(2/\epsilon)) = \eta/2$ and solving for η, we have that $\mathbb{P}(E_t) \ge 1 - \eta/2$.

Step 2. Putting together. Recall the event $A_t = \{\beta^\gamma \in C\}$. Now let events A_t and E_t hold. Note $\mathbb{P}(A_t \cap E_t) \ge 1 - \eta$ if $t \ge 2\bar{v}^2 \log(4n/\eta)$. Then

$$\begin{aligned}
&\Big| \sup_{\beta \in \mathbb{R}^n_{++}} H_t(\beta) - \sup_{\beta \in \mathbb{R}^n_{++}} H(\beta) \Big| \\
&= \Big| \sup_{\beta \in C} H_t(\beta) - \sup_{\beta \in C} H(\beta) \Big| \\
&\le \sup_{\beta \in C} |H_t(\beta) - H(\beta)| \\
&\le 2(\bar{v} + 2n)\epsilon + \iota \,,
\end{aligned} \tag{12}$$

where the first equality is due to event A_t and the last inequality is due to event E_t defined in Eq. (11). Now we choose the discretization error as $\epsilon = \frac{1}{\sqrt{t}(\bar{v}+2n)}$. Then, the expression in Eq. (12) can be upper bounded as follows.

$$\begin{aligned}
&2(\bar{v} + 2n)\epsilon + \iota \\
&= \frac{2}{\sqrt{t}} + \frac{2\bar{v}}{\sqrt{t}} \sqrt{\log(4/\eta) + n\log(2\sqrt{t}(\bar{v} + 2n))} \,.
\end{aligned}$$

This completes the proof. □

J Proof of Theorem 3

Proof of Theorem 3. The proof idea of this theorem closely follows Section 5.3 of Shapiro et al. (2021).

We first need some additional notations. Define the approximate solutions sets of surrogate problems as follows: For a closed set $A \subset \mathbb{R}^n_{++}$, let

$$\begin{aligned}
\mathscr{B}^*_A(\epsilon) &:= \{\beta \in A : H(\beta) \le \min_A H + \epsilon\} \,, \\
\mathscr{B}^\gamma_A(\epsilon) &:= \{\beta \in A : H_t(\beta) \le \min_A H_t + \epsilon\} \,.
\end{aligned}$$

In words, they solve the surrogate optimization problems which are defined with a new constraint set A. Note that if $\beta^* \in A$ then $\mathscr{B}^*_A(\epsilon) = A \cap \mathscr{B}^*(\epsilon)$. Recall on the compact set C, both H_t and H are L-Lipschitz and λ-strongly convex w.r.t the ℓ_∞-norm, where $L = (\bar{v} + 2n)$ and $\lambda = \underline{b}/4$.

Let $r := \sup\{H(\beta) - H^* : \beta \in C\}$. Then if $\epsilon \ge r$ then $C \subset \mathscr{B}^*(\epsilon)$ and the claim is trivial. Now we assume $\epsilon < r$.

Define $a = \min\{2\epsilon, (r + \epsilon)/2\}$. Note $\epsilon < a < r$. Define $S = C \cap \mathscr{B}^*(a)$. The role of S will be evident as follows. We will show that, with high probability, the following chain of inclusions holds

$$\mathscr{B}^\gamma_C(\delta) \overset{(1)}{\subset} \mathscr{B}^\gamma_S(\delta) \overset{(2)}{\subset} \mathscr{B}^*_S(\epsilon) \overset{(3)}{\subset} \mathscr{B}^*_C(\epsilon) \,.$$

Step 1. Reduction to discretized problems. We let S' be a ν-cover of the set $S = \mathscr{B}^*(a) \cap C$. Let $X = S' \cup \{\beta^*\}$. In this part the goal is to show

$$\mathbb{P}\big(\mathscr{B}^\gamma_C(\delta) \subset \mathscr{B}^*_C(\epsilon)\big) \ge \mathbb{P}\big(\mathscr{B}^\gamma_X(\delta') \subset \mathscr{B}^*_X(\epsilon')\big)$$

where

$$\nu = (\epsilon' - \delta')/4 > 0 \,, \quad \delta' = \delta + L\nu > 0 \,, \quad \epsilon' = \epsilon - L\nu > 0 \,.$$

First, we claim

Claim 1. *It holds* $\mathscr{B}_X^\gamma(\delta') \subset \mathscr{B}_X^*(\epsilon') \implies \mathscr{B}_S^\gamma(\delta) \subset \mathscr{B}_S^*(\epsilon)$ *(Inclusion (2)).*

Next, we show

Claim 2. *Inclusion (2) implies Inclusion (1):* $\mathscr{B}_S^\gamma(\delta) \subset \mathscr{B}_S^*(\epsilon) \implies \mathscr{B}_C^\gamma(\delta) \subset \mathscr{B}_S^\gamma(\delta)$.

Proofs of Claim 1 and Claim 2 are deferred after the proof of Theorem 3. At a high level, Claim 1 uses the covering property of the set X. Claim 2 exploits convexity of the problem.

Finally, we show Inclusion (3) $\mathscr{B}_S^*(\epsilon) \subset \mathscr{B}_C^*(\epsilon)$. Note that β^* belongs to both C and S. And thus for any $\beta \in \mathscr{B}_S^*(\epsilon)$, it holds $H(\beta) \leq \min_X H + \epsilon = H^* + \epsilon = \min_S H + \epsilon$. We obtain $\beta \in \mathscr{B}_C^*(\epsilon)$.

To summarize, Claim 1 shows that $\mathscr{B}_X^\gamma(\delta') \subset \mathscr{B}_X^*(\epsilon')$ implies Inclusion (2). Inclusion (3) holds automatically. By Claim 2 we know Inclusion (2) implies Inclusion (1). So it holds deterministically that

$$\{\mathscr{B}_X^\gamma(\delta') \subset \mathscr{B}_X^*(\epsilon')\} \subset \{\mathscr{B}_C^\gamma(\delta) \subset \mathscr{B}_C^*(\epsilon)\} .$$

Step 2. Probability of inclusion for discretized problems. Now we bound the probability $\mathbb{P}(\mathscr{B}_X^\gamma(\delta') \subset \mathscr{B}_X^*(\epsilon'))$.

For now, we forget the construction $X = S' + \{\beta^*\}$ where S' is a ν-cover of S. Let $X \subset C$ be any discrete set with cardinality $|X|$.

Let $\beta_X^* \in \arg\min_X H$ be a minimizer of H over the set X. For $\beta \in X$ define the random variable $Y_\beta^\tau := F(\beta_X^*, \theta^\tau) - F(\beta, \theta^\tau)$. Also let $\mu_\beta := \mathbb{E}[Y_\beta^\tau]$, which is well-defined by the i.i.d. item assumption. Let $D := \sup_{\beta \in X} \|\beta - \beta_X^*\|_\infty$.

Consider any $0 \leq \delta' < \epsilon'$. If $X - \mathscr{B}_X^*(\epsilon')$ is empty, then all elements in X are ϵ'-optimal for the problem $\min_X H$. Next assume $X - \mathscr{B}_X^*(\epsilon')$ is not empty. We upper bound the probability of the event $\mathscr{B}_X^\gamma(\delta') \not\subset \mathscr{B}_X^*(\epsilon')$.

$$\begin{aligned}
&\mathbb{P}\big(\mathscr{B}_X^\gamma(\delta') \not\subset \mathscr{B}_X^*(\epsilon')\big) \\
&= \mathbb{P}\big(\text{there exists } \beta \in X - \mathscr{B}_X^*(\epsilon'),\ H_t(\beta) \leq H_t(\beta_X^*) + \delta'\big) \\
&\leq \sum_{\beta \in X - \mathscr{B}_X^*(\epsilon')} \mathbb{P}\big(H_t(\beta) \leq H_t(\beta_X^*) + \delta'\big) \\
&= \sum_{\beta \in X - \mathscr{B}_X^*(\epsilon')} \mathbb{P}\left(\frac{1}{t}\sum_{\tau=1}^t Y_\beta^\tau \geq -\delta'\right) \\
&\leq \sum_{\beta \in X - \mathscr{B}_X^*(\epsilon')} \mathbb{P}\left(\frac{1}{t}\sum_{\tau=1}^t Y_\beta^\tau - \mu_\beta \geq \epsilon' - \delta'\right) && \text{(A)} \\
&\leq \sum_{\beta \in X - \mathscr{B}_X^*(\epsilon')} \exp\Big(-\frac{2t(\epsilon' - \delta')^2}{L^2\|\beta - \beta^*\|_\infty^2}\Big) && \text{(B)} \\
&\leq |X| \exp\Big(-\frac{2t(\epsilon' - \delta')^2}{L^2\|\beta - \beta_X^*\|_\infty^2}\Big) . && (13)
\end{aligned}$$

Here in (A) we use the fact that $\mu_\beta = H(\beta_X^*) - H(\beta) > -\epsilon'$ for $\beta \in X - \mathscr{B}_X^*(\epsilon')$. In (B), using L-Lipschitzness of H on the set C, we obtain $|Y_\beta^\tau| \leq L\|\beta - \beta_X^*\|_\infty$ and then apply Hoeffding's inequality for bounded random variables. Setting Eq. (13) equal to α and solving for t, we have that if

$$t \geq \frac{L^2 D^2}{2(\epsilon' - \delta')^2}\Big(\log|X| + \log\frac{1}{\alpha}\Big) , \tag{14}$$

then $\mathbb{P}\big(\mathscr{B}_X^\gamma(\delta') \not\subset \mathscr{B}_X^*(\epsilon')\big) \leq \alpha$. Note the above derivation applies to any finite set $X \subset S$.

