Statistical Inference and A/B Testing for First-Price Pacing Equilibria

Luofeng Liao [1] **Christian Kroer** [1]

Abstract

We initiate the study of statistical inference and A/B testing for first-price pacing equilibria (FPPE). The FPPE model captures the dynamics resulting from large-scale first-price auction markets where buyers use pacing-based budget management. Such markets arise in the context of internet advertising, where budgets are prevalent.

We propose a statistical framework for the FPPE model, in which a limit FPPE with a continuum of items models the long-run steady-state behavior of the auction platform, and an observable FPPE consisting of a finite number of items provides the data to estimate primitives of the limit FPPE, such as revenue, Nash social welfare (a fair metric of efficiency), and other parameters of interest. We develop central limit theorems and asymptotically valid confidence intervals. Furthermore, we establish the asymptotic local minimax optimality of our estimators. We then show that the theory can be used for conducting statistically valid A/B testing on auction platforms. Numerical simulations verify our central limit theorems, and empirical coverage rates for our confidence intervals agree with our theory.

1. Introduction

A/B testing is a form of randomized controlled experiment, where each sample is assigned to one of two groups: the 'A' group or 'B' group, and a different treatment is applied to each group. For example, say a social media site wants to test whether a new layout will increase user engagement. A subset of users are sampled, and each user in the subset is randomly assigned the current layout (group A), or the new layout (group B). Now, if we ignore network effects, then we can measure whether the new layout increases user engagement by checking whether engagement in group B is higher than in group A with statistical significance. As of 2017, large internet companies such as Google and Microsoft each conduct more than 10,000 A/B tests annually (Kohavi & Thomke, 2017).

However, now consider a setting where advertisers also bid on their ads being shown to users. If we randomly assign users to groups A and B, then we get interference because an advertiser's outcomes in group A affect their behavior when buying ads in group B. This is especially problematic if we are trying to measure something that directly pertains to ads (e.g. revenue changes, or user interest in the ads they are shown). In practice, a popular solution to this issue is to create two separate markets, one for group A and one for group B. Then, each advertiser participates in both markets, with half of their budget assigned to each market, and those budgets are then treated as separate budget constraints. We will refer to this as *budget splitting*. Despite the practical popularity of budget splitting, its statistical properties are not well-understood. A major obstacle to statistical inference with budget splitting is that we can no longer think of the mean user behavior as a sum of independent samples. Instead, we essentially have only two samples: a sample of a market under condition A, and a sample of a market under condition B. Thus, we need to understand when randomized assignment of users (which act as the supply of impressions) into separate markets can be used to make statistical inferences about market outcomes, given the effects of competition.

As stated at the beginning, we study this phenomenon in two important contexts: first-price pacing equilibria (FPPE), and Fisher markets. We will focus the majority of our writing on FPPE, because FPPE model the advertising auction setting faced by large internet companies. All our results carry over to Fisher markets, except results on revenue, which are not meaningful in the standard Fisher market model.

In an FPPE, a set of buyers compete in a set of first-price auctions, and each buyer has a budget. This models how impressions are sold in practice, where first or second-price auction generalizations are used: When a user shows up, an auction is run in order to determine which ads to show, before the page is returned to the user. This auction must run extremely fast. This is typically achieved by having each

[1]IEOR, Columbia University. Correspondence to: Luofeng Liao <ll3530@columbia.edu>.

Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

advertiser specify their target audience, their willingness-to-pay for an impression (or values per click, which are then multiplied by platform-supplied *click-through-rate* estimates), and a budget ahead of time. The control of the bids for individual impressions is then ceded to proxy bidders that are controlled by the ad platform. As a concrete example, to create an ad campaign on Meta Ads Manager, advertisers need to specify the following parameters: (1) the conversion location (how do you want people to reach out to you, via say website, apps, Messenger and so on), (2) optimization and delivery (target your ads to users with specific behavior patterns, such as those who are more likely to view the ad or click the ad link), (3) audience (age, gender, demographics, interests and behaviors), and (4) how much money do you want to spend (budget).

Given the above parameters reported by the advertiser, the (algorithmic) proxy bidder supplied by the platform is then responsible for bidding in individual auctions so as to maximize advertiser utility, while respecting the budget constraint. Two prevalent budget management methods are *throttling* and *pacing*. Throttling tries to enforce budget constraints by adaptively selecting which auctions the advertiser should participate in. Pacing, on the other hand, modifies the advertiser's bids by applying a shading factor, referred to as a *multiplicative pacing multiplier*. Tuning the pacing multiplier changes the spending rate: the larger the pacing multiplier, the more aggressive the bids. The goal of the proxy bidder is to choose this pacing multiplier such that the advertiser exactly exhausts their budget (or alternatively use a multiplier of one in the case where their budget is not exhausted by using unmodified bids). In this paper we focus on pacing-based budget management systems.

In the case where each individual auction is a first-price auction, FPPE capture the outcomes of pacing-based budget-management systems. Conitzer et al. (2022a) introduced the FPPE notion, and showed that FPPE always exists and is unique. Moreover, FPPE enjoys lots of nice properties such as being revenue-maximizing among all budget-feasible pacing strategies, shill-proof (the platform does not benefit from adding fake bids under first-price auction mechanism) and revenue-monotone (revenue weakly increases when adding bidders, items or budget). Crucially for us, FPPE are fully characterized by a quasi-linear Eisenberg-Gale convex program (Conitzer et al., 2022a; Chen et al., 2007).

We remark that all the theory in the paper (CLT, inferential theory, and local asymptotic minimax theory) can be extended to Fisher market with quasilinear utility (Cole et al., 2017) given its equivalence to FPPE.

Given the above motivation, we study the question:

> *Suppose the auction platform operates at FPPE, i.e., at a market equilibrium. How can we quantify the variability in quantities of interest, and use this to perform A/B testing?*

Our contributions are as follows.

A statistical model for first-price pacing equilibrium and A/B testing in auction markets. We leverage the FPPE model of Conitzer et al. (2022a) and the infinite-dimensional Fisher market model of Gao & Kroer (2022) in order to propose a statistical model for first-price auction markets. In this model, we observe market equilibria formed with a finite number of items that are i.i.d. draws from some distribution, and aim to make inferences about several primitives of the limit market, such as revenue, Nash social welfare (a fair metric of efficiency), and other quantities of interest. More importantly, we lay the theoretical foundations for A/B testing in auction markets, which is a difficult statistical problem because buyers interfere with each other through the supply and the budget constraints, the first-price auction allocation, and so on. With the presence of equilibrium effects, traditional statistical approaches which rely on the i.i.d. or the SUTVA (stable unit treatment value assumption, Imbens & Rubin (2015)) assumption fail. The key lever we use to approach this problem is a convex program characterization of the first-price pacing equilibrium, called the Eisenberg-Gale (EG) program. With the EG program, the inference problem reduces to an M-estimation problem (Shapiro et al., 2021; Van der Vaart, 2000) on a constrained non-smooth convex optimization problem.

Convergence and inference results for the limit market. The technical challenges for developing inferential theory for FPPE are two-fold: (1) Nonsmoothness. The sample function in the convex program is non-differentiable on the constraint set almost surely as it involves the max operator (cf. Eq. (P-DualEG)). Such nonsmoothness results from the fact that the allocation produced by first-price auction is highly nonsmooth w.r.t. buyer's pacing strategy. (2) The parameter-on-boundary issue: the optimal population solution might be on the boundary of the constraint set. Asymptotic convergence is established by showing that the EG convex program satisfies a set of regularity conditions from Theorem 3.3 by Shapiro (1989). The hardest condition to verify is stochastic equicontinuity (Cond. 8.c), which we establish by leveraging empirical process theory (Vaart & Wellner, 1996; Kosorok, 2008). For constrained M-estimators, the asymptotic distribution might not be normal, causing challenges for inference. We discover sufficient conditions to ensure normality of the limit distributions. We also establish that the observed market is an optimal estimator of the limit market in the asymptotic local

minimax sense (Van der Vaart, 2000; Le Cam et al., 2000; Duchi & Ruan, 2021). Finally, we provide consistent variance estimators, whose consistency is proved by a uniform law-of-large-numbers over certain function classes.

Statistically-valid inference for A/B testing. Applying our theory, we develop an A/B testing design for item-side randomization that resembles practical A/B testing methodology. In the proposed design treatment and control markets are formed, and buyer's budgets are split proportionally between them, while items are randomly assigned. Then, based on the equilibrium outcomes, we construct estimators and confidence intervals that enable statistical inference. A recipe for applying our theory is presented in Algorithm 1.

Notations. For a measurable space $(\Theta, \mathrm{d}\theta)$, 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 $\mathrm{d}\theta$ for any $p \in [1, \infty]$ (including $p = \infty$). Given $x \in L^\infty$ and $v \in L^1$, we let $\langle v, x\rangle = \int_\Theta v(\theta)x(\theta)\,\mathrm{d}\theta$. We treat all functions that agree on all but a measure-zero set as the same. Denote by e_j the j-th unit vector. For a sequence of random variables $\{X_n\}$, we say $X_n = O_p(1)$ if for any $\epsilon > 0$ 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 $X_n = o_p(1)$ if X_n converges to zero in probability. We say $X_n = O_p(a_n)$ (resp. $o_p(a_n)$) if $X_n/a_n = O_p(1)$ (resp. $o_p(1)$). The subscript i is for indexing buyers and superscript τ is for items. Furthermore, we let $A^\dagger$ be the Moore-Penrose pseudo inverse of a matrix A. Given vectors a and b, let $a \odot b$ be the element-wise product.

In App. B we survey related works on A/B testing in two-sided markets, pacing equilibrum, M-estimation when the parameter is on the boundary, and statistical inference with equilibrium effects.

2. Statistical Model for First-Price Pacing Equilibrium

Following Gao & Kroer (2022); Conitzer et al. (2022a), we consider a single-slot auction market with n buyers and a possibly continuous set of items Θ with an integrating measure $\mathrm{d}\theta$. For example, one could take $\Theta = [0, 1]$ and $\mathrm{d}\theta =$ the Lebesgue measure on $[0, 1]$. Defining first price pacing equilibrium requires the following elements.

- The *budget* b_i of buyer i. Let $b = (b_1, \ldots, b_n)$.
- The *valuation* for buyer i is a function $v_i \in L^1_+$, i.e., buyer i has valuation $v_i(\theta)$ (value per unit supply) of item $\theta \in \Theta$. Let $v : \Theta \to \mathbb{R}^n$, $v(\theta) = [v_1(\theta), \ldots, v_n(\theta)]$. We assume $\bar{v} = \max_i \sup_\theta v_i(\theta) < \infty$.
- The *supplies* of items are given by a function $s \in L^\infty_+$, i.e., item $\theta \in \Theta$ has $s(\theta)$ unit of supply. Without loss of generality, we assume a unit total supply $\int_\Theta s\,\mathrm{d}\theta = 1$. Given $g : \Theta \to \mathbb{R}$, we let $\mathbb{E}[g] = \int g(\theta)s(\theta)\,\mathrm{d}\theta$ and $\mathrm{Var}[g] = \mathbb{E}[g^2] - (\mathbb{E}[g])^2$.

For buyer i, an *allocation* of items $x_i \in L^\infty_+$ gives a utility of $\langle v_i, sx_i\rangle$. Let $x = (x_1, \ldots, x_n)$. The *prices* of items are modeled as $p \in L^1_+$; the price of item $\theta \in \Theta$ is $p(\theta)$.

2.1. Definition and Interpretation of limit FPPE

Central to the notion of an FPPE is the *pacing multiplier*, which is a scalar $\beta_i \in [0, 1]$ such that buyer i bids their "paced" value $\beta_i \cdot v_i(\theta)$ on a given item θ.