Now we use the construction $X = S' + \{\beta^*\}$. Then the cardinality of X can be upper bounded by $(4/\nu)^n$. Note since $\beta^* \in X$ it holds $\beta^* = \beta_X^*$. We apply the result in Eq. (14) with the following

parameters

$$\nu = (\epsilon' - \delta')/(4L)\,, \quad \delta' = \delta + L\nu\,, \quad \epsilon' = \epsilon - L\nu\,, \quad \epsilon' - \delta' = \frac{1}{2}(\epsilon - \delta)\,,$$
$$D = \min\{\sqrt{2a/\lambda}, 2\}\,, \quad |X| \le \left(\frac{16L}{\epsilon - \delta}\right)^n.$$

We justify the choice of D. First, $S \subset C$ implies $D \le 2$. By the λ-strong convexity of H on C: for all $\beta \in X \subset S \subset \mathscr{B}^*(a)$, it holds

$$(1/2)\lambda\|\beta - \beta^*_X\|_\infty^2 = (1/2)\lambda\|\beta - \beta^*\|_\infty^2 \le H(\beta) - H^* \le a$$
$$\implies D = \sup_{\beta \in X} \|X - \beta^*_X\|_\infty \le \sqrt{2a/\lambda}\,.$$

Substituting these quantities into the bound Eq. (14) the expression becomes

$$t \ge c' \cdot \frac{L^2}{(\epsilon - \delta)^2} \cdot \min\left\{\frac{2a}{\lambda}, 4\right\} \cdot \left(n \log\left(\frac{16L}{\epsilon - \delta}\right) + \log\frac{1}{\alpha}\right).$$

Here c' is an absolute constant that changes from line to line. Moreover, noting that $a \le 2\epsilon$ and $\delta \le \epsilon/2$ implies $a/(\epsilon - \delta)^2 \le 8/\epsilon$, we know that if

$$t \ge c' \cdot L^2 \min\left\{\frac{1}{\lambda\epsilon}, \frac{1}{\epsilon^2}\right\} \cdot \left(n \log\left(\frac{16L}{\epsilon - \delta}\right) + \log\frac{1}{\alpha}\right), \tag{15}$$

then $\mathbb{P}\big(\mathscr{B}^\gamma_X(\delta') \subset \mathscr{B}^*_X(\epsilon')\big) \ge 1 - \alpha$. By plugging in $L = (2n + \bar{v})$ and $\lambda = \underline{b}/4$, we know $\mathbb{P}\Big(\mathscr{B}^\gamma_S(\delta) \subset \mathscr{B}^*_S(\epsilon)\Big) \ge 1 - \alpha$ as long as

$$t \ge c' \cdot (2n + \bar{v})^2 \min\left\{\frac{1}{\underline{b}\epsilon}, \frac{1}{\epsilon^2}\right\} \cdot \left(n \log\left(\frac{16(2n + \bar{v})}{\epsilon - \delta}\right) + \log\frac{1}{\alpha}\right).$$

Step 3. Putting together. By Lemma 1, if $t \ge 2\bar{v}^2\log(2n/\alpha)$ then $\beta^\gamma \in C$ with probability $\ge 1 - \alpha$. Under the event $\beta^\gamma \in C$, it holds $\mathscr{B}^\gamma_C(\delta) = C \cap \mathscr{B}^\gamma(\delta)$. Since $\beta^* \in C$ it holds that $\mathscr{B}^*_C(\epsilon) = C \cap \mathscr{B}^*(\epsilon)$. Moreover, if t satisfies the bound in Eq. (15), we know Inclusion (2) holds with probability $\ge 1 - \alpha$, which then implies Inclusion (1). So if t satisfies the two requirements, $t \ge 2\bar{v}^2\log(2n/\alpha)$ and Eq. (15), then with probability $\ge 1 - 2\alpha$,

$$C \cap \mathscr{B}^\gamma(\delta) = \mathscr{B}^\gamma_C(\delta) \subset \mathscr{B}^*_C(\epsilon) = C \cap \mathscr{B}^*(\epsilon)\,.$$

□

Proof of Claim 1. To see this, for $\beta \in \mathscr{B}^\gamma_S(\delta)$ let $\beta' \in X$ be such that $\|\beta - \beta'\|_\infty \le \nu$. By Lipschitzness of H_t on C, we know

$$\begin{aligned}
H_t(\beta') &\le H_t(\beta) + L\nu && \text{(Lipschitzness of } H_t\text{)}\\
&\le \min_S H_t + \delta + L\nu && (\beta \in \mathscr{B}^\gamma_S(\delta))\\
&\le \min_X H_t + \delta + L\nu && (X \subset S)\\
&= \min_X H_t + \delta'\,.
\end{aligned}$$

This implies the membership $\beta' \in \mathscr{B}^\gamma_X(\delta')$. Furthermore, we have

$$\mathscr{B}^\gamma_X(\delta') \subset \mathscr{B}^*_X(\epsilon') \subset \mathscr{B}^*_C(\epsilon')\,.$$

Here the first inclusion is simply the assumption that $\mathscr{B}^\gamma_X(\delta') \subset \mathscr{B}^*_X(\epsilon')$. The second inclusion follows by the construction of X; since $\beta^* \in X$, we know $\mathscr{B}^*_X(\epsilon') \subset \mathscr{B}^*_C(\epsilon')$ and thus $\min_X H = \min_X H = H^*$. We now obtain

$$\beta' \in \mathscr{B}^*_C(\epsilon')\,.$$

Using the Lipschitzness of H on C, we have for all $\beta \in \mathscr{B}_S^\gamma(\delta)$

$$\begin{aligned} H(\beta) &\leq H(\beta') + L\nu && \text{(Lipschitzness of } H\text{)} \\ &\leq \min_C H + \epsilon' + L\nu && (\beta' \in \mathscr{B}_C^*(\epsilon')) \\ &= \min_C H + \epsilon\,. \end{aligned}$$

So we conclude $\beta \in \mathscr{B}_C^*(\epsilon)$, implying $\mathscr{B}_S^\gamma(\delta) \subset \mathscr{B}_C^*(\epsilon)$. This completes the proof of Claim 1. □

Proof of Claim 2. This claim relies on convexity of the problem.

Assume, for the sake of contradiction, there exists $\beta^\diamond \in \mathscr{B}_C^\gamma(\delta)$ but $\beta^\diamond \notin \mathscr{B}_S^\gamma(\delta)$. The only possibility this can happen is $\beta^\diamond \in C$ but $\beta^\diamond \notin S = C \cap \mathscr{B}^*(a)$. So $\beta^\diamond \notin \mathscr{B}^*(a)$ (note $a < r$ implies the set $C - \mathscr{B}^*(a)$ is not empty), which by definition means

$$H(\beta^\diamond) - H^* > a\,. \tag{16}$$

Now define

$$\bar{\beta} = \arg\min_{\beta \in S} H_t(\beta) \in \mathscr{B}_S^\gamma(\delta)\,.$$

By the assumption $\mathscr{B}_S^\gamma(\delta) \subset \mathscr{B}_S^*(\epsilon)$, we know $\bar{\beta} \in \mathscr{B}_S^*(\epsilon)$ and so

$$H(\bar{\beta}) - H^* \leq \epsilon\,. \tag{17}$$

Next, let $\beta^c = c\bar{\beta} + (1-c)\beta^\diamond$ with $c \in [0,1]$, which is a point lying on the line segment joining the two points $\bar{\beta}$ and $\beta^\diamond$. By the optimality of $\beta^\diamond \in \mathscr{B}_C^\gamma(\delta)$ and $\bar{\beta} \in C$, we know $H_t(\beta^\diamond) \leq H(\bar{\beta}) + \delta$. By convexity of H_t, we have for all $c \in [0,1]$,

$$H_t(\beta^c) \leq \max\{H_t(\bar{\beta}), H_t(\beta^\diamond)\} \leq H_t(\bar{\beta}) + \delta\,. \tag{18}$$

Now consider the map $K : [0,1] \to \mathbb{R}_+, c \mapsto H(\beta^c) - H^*$. Since any convex function is continuous on its effective domain (Rockafellar, 1970, Corollary 10.1.1), we know H is continuous. Continuity of H implies continuity of K. Note $K(0) = H(\beta^\diamond) - H^* > a$ by Eq. (16) and $K(1) = H(\bar{\beta}) - H^* \leq \epsilon$ by Eq. (17). By intermediate value theorem, there is $c^* \in [0,1]$ such that $\epsilon < H(\beta^{c*}) - H^* < a$. Moreover, by $H(\beta^{c*}) - H^* < a$ and $\beta^{c*} \in C$ we obtain $\beta^{c*} \in S = \mathscr{B}^*(a) \cap C$. In addition, recalling $H_t(\beta^{c*}) \leq H_t(\bar{\beta}) + \delta$ (Eq. (18)), we conclude by definition $\beta^{c*} \in \mathscr{B}_S^\gamma(\delta)$.