Definition 1 (Limit FPPE)**.** *Given (b, v, s), a limit FPPE (denoted FPPE(b, v, s)) is a tuple $(\beta, p, x_1, \ldots, x_n) \in [0, 1]^n \times L^1_+ \times (L^\infty_+)^n$ such that*

1.1 (First-price) For all item $\theta \in \Theta$, $p(\theta) = \max_i \beta_i v_i(\theta)$. Moreover, $x_i(\theta) > 0$ implies $\beta_i v_i(\theta) = \max_k \beta_k v_k(\theta)$ for all i and θ.

1.2 (Supply and budget feasible) For all i, $\int x_i(\theta)p(\theta)s(\theta)\,\mathrm{d}\theta \leq b_i$. For all θ, $\sum_{i=1}^n x_i(\theta) \leq 1$.

1.3 (Market clearing) For all i, $\int x_i(\theta)p(\theta)s(\theta)\,\mathrm{d}\theta < b_i$ implies $\beta_i = 1$. For all θ, $p(\theta) > 0$ implies $\sum_{i=1}^n x_i(\theta) = 1$.

By Conitzer et al. (2022a), the limit FPPE exists and is unique. Next we unpack Def. 1. The $\beta_i \in [0, 1]$ constraint is natural since a rational buyer should not bid more than their value. Cond. 1.1 captures the fact that FPPE is a model for first-price auctions, where pacing is used to manage budgets. The scalar β_i controls the expenditure of buyer i, and the constraint ensures that the price is equal to the highest bid, and that only buyers tied for highest are allocated a non-zero amount. Cond. 1.2 ensures that the budget and supply constraints are satisfied. Cond. 1.3 ensures that the solution satisfies *no unnecessary pacing*, meaning that we should only scale down a buyer's bids in case their budget constraint is binding. Secondly, it ensures that if a good is demanded by any buyers, then it must be fully allocated.

Fact 1 (Buyer's satisfaction, Theorem 2 from Conitzer et al. (2022a))**.** *Let $(\beta, p, x) \in$ FPPE(b, v, s). For all i, it holds $x_i \in \arg\max_x \{u_i(x) : 0 \leq x \leq 1, \langle p, sx\rangle \leq b_i\}$ where the utility of a buyer is*

$$u_i(x_i) \equiv \langle v_i, sx_i\rangle + (b_i - \langle p, sx_i\rangle)\ ,$$

where the first term is utility from the allocated items, the second term is the leftover budget.

This means FPPE is a competitive equilibrium. In the first-price auction context, each buyer's allocated items maximize their utility (item utility + leftover budget) among all budget-feasible allocations, given the price.

From now on we use $(.)^*$ to denote limit FPPE quantities. Given an FPPE (β^*, p^*, x^*), we define by

$$\delta_i^* \equiv b_i - \langle p^*, sx_i^* \rangle \ , \quad \bar{\mu}_i^* \equiv \langle v_i, sx_i^* \rangle \ ,$$
$$u_i^* \equiv \bar{\mu}_i^* + \delta_i^* = u_i(x_i^*) \ . \tag{1}$$

the *leftover budget*, the *item utility* and the *total utility* of buyer i. Let $\delta^*, \bar{\mu}^*, u^*$ be the vectors that collect these quantities for all buyers. It is well-known that (p^*, u^*, β^*) are unique in equilibrium, but $(x^*, \delta^*, \bar{\mu}^*)$ might not be unique. Later we will see that for statistical inference we need conditions to ensure uniqueness of x^*. The following equations about limit FPPE (Gao & Kroer, 2022) are important.

$$u_i^* = b_i/\beta_i^* \ , \quad (\beta_i^* - 1)\delta_i^* = 0 \ . \tag{2}$$

We want to estimate the following quantities in the limit FPPE. **(1) Revenue.** The revenue in the limit FPPE is $\text{REV}^* \equiv \int p^*(\theta)s(\theta)\,\mathrm{d}\theta$. It measures the profitability of the auction platform. When the platform operates at the limit FPPE, REV^* is the maximum revenue the platform could extract from the buyers over the space of budget-feasible pacing strategies Conitzer et al. (2022a). **(2) Nash social welfare (NSW).** The (logarithm of) NSW is defined as $\text{NSW}^* \equiv \sum_{i=1}^n b_i \log u_i^*$. The NSW at equilibrium measures total utility of the buyers and, when used as a summary metric of the efficiency of the auction platform, is able to promote fairness better than the utilitarian social welfare, that is, the sum of buyer utilities (Bertsimas et al., 2012; Caragiannis et al., 2019). **(3) Individual utilities** at limit FPPE, u_i^*. **(4) Pacing multipliers** β_i^*. Pacing multiplier has a two-fold interpretation. First, through the equation $\beta_i^* = b_i/u_i^*$, it is the ratio of budget and utility. Second, β^* is the pacing policy employed by the buyers in first-price auctions, a quantity of natural interest.

As we will see next, counterparts of these quantities in the observed market are good estimators of the limit quantities.

2.2. The Observed FPPE

Let $\gamma = (\theta^1, \ldots, \theta^t)$ be a sequence of observed items drawn from the distribution s in an i.i.d. manner. Assume each item has the same supply of $\sigma \in \mathbb{R}_+$ units. Most of the time we take $\sigma = \frac{1}{t}$ to ensure total supply agrees with the limit market.

Definition 2 (Observed FPPE, informal)**.** *Given (b, v, σ, γ), the observed FPPE $\widehat{\text{FPPE}}(b, v, \sigma, \gamma)$ contains tuples $(\beta, p, x_1, \ldots, x_n) \in [0,1]^n \times \mathbb{R}_+^t \times ([0,1]^t)^n$ such that the conditions in Def. 1 hold with the following modifications: limit item space $\Theta \to$ observed item set γ and limit supply distribution $s(\cdot) \to$ weighted observed item distribution $\sigma \sum_{\tau=1}^t \delta_{\theta^\tau}(\cdot)$ where $\delta_\theta(\cdot)$ is a point mass on θ.*

A formal definition and further properties of FPPE can be found in App. C. Here $x_i^\tau \in [0,1]$ is the fractional allocation of item θ^τ to buyer i. The mechanism of forming the observed FPPE is exactly the same as the limit FPPE in Def. 1, except now the price p, the supply s and the allocation x_i reduce to vectors in $\mathbb{R}^t$ as opposed to functions.

To emphasize dependence on the item sequence γ, we use $(.)^\gamma$ to denote equilibrium quantities in $\widehat{\text{FPPE}}(b, v, t^{-1}, \gamma)$. We let $(\beta^\gamma, p^\gamma, x^\gamma)$ be an observed FPPE with $x^\gamma = (x_1^\gamma, \ldots, x_n^\gamma)$. The leftover budget $\delta_i^\gamma \equiv b_i - \sigma\langle p^\gamma, x_i^\gamma \rangle$, item utility $\bar{\mu}_i^\gamma \equiv \sigma\langle v_i, x_i^\gamma \rangle$ and total utility $u_i^\gamma \equiv \delta_i^\gamma + \bar{\mu}_i^\gamma$ are defined similarly. Let $\delta^\gamma, \bar{\mu}^\gamma, u^\gamma$ be the vectors that collect these quantities for all buyers. The observed revenue is $\text{REV}^\gamma \equiv \sigma \sum_{\tau=1}^t p^\gamma(\theta^\tau)$, and NSW is $\text{NSW}^\gamma \equiv \sum_{i=1}^n b_i \log u_i^\gamma$.

Having observed an FPPE with a finite number of items, our goal is to estimate the quantities of interest in the limit FPPE.

2.3. Convex Programs for FPPE

Before we present the main statistical theories for FPPE, we review convex program characterization of FPPE, which are at the core of the paper. These convex characterization results reduce the FPPE inference problem to the one on a constrained nonsmooth convex program.

The limit FPPE pacing multiplier β can be recovered through the population dual Eisenberg-Gale (EG) program

$$\min_B H(\beta) \equiv \mathbb{E}[F(\theta, \beta)] \tag{P-DualEG}$$

where $B \equiv (0,1]^n$, $F \equiv f + \Psi$,

$$f(\theta, \beta) \equiv \max_{i \in [n]} \beta_i v_i(\theta) \ , \ \Psi(\beta) \equiv -\sum_{i=1}^n b_i \log \beta_i \ .$$

It is known that the FPPE β^* is the unique solution to Eq. (P-DualEG), and any FPPE (x^*, u^*, δ^*) belongs to the set of optimal solutions to the population primal EG program (to be presented in App. C).

The observed FPPE has a similar convex program characterization. By taking $\sigma = \frac{1}{t}$ and replacing s with the empirical measure $\frac{1}{t}\sum_{\tau=1}^t \delta_{\theta^\tau}$ in Eq. (P-DualEG), it can be shown that the observed FPPE pacing multiplier solves the sample dual EG program

$$\min_{\beta \in B} H_t(\beta) \equiv \frac{1}{t}\sum_{\tau=1}^t F(\theta^\tau, \beta) \ . \tag{S-DualEG}$$

To develop inferential theory for FPPE, we study the concentration of the *dual* EG programs. The study of the convergence $\widehat{\mathsf{FPPE}}(b, v, t^{-1}, \gamma)$ " $\Longrightarrow$ " $\mathsf{FPPE}(b, v, s)$ reduces to the one of

$$\min_{\beta \in B} H_t(\beta) \quad \text{“}\Longrightarrow\text{”} \quad \min_{\beta \in B} H(\beta) \ .$$

As mentioned previously, the difficulty of analyzing the above convex programs lies in the nonsmoothness of the sample function F and that the population optimum could lie on the boundary of B. We define the set of constraints that are active/inactive at β^* by

$$I \equiv \{i : \beta_i^* = 1\} \ , \quad I^c \equiv \{i : \beta_i^* < 1\} \ . \tag{3}$$

3. Statistical Inference

For statistical inference we need the limit market to behave smoothly around the the optimal pacing multipliers β^*. To that end, we make the following assumption.

Assumption 1 (SMO)**.** *Assume the map $\beta \mapsto \mathbb{E}_s[\max_i \beta_i v_i(\theta)]$ is C^2 in a neighborhood of the limit FPPE pacing multiplier β^*.*

For a given $\beta \in \mathbb{R}_+^n$, the quantity $\max_i \beta_i v_i(\theta)$ is the price of item θ in the first-price auction. The assumption requires that the revenue, $\mathbb{E}_s[\max_i \beta_i v_i(\theta)]$, when viewed as a function of β, changes smoothly around β^*. The assumption implies that H, defined in Eq. (P-DualEG), is also C^2 at β^*.

Assumption 1 implies a number of nice regularity conditions. One is that the set of items that are tied at the limit FPPE is s-measure zero. The set of tied items is $\Theta_{\text{tie}} \equiv \{\theta \in \Theta : \beta_i^* v_i(\theta) = \beta_k^* v_k(\theta) \text{ for some } i \neq k\}$.

Lemma 1. *Under SMO, the set Θ_{tie} is s-zero measure (up to a measure-zero set), and the equilibrium allocation x^*, the leftover budget δ^* and the item utility $\bar{\mu}^*$ are all unique. Moreover, there is a neighborhood N_{diff} of β^* such that each pacing strategy in this neighborhood results in no tie. Proof in App. C.*

SMO is a joint assumption on value functions v and the supply function s. Lower level conditions on v and s that imply SMO were derived by Liao et al. (2022b). For example, if the distribution of the values v_i is smooth, then SMO holds. If we impose functional structure on v, such as $\Theta = [0, 1]$ and $v_i = a_i\theta + b_i$ with $\mathbb{E}[v_i] = 1$, b_i's distinct, and s is uniform, then SMO also holds. If the gap between the highest and the second-highest bid is large for most items in the limit FPPE, SMO also holds. For a precise statement, we refer readers to Theorem 7 from Liao et al. (2022b).

In order to state our main CLT results, we define

$$\mu^*(\theta) \equiv x^*(\theta) \odot v(\theta) \ , \quad \mathcal{H} \equiv \nabla^2 H(\beta^*) \ . \tag{4}$$

Under SMO, $\mu^*(\cdot)$ is unique and well-defined. Clearly $\bar{\mu}^* = \mathbb{E}[\mu^*(\theta)]$.

In the unconstrained case, classical M-estimation theory says that, under regularity conditions, an M-estimator is asymptotically normal with covariance matrix $\mathcal{H}^{-1}\mathrm{Var}(\text{gradient})\mathcal{H}^{-1}$ (Van der Vaart, 2000, Chap. 5). However, in the case of FPPE which is characterized by a constrained convex problem, the Hessian matrix needs to be adjusted to take into account the geometry of the constraint set $B = (0, 1]^n$ at the optimum β^*. We let $\mathcal{P} \equiv \mathrm{diag}(\mathbb{1}(i \in I^c))$ be an "indicator matrix" of buyers whose $\beta_i^* < 1$, and define the projected Hessian

$$\mathcal{H}_B \equiv \mathcal{P}\mathcal{H}\mathcal{P} \ . \tag{5}$$

It will be shown that the asymptotic variance of β^γ is $\mathcal{H}_B^\dagger \mathrm{Var}(\text{gradient})\mathcal{H}_B^\dagger$.

Assumption 2 (SCS)**.** *Strict complementary slackness holds: $\beta_i^* = 1$ implies $\delta_i^* > 0$.*

SCS can be viewed as a non-degeneracy condition from a convex programming perspective, since δ_i corresponds to a Lagrange multiplier on $\beta_i \leq 1$. From a market perpective, SCS requires that if a buyer's bids are not paced ($\beta_i^* = 1$), then the leftover budget δ_i^* must be strictly positive. This can again be seen as a market-based non-degeneracy condition: if $\delta_i^* = 0$ then the budget constraint of buyer i is binding, yet $\beta_i^* = 1$ would imply that they have no use for additional budget. If SCS fails, one could slightly increase the budgets of buyers for which SCS fails, i.e., those who do not pace yet have exactly zero leftover budget, and obtain a market instance with the same equilibrium, but where SCS holds.

From a technical viewpoint, SCS is a stronger form of the first-order optimality condition. Note $\nabla H(\beta^*) = -\delta^*$ (cf. App. C). The usual first-order optimality condition is

$$-\nabla H(\beta^*) \in \mathcal{N}_B(\beta^*) \ , \tag{6}$$

where $\mathcal{N}_B(\beta) = \prod_{i=1}^n J_i(\beta)$ is the normal cone with $J_i(\beta) = [0, \infty)$ if $\beta_i = 1$ and $J_i(\beta) = \{0\}$ if $\beta_i < 1$ for $\beta \in \mathbb{R}_{++}^n$. Then Eq. (6) translates to the condition that $\beta_i^* = 1$ implies $\delta_i^* \geq 0$. On the other hand, when written in the form that resembles optimality condition, SCS is equivalent to

$$-\nabla H(\beta^*) \in \mathrm{relint}(\mathcal{N}_B(\beta^*)) \ .^{1}$$

Given that $\mathrm{relint}(\mathcal{N}_B(\beta^*)) \subset \mathcal{N}_B(\beta^*)$, SCS is obviously a stronger form of first-order condition. The SCS condition is commonly seen in the study of statistical properties of constrained M-estimators (Duchi & Ruan (2021, Assumption

[1] The relative interior of a set is $\mathrm{relint}(S) \equiv \{x \in S : \text{there exists } \epsilon > 0 \text{ such that } N_\epsilon(x) \cap \mathrm{aff}(S) \subseteq S\}$ where $\mathrm{aff}(S)$ is the affine hull of S, and $N_\epsilon(x)$ is a ball of radius ϵ centered on x.

B) and Shapiro (1989)). In the proof of Thm. 1, SCS forces the critical cone to reduce to a hyperplane and thus ensures asymptotic normality of the estimates. Without SCS, the asymptotic distribution of β^γ could be non-normal.

3.1. Central Limit Theorems

We now show that the observed pacing multipliers β^γ and the observed revenue REV^γ converge to the limit market quantities in probability, and satisfy central limit theorems. Define the *influence functions*

$$\begin{aligned} D_\beta(\theta) &\equiv -(\mathcal{H}_B)^\dagger(\mu^*(\theta) - \bar{\mu}^*), \\ D_{\mathrm{REV}}(\theta) &\equiv p^*(\theta) - \mathrm{REV}^* + (\bar{\mu}^*)^\top D_\beta(\theta). \end{aligned} \tag{7}$$

Recall μ^* is defined in Eq. (4), $\mathcal{H}_B$ in Eq. (5). Clearly $\mathbb{E}[D_\beta] = 0$ and $\mathbb{E}[D_{\mathrm{REV}}] = 0$.

Theorem 1. *It holds that $\beta^\gamma \xrightarrow{p} \beta^*$ and $\mathrm{REV}^\gamma \xrightarrow{p} \mathrm{REV}^*$. Furthermore, if SCS and SMO hold, then*

$$\sqrt{t}(\beta^\gamma - \beta^*) = t^{-1/2}\sum_{\tau=1}^{t} D_\beta(\theta^\tau) + o_p(1)\ ,$$

$$\sqrt{t}(\mathrm{REV}^\gamma - \mathrm{REV}^*) = t^{-1/2}\sum_{\tau=1}^{t} D_{\mathrm{REV}}(\theta^\tau) + o_p(1)\ .$$

Consequently, $\sqrt{t}(\beta^\gamma - \beta^*)$ and $\sqrt{t}(\mathrm{REV}^\gamma - \mathrm{REV}^*)$ are asymptotically normal with means zero and variances $\Sigma_\beta \equiv \mathbb{E}[D_\beta^{\otimes 2}] = (\mathcal{H}_B)^\dagger \mathrm{Var}(\mu^*)(\mathcal{H}_B)^\dagger$ and $\sigma^2_{\mathrm{REV}} \equiv \mathbb{E}[D_{\mathrm{REV}}(\theta)^2]$. Proof in App. F.1.

The functions D_β and D_{REV} are called the influence functions of the estimates β^γ and REV^γ because they measure the change in the estimates caused by adding a new item to the market (asymptotically).

Thm. 1 implies fast convergence rate of β_i^γ for i whose constraint is tight in the limit market. To see this, we suppose wlog. that $I = [k]$, i.e. the first k buyers are the ones with $\beta_i = 1$. Then the pseudo-inverse of projected Hessian $(\mathcal{H}_B)^\dagger = \mathrm{diag}(0_{k\times k}, (\mathcal{H}_{I^cI^c})^{-1})$ where $\mathcal{H}_{I^cI^c}$ is the lower right $(n-k)\times(n-k)$ block of $\mathcal{H}$. Consequently, entries of Σ_u (resp. Σ_β) are zeros except those on the lower right $(n-k)\times(n-k)$ blocks $(\Sigma_u)_{I^cI^c}$ (resp. $(\Sigma_\beta)_{I^cI^c}$). This result shows that the constraint set B "improves" the covariance by zeroing out the entries corresponding to the active constraints I. Consequently, $\sqrt{t}(\beta_i^\gamma - \beta_i^*)$ and $\sqrt{t}(u_i^\gamma - u_i^*)$ are of order $o_p(1)$ for $i \in I$, and thus converging faster than the usual $O_p(1)$ rate. The fast rate phenomenon is empirically investigated in App. G.1.

By SCS we have $I = \{i : \delta_i^* > 0\}$, i.e., I is the set of buyers with positive leftover budgets, and $I^c = \{i : \delta_i^* = 0\}$, i.e., I^c is the set of buyers who exhaust their budgets. [2]

[2] Without SCS, it only holds $I^c \subset \{i : \delta_i^* = 0\}$ and $\{i : \delta_i^* > 0\} \subset I$ by complementary slackness Eq. (2).

In the context of first-price auctions, the fast rate $o_p(t^{-1/2})$ implies that the platform can identify buyers that are unpaced in the limit FPPE even when the market size is small.

The proof of Thm. 1 proceeds by showing that FPPE satisfy a set of regularity conditions that are sufficient for asymptotic normality (Shapiro, 1989, Theorem 3.3); the conditions are stated in Lemma 8 in the appendix. Maybe the hardest condition to verify is the so called stochastic equicontinuity condition (Cond. 8.c), which we establish with tools from the empirical process literature. In particular, we show that the function class whose functions map an item to the first-price auction allocation of items, is VC-subgraph, which implies stochastic equicontinuity. SCS is used to ensure normality of the limit distribution.

Finally, we remark that the CLT result for revenue holds true even if $I = \emptyset$. If $\beta_i^\gamma < 1$ for all i, then all buyers' budgets are exhausted in the observed FPPE, and so we have $\mathrm{REV}^\gamma = \mathrm{REV}^* = \sum_{i=1}^n b_i$ if $I = \emptyset$. By the convergence $\beta^\gamma \xrightarrow{p} \beta^*$, we know that $\mathrm{REV}^\gamma = \mathrm{REV}^*$ with high probability for all large t if $I = \emptyset$. In that case, it must be that the asymptotic variance of revenue equals zero. Our result covers this case because one can show $\sigma^2_{\mathrm{REV}} = 0$ using Euler's identify for homogenous functions and that $\mathcal{H}\beta^* = \mu^*$ if $I = \emptyset$; see Lemma 6.

By applying the delta method based on Thm. 1, we can derive a CLT for individual utilities u^γ, leftover budget δ^γ and Nash social welfare NSW^γ since they are smooth functions of β^γ; see Lemma 4.

Corollary 1. *Under the same conditions as Thm. 1, $\sqrt{t}(u^\gamma - u^*)$, $\sqrt{t}(\delta^\gamma - \delta^*)$ and $\sqrt{t}(\mathrm{NSW}^\gamma - \mathrm{NSW}^*)$ are asymptotically normal with means zero and (co)variances defined in App. F.2.*

3.2. Asymptotic Local Minimax Optimality

Given the asymptotic normality of observed FPPE, it is desirable to understand the best possible statistical procedure for estimating the limit FPPE. One way to discuss the optimality is to measure the difficulty of estimating the limit FPPE when the supply distribution varies over small neighborhoods of the true supply s, asymptotically. When an estimator achieves the best worst-case risk over these small neighborhoods, we say it is asymptotically locally minimax optimal. For general references, see Vaart & Wellner (1996); Le Cam et al. (2000). More recently Duchi & Ruan (2021, Sec. 3.2) develop asymptotic local minimax theory for constrained convex optimization, and we rely on their results.