At this point we have shown the existence of a point β^{c*} such that

$$\beta^{c*} \in \mathscr{B}_S^\gamma(\delta)\,, \quad \beta^{c*} \notin \mathscr{B}^*(\epsilon)\,.$$

This clearly contradicts the assumption $\mathscr{B}_S^\gamma(\delta) \subset \mathscr{B}_S^*(\epsilon) = \mathscr{B}^*(\epsilon) \cap S$. This completes the proof of Claim 2.

□

Proof of Corollary 1. Under the event $\{\beta^\gamma \in C\}$, the set $C \cap \mathscr{B}^\gamma(0) = \{\beta^\gamma\}$. Moreover, $\beta^\gamma \in C \cap \mathscr{B}^*(\epsilon)$ implies $H(\beta^\gamma) \leq H(\beta^*) + \epsilon$. This completes the proof. □

Proof of Corollary 2. Under the event $\{\beta^\gamma \in C\}$, we use strong convexity of H over C w.r.t. ℓ_2-norm and obtain $\frac{\lambda}{2}\|\beta^\gamma - \beta^*\|_2^2 \leq H(\beta^\gamma) - H(\beta^*)$ where $\lambda = \underline{b}/4$ is the strong-convexity parameter.

For the second claim we use the equality $\beta_i^\gamma = b_i/u_i^\gamma$ and $\beta_i^* = b_i/u_i^*$. For $\beta, \beta' \in C$, it holds $|\frac{1}{\beta_i} - \frac{1}{\beta_i'}| \leq \frac{4}{{b_i}^2}|\beta_i - \beta_i'|$. And so $\|u^\gamma - u^*\|_2 = \sum_i (b_i)^2 (\frac{1}{\beta_i^\gamma} - \frac{1}{\beta_i^*})^2 \leq \sum_i \frac{16}{(b_i)^2}|\beta_i^\gamma - \beta_i^*|^2 \leq \frac{16}{(\underline{b})^2}\|\beta^\gamma - \beta^*\|_2^2$. So we obtain $\|u^\gamma - u^*\|_2 \leq \frac{4}{\underline{b}}\|\beta^\gamma - \beta^*\|_2$. We complete the proof. □

K PROOF OF THEOREMS 4 AND 5

Proof of Theorem 4. We aim to apply Lemma 7 to our problem. To do this we first introduce surrogate problems

$$H_C^\gamma := \min_{\beta \in C} H_t(\beta) \ , \quad H_C^* := \min_{\beta \in C} H(\beta) \ .$$

Since $\beta^* \in C$ we know $H_C^* = H^*$. We write down the decomposition

$$\sqrt{t}(H^\gamma - H^*) = \sqrt{t}(H^\gamma - H_C^\gamma) + \sqrt{t}(H_C^\gamma - H_C^*) \ .$$

For the first term we show that $\sqrt{t}(H^\gamma - H_C^\gamma) \xrightarrow{\text{p}} 0$. Choose any $\epsilon > 0$, define the event $A_t^\epsilon = \{\sqrt{t}|H^\gamma - H_C^\gamma| \geq \epsilon\}$. By Lemma 1 we know that with probability 1, $\beta^\gamma \in C$ eventually and so $H^\gamma - H_C^\gamma = 0$ eventually. This implies $\mathbb{P}((A_t^\epsilon)^c \text{ eventually}) = 1 \Leftrightarrow \mathbb{P}(A_t^\epsilon \text{ infinitely often}) = 0$. By Fatou's lemma, $\limsup_{t\to\infty} \mathbb{P}(A_t^\epsilon) \leq \mathbb{P}(\limsup_{t\to\infty} A_t^\epsilon) = 0$. We conclude for all $\epsilon > 0$, $\lim_{t\to\infty} \mathbb{P}(\sqrt{t}|H^\gamma - H_C^\gamma| > \epsilon) = 0$.

For the second term, we invoke Lemma 7 and obtain $\sqrt{t}(H_C^\gamma - H_C^*) \xrightarrow{\text{d}} \mathcal{N}(0, \text{var}[F(\beta^*, \theta)])$, where we recall $F(\beta, \theta) = \max_i \beta_i v_i(\theta) - \sum_{i=1}^n b_i \log \beta_i$. To do this we verify all hypotheses in Lemma 7.

- The set C is compact and therefore Condition 7.a is satisfied.
- The function F is finite for all $\beta \in \mathbb{R}_{++}^n$ and thus Condition 7.b holds.
- The function $F(\cdot, \theta)$ is $(2n + \bar{v})$-Lipschitz on C for all θ, and thus Condition 7.c holds.
- Condition 7.d holds because the function H has a unique minimizer over C.

Now we calculate the variance term.

$$\begin{aligned}
\text{var} F(\beta^*, \theta) &= \text{var} \max_i \{v_i(\theta)\beta_i^*\} \\
&= \text{var}\, p^*(\theta) \\
&= \int_\Theta (p^*)^2 \, \text{d}S(\theta) - \left(\int_\Theta p^* \, \text{d}S(\theta) \right)^2 \\
&= \int_\Theta (p^*)^2 \, \text{d}S(\theta) - 1 \ ,
\end{aligned}$$

where in the last equality we used $\int p^* = \sum_{i=1}^n b_i = 1$. By Slutsky's theorem, we obtain the claimed result.

□

Proof of Theorem 5. We verify all the conditions in Theorem 2.1 from Hjort and Pollard (2011).

Because H is C^2 at β^*, there exists a neighborhood N of β^* such that H is continuously differentiable on N. By Theorem 6 this implies that the random variable $\epsilon(\beta, \cdot)^{-1}$ is finite almost surely for each $\beta \in N$. This implies $I(\beta, \theta)$ is single valued a.s. for $\beta \in N$.

Define

$$D(\theta) := \nabla F(\beta^*, \theta) = G(\beta^*, \theta) - \nabla \Psi(\beta^*)$$

where we recall the subgradient $G(\beta^*, \theta) = e_{i(\beta^*, \theta)} v_{i(\beta^*, \theta)}$ and $i(\beta^*, \theta) = \arg\max_i \beta_i^* v_i(\theta)$ is the winner of item θ when the pacing multiplier of buyers is β^*. Let

$$R(h, \theta) := F(\beta^* + h, \theta) - F(\beta^*, \theta) - D(\theta)^\top h$$

measure the first-order approximation error. By optimality of β^* we know $\nabla H(\beta^*) = \mathbb{E}[D(\theta)] = 0$. Moreover, by twice differentiability of H at β^*, the following expansion holds:

$$H(\beta^* + h) - H(\beta^*) = \frac{1}{2} h^\top \big(\nabla^2 H(\beta^*)\big) h + o(\|h\|_2^2) \ .$$

To invoke Theorem 2.1 from Hjort and Pollard (2011), we check the following stochastic version of differentiability condition holds

$$\mathbb{E}[R(h, \theta)^2] = o(\|h\|_2^2) \quad \text{as } \|h\|_2 \downarrow 0 \ . \tag{19}$$

As a quick remark, by $\mathbb{E}[D(\theta)] = 0$ being the gradient of H at β^*, we know $\mathbb{E}[R(h,\theta)] = o(\|h\|_2)$. Therefore, the above condition is a slightly stronger condition. Recall $\mathbb{E}[F(\beta,\theta)] = \mathbb{E}[f(\beta,\theta)] + \Psi(\beta)$, where $f(\beta,\theta) = \max_i \beta_i v_i(\theta)$. We discuss the f part and the Ψ part separately. Note

$$(R(h,\theta))^2 \leq \underbrace{2\big(f(\beta^*+h,\theta) - f(\beta^*,\theta) - (e_{i(\beta^*,\theta)} v_{i(\beta^*,\theta)})^\top h\big)^2}_{=:\mathrm{I}(h,\theta)} + \underbrace{2\big(\Psi(\beta^*+h) - \Psi(\beta^*) - \nabla\Psi(\beta^*)^\top h\big)^2}_{=:\mathrm{II}(h)} .$$