Given the central limit results for β, u, NSW and REV, we will show that the observed FPPE estimates are optimal in a asymptotic local minimax sense. To make this precise we introduce a few more notations to parametrize neigh-

borhoods of the supply s. Let $g \in G_d = \{g : \Theta \to \mathbb{R}^d : \mathbb{E}[g] = 0, \mathbb{E}[\|g\|^2] < \infty\}$ be a direction along which we wish to perturb the supply s. Given a vector $\alpha \in \mathbb{R}^d$ signifying the magnitude of perturbation, we want to scale the original supply of item θ by $\exp(\alpha^\top g(\theta))$ and then obtain a perturbed supply distribution by appropriate normalization. To do this we define the perturbed supply by [3]

$$s_{\alpha,g}(\theta) \equiv C^{-1}[1 + \alpha^\top g(\theta)]s(\theta) \tag{8}$$

with a normalizing constant $C = 1 + \int \alpha^\top g(\theta)s(\theta)\,\mathrm{d}\theta$. As $\alpha \to 0$, the perturbed supply $s_{\alpha,g}$ effectively approximates $s_{\alpha,g}(\theta) \propto \exp(\alpha^\top g(\theta))s(\theta)$.

Asymptotic local minimax optimality for β. We first focus on estimation of pacing multipliers. For a given perturbation (α, g), we let $\beta^*_{\alpha,g}$, $p^*_{\alpha,g}$ and $\mathrm{REV}^*_{\alpha,g}$ be the limit FPPE pacing multiplier, price and revenue under supply distribution $s_{\alpha,g}$. Clearly $\beta^* = \beta^*_{0,g}$ for any g and similarly for $p^*_{\alpha,g}$ and $\mathrm{REV}_{\alpha,g}$. Let $L : \mathbb{R}^n \to \mathbb{R}$ be any symmetric quasi-convex loss. [4] In asymptotic local minimax theory we are interested in the local asymptotic risk: given a sequence of estimators $\{\hat\beta_t : \Theta^t \to \mathbb{R}^n\}_t$, $\mathrm{LAR}_\beta(\{\hat\beta_t\}_t) \equiv$

$$\sup_{g\in G_d, d\in\mathbb{N}} \lim_{c\to\infty} \liminf_{t\to\infty} \sup_{\|\alpha\|_2 \le \frac{c}{\sqrt{t}}} \mathbb{E}_{s^{\otimes t}_{\alpha,g}}[L(\sqrt{t}(\hat\beta_t - \beta^*_{\alpha,g}))] \,.$$

If we ignore the limits and consider a fixed t, then LAR_β roughly measures the worst-case risk for the estimators $\{\hat\beta_t\}$. Note that α is a d-dimensional vector, and thus the shrinking norm-balls depend on the choice of d, and the expectation is taken w.r.t. the t-fold product of the perturbed supply. As an immediate application of Theorem 1 from Duchi & Ruan (2021), it holds that

$$\inf_{\{\hat\beta_t\}_t} \mathrm{LAR}_\beta(\{\hat\beta_t\}_t) \ge \mathbb{E}[L(\mathcal{N}(0, (\mathcal{H}_B)^\dagger \mathrm{Var}(\mu^*)(\mathcal{H}_B)^\dagger))] \,.$$

where the expectation is taken w.r.t. a normal specified above. Moreover, the lower bound is achieved by the observed FPPE pacing multipliers β^γ according to the CLT result in Thm. 1.

Asymptotic local minimax optimality for revenue estimation. More interesting is the asymptotic minimax result of revenue estimation. Given a symmetric quasi-convex loss $L : \mathbb{R} \to \mathbb{R}$, we define the local asymptotic risk for any procedure $\{\hat r_t : \Theta^t \to \mathbb{R}\}$ that aims to estimate the revenue: $\mathrm{LAR}_{\mathrm{REV}}(\{\hat r_t\}) \equiv$

$$\sup_{g\in G_d, d\in\mathbb{N}} \lim_{c\to\infty} \liminf_{t\to\infty} \sup_{\|\alpha\|_2 \le \frac{c}{\sqrt{t}}} \mathbb{E}_{s^{\otimes t}_{\alpha,g}}[L(\sqrt{t}(\hat r_t - \mathrm{REV}^*_{\alpha,g}))] \,.$$

[3] In Duchi & Ruan (2021) they allow more general classes of perturbations, we specialize their results for our purposes.

[4] A function is quasi-convex if its sublevel sets are convex.

Theorem 2 (Asymptotic local minimaxity for revenue)**.** *If SMO and SCS hold, then*

$$\inf_{\{\hat r_t\}} \mathrm{LAR}_{\mathrm{REV}}(\{\hat r_t\}) \ge \mathbb{E}[L(\mathcal{N}(0, \sigma^2_{\mathrm{REV}}))] \,.$$

Proof in App. F.3. In the proof we calculate the derivative of revenue w.r.t. α, which in turns uses a perturbation result for constrained convex program from Duchi & Ruan (2021); Shapiro (1989). Again, the lower bound is achieved by the observed FPPE revenue REV^γ according to the CLT result in Thm. 1. Similar optimality statements can be made for u and NSW by finding the corresponding derivative expressions.

3.3. Inference

In order to perform inference, we need to construct estimators for the influence functions Eq. (7). We show how each component can be estimated by the observed FPPE.

Given a sequence of smoothing parameters $\varepsilon_t = o(1)$, we estimate $\mathcal{P}$ by $\hat\mathcal{P} \equiv \mathrm{diag}(\mathbb{1}(\beta^\gamma_i < 1 - \varepsilon_t))$. For the same sequence ε_t, we introduce a numerical difference estimator $\hat\mathcal{H}$ for the Hessian matrix $\mathcal{H}$, whose (i,j)-th entry is $\hat\mathcal{H}_{ij} \equiv [H_t(\beta^\gamma_{++}) - H_t(\beta^\gamma_{+-}) - H_t(\beta^\gamma_{-+}) + H_t(\beta^\gamma_{--})]/4\varepsilon_t^2$ with $\beta^\gamma_{\pm\pm} \equiv \beta^\gamma \pm e_i\varepsilon_t \pm e_j\varepsilon_t$, and H_t is defined in Eq. (S-DualEG). Then $\hat\mathcal{H}_B = \hat\mathcal{P}\hat\mathcal{H}\hat\mathcal{P}$ is a natural estimator of $\mathcal{H}_B$. Also, x^*, p^* will be estimated by x^γ, p^γ. Let $x^\gamma_i(\theta^\tau) \equiv (x^\gamma_i)^\tau \in [0,1]$ be the proportion of θ^τ allocated to buyer i and $p^\gamma(\theta^\tau) \equiv (p^\gamma)^\tau$ be the price.

Mirroring the definitions in Eq. (7), we define the following influence function estimators

$$\hat D^\tau_\beta \equiv -(\hat\mathcal{H}_B)^\dagger(x^\gamma(\theta^\tau) \odot v(\theta^\tau) - \bar\mu^\gamma) \,,$$
$$\hat D^\tau_{\mathrm{REV}} \equiv p^\gamma(\theta^\tau) - \mathrm{REV}^\gamma + (\bar\mu^\gamma)^\top \hat D^\tau_\beta \,.$$

Given that the asymptotic variance of β^γ (resp. REV^γ) is $\mathbb{E}[D^{\otimes 2}_\beta]$ (resp. $\mathbb{E}[D^2_{\mathrm{REV}}]$), plug-in estimators of the (co)variance are naturally

$$\hat\Sigma_\beta \equiv \frac{1}{t}\sum_{\tau=1}^t \hat D^\tau_\beta (\hat D^\tau_\beta)^\top \,, \quad \hat\sigma^2_{\mathrm{REV}} \equiv \frac{1}{t}\sum_{\tau=1}^t (\hat D^\tau_{\mathrm{REV}})^2 \,. \tag{9}$$

For any $\alpha \in (0,1)$, the $(1-\alpha)$-confidence region/interval for β^* and REV^* are

$$\mathrm{CR}_\beta \equiv \beta^\gamma + (\chi_{n,\alpha}/\sqrt{t})\hat\Sigma^{1/2}_\beta \mathbb{B} \,, \tag{10}$$
$$\mathrm{CI}_{\mathrm{REV}} \equiv [\mathrm{REV}^\gamma \pm z_{\alpha/2}\hat\sigma_{\mathrm{REV}}/\sqrt{t}] \,. \tag{11}$$

where $\chi_{n,\alpha}$ is the $(1-\alpha)$-th quantile of a chi-square distribution with degree n, $\mathbb{B}$ is the unit ball in $\mathbb{R}^n$, and $z_{\alpha/2}$ is the $(1-\frac{\alpha}{2})$-th quantile of a standard normal distribution. The coverage rate of $\mathrm{CI}_{\mathrm{REV}}$ is empirically verified in App. G.3.

Theorem 3. *Under the conditions of Thm. 1, let $\varepsilon_t\sqrt{t} \to \infty$ and $\varepsilon_t \downarrow 0$. Then $\hat{\Sigma}_\beta \xrightarrow{p} \Sigma_\beta$ and $\hat{\sigma}^2_{\text{REV}} \xrightarrow{p} \sigma^2_{\text{REV}}$. Consequently, for any $\alpha \in (0,1)$, $\mathbb{P}(\beta^* \in \text{CR}_\beta) \to 1-\alpha$ and $\mathbb{P}(\text{REV}^* \in \text{CI}_{\text{REV}}) \to 1-\alpha$. Proof in App. F.4.*

The theorem suggests choosing smoothing parameter $\varepsilon_t = t^{-d}$ for $0 < d < \frac{1}{2}$; see App. G.2 for a numerical study on how d affects Hessian estimation. Variance estimators for u, δ and NSW can be constructed similarly.

4. A/B Testing for First-Price Auction Platforms

Consider an auction market with n buyers with a continuum of items Θ with supply function s. To model treatment application we introduce the *potential value functions*

$$v(0) \equiv (v_1(0,\cdot),\dots,v_n(0,\cdot)),\ v(1) \equiv (v_1(1,\cdot),\dots,v_n(1,\cdot)).$$

If item θ is exposed to treatment $w \in \{0,1\}$, then its value to buyer i will be $v_i(w,\theta)$.

Suppose we are interested in estimating the change in the auction market when treatment 1 is deployed to the entire item set Θ. In this section we describe how to do this using A/B testing, specifically for estimating the treatment effect on revenue. We discuss other quantities like Nash social welfare in App. D. Formally, we wish to look at the difference in revenues between the markets

$$\mathsf{FPPE}(b, v(0), s) \text{ and } \mathsf{FPPE}(b, v(1), s),$$

where $\mathsf{FPPE}(b, v(0), s)$ is the market with treatment 1, and $\mathsf{FPPE}(b, v(1), s)$ is the one with treatment 0. The treatment effects on revenue is defined as

$$\tau_{\text{REV}} \equiv \text{REV}^*(1) - \text{REV}^*(0)\ ,$$

where $\text{REV}^*(w)$ is revenue in the equilibrium $\mathsf{FPPE}(b, v(w), s)$.

We will refer to the experiment design as *budget splitting with item randomization*. **Step 1. Budget splitting.** We replicate buyers by splitting their budgets and form two markets with the same set of buyers. For each buyer i we allocate πb_i of their budget to the market with treatment $w = 1$, and the remaining budget, $(1-\pi)b_i$, to the market with treatment $w = 0$. **Step 2. Item randomization.** Let $(\theta^1, \theta^2, \dots)$ be i.i.d. draws from the supply distribution s. For each sampled item, it is applied treatment 1 with probability π and treatment 0 with probability $1-\pi$. The total A/B testing horizon is t. When the end of horizon is reached, two observed FPPEs are formed. Assume each item has a supply of π/t_1 in the 1-treated market and $(1-\pi)/t_0$ in the 0-treated market. The $1/t_1$ is the scaling required for our CLTs and the π factor ensures the budget-supply ratio agrees with the limit market; see Lemma 5 regarding scale-invariance of FPPE.