By the assumption that H is twice differentiable at β^*, we know Eq. (NO-TIE) holds at β^*, and so $\epsilon(\beta^*,\theta) > 0$. By Lemma 8, for h small enough so that $\|h\|_\infty \leq \epsilon(\beta^*,\theta)/(3\bar{v})$ we have $\arg\max_i \beta_i^* v_i = \arg\max_i (\beta_i^* + h_i) v_i$ and thus $f(\beta^*+h) = v_{i(\beta^*,\theta)}(\beta_{i(\beta^*,\theta)} + h_{i(\beta^*,\theta)})$. And so $f(\beta^*+h) - f(\beta^*,\theta) = v_{i(\beta^*,\theta)} h_{i(\beta^*,\theta)} = G(\theta)^\top h$. Again, since $\epsilon(\beta^*,\theta)^{-1}$ is finite almost surely, we know for almost every θ, the term $\mathrm{I}(h,\theta)$ is exactly zero as $h \downarrow 0$. So $\mathrm{I}(h,\theta)/\|h\|_2^2 \to 0$. By smoothness of Ψ we know $\mathrm{II}(h)/\|h\|_2^2 = (o(\|h\|_2)^2)/\|h\|_2^2 \to 0$. By bounded convergence theorem we can exchange limit and expectation and obtain

$$0 \leq \lim_{\|h\|_2 \downarrow 0} \mathbb{E}\left[\frac{R(h,\theta)^2}{\|h\|_2^2}\right] \leq \mathbb{E}\left[\lim_{\|h\|_2\downarrow 0} (\mathrm{I}(h,\theta) + \mathrm{II}(h))/\|h\|_2^2\right] = 0 .$$

We conclude Eq. (19) holds true.

At this stage we have verified all the conditions in Theorem 2.1 from Hjort and Pollard (2011). Invoking the theorem we obtain

$$\sqrt{t}(\beta^\gamma - \beta^*) = -[\nabla^2 H(\beta^*)]^{-1}\left(\frac{1}{\sqrt{t}}\sum_{\tau=1}^{t} D(\theta^\tau)\right) + o_p(1) .$$

In particular, $\sqrt{t}(\beta^\gamma - \beta^*) \xrightarrow{\mathrm{d}} \mathcal{N}(0, [\nabla^2 H(\beta^*)]^{-1}\mathrm{var}(D)[\nabla^2 H(\beta^*)]^{-1})$. We calculate the variance matrix of D:

$$\begin{aligned}
\mathrm{var}(D(\theta)) &= \mathrm{var}\left(\sum_{i=1}^{n} \mathbb{1}(\Theta_i^*) v_i(\theta) e_i\right) \\
&= \sum_{i=1}^{n} \Omega_i^2 (e_i e_i^\top) \qquad \text{(A)} \\
&= \mathrm{Diag}(\{\Omega_i^2\}_i) ,
\end{aligned}$$

where in (A) we have used $\Omega_i^* \cap \Omega_j^* = \emptyset$ for $j \neq i$ when Eq. (NO-TIE) holds at β^*.

Proof of CLT for β This follows from the discussion above.

Proof of CLT for u. We use the delta method. Take $g(\beta) = [b_1/\beta_1, \ldots, b_n/\beta_n]$. Then the asymptotic variance of $\sqrt{t}(g(\beta^\gamma) - g(\beta^*))$ is $\nabla g(\beta^*)^\top \Sigma_\beta \nabla g(\beta^*)$. Note $\nabla g(\beta^*)$ is the diagonal matrix $\mathrm{Diag}(\{-b_i/\beta_i^{*2}\})$. From here we obtain the expression for Σ_u.

□

L Proofs for Analytical Properties of the Dual Objective

Remark 1 (Comment on Theorem 6)**.** *We briefly discuss why differentiability is related to the gap in buyers' bids. Recall $\bar{f}(\beta) = \mathbb{E}[\max_i \beta_i v_i(\theta)]$. Let $\delta \in \mathbb{R}_+^n$ be a direction with positive entries, and let $I(\beta,\theta) = \arg\max_i \beta_i v_i(\theta)$ be the set of winners of item θ which could be multivalued. Consider the directional derivative of $\bar{f}$ at β along the direction δ:*

$$\begin{aligned}
&\lim_{t\downarrow 0} \mathbb{E}\left[\frac{\max_i(\beta_i + t\delta_i)v_i(\theta) - \max_i \beta_i v_i(\theta)}{t}\right] \\
&= \mathbb{E}\left[\lim_{t\downarrow 0} \frac{\max_i(\beta_i + t\delta_i)v_i(\theta) - \max_i \beta_i v_i(\theta)}{t}\right] \\
&= \mathbb{E}\left[\max_{i\in I(\beta,\theta)} v_i(\theta)\delta_i\right] ,
\end{aligned}$$

where the exchange of limit and expectation is justified by the dominated convergence theorem. Similarly, the left limit is

$$\lim_{t\uparrow 0}\mathbb{E}\left[\frac{\max_i(\beta_i+t\delta_i)v_i(\theta)-\max_i\beta_i v_i(\theta)}{t}\right]=\mathbb{E}\Big[\min_{i\in I(\beta,\theta)}v_i(\theta)\delta_i\Big]\ .$$

If there is a tie at β with positive probability, i.e., the set $I(\beta,\theta)$ is multivalued for a non-zero measure set of items, then the left and right directional derivatives along the direction δ do not agree. Since differentiability at a point β implies existence of directional derivatives, we conclude differentiability implies Eq. (NO-TIE).

Proof of Theorem 6. Recall $f(\beta,\theta)=\max_i\beta_i v_i(\theta)$. Note f is differentiable at β if and only if $\epsilon(\beta,\theta)>0$. Let $\Theta_{\text{diff}}(\beta):=\{\theta: f(\beta,\theta)\text{ is continuously differentiable at }\beta\}$. Then

$$\Theta_{\text{diff}}(\beta)=\left\{\theta:\frac{1}{\epsilon(\beta,\theta)}<\infty\right\}=\{\theta: I(\beta,\theta)\text{ is single-valued}\}\ .$$

By Proposition 2.3 from Bertsekas (1973) we know $\bar{f}(\beta)=\mathbb{E}[f(\beta,\theta)]=\int_\Theta f(\beta,\theta)\,\mathrm{d}S(\theta)$ is differentiable at β if and only if $S(\Theta_{\text{diff}}(\beta))=1$. From here we obtain Theorem 6.

□

Remark 2. *Suppose Eq.* (NO-TIE) *holds in a neighborhood N of β^*, i.e., $\frac{1}{\epsilon(\beta,\theta)}$ is finite a.s. for each $\beta\in N$, then by Theorem 6 we know H is differentiable on N. In fact, a stronger statement holds: H is* ***continuously*** *differentiable on N. See Proposition 2.1 from Shapiro (1989).*

Proof of Theorem 10. Eq. (INT) holds in a neighborhood N of β^* implies the Eq. (NO-TIE) holds on N, and thus H is differentiable on N with gradient $\nabla H(\beta)=\mathbb{E}[v_{i(\beta,\theta)}e_{i(\beta,\theta)}]+\nabla\Psi=\mathbb{E}[G(\beta,\theta)]+\nabla\Psi$. To compute the Hessian w.r.t. the first term, we look at the limit

$$\lim_{\|h\|\downarrow 0}\mathbb{E}\left[\frac{G(\beta^*+h,\theta)-G(\beta^*,\theta)}{\|h\|}\right]\ . \tag{20}$$

Lemma 8 ($\epsilon(\beta,\theta)$ as Lipschitz parameter of G)**.** *Suppose for some $\beta\in\mathbb{R}^n_{++}$ the gap function $\epsilon(\beta,\theta)>0$. Let $\beta'=\beta+h$.*

- *If $\|h\|_\infty\le\epsilon(\beta,\theta)/\bar{v}$ then $I(\beta',\theta)$ is single-valued and moreover $i(\beta',\theta)=i(\beta,\theta)$, implying $G(\beta',\theta)=G(\beta,\theta)$.*
- *It holds $\|G(\beta+h,\theta)-G(\beta,\theta)\|_2\le 6\bar{v}^2\cdot\frac{1}{\epsilon(\beta,\theta)}\|h\|_2$ for all $\beta+h\in\mathbb{R}^n_+$.*

Suppose we could exchange expectation and limit in Eq. (20), then the above expression would become zero: for a fixed θ, since Eq. (NO-TIE) holds at β^*, i.e, $\epsilon(\beta^*,\theta)>0$, we apply Lemma 8 and obtain $\lim_{\|h\|\downarrow 0}(G(\beta^*+h,\theta)-G(\beta^*,\theta))/\|h\|=0$. This implies that H is twice differentiable at β^* with Hessian $\nabla^2H(\beta^*)=\nabla^2\Psi(\beta^*)$. It is then natural to ask for sufficient conditions for exchanging limit and expectation.