Let t_0 be the number of 0-treated items, and t_1 be the number of 1-treated items. Conditional on the total number of items $t = t_1 + t_0$, the random variable t_1 is a binomial random variable with mean πt. Let $\gamma(0) = (\theta^{1,1},\dots,\theta^{1,t_1})$ be the set of 0-treated items, and similarly $\gamma(1) = (\theta^{0,1},\dots,\theta^{0,t_0})$. The total item set $\gamma = \gamma(0) \cup \gamma(1)$. Compactly, the observables in the described A/B testing experiment are

$$\widehat{\mathsf{FPPE}}\big(\pi b, v(1), \tfrac{\pi}{t_1}, \gamma(1)\big),\ \widehat{\mathsf{FPPE}}\big((1-\pi)b, v(0), \tfrac{1-\pi}{t_0}, \gamma(0)\big),$$

both defined in Def. 2. Let $\text{REV}^\gamma(w)$ denote the observed revenue in the w-treated market. The estimator of revenue treatment effect is

$$\hat{\tau}_{\text{REV}} \equiv \text{REV}^\gamma(1) - \text{REV}^\gamma(0).$$

For fixed (b, s), the variance σ^2_{REV} in Thm. 1 is a functional of the value functions. We will use $\sigma^2_{\text{REV}}(w)$ to represent the revenue variance in the equilibrium $\mathsf{FPPE}(b, v(w), s)$. Each variance can be estimated using Eq. (9).

Theorem 4 (Revenue treatment effects CLT)**.** *Suppose SMO and SCS hold in the limit markets $\mathsf{FPPE}(b, v(1), s)$ and $\mathsf{FPPE}(b, v(0), s)$. Then $\sqrt{t}(\hat{\tau}_{\text{REV}} - \tau_{\text{REV}}) \xrightarrow{d} \mathcal{N}\big(0, \frac{\sigma^2_{\text{REV}}(1)}{\pi} + \frac{\sigma^2_{\text{REV}}(0)}{(1-\pi)}\big)$. Proof in App. F.5.*

Based on the theorem, an A/B testing procedure is the following. Compute the revenue variance as Eq. (9) for each market, obtaining $\hat{\sigma}^2_{\text{REV}}(1)$ and $\hat{\sigma}^2_{\text{REV}}(0)$, and form the confidence interval

$$\hat{\tau}_{\text{REV}} \pm z_{\alpha/2}\left(\frac{\hat{\sigma}^2_{\text{REV}}(1)}{\pi} + \frac{\hat{\sigma}^2_{\text{REV}}(0)}{(1-\pi)}\right) \tag{12}$$

If zero is on the left (resp. right) of the CI Eq. (17), we conclude that the new feature increases (resp. decreases) revenue with $(1-\alpha)\times 100\%$ confidence. If zero is inside the interval, the effect of new feature is undecided. See App. G.4 for a numerical study verifying the validity of this procedure. Algorithm 1 presents a step-by-step procedure for using the above results.

5. Experiment

In App. G we conduct simulations to investigate asymptotic normality for $i \notin I$ and fast convergence rate for $i \in I$, the effect of smoothing parameter on Hessian estimation, the coverage rate of the CI for revenue, and the coverage rate of the treatment effect CI for revenue. Through these experiments, we confirm that the finite sample of β converges to a normal distribution, with fast convergence in the entries whose constraints are tight. We confirm that for Hessian estimation, a suitable choice of ε_t of smoothing parameter sequence is t^{-d} for $d \in (.3, .5)$. Finally, the revenue confidence intervals in Thm. 3 and Eq. (17) attain the nominal coverage rate.

Acknowledgements

This research was supported by the Office of Naval Research under grants N00014-22-1-2530 and N00014-23-1-2374, and the National Science Foundation award IIS-2147361.

References

Andrews, D. W. Asymptotics for semiparametric econometric models via stochastic equicontinuity. *Econometrica: Journal of the Econometric Society*, pp. 43–72, 1994.

Andrews, D. W. Estimation when a parameter is on a boundary. *Econometrica*, 67(6):1341–1383, 1999.

Andrews, D. W. Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica*, 69(3): 683–734, 2001.

Arcones, M. A. and Yu, B. Central limit theorems for empirical and u-processes of stationary mixing sequences. *Journal of Theoretical Probability*, 7(1):47–71, 1994.

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

Athey, S. and Haile, P. A. Nonparametric approaches to auctions. *Handbook of econometrics*, 6:3847–3965, 2007.

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

Bajari, P., Burdick, B., Imbens, G. W., Masoero, L., McQueen, J., Richardson, T., and Rosen, I. M. Multiple randomization designs. *arXiv preprint arXiv:2112.13495*, 2021.

Balseiro, S., Kim, A., Mahdian, M., and Mirrokni, V. Budget-management strategies in repeated auctions. *Operations research*, (3):859–876, 2021.

Balseiro, S. R. and Gur, Y. Learning in repeated auctions with budgets: Regret minimization and equilibrium. *Management Science*, 65(9):3952–3968, 2019.

Balseiro, S. R., Besbes, O., and Weintraub, G. Y. Repeated auctions with budgets in ad exchanges: Approximations and design. *Management Science*, 61(4):864–884, 2015.

Basse, G. W., Soufiani, H. A., and Lambert, D. Randomization and the pernicious effects of limited budgets on auction experiments. In *Artificial Intelligence and Statistics*, pp. 1412–1420. PMLR, 2016.

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

Bertsimas, D., Farias, V. F., and Trichakis, N. On the efficiency-fairness trade-off. *Management Science*, 58 (12):2234–2250, 2012.

Blake, T. and Coey, D. Why marketplace experimentation is harder than it seems: The role of test-control interference. In *Proceedings of the fifteenth ACM conference on Economics and computation*, pp. 567–582, 2014.

Bojinov, I. and Gupta, S. Online experimentation: Benefits, operational and methodological challenges, and scaling guide. *Harvard Data Science Review*, 4(3), 2022.

Bojinov, I. and Shephard, N. Time series experiments and causal estimands: exact randomization tests and trading. *Journal of the American Statistical Association*, 114 (528):1665–1682, 2019.

Bojinov, I., Simchi-Levi, D., and Zhao, J. Design and analysis of switchback experiments. *Management Science*, 2022.

Borgs, C., Chayes, J., Immorlica, N., Jain, K., Etesami, O., and Mahdian, M. Dynamics of bid optimization in online advertisement auctions. In *Proceedings of the 16th international conference on World Wide Web*, pp. 531–540, 2007.

Caragiannis, I., Kurokawa, D., Moulin, H., Procaccia, A. D., Shah, N., and Wang, J. The unreasonable fairness of maximum nash welfare. *ACM Transactions on Economics and Computation (TEAC)*, 7(3):1–32, 2019.

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

Chen, L., Ye, Y., and Zhang, J. A note on equilibrium pricing as convex optimization. In *International Workshop on Web and Internet Economics*, pp. 7–16. Springer, 2007.

Chen, X., Linton, O., and Van Keilegom, I. Estimation of semiparametric models when the criterion function is not smooth. *Econometrica*, 71(5):1591–1608, 2003.

Clarke, F. H. *Optimization and Nonsmooth Analysis*. Society for Industrial and Applied Mathematics, January 1990. doi: 10.1137/1.9781611971309. URL https://doi.org/10.1137/1.9781611971309.

Cole, R., Devanur, N. R., Gkatzelis, V., Jain, K., Mai, T., Vazirani, V. V., and Yazdanbod, S. 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.

Conitzer, V., Kroer, C., Panigrahi, D., Schrijvers, O., Stier-Moses, N. E., Sodomka, E., and Wilkens, C. A. Pacing equilibrium in first price auction markets. *Management Science*, 2022a.

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

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

Duchi, J. C. and Ruan, F. Asymptotic optimality in stochastic optimization. *The Annals of Statistics*, 49(1):21–48, 2021.

Dupačová, J. On non-normal asymptotic behavior of optimal solutions for stochastic programming problems and on related problems of mathematical statistics. *Kybernetika*, 27(1):38–52, 1991.

Dupacová, J. and Wets, R. Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. *The annals of statistics*, 16(4):1517–1549, 1988.

Durrett, R. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

Fradkin, A. A simulation approach to designing digital matching platforms. *Boston University Questrom School of Business Research Paper Forthcoming*, 2019.

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

Gao, Y., Kroer, C., and Peysakhovich, A. Online market equilibrium with application to fair division. *arXiv preprint arXiv:2103.12936*, 2021.

Geyer, C. J. On the asymptotics of constrained m-estimation. *The Annals of statistics*, pp. 1993–2010, 1994.

Giné, E. and Nickl, R. *Mathematical foundations of infinite-dimensional statistical models*. Cambridge university press, 2021.

Glynn, P. W., Johari, R., and Rasouli, M. Adaptive experimental design with temporal interference: A maximum likelihood approach. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H. (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 15054–15064. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/abd987257ff0eddc2bc6602538cb3c43-Paper.pdf.

Guo, W., Kandasamy, K., Gonzalez, J. E., Jordan, M. I., and Stoica, I. Online learning of competitive equilibria in exchange economies. *arXiv preprint arXiv:2106.06616*, 2021.

Hong, H. and Li, J. The numerical bootstrap. *The Annals of Statistics*, 48(1):397–412, 2020.

Hsieh, Y.-W., Shi, X., and Shum, M. Inference on estimators defined by mathematical programming. *Journal of Econometrics*, 226(2):248–268, 2022.

Hu, Y. and Wager, S. Switchback experiments under geometric mixing. *arXiv preprint arXiv:2209.00197*, 2022.

Hu, Y., Li, S., and Wager, S. Average direct and indirect causal effects under interference. *Biometrika*, 2022.

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

Imbens, G. W. and Rubin, D. B. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, 2015.

Jagadeesan, M., Wei, A., Wang, Y., Jordan, M., and Steinhardt, J. Learning equilibria in matching markets from bandit feedback. *Advances in Neural Information Processing Systems*, 34:3323–3335, 2021.

Johari, R., Li, H., Liskovich, I., and Weintraub, G. Y. Experimental design in two-sided platforms: An analysis of bias. *Management Science*, 2022.

Knight, K. Epi-convergence in distribution and stochastic equi-semicontinuity. *Unpublished manuscript*, 37(7):14, 1999.

Knight, K. Limiting distributions of linear programming estimators. *Extremes*, 4(2):87–103, 2001.

Knight, K. Asymptotic theory for m-estimators of boundaries. In *The Art of Semiparametrics*, pp. 1–21. Springer, 2006.

Knight, K. On the asymptotic distribution of the analytic center estimator. *Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in honor of Professor Jana Jurecková*, pp. 123, 2010.

Kohavi, R. and Thomke, S. The surprising power of online experiments. *Harvard business review*, 95(5):74–82, 2017.

Kosorok, M. R. *Introduction to empirical processes and semiparametric inference.* Springer, 2008.

Larsen, N., Stallrich, J., Sengupta, S., Deng, A., Kohavi, R., and Stevens, N. Statistical challenges in online controlled experiments: A review of a/b testing methodology. *arXiv preprint arXiv:2212.11366*, 2022.

Le Cam, L., LeCam, L. M., and Yang, G. L. *Asymptotics in statistics: some basic concepts*. Springer Science & Business Media, 2000.

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

Li, H., Zhao, G., Johari, R., and Weintraub, G. Y. Interference, bias, and variance in two-sided marketplace experimentation: Guidance for platforms. In *Proceedings of the ACM Web Conference 2022*, pp. 182–192, 2022.

Li, J. The proximal bootstrap for constrained estimators. 2022.

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

Liao, L., Gao, Y., and Kroer, C. Nonstationary dual averaging and online fair allocation. *arXiv preprint arXiv:2202.11614v1*, 2022a.