By Lemma 8, we know the ratio $(G(\beta^*+h,\theta)-G(\beta^*,\theta))/\|h\|$ is dominated by $6\bar{v}\epsilon(\beta^*,\theta)^{-1}$, which by Eq. (INT) is integrable. By dominated convergence theorem, we can exchange limit and expectation, and the claim follows. □

Proof of Lemma 8. Note that for any β and θ with $\epsilon(\beta,\theta)>0$, and any $\beta'=\beta+h$,

$$\frac{\|G(\beta+h,\theta)-G(\beta,\theta)\|_2}{\|h\|_2}\le 6\bar{v}^2\cdot\frac{1}{\epsilon(\beta,\theta)}\ . \tag{21}$$

To see this, we notice that on one hand, if $\|h\|_\infty\le\epsilon/(3\bar{v})$ where $\epsilon=\epsilon(\beta,\theta)$, then for $i=i(\beta,\theta)$ and all $\theta\in\Theta_i(\beta)$,

$$\begin{aligned}\beta'_i v_i(\theta)&=(\beta_i+h_i)v_i(\theta)\\&\ge\beta_i v_i(\theta)-\epsilon/3 &\text{(A)}\\&\ge\beta_k v_k(\theta)+\epsilon-\epsilon/3 &\text{(B)}\\&\ge\beta'_k v_k(\theta)-\epsilon/3+\epsilon-\epsilon/3\ , &\text{(C)}\end{aligned}$$

where (A) and (C) use the fact $\|h\|_\infty \le \epsilon/(3\bar{v})$, and (B) uses the definition of ϵ. This implies $\arg\max_i \beta_i' v_i(\theta) = \arg\max_i \beta_i v_i(\theta)$ and thus $G(\beta+h,\theta) - G(\beta,\theta) = 0$. On the other hand, if $\|h\|_\infty > \epsilon/(3\bar{v})$, then $\|h\|_2 \ge \|h\|_\infty > \epsilon/(3\bar{v})$. Using the bound $\|G\|_2 \le \bar{v}$, we obtain Eq. (21). This completes proof of Lemma 8. □

Proof of Theorem 11. By Lemma 5 from Gao and Kroer (2022), we know that at equilibrium the there exists unique breakpoints $0 = a_0^* < a_1^* < \cdots < a_n^* = 1$ such that buyer i receives the item set $[a_{i-1}^*, a_i^*] \subset \Theta$. Moreover, it holds

$$\beta_1^* d_1 > \beta_2^* d_2 > \cdots > \beta_n^* d_n \ ,$$
$$\beta_1^* c_1 < \beta_2^* c_2 < \cdots < \beta_n^* c_n \ .$$

Now we consider a small enough neighborhood N of β^*. For each $\beta \in N$, we define the breakpoint $a_i^*(\beta)$ by solving for θ through $\beta_i(c_i\theta + d_i) = \beta_{i+1}(c_{i+1}\theta + d_{i+1})$ for $i \in [n-1]$, $a_0^*(\beta) = 0$, and $a_n^*(\beta) = 1$. As a sanity check, note $a_i^*(\beta^*) = a_i^*$ for all i. Recall $\Theta_i(\beta) = \{\theta \in \Theta : v_i(\theta)\beta_i \ge v_k(\theta)\beta_k, \forall k \ne i\}$ is the winning item set of buyer i when pacing multiplier equal β. We will show later $\Theta_i(\beta) = [a_{i-1}^*(\beta), a_i^*(\beta)]$ for $\beta \in N$ when N is appropriately constructed.

Now we recall the gradient expression

$$\begin{aligned}
\nabla \bar{f}(\beta) &= \mathbb{E}[e_{i(\beta,\theta)} v_{i(\beta,\theta)}] \\
&= \sum_{i=1}^n e_i \int_{a_{i-1}^*(\beta)}^{a_i^*(\beta)} c_i\theta + d_i \,\mathrm{d}\theta \\
&= \sum_{i=1}^n e_i \left(\frac{c_i}{2}\big([a_i^*(\beta)]^2 - [a_{i-1}^*(\beta)]^2\big) + d_i\big(a_i^*(\beta) - a_{i-1}^*(\beta)\big)\right) .
\end{aligned}$$

On N, the breakpoints $a_i^*(\beta) = (-\beta_i d_i + \beta_{i+1} d_{i+1})/(\beta_i c_i - \beta_{i+1} c_{i+1})$ is C^1 in the parameter β. We conclude $\nabla \bar{f}(\beta)$ is continuously differentiable.

What remains is to construct such a neighborhood N. Define

$$\delta = \min\left\{ \frac{1}{2}\underline{\Delta}_{\beta d}/\bar{\Delta}_d, \frac{1}{2}\underline{\Delta}_{\beta c}/\bar{\Delta}_c, \frac{1}{4}\underline{\Delta}_a \underline{\Delta}_{\beta c}/\bar{v} \right\} ,$$

where $\underline{\Delta}_a := \min |a_i - a_{i-1}|$, $\underline{\Delta}_{\beta c} := \min\{\beta_{i-1}^* c_{i-1} - \beta_i^* c_i\} > 0$, $\bar{\Delta}_c := \max_i\{c_{i-1} - c_i\} > 0$, and $\underline{\Delta}_{\beta d} > 0$ and $\bar{\Delta}_d > 0$ are similarly defined. Let $N = \{\beta : \|\beta - \beta^*\|_\infty \le \delta\}$. The neighborhood N is constructed so that on N it holds

$$\begin{aligned}
&\beta_1 d_1 > \beta_2 d_2 > \cdots > \beta_n d_n \ , \\
&\beta_1 c_1 < \beta_2 c_2 < \cdots < \beta_n c_n \ , \\
&0 = a_0^*(\beta) < a_1^*(\beta) < \cdots < a_n^*(\beta) = 1 \ ,
\end{aligned}$$

where the first inequality follows from $\delta \le \frac{1}{2}\underline{\Delta}_{\beta d}/\bar{\Delta}_d$, the second inequality from $\delta \le \frac{1}{2}\underline{\Delta}_{\beta c}/\bar{\Delta}_c$, and the third inequality follows from the $\delta \le \underline{\Delta}_a \underline{\Delta}_{\beta c}/(4\bar{v})$, where $\bar{v} = \max_i \sup_{\theta\in[0,1]} c_i\theta + d_i$. From here we can see that the partition of item set Θ induced by these breakpoints corresponds to exactly the winning item sets when buyers' pacing multiplier is β. So we have shown $\Theta_i(\beta) = [a_{i-1}^*(\beta), a_i^*(\beta)]$ for $\beta \in N$. This finishes the proof of Theorem 11. □

Proof of Theorem 12. We need the following technical lemma on the continuous differentiability of integral functions.

Lemma 9 (Adapted from Lemma 2.5 from Wang (1985))**.** *Let $u = [u_1, \ldots, u_n] \in \mathbb{R}_{++}^n$, and*

$$I(u) = \int_0^{u_1} \mathrm{d}t_1 \cdots \int_0^{u_n} h(t_1, t_2, \ldots, t_n)\,\mathrm{d}t_n \ ,$$

where h is a continuous density function of a probabilistic distribution function on $\mathbb{R}_+^n$ and such that all lower dimensional density functions are also continuous. Then the integral $I(u)$ is continuously differentiable.