Liao, L., Gao, Y., and Kroer, C. Statistical inference for fisher market equilibrium. In *arXiv preprint arXiv:2209.15422v1*, 2022b. URL https://arxiv.org/abs/2209.15422.

Liu, L. T., Ruan, F., Mania, H., and Jordan, M. I. Bandit learning in decentralized matching markets. *J. Mach. Learn. Res.*, 22:211–1, 2021a.

Liu, M., Mao, J., and Kang, K. Trustworthy and powerful online marketplace experimentation with budget-split design. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 3319–3329, 2021b.

Liu, Z., Lu, M., Wang, Z., Jordan, M., and Yang, Z. Welfare maximization in competitive equilibrium: Reinforcement learning for markov exchange economy. In *International Conference on Machine Learning*, pp. 13870–13911. PMLR, 2022.

Min, Y., Wang, T., Xu, R., Wang, Z., Jordan, M. I., and Yang, Z. Learn to match with no regret: Reinforcement learning in markov matching markets. *arXiv preprint arXiv:2203.03684*, 2022.

Munro, E., Wager, S., and Xu, K. Treatment effects in market equilibrium. *arXiv preprint arXiv:2109.11647*, 2021.

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

Pakes, A. and Pollard, D. Simulation and the asymptotics of optimization estimators. *Econometrica: Journal of the Econometric Society*, pp. 1027–1057, 1989.

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

Self, S. G. and Liang, K.-Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398):605–610, 1987.

Shapiro, A. Sensitivity analysis of nonlinear programs and differentiability properties of metric projections. *SIAM Journal on Control and Optimization*, 26(3):628–645, 1988.

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

Shapiro, A. On differential stability in stochastic programming. *Mathematical Programming*, 47(1):107–116, 1990.

Shapiro, A. Asymptotic analysis of stochastic programs. *Annals of Operations Research*, 30(1):169–186, 1991.

Shapiro, A. Asymptotic behavior of optimal solutions in stochastic programming. *Mathematics of Operations Research*, 18(4):829–845, 1993.

Shapiro, A. On the asymptotics of constrained local m-estimators. *Annals of statistics*, pp. 948–960, 2000.

Shapiro, A., Dentcheva, D., and Ruszczynski, A. *Lectures on stochastic programming: modeling and theory*. SIAM, 2021.

Sneider, C., Tang, Y., and Tang, Y. Experiment rigor for switchback experiment analysis. *URL: https://doordash. engineering/2019/02/20/experiment-rigor-for-switchback-experiment-analysis*, 2018.

Vaart, A. W. and Wellner, J. A. Weak convergence. In *Weak convergence and empirical processes*, pp. 16–28. Springer, 1996.

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

Wager, S. and Xu, K. 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.

Wainwright, M. J. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.

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

A. Notations

Symbol	Meaning
Notations in First-Price Pacing Equilibrium	
β^*, β^γ	pacing multiplier
$p^*(\cdot), p^\gamma$, REV^*, REV^γ	price and revenue
$\mu^*(\cdot), \mu^\gamma, \bar{\mu}^*, \bar{\mu}^\gamma$	utility generated from items
δ^*, δ^γ	leftover budget
u^*, u^γ	total utility = utility from items + leftover budgets
$x^*(\cdot), x^\gamma$	allocation
$s(\cdot)$	supply (a probability density)
b	budget
v, v_i	valuations
Notations in Eisenberg-Gale Grogram	
$F(\theta, \beta)$, $\Psi(\beta)$, and $f(\theta, \beta)$	see Eq. (P-DualEG)
H, H_t	objective function of Eisenberg-Gale convex programs
B	$B = (0, 1]^n$ the domain of H and H_t
$\mathcal{H}$	the Hessian matrix of H at β^*
$\mathcal{P}$	matrix whose diagonal = $\mathbb{1}(\beta_i^* < 1)$
$\mathcal{H}_B$	$= \mathcal{P}\mathcal{H}\mathcal{P}$
I, I^c	partition of buyers into those with $\beta_i^* = 1$ (I) and those not (I^c)

B. Related Works

A/B testing in two-sided markets. Empirical studies by Blake & Coey (2014); Fradkin (2019) demonstrate bias in experiments due to marketplace interference. Basse et al. (2016) study the bias and variance of treatment effects under two randomization schemes for auction experiments. Bojinov & Shephard (2019) study the estimation of causal quantities in time series experiments. Some recent state-of-the-art designs are the multiple randomization designs (Liu et al., 2021b; Johari et al., 2022; Bajari et al., 2021) and the switch-back designs (Sneider et al., 2018; Hu & Wager, 2022; Li et al., 2022; Bojinov et al., 2022; Glynn et al., 2020). The surveys by Kohavi & Thomke (2017); Bojinov & Gupta (2022) contain detailed accounts of A/B testing in internet markets. See Larsen et al. (2022) for an extensive survey on statistical challenges in A/B testing. Our paper focuses on A/B testing in first-price auction markets with the consideration of equilibrium effects, to the best of our knowledge this is the first work to consider market equilibrium effects in A/B testing.

Pacing equilibrium. Pacing and throttling are two prevalent budget-management methods on ad auction platforms. Here we focus on pacing methods since that is our setting. In the first-price setting, Borgs et al. (2007) study first price auctions with budget constraints in a perturbed model, whose limit prices converge to those of an FPPE. Building on the work of Borgs et al. (2007), Conitzer et al. (2022a) introduce the FPPE model and discover several properties of FPPE such as shill-proofness and monotonicity in buyers, budgets and goods. There it is also established that FPPE is closely related to the quasilinear Fisher market equilibrium (Chen et al., 2007; Cole et al., 2017). Gao & Kroer (2022) propose an infinite-dimensional variant of the quasilinear Fisher market, which lays the probability foundation of the current paper. Gao et al. (2021); Liao et al. (2022a) study online computation of the infinite-dimensional Fisher market equilibrium. In the second-price setting, Balseiro et al. (2015) investigate budget-management in second-price auctions through a fluid mean-field approximation; Balseiro & Gur (2019) study adaptive pacing strategy from buyers' perspective in a stochastic continuous setting; Balseiro et al. (2021) study several budget smoothing methods including multiplicative pacing in a stochastic context; Conitzer et al. (2022b) study second price pacing equilibrium, and shows that the equilibria exist under fractional allocations.

M-estimation when the parameter is on the boundary There is a long literature on the statistical properties of M-estimators when the parameter is on the boundary (Geyer, 1994; Shapiro, 1990; 1988; 1989; 1991; 1993; 2000; Andrews, 1999; 2001; Knight, 1999; 2001; 2006; 2010; Dupacová & Wets, 1988; Dupačová, 1991; Self & Liang, 1987). Some recent works on the statistical inference theory for constrained M-estimation include Li (2022); Hong & Li (2020); Hsieh et al. (2022). Our work leverages Shapiro (1989), which develops a general set of conditions for CLTs of constrained M-estimators when the sample function is nonsmooth. Working under the specific model of FPPE, we build on and go beyond these contributions by deriving sufficient condition for asymptotic normality in FPPE, establishing local asymptotic

minimax theory and developing valid inferential procedures.

Statistical learning and inference with equilibrium effects Wager & Xu (2021); Munro et al. (2021); Sahoo & Wager (2022) take a mean-field game modeling approach and perform policy learning with a gradient descent method. Liao et al. (2022b) consider statistical inference in the Fisher market equilibrium which is useful for fair and efficient resource allocations. 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 & Shah, 2022; Dai & Jordan, 2021; Liu et al., 2021a; Jagadeesan et al., 2021; Min et al., 2022). Our work is also related to the rich literature of inference under interference (Hudgens & Halloran, 2008; Aronow & Samii, 2017; Athey et al., 2018; Leung, 2020; Hu et al., 2022; Li & Wager, 2022). In the FPPE model, the interference among buyers is caused by the supply and budget constraint and the revenue-maximizing incentive of the platform. In the economic literature, researchers have studied how to estimate auction market primitives from bid data; see (Athey & Haile, 2007) for a survey.

C. Omitted properties of FPPE

Definition 3 (Observed FPPE, formal)**.** *Given (b, v, σ, γ), an observed FPPE is a tuple $(\beta, x_1, \ldots, x_n) \in [0,1]^n \times ([0,1]^t)^n$ such that*

- *(First-price) For all θ^τ, $p_j(\theta^\tau) = \max_i \beta_i v_i(\theta^\tau)$. Moreover, $x_i(\theta^\tau) > 0$ implies $\beta_i v_i(\theta) = \max_i \beta_i x_i(\theta^\tau)$ for all i and θ^τ.*
- *(Supply and budget feasible) For all i, $\sigma \sum_{\tau=1}^t x_i(\theta^\tau) p(\theta^\tau) \le b_i$. For all θ^τ, $\sum_{i=1}^n x_i(\theta^\tau) \le 1$.*
- *(Revenue maximizing and market clearing) For all i, $\sigma \sum_{\tau=1}^t x_i(\theta^\tau) p(\theta^\tau) < b_i$ implies $\beta_i = 1$. For all θ^τ, $p(\theta^\tau) > 0$ implies $\sum_{i=1}^n x_i(\theta^\tau) = 1$.*

We start by listing a number of known properties of the limit FPPE that we will use in our proofs.

A limit FPPE allocation x and the limit FPPE pacing multiplier β can be recovered through the convex programs, the population primal EG

$$\max_{x \in L_+^\infty(\Theta), u, \delta \ge 0} \sum_{i=1}^n (b_i \log(u_i) - \delta_i) \tag{P-EG}$$
$$\text{s.t. } u_i \le \langle v_i, s x_i \rangle + \delta_i, \ \sum_{i=1}^n x_i(\theta) \le 1$$

and the population dual EG

$$\min_{0 < \beta_i \le 1, i \in [n]} H(\beta) = \mathbb{E}[F(\theta, \beta)] = \mathbb{E}[f(\beta, \theta)] + \Psi(\theta) \tag{P-DualEG}$$

where $F(\theta, \beta) = f(\beta, \theta) + \Psi(\theta)$, $f(\theta, \beta) = \max_{i \in [n]} \beta_i v_i(\theta)$ and $\Psi(\beta) = -\sum_{i=1}^n b_i \log \beta_i$.

Lemma 2 (FPPE $\Leftrightarrow$ EG)**.** *The limit FPPE β^* is the unique solution to Eq.* (P-DualEG)*, and any limit FPPE (x^*, u^*, δ^*) belongs to the set of optimal solutions to Eq.* (P-EG) *(Gao & Kroer, 2022).*

Lemma 3 (First-order conditions of limit FPPE, Theorem 10 from Gao & Kroer (2022))**.** *Given (b, v, s), the limit FPPE satisfies the following.*

- $u_i^* = b_i / \beta_i^*$ *all* i.
- $\beta_i^* v_i(\theta) \le p^*(\theta)$ *all* i, θ.
- $x_i^*(\theta), \delta_i^*, \beta_i^*, p^*(\theta) \ge 0$ *all* i, θ.
- $p^*(\theta) > 0 \implies \sum_{i=1}^n x_i^*(\theta) = 1$ *all* θ.
- $\delta_i^* > 0 \implies \beta_i^* = 1$ *all* i.
- $x_i^*(\theta) > 0 \implies \beta_i^* = p^*(\theta)/v_i(\theta)$ *all* i, θ.