Remark 3. *The difference between the above lemma and the original statement is that the original theorem works with density h and integral function $I(u)$ both defined on $\mathbb{R}^n$, while the adapted version works with density h and integral $I(u)$ defined only on $\mathbb{R}^n_{++}$.*

Recall the gradient expression

$$\nabla \bar{f}(\beta) = \mathbb{E}[e_{i(\beta,\theta)} v_{i(\beta,\theta)}] = \sum_{i=1}^{n} e_i \int v_i \mathbb{1}(V_i(\beta)) f(v)\,\mathrm{d}v \ ,$$

where the set $V_i(\beta) = \{v \in \mathbb{R}^n_+ : v_i\beta_i \geq v_k\beta_k, k \neq i\}, i \in [n]$, is the values for which buyer i wins. For now, we focus on the first entry of the gradient, i.e., $\int v_1 \mathbb{1}(V_1(\beta)) f(v)\,\mathrm{d}v$. We write the integral more explicitly as follows. By Fubini's theorem,

$$\int v_1 \mathbb{1}(V_1(\beta)) f(v)\,\mathrm{d}v \tag{22}$$

$$= \int_0^\infty \mathrm{d}v_1 \int_0^{\frac{\beta_1 v_1}{\beta_2}} \mathrm{d}v_2 \int_0^{\frac{\beta_1 v_1}{\beta_3}} \mathrm{d}v_3 \cdots \int_0^{\frac{\beta_1 v_1}{\beta_n}} \underbrace{(v_1 f(v_1, \ldots, v_n))}_{=:A_1(v)} \mathrm{d}v_n \ . \tag{23}$$

To apply the lemma we use a change of variable. Let $t = T(v) = [v_1, \frac{v_2}{v_1}, \ldots, \frac{v_n}{v_1}]$ and $v = T^{-1}(t) = [t_1, t_2 t_1, \ldots, t_n t_1]$. Then Eq. (23) is equal to

$$\int_0^\infty \mathrm{d}t_1 \int_0^{\frac{\beta_1}{\beta_2}} \mathrm{d}t_2 \int_0^{\frac{\beta_1}{\beta_3}} \mathrm{d}t_3 \cdots \int_0^{\frac{\beta_1}{\beta_n}} \underbrace{\left(t_1^n f(t_1, t_2 t_1, \ldots, t_n t_1)\right)}_{=:A_2(t)} \mathrm{d}t_n \ . \tag{24}$$

Note $\mathbb{E}[v_1(\theta)] = \int_{\mathbb{R}^n_{++}} A_1(v)\,\mathrm{d}v = \int_{\mathbb{R}^n_{++}} A_2(t)\,\mathrm{d}t = 1$. We use Fubini's theorem and obtain

$$Eq.\ (24) = \int_0^{\frac{\beta_1}{\beta_2}} \mathrm{d}t_2 \int_0^{\frac{\beta_1}{\beta_3}} \mathrm{d}t_3 \cdots \int_0^{\frac{\beta_1}{\beta_n}} h(t_{-1})\,\mathrm{d}t_{-1} \ ,$$

where we have defined $h(t_{-1}) = \int_{\mathbb{R}_+} t_1^n f(t_1, t_2 t_1, \ldots, t_n t_1)\,\mathrm{d}t_1$. By the smoothness assumption on h and Lemma 9, we know that the map $u_{-1} \mapsto \int_0^{u_2} \mathrm{d}t_2 \cdots \int_0^{u_n} h(t_{-1})\,\mathrm{d}t_n$ is C^1 for all $u_{-1} \in \mathbb{R}^{n-1}_{++}$. Moreover, the map $\beta \mapsto [\frac{\beta_1}{\beta_2}, \ldots, \frac{\beta_1}{\beta_n}]$ is C^1. We conclude the first entry of $\nabla \bar{f}(\beta)$ is C^1 in the parameter β. A similar argument applies to other entries of the gradient. We complete the proof of Theorem 12. □

M Proof of Theorem 8

Proof of Theorem 8. Define the functions

$$\hat{\sigma}^2(\beta) := \frac{1}{t} \sum_{\tau=1}^{t} \Big(F(\beta, \theta^\tau) - H_t(\beta) \Big)^2 \ ,$$

$$\sigma^2(\beta) := \mathrm{var}(F(\beta, \theta)) = \mathbb{E}\big[(F(\beta, \theta) - H(\beta))^2 \big] \ .$$

We will show uniform convergence of $\hat{\sigma}^2$ to σ^2 on C, i.e., $\sup_{\beta \in C} |\hat{\sigma}^2 - \sigma^2| \xrightarrow{\text{a.s.}} 0$. We first rewrite $\hat{\sigma}^2$ as follows

$$\hat{\sigma}^2(\beta) = \underbrace{\frac{1}{t} \sum_{\tau=1}^{t} \big(F(\beta, \theta^\tau) - H(\beta) \big)^2}_{=:\mathrm{I}(\beta)} - \underbrace{(H_t(\beta) - H(\beta))^2}_{=:\mathrm{II}(\beta)} \ .$$

By Theorem 7.53 of Shapiro et al. (2021), the following uniform convergence results hold

$$\sup_{\beta \in C} |\mathrm{I}(\beta) - \sigma^2(\beta)| \xrightarrow{\text{a.s.}} 0 \ , \quad \sup_{\beta \in C} |\mathrm{II}(\beta)| \xrightarrow{\text{a.s.}} 0 \ .$$

The above two inequalities imply $\sup_{\beta\in C}|\hat{\sigma}^2-\sigma^2|\xrightarrow{\text{a.s.}}0$. Note the variance estimator $\hat{\sigma}^2_{\text{NSW}}=\hat{\sigma}^2(\beta^\gamma)$ and the asymptotic variance $\sigma^2_{\text{NSW}}=\sigma^2(\beta^*)$. By $\beta^\gamma\xrightarrow{\text{a.s.}}\beta^*$ we know,

$$\begin{aligned}&|\hat{\sigma}^2_{\text{NSW}}-\sigma^2_{\text{NSW}}|\\&=|\hat{\sigma}^2(\beta^\gamma)-\sigma^2(\beta^*)|\\&\le|\hat{\sigma}^2(\beta^\gamma)-\sigma^2(\beta^\gamma)|+|\sigma^2(\beta^\gamma)-\sigma^2(\beta^*)|\\&\to 0\quad\text{a.s.}\end{aligned}$$

where the first term vanishes by the uniform convergence just established, the second term by continuity of $\sigma^2(\cdot)$ at β^*. Now we have shown $\hat{\sigma}^2_{\text{NSW}}$ is a consistent variance estimator for the asymptotic variance. Then note

$$\frac{\sqrt{t}(\text{NSW}^\gamma-\text{NSW}^*)}{\hat{\sigma}_{\text{NSW}}}=\frac{\sqrt{t}(\text{NSW}^\gamma-\text{NSW}^*)}{\sigma_{\text{NSW}}}\cdot\frac{\sigma_{\text{NSW}}}{\hat{\sigma}_{\text{NSW}}}\ .$$

Since $\frac{\sqrt{t}(\text{NSW}^\gamma-\text{NSW}^*)}{\sigma_{\text{NSW}}}\xrightarrow{\text{d}}\mathcal{N}(0,1)$ by Theorem 4, and $\frac{\sigma_{\text{NSW}}}{\hat{\sigma}_{\text{NSW}}}\xrightarrow{\text{p}}1$ which is a constant, by Slutsky's theorem we know $\frac{\sqrt{t}(\text{NSW}^\gamma-\text{NSW}^*)}{\hat{\sigma}_{\text{NSW}}}\xrightarrow{\text{d}}\mathcal{N}(0,1)$. This completes the proof of Theorem 8. □

Proof of Theorem 13. By assumption, there exists a neighborhood N of β^* such that the set $I(\beta,\theta)$ is single-valued for all $\beta\in N$ and almost all θ. So $G(\beta,\theta)=v_{i(\beta,\theta)}e_{i(\beta,\theta)}$ for $\beta\in N$. Let $G_i(\beta,\theta)=v_i\mathbb{1}\{i=i(\beta,\theta)\}$ be the i-th entry of the vector $G(\beta,\theta)$. For any $\beta'\in N$ define

$$\hat{\Omega}_i^2(\beta'):=\frac{1}{t}\sum_{\tau=1}^t\left(G_i(\beta',\theta^\tau)-\left(\frac{1}{t}\sum_{\tau=1}^t G_i(\beta',\theta^\tau)\right)\right)^2\ .$$

By $\beta^\gamma\xrightarrow{\text{a.s.}}\beta^*$, we know that for large enough t, $\beta^\gamma\in N$. Moreover, we claim

$$\beta^\gamma\in N\implies G_i(\beta^\gamma,\theta^\tau)=tu_i^{\gamma,\tau}\ .$$

To see this, $\beta^\gamma\in N$ implies that the set of items that incurs a tie is zero-measure, i.e, $S(\theta: I(\beta^\gamma,\theta)\text{ multivalued})=0$. By the first-order condition of finite sample EG (Fact 2), the equilibrium allocation in the observed market is then unique and pure (no splitting of items, $x_i^{\gamma,\tau}\in\{0,\frac{1}{t}\}$), in which case $G_i(\beta^\gamma,\theta)=\mathbb{1}\{i=i(\beta,\theta)\}v_i(\theta^\tau)=tx_i^{\gamma,\tau}v_i(\theta^\tau)=tu_i^{\gamma,\tau}$ and thus $\hat{\Omega}_i^2(\cdot)|_{\beta^\gamma}=\hat{\Omega}_i^2$. By Theorem 7.53 of Shapiro et al. (2021) one can show the following uniform convergence result

$$\sup_{\beta\in N}|\hat{\Omega}_i^2(\beta)-\text{var}[G_i(\beta,\theta)]|\xrightarrow{\text{a.s.}}0\ .$$

Noting $\Omega_i^2=\text{var}[G_i(\beta^*,\theta)]$, the uniform convergence result implies $\hat{\Omega}_i^2(\beta^\gamma)-\text{var}[G_i(\beta^*,\theta)]\xrightarrow{\text{p}}0$, which is equivalent to $\hat{\Omega}_i^2\xrightarrow{\text{p}}\Omega_i^2$.

This completes the proof of Theorem 13. □

N Proof of Theorem 14

We start by formally defining quasilinear market equilibrium.

Definition 6. *The long-run QME, $\mathcal{QME}(b,v,s)$ is an allocation-utility-price tuple $(x^*,u^*,p^*)\in(L_+^\infty)^n\times\mathbb{R}_+^n\times L_+^1$ such that conditions in Definition 1 hold, except that buyer optimality is now defined as*

2' $x_i^*\in D_i(p^*)$ and $u_i^*=\langle v_i-p^*,x_i\rangle$ for all i where the demand D_i of buyer i is its set of utility-maximizing allocations given the prices and budget:

$$D_i(p):=\arg\max\{\langle v_i-p,x_i\rangle: x_i\in L_+^\infty,\ \langle p,x_i\rangle\le b_i\}\ .$$

Definition 7. *The observed QME, $\mathcal{QME}^\gamma(b,v,\mathsf{s})$, given an item sequence γ, is an allocation-utility-price tuple $(x^\gamma,u^\gamma,p^\gamma)\in(\mathbb{R}_+^t)^n\times\mathbb{R}_+^n\times\mathbb{R}_+^t$ such that Definition 2 holds except that buyer optimality is now defined as*

2' $x_i^\gamma \in D_i(p^\gamma)$ *and* $u_i^\gamma = \langle v_i(\gamma) - p^\gamma, x_i \rangle$ *for all i, where (overloading notations)*

$$D_i(p) := \arg\max\{\langle v_i(\gamma) - p, x_i \rangle : x_i \geq 0,\ \langle p, x_i \rangle \leq b_i\}\ .$$

We will need the following convex program characterizations of infinite-dimensional QME introduced in Secion 6 of Gao and Kroer (2022). First we state the primal and dual convex programs.

$$\begin{aligned} \sup \sum_{i=1}^{n} (b_i \log \mu_i - \delta_i) \\ \text{s.t. } \mu_i \leq \langle v_i, x_i \rangle + \delta_i, \forall i \in [n] \\ \sum_{i=1}^{n} x_i \leq s \\ \mu_i \geq 0, \delta_i \geq 0, x_i \in L_1(\Theta)_+, \forall i \in [n] \end{aligned} \tag{P-QEG}$$

$$\begin{aligned} \inf \langle p, 1 \rangle - \sum_{i=1}^{n} b_i \log \beta_i \\ \text{s.t. } p \geq \beta_i v_i, \beta_i \leq 1, \forall i \in [n] \\ p \in L_1(\Theta)_+, \beta \in \mathbb{R}_+^d \end{aligned} \tag{P-DQEG}$$

Fact 3 (Theorem 10 from Gao and Kroer (2022) and Appendix C in Gao et al. (2021))**.** *The following holds.*

1. First-order conditions. For any feasible solutions $(x^*_{\text{QEG}}, \mu^*, \delta^*)$ *to Eq.* (P-QEG) *and a feasible solution* $(p^*_{\text{QEG}}, \beta^*)$ *to Eq.* (P-DQEG)*. They are optimal both to the respective convex programs if and only if*

$$\begin{aligned} & p^*_{\text{QEG}} = \max_i \beta_i^* v_i \\ & \left\langle p^*_{\text{QEG}}, 1 - \textstyle\sum_i x^*_{\text{QEG},i} \right\rangle = 0 \\ & \mu_i^* = \frac{b_i}{\beta_i^*}, \forall i \\ & \delta_i^* (1 - \beta_i^*) = 0, \forall i \\ & \left\langle p^*_{\text{QEG}} - \beta_i^* v_i, x_i^* \right\rangle = 0, \forall i \end{aligned}$$

2. Equilibrium. A pair of allocations and prices (x^*, p^*) *is a QME if and only if there exists a* δ^* *and* β^* *such that* (x^*, δ^*) *and* (p^*, β^*) *are optimal solutions to Eq.* (P-QEG) *and Eq.* (P-DQEG)*, respectively.*

3. Bounds on β^**. It holds* $\frac{b_i}{\nu_i + b_i} \leq \beta_i^* \leq 1$ *for all i, where* $\nu_i = \int v_i(\theta)\, \mathrm{d}S(\theta)$*.*

Note the variable μ in Eq. (P-QEG) doest not correspond to the equilibrium utility of buyer i at optimality. The equilibrium utility of buyer i is $\langle v_i - p^*, x_i^* \rangle$. By the discussion in Section 6 of Gao and Kroer (2022), if $\beta_i^* < 1$, then $\langle v_i - p^*, x_i^* \rangle = (1 - \beta_i^*)\mu_i^*$. If $\beta_i^* = 1$, then $\langle v_i - p^*, x_i^* \rangle = 0$. Moreover, δ_i^* represents the leftover budget in equilibrium (Conitzer et al., 2022a).

Proof of Theorem 14. Given the above equivalence results, we use (p^*, x^*) to denote both the equilibrium prices and allocations, as well as the optimal p and x variables in the quasilinear EG programs.

Now the study of the convergence of revenue is reduced to that of the convergence behavior of convex programs

$$\min_{0<\beta\leq 1_n} H_t(\beta) \quad \text{“}\Longrightarrow\text{”} \quad \min_{0<\beta\leq 1_n} H(\beta)$$

Note that in contrast to the EG programs for linear utilities (Eq. (P-DEG) and Eq. (S-DEG)), we now have an upper bound on the variables β.

By repeating the proof of Theorem 9 we obtain $\beta^\gamma \xrightarrow{\text{a.s.}} \beta^*$. To show almost sure convergence of revenue, we note

$$\begin{aligned}
&\Big|\frac{1}{t}\sum_{\tau=1}^{t} p^\tau - \int_\Theta p^*(\theta)s(\theta)\,\mathrm{d}\mu(\theta)\Big| \\
&\le \frac{1}{t}\sum_{\tau=1}^{t}\big|\max_i\{v_i(\theta^\tau)\beta_i^\gamma\} - \max_i\{v_i(\theta^\tau)\beta_i^*\}\big| + \Big|\frac{1}{t}\sum_{\tau=1}^{t}\max_i\{v_i(\theta^\tau)\beta_i^*\} - \int_\Theta p^*(\theta)s(\theta)\,\mathrm{d}\mu(\theta)\Big| \\
&\le \bar{v}\|\beta^\gamma - \beta^*\|_\infty + \Big|\frac{1}{t}\sum_{\tau=1}^{t}\max_i\{v_i(\theta^\tau)\beta_i^*\} - \int_\Theta p^*(\theta)\,\mathrm{d}S(\theta)\Big| \xrightarrow{\text{a.s.}} 0\,.
\end{aligned}$$

Here the first term converges to zero a.s. by $\beta^\gamma \xrightarrow{\text{a.s.}} \beta^*$, and the second term converges to 0 a.s. by strong law of large numbers and noting $\mathbb{E}[\max_i\{v_i(\theta)\beta_i^*\}] = \mathbb{E}[p^*(\theta)]$. This proves the first part of the statement.

Define $\underline{\beta}_{Q,i} := \frac{b_i}{\nu_i + b_i}$ and $\bar{\beta}_Q := 1$. We know $\underline{\beta}_{Q,i} \le \beta_i^* \le \bar{\beta}_Q$ from Fact 3. Here we use subscript Q to denote quantities related to the quasilinear market. Define the set

$$C_Q := \prod_{i=1}^{n}\left[\frac{b_i}{2\nu_i + b_i}, 1\right].$$

Clearly we have $\beta^* \in C_Q$. Furthermore, for t large enough $\beta^\gamma \in C_Q$ with high probability. To see this, if t satisfies $t \ge 2(\bar{v}/\min_i \nu_i)^2 \log(2n/\eta)$, then $\frac{1}{t}\sum_{\tau=1}^{t} v_i(\theta^\tau) \le 2\mathbb{E}[v_i(\theta)]$ for all i with probability $\ge 1-\eta$. By a bound on β^γ in the QME

$$\beta_i^\gamma \ge \frac{b_i}{b_i + \frac{1}{t}\sum_{\tau=1}^{t} v_i(\theta^\tau)},$$

(see Section 6 in Gao and Kroer (2022)), we obtain $\beta_i^\gamma \ge \frac{b_i}{b_i + 2\nu_i}$ (recall $\nu_i = \mathbb{E}[v_i(\theta)]$).