Lemma 4 (From EG to FPPE)**.** *Recall the definition of H in Eq.* (P-DualEG)*. Under SMO*

$$\begin{aligned}
&\delta^* = -\nabla H(\beta^*),\\
&p^*(\theta) = f(\theta, \beta^*),\\
&\mu^*(\theta) = x^*(\theta) \odot v(\theta) = \nabla_\beta f(\theta, \beta^*),\\
&\bar{\mu}^* = \mathbb{E}[\nabla f(\theta, \beta^*)] = \nabla \bar{f}(\beta^*),\\
&\text{REV}^* = \bar{f}(\beta^*), \textit{ and}\\
&\text{NSW}^* = \Psi(\beta^*) + \sum_{i=1}^n b_i \log b_i.
\end{aligned}$$

Lemma 5 (Scale-invariance)**.** *Scaling the budget and values at the same time does not change the market equilibrium. Scaling the value and the supply inversely does not change the market equilibrium. That is, given a positive scalar α, if $(x, \beta, p) \in$ FPPE(b, v, s), then*

$$\begin{aligned}
&(x, \beta, \alpha p) \in \textsf{FPPE}(\alpha b, \alpha v, s)\ ,\\
&(x, \beta, p) \in \textsf{FPPE}(b, \alpha v, \tfrac{1}{\alpha} s)\ .
\end{aligned}$$

Similarly, for a given observed FPPE $\widehat{\textsf{FPPE}}(b, v, \sigma, \gamma)$ defined in Def. 2, and any positive scalar α, if $(x, \beta, p) \in \widehat{\textsf{FPPE}}(b, v, \sigma, \gamma)$, then

$$(x, \beta, \alpha p) \in \widehat{\textsf{FPPE}}(\alpha b, \alpha v, \sigma, \gamma), \tag{13}$$

$$(x, \beta, p) \in \widehat{\textsf{FPPE}}(b, \alpha v, \tfrac{1}{\alpha}\sigma, \gamma) \tag{14}$$

Lemma 6. *Let $(x^*, \beta^*, p^*) =$ FPPE(b, v, s). Assume SMO holds. Then $\beta_i^* < 1$ for all i implies $p^*(\theta) = (\bar{\mu}^*)^\top \mathcal{H}^{-1} \mu^*(\theta)$.*

Proof. If $\beta_i^* < 1$ for all i, then $\bar{\mu}^* = u^*$. By definition and SMO, $\mu_i^*(\theta) = x_i^*(\theta) v_i(\theta)$ and $p^*(\theta) = \max_i \beta_i^* v_i(\theta) = \sum_{i=1}^n x_i^*(\theta)\beta_i^* v_i(\theta)$. It suffices to show $(\bar{\mu}^*)^\top \mathcal{H}^{-1} = (\beta^*)^\top$, or equivalently $\mathcal{H}\beta^* = u^*$. Recall $\bar{f}(\beta) = \mathbb{E}[\max_i \beta_i v_i(\theta)]$ is homogenous, i.e., $\bar{f}(\alpha\beta) = \alpha \bar{f}(\beta)$ for any positive scalar $\alpha \geq 0$. By the Euler's identity for homogenous functions, we have

$$\nabla \bar{f}(\beta) = \sum_{i=1}^n \beta_i \nabla_i \bar{f}(\beta)$$

Taking derivative again, we have $\nabla^2 \bar{f}(\beta)\beta = 0$ for all β. Finally, note $\mathcal{H} = \nabla^2 \bar{f}(\beta^*) + \nabla^2 \Psi(\beta^*)$, we have $\mathcal{H}\beta^* = \nabla^2 \Psi(\beta^*)\beta^* = u^* = \bar{\mu}^*$, where the second equality holds by the first-order condition that $u_i^* = b_i/\beta_i^*$, and the third by the fact that $\beta_i^* < 1$ for all i implies $\delta_i^* = 0$. So $(\bar{\mu}^*)^\top \mathcal{H}^{-1} \mu^*(\theta) = (\beta^*)^\top \mu^*(\theta) = \sum_{i=1}^n \beta_i^* \mu_i^*(\theta) = p^*(\theta)$. □

As for the limit FPPE, the observed FPPE $\widehat{\textsf{FPPE}}(b, v, \sigma, \gamma)$ is characterized by primal and dual convex programs:

$$\max_{x, \delta, u \geq 0} \left\{ \sum_{i=1}^n (b_i \log(u_i) - \delta_i) \mid u_i \leq \sigma \langle v_i(\gamma), x_i \rangle + \delta_i\ \forall i, \quad \sum_{i=1}^n x_i^\tau \leq 1\ \forall \tau \right\}, \tag{15}$$

$$\min_{1_t \geq \beta > 0} \left\{ H_t(\beta) \equiv \sigma \sum_{\tau=1}^t \max_{i \in [n]} \beta_i v_i(\theta^\tau) - \sum_{i=1}^n b_i \log \beta_i \right\}. \tag{16}$$

See Conitzer et al. (2022a, Sec. 5) and Gao & Kroer (2022) for more details on the convex program characterization of observed FPPE. As with the limit FPPE, the observed FPPE has an analogous set of properties.

Lemma 7 (First-order conditions of observed FPPE, from EG to FPPE)**.** *Given (b, v, γ), the observed FPPE $\widehat{\textsf{FPPE}}(b, v, \frac{1}{t}, \gamma)$ satisfies the following.*

- $u_i^\gamma = b_i/\beta_i^\gamma$ *all* i.
- $\beta_i^\gamma v_i(\theta) \leq p^\gamma(\theta^\tau)$ *all* i, θ^τ.

- $x_i^\gamma(\theta^\tau), \delta_i^\gamma, \beta_i^\gamma, p^\gamma(\theta^\tau) \geq 0$ *all* i, θ^τ.
- $p^\gamma(\theta^\tau) > 0 \implies \sum_{i=1}^n x_i^\gamma(\theta^\tau) = 1$ *all* θ^τ.
- $\delta_i^\gamma > 0 \implies \beta_i^\gamma = 1$ *all* i.
- $x_i^\gamma(\theta^\tau) > 0 \implies \beta_i^\gamma = p^\gamma(\theta^\tau)/v_i(\theta^\tau)$ *all* i, θ^τ.

Moreover, recall the definition of H_t *in Eq.* (S-DualEG). *Then*

$$\delta^\gamma \in -\partial H_t(\beta^\gamma), \bar{\mu}^\gamma \in \partial\left(\frac{1}{t}\sum_{\tau=1}^t f(\theta^\tau, \beta^\gamma)\right), p^\gamma(\theta^\tau) = f(\theta^\tau, \beta^\gamma),$$
$$\mu^\gamma(\theta^\tau) = x^\gamma(\theta^\tau) \odot v(\theta^\tau) \in \partial_\beta f(\theta^\tau, \beta^\gamma)$$
$$\text{REV}^\gamma = \frac{1}{t}\sum_{\tau=1}^t f(\theta^\tau, \beta^\gamma), \text{NSW}^\gamma = \Psi(\beta^\gamma) + \sum_{i=1}^n b_i \log b_i.$$

C.1. Proof of Lemma 1

Proof of Lemma 1. The proof follows a similar argument as in Liao et al. (2022b) for non-quasilinear Fisher market (recall that an FPPE is a quasilinear Fisher market). Recall $f(\theta, \beta) = \max_i \beta_i v_i(\theta)$. We define

$$\epsilon(\theta, \beta) = \max_i\{\beta_i v_i(\theta)\} - \operatorname{secondmax}_i\{\beta_i v_i(\theta)\}$$

where secondmax is the second-highest entry (which could possibly be equal to the highest entry). For example, $\operatorname{secondmax}(1, 1, 2) = 1$.

Note $f(\theta, \cdot)$ is differentiable at β if and only if $\epsilon(\theta, \beta) > 0$, since this holds if and only if the subdifferential is a singleton at β. Let $\Theta_{\text{diff}}(\beta) \equiv \{\theta : f(\theta, \beta) \text{ differentiable at } \beta\}$ and $\Theta_{\text{tie}}(\beta) = \{\theta : \beta_i v_i(\theta) = \beta_k v_k(\theta) \text{ for some } k \neq i\}$. Then

$$\Theta_{\text{diff}}(\beta) = \{\theta : \epsilon(\theta, \beta) > 0\}, \Theta_{\text{tie}}(\beta) = \{\theta : \epsilon(\theta, \beta) = 0\}.$$

By Proposition 2.3 from Bertsekas (1973) we know $\bar{f}(\beta) = \mathbb{E}[f(\theta, \beta)]$ is differentiable at β if and only if $\Theta_{\text{diff}}(\beta)$ is measure one.

Proof of Θ_{tie} *is measure zero.* Recall $\Theta_{\text{tie}} = \Theta_{\text{tie}}(\beta^*)$. The claim follows from the fact that the SMO assumption implies differentiability, which implies that the complement of Θ_{tie} is measure one.

Proof of uniqueness of x^*, μ^*, δ^*. By Lemma 3, we know $x_i^*(\theta) > 0$ only if $\beta_i^* v_i(\theta) \geq \beta_k^* v_k(\theta)$ for all other k. Under the pacing profile β^*, for all (but a measure-zero set of) items there is only one winning bidder, i.e., for almost all θ there is a unique i such that $x_i^*(\theta) > 0$ and $x_k^*(\theta) = 0$ for all $k \neq i$. Coupled with the limit FPPE first-order condition that $\sum_{i=1}^n x_i^*(\theta) = 1$ we know x^* is unique. This immediately implies uniqueness of μ^*, δ^* as well.

Proof of existence of N_{diff}. By the assumption that $\bar{f}$ is twice continuously differentiable at β^*, there is a neighborhood N_{diff} such that $\bar{f}$ is continuously differentiable on N_{diff}. By the same argument as for $\Theta_{\text{tie}}(\beta^*)$ being measure zero, $\Theta_{\text{tie}}(\beta)$ is measure zero for each $\beta \in N_{\text{diff}}$.

□

D. More A/B testing estimands

We remark that the potential value functions are suitable for modeling either item-side or buyer-side treatments. In the context of ad auctions, item-size treatment are, for example, positions of the ads in the browser, whether links are attached to the ads and so on. Buyer-side treatments are, for example, a new layout of the ad campaign setup portal for the advertisers. The following discussion centers around item-side treatments since they are more prominent in practice, but readers should keep in mind that our theory extends to buyer-side treatments.

As discussed in Sec. 2.3, each FPPE has a convex program characterization. If the market is given treatment $w \in \{0, 1\}$, then the limit FPPE pacing multipliers can be recovered by $\max_{\beta \in B} \int \max_i \beta_i v_i(w, \theta) s(\theta)\, \mathrm{d}\theta - \sum_{i=1}^n b_i \log \beta_i$. Let

$\beta^*(w)$ be the unique solution to the above program and also the unique FPPE pacing multiplier. The FPPE prices and revenue are $p^*(w,\theta) \equiv \max_i v_i(w,\theta)\beta_i^*(w)$ and $\text{REV}^*(w) \equiv \int p^*(w,\theta)s(\theta)\,\mathrm{d}\theta$. The utility vector under treatment w is $u_i^*(w) \equiv b_i/\beta^*(w)$. The NSW is $\text{NSW}^*(w) \equiv \sum_{i=1}^n b_i \log u_i^*(w)$.

Other metrics of treatment effects could be (i) treatment effects on revenue: $\tau_{\text{REV}} \equiv \text{REV}^*(1) - \text{REV}^*(0)$, (ii) treatment effects on Nash social welfare: $\tau_{\text{NSW}} \equiv \sum_{i=1}^n b_i \log u_i^*(1) - \sum_{i=1}^n b_i \log u_i^*(0)$, (iii) treatment effects on pacing multiplier: $\tau_\beta \equiv \beta^*(1) - \beta^*(0)$, and (vi) treatment effects on utilities: $\tau_u \equiv u^*(1) - u^*(0)$. The estimators of treatment effects are $\hat{\tau}_{\text{REV}} \equiv \text{REV}^\gamma(1) - \text{REV}^\gamma(0)$, $\hat{\tau}_\beta \equiv \beta^\gamma(1) - \beta^\gamma(0)$, $\hat{\tau}_{\text{NSW}} \equiv \text{NSW}^\gamma(1) - \text{NSW}^\gamma(0)$, and $\hat{\tau}_u \equiv u^\gamma(1) - u^\gamma(0)$.