To obtain the convergence rate, we simply repeat the proof of Theorem 3. Let L_Q and λ_Q be the Lipschitz constant and strong convexity constants of H and H_t w.r.t ℓ_∞-norm on C_Q. We obtain from Eq. (15) that with probability $\ge 1-2\alpha$, there exists a constant c' such that as long as

$$t \ge c' \cdot L_Q^2 \min\left\{\frac{1}{\lambda_Q \epsilon}, \frac{1}{\epsilon^2}\right\} \cdot \left(n \log\left(\frac{16 L_Q}{\epsilon - \delta}\right) + \log\frac{1}{\alpha}\right), \tag{25}$$

it holds $|H(\beta^\gamma) - H(\beta^*)| < \epsilon$ and that $\beta^\gamma \in C_Q$ (see Corollary 1).

Next we calculate L_Q and λ_Q. Note on C_Q, the minimum eigenvalue of $\nabla^2\Psi(\beta) = \mathrm{Diag}\{\frac{b_i}{(\beta_i)^2}\}$ can be lower bounded by $\underline{b}$. So we conclude $\lambda_Q = \underline{b}$. And the Lipschitzness constant can be seen by the following. For $\beta, \beta' \in C_Q$,

$$\begin{aligned}
&|H_t(\beta) - H_t(\beta')| \\
&\le \frac{1}{t}\sum_{\tau=1}^{t}\big|\max_i\{v_i(\theta^\tau)\beta_i\} - \max_i\{v_i(\theta^\tau)\beta_i'\}\big| + \sum_{i=1}^{n} b_i\big|\log\beta_i - \log\beta_i'\big| \\
&\le \bar{v}\|\beta - \beta'\|_\infty + \sum_{i=1}^{n} b_i \cdot \frac{1}{b_i/(2\nu_i + b_i)}|\beta_i - \beta_i'| \\
&\le \big(\bar{v} + 2\bar{\nu}n + 1\big)\|\beta - \beta'\|_\infty\,.
\end{aligned}$$

Similar argument shows that H is also $(\bar{v} + 2\bar{\nu}n + 1)$-Lipschitz on C_Q. We conclude $L_Q = (\bar{v} + 2\bar{\nu}n + 1)$. Now Eq. (25) shows that for $t = \Omega(\underline{b}^{-1})$ (so that the $1/(\lambda_Q\epsilon)$ term in the min becomes dominant) we have

$$|H(\beta^\gamma) - H(\beta^*)| = \tilde{O}_p\left(\frac{n\big(\bar{v} + 2\bar{\nu}n + 1\big)^2}{\underline{b}t}\right),$$

where we use $\tilde{O}_p$ to ignore logarithmic factors of t. Moreover,

$$\|\beta^\gamma - \beta^*\|_\infty \leq \sqrt{2|H(\beta^\gamma) - H(\beta^*)|/\lambda_Q} = \tilde{O}_p\left(\frac{\sqrt{n}\big(\bar{v} + 2\bar{\nu}n + 1\big)}{\underline{b}\sqrt{t}}\right).$$

From here we obtain

$$\begin{aligned}
&|\mathrm{REV}^\gamma - \mathrm{REV}^*| \\
&\leq \bar{v}\|\beta^\gamma - \beta^*\|_\infty + \left|\tfrac{1}{t}\sum_{\tau=1}^{t} \max_i\{v_i(\theta^\tau)\beta_i^*\} - \int_\Theta p^*(\theta)\,\mathrm{d}S(\theta)\right| \\
&= \tilde{O}_p\left(\frac{\bar{v}\sqrt{n}\big(\bar{v} + 2\bar{\nu}n + 1\big)}{\underline{b}\sqrt{t}}\right) + O_p\left(\frac{\bar{v}}{\sqrt{t}}\right) \\
&= \tilde{O}_p\left(\frac{\bar{v}\sqrt{n}\big(\bar{v} + 2\bar{\nu}n + 1\big)}{\underline{b}\sqrt{t}}\right).
\end{aligned}$$

We conclude $|\mathrm{REV}^\gamma - \mathrm{REV}^*| = \tilde{O}_p\left(\frac{\bar{v}\sqrt{n}(\bar{v}+2\bar{\nu}n+1)}{\underline{b}\sqrt{t}}\right)$. This completes the proof of Theorem 14. □

O Omitted Figures

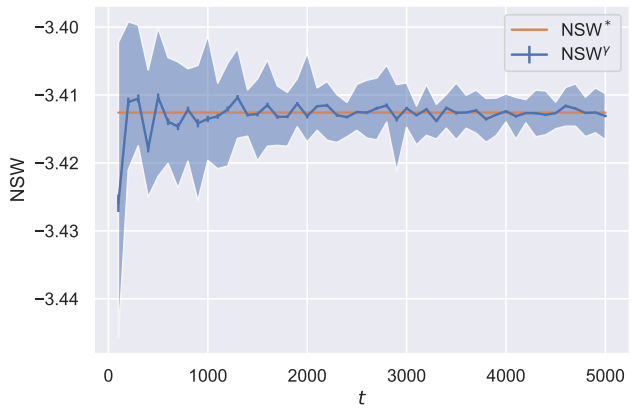

Figure 2: Mean and standard errors of NSW^γ of observed markets of sizes $t = 100, 200, \ldots, 5000$ ($k = 10$ repeats) sampled from the infinite-dimensional market $\mathscr{M}_1$ with linear valuations $v_i(\theta) = a_i\theta + c_i$.

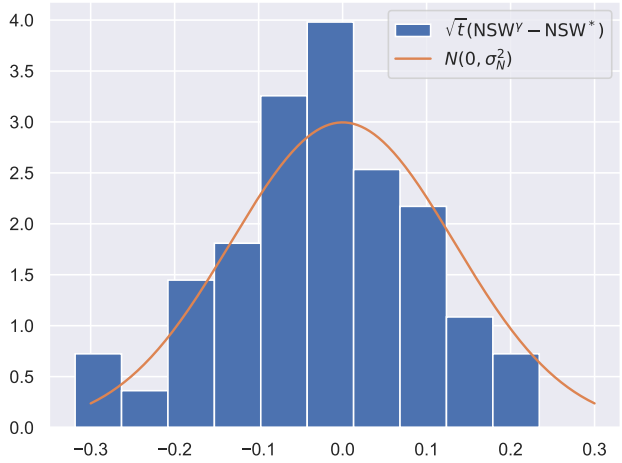

Figure 3: Empirical distribution of $\sqrt{t}(\mathrm{NSW}^{\gamma} - \mathrm{NSW}^{*})$ and $\mathcal{N}(0, \sigma^2_{\mathrm{NSW}})$. Kolmogorov-Smirnov test null hypothesis: $\sqrt{t}(\mathrm{NSW}^{\gamma} - \mathrm{NSW}^{*})$ values are sampled i.i.d. from $\mathcal{N}(0, \sigma^2_{\mathrm{NSW}})$; alternative hypothesis: they are not sampled i.i.d. from $\mathcal{N}(0, \sigma^2_{\mathrm{NSW}})$; test statistic: 0.1256; p-value: 0.3779.

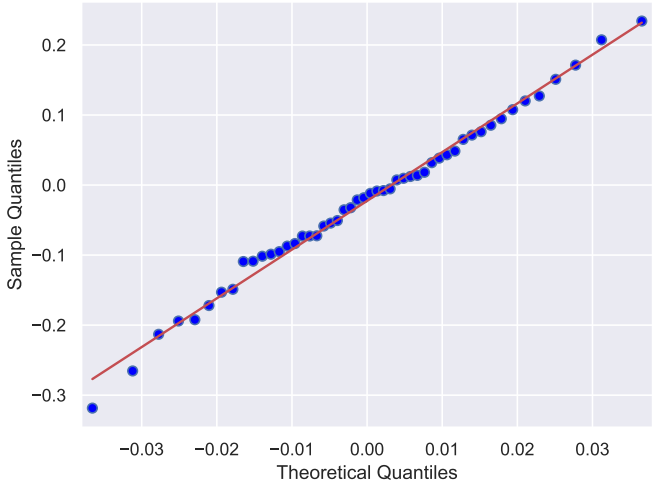

Figure 4: Q-Q Plot of $\sqrt{t}(\mathrm{NSW}^{\gamma} - \mathrm{NSW}^{*})$ values against theoretical quantiles of $\mathcal{N}(0, \sigma^2_{\mathrm{NSW}})$; a (near) straight line indicates that $\sqrt{t}(\mathrm{NSW}^{\gamma} - \mathrm{NSW}^{*})$ values appear to be normal.

Figure 5: Mean and standard errors of NSW^γ of observed markets of sizes $t = 100, 200, \ldots, 5000$ ($k = 10$ repeats) sampled from the infinite-dimensional market $\mathscr{M}_2$ with linear valuations $v_i(\theta) = a_i^\top \theta + c_i$, $a_i \in \mathbb{R}^{10}$.