For given (b,s), the (co)variances $\Sigma_\beta, \Sigma_u, \sigma^2_{\text{NSW}}, \sigma^2_{\text{REV}}$ in Thm. 1 and Corollary 1 are functionals of the value functions. We will use $\Sigma_\beta(w), \Sigma_u(w), \sigma^2_{\text{NSW}}(w), \sigma^2_{\text{REV}}(w)$ to represents the (co)variances in the market formed with value functions $\{v_i(w,\cdot)\}_i$. Each variance can be estimated the same way as in Sec. 3.3. Let $\beta^\gamma(w), u^\gamma(w)$, $\text{REV}^\gamma(w)$ and $\text{NSW}^\gamma(w)$ denote the observed FPPE quantities for treatment $w \in \{0,1\}$.

Theorem 5 (Treatment effects CLT)**.** *Suppose SMO and SCS hold in the limit markets FPPE$(b, v(1), s)$ and FPPE$(b, v(0), s)$. Then* $\sqrt{t}(\hat{\tau}_{\text{REV}} - \tau_{\text{REV}}) \xrightarrow{d} \mathcal{N}\big(0, \frac{\sigma^2_{\text{REV}}(1)}{\pi} + \frac{\sigma^2_{\text{REV}}(0)}{(1-\pi)}\big)$, $\sqrt{t}(\hat{\tau}_{\text{NSW}} - \tau_{\text{NSW}}) \xrightarrow{d} \mathcal{N}\big(0, \frac{\sigma^2_{\text{NSW}}(1)}{\pi} + \frac{\sigma^2_{\text{NSW}}(0)}{(1-\pi)}\big)$, $\sqrt{t}(\hat{\tau}_\beta - \tau_\beta) \xrightarrow{d} \mathcal{N}\big(0, \frac{1}{\pi}\Sigma_\beta(1) + \frac{1}{(1-\pi)}\Sigma_\beta(0)\big)$, *and* $\sqrt{t}(\hat{\tau}_u - \tau_u) \xrightarrow{d} \mathcal{N}\big(0, \frac{1}{\pi}\Sigma_u(1) + \frac{1}{(1-\pi)}\Sigma_u(0)\big)$. *Proof in App. F.5.*

Algorithm 1 A/B test effect of a new feature on revenue

Step 1. Experiment. Choose the new feature assignment probability π. Perform A/B testing with budget splitting and item randomization. Form two first-price pacing equilibria.

Step 2. Collect data. Observe the equilibrium data from the two markets, including prices, item allocations, pacing multipliers, leftover budgets, and values of the observed items.

Step 3. Compute CI. Compute the revenue variance as Eq. (11) for each market, obtaining $\hat{\sigma}^2_{\text{REV}}(1)$ and $\hat{\sigma}^2_{\text{REV}}(0)$, and form the confidence interval

$$\hat{\tau}_{\text{REV}} \pm z_{\alpha/2}\left(\frac{\hat{\sigma}^2_{\text{REV}}(1)}{\pi} + \frac{\hat{\sigma}^2_{\text{REV}}(0)}{(1-\pi)}\right) \tag{17}$$

E. Technical Lemmas

E.1. A CLT for constrained M-estimator

We introduce a CLT result from (Shapiro, 1989) that handles M-estimation when the true parameter is on the boundary of the constraint set. Throughout this section, when we refer to assumptions A1, A2, B2, etc, we mean those assumptions in Shapiro (1989).

Let (Θ, P) be a probability space. Consider $f : \Theta \times \mathbb{R}^n \to \mathbb{R}$ and a set $B \subset \mathbb{R}^n$. Let $\theta_1, \ldots, \theta_t$ be a sample of independent random variables with values in Θ having the common probability distribution P. Let $\phi(\beta) = Pf(\cdot,\beta) = \mathbb{E}[f(\theta,\beta)]$, and $\psi_t(\beta) = P_t f(\cdot,\beta) = \frac{1}{t}\sum_{i=1}^t f(\theta_i,\beta)$. Let β_0 be the unique minimizer of ϕ over B (Assumption A4 in Shapiro (1989)). Let $\vartheta_t = \inf_B \psi_t$ and $\hat{\beta}$ be an optimal solution.

We begin with some blanket assumptions. Suppose the geometry of B at β_0 is given by functions $g_i(\beta)$ (Assumption B1), i.e., there exists a neighborhood N such that

$$B \cap N = \{\beta \in N : g_i(\beta) = 0, i \in K; g_i(\beta) \le 0, i \in J\},$$

where K and J are finite index sets and the constraints in J are active at β_0, meaning $g_i(\beta_0) = 0$ for all $i \in J$. Assume the functions $g_i, i \in K \cup J$, are twice continuously differentiable in a neighborhood of β_0 (Assumption B2). Define the Lagrangian function by $l(\beta,\lambda) = \phi(\beta) + \sum_{i \in K\cup J} \lambda_i g_i(\beta)$. Let Λ_0 be the set of optimal Lagrange multipliers, i.e., $\lambda \in \Lambda_0$ iff $\nabla l(\beta_0,\lambda) = 0$ (assuming differentiability) and $\lambda_i \ge 0, i \in J$.

Lemma 8 (Theorems 3.1 and 3.2 from Shapiro (1989))**.** *Assume there exists a neighborhood N of β_0 such that the following holds.*

8.a Conditions on the sample function f and the distribution P.

- *(Assumption A1 in the original paper) For almost every θ, $f(\theta, \beta)$ is a continuous function of β, and for all $\beta \in B$, $f(\theta, \beta)$ is a measurable function of θ.*
- *(Assumption A2) The family $\{f(\theta, \beta)\}, \beta \in B$, is uniformly integrable.*
- *(Assumption A4) For all θ, there exist a positive constant $K(\theta)$ such that $|f(\theta, w) - f(\theta, \beta)| \leq K(\theta)\|w - \beta\|$ for all $\beta, w \in N$.*
- *(Assumption A5) For each fixed $\beta \in N$, $f(\theta, \cdot)$ is continuously differentiable at β for almost every θ.*
- *(Assumption A6) The family $\{\nabla f(\theta, \beta)\}_{\beta \in N}$, is uniformly integrable.*
- *(Assumption D) The expectation $\mathbb{E}[\|\nabla f(\theta, \beta_0)\|^2]$ is finite.*
- *(Assumption B4) The function ϕ is twice continuously differentiable in a neighborhood of β_0.*

8.b Conditions on the optimal solution.

- *(Assumption B3) A constraint qualification, the Mangasarian-Fromovitz condition: The gradient vectors $\nabla g_i(\beta_0), i \in K$, are linearly independent, and there exists a vector w such that $w^\top \nabla g_i(\beta_0) = 0, i \in K$ and $w^\top \nabla g_i(\beta_0) < 0, i \in J$.*
- *(Assumption B5) Second-order sufficient conditions: Let C be the cone of critical directions*

$$C = \left\{w : w^\top \nabla g_i(\beta_0) = 0, i \in K; w^\top \nabla g_i(\beta_0) \leq 0, i \in J; w^\top \nabla \phi(\beta_0) \leq 0\right\}. \tag{18}$$

The assumption requires that for all nonzero $w \in C$, $\max_{\lambda \in \Lambda_0} w^\top \nabla^2 l(\beta_0, \lambda) w > 0$,

8.c Stochastic equicontinuity, a modified version of Assumption C1 in the original paper. For any sequence $\delta_t = o(1)$, the variable

$$\sup_{\beta : \|\beta - \beta_0\| \leq \delta_t} \frac{\|(\nabla \psi_t - \nabla \phi)(\beta) - (\nabla \psi_t - \nabla \phi)(\beta_0)\|}{t^{-1/2} + \|\beta - \beta_0\|} = o_p(1) \tag{19}$$

as $t \to \infty$. Here the supremum is taken over β such that $\nabla \psi_t(\beta)$ exists.

Then it holds that $\hat{\beta} \xrightarrow{p} \beta_0$. Let

$$\zeta_t = \nabla \psi_t(\beta_0) - \nabla \phi(\beta_0), \tag{20}$$

and

$$q(w) = \max_{\lambda \in \Lambda_0} \{w^\top \nabla^2 l(\beta_0, \lambda) w\}. \tag{21}$$

Then

$$\vartheta_t \equiv \inf_B \psi_t = \psi_t(\beta_0) + \min_{w \in C} \{w^\top \zeta_t + \frac{1}{2} q(w)\} + o_p(t^{-1}).$$

Furthermore, suppose for all ζ the function $w \mapsto w^\top \zeta + \frac{1}{2} q(w)$ has a unique minimizer $\bar{\omega}(\zeta)$ over C. Then

$$\|\hat{\beta} - \beta_0 - \bar{\omega}(\zeta_t)\| = o_p(t^{-1/2}).$$

Remark 1 (The stochastic equicontinuity condition)**.** *By inspecting the proof, the original Assumption C1, $\sup_{\beta \in B \cap N} \|\nabla \psi_t(\beta) - \nabla \phi(\beta) - \nabla \psi_t(\beta_0) + \nabla \phi(\beta_0)\| / [t^{-1/2} + \|\beta - \beta_0\|] = o_p(1)$, which requires uniform convergence over a fixed neighborhood N, can be relaxed to the uniform convergence in a shrinking neighborhood of β_0. The shrinking neighborhood condition is in fact standard, see, e.g., Pakes & Pollard (1989); Newey & McFadden (1994).*

Remark 2. *The limit distribution of the minimizer is characterized by three objects: the limit distribution of ζ_t defined in Eq. (20), the critical cone C defined in Eq. (18) and the piecewise quadratic function q defined in Eq. (21).*

Hessian matrix estimation at the optimum β_0 can be done via the numerical difference method.

Lemma 9 (Hessian estimation via numerical difference, adapted from Theorem 7.4 from Newey & McFadden (1994))**.** *Recall $\phi(\beta) = Pf(\cdot, \beta)$, $\psi_t(\beta) = P_t f(\cdot, \beta)$ and $\zeta_t = \nabla \psi_t(\beta_0) - \nabla \phi(\beta_0)$. We are interested in the Hessian matrix $H = \nabla^2 \phi(\beta_0)$. Let β_0 be any point and let $\hat{\beta}$ be an estimate of β_0. Assume*

9.a $\hat{\beta} - \beta_0 = O_p(t^{-1/2})$;

9.b ϕ *is twice differentiable at* β_0 *with non-singular Hessian matrix* H;

9.c $\sqrt{t}\zeta_t \xrightarrow{d} N(0, \Omega)$ *for some matrix* Ω;

9.d for any positive sequence $\delta_t = o(1)$, *the stochastic equicontinuity condition Eq.* (19) *holds.*

Suppose $\varepsilon_t \to 0$ *and* $\varepsilon_t\sqrt{t} \to \infty$. *Then* $\hat{H} \xrightarrow{p} H$, *where* $\hat{H}$ *is the numerical difference estimator whose* (i, j)*-th entry is*

$$\begin{aligned}\hat{H}_{ij} =&[\psi_t(\hat{\beta} + e_i\varepsilon_t + e_j\varepsilon_t) - \psi_t(\hat{\beta} - e_i\varepsilon_t + e_j\varepsilon_t) - \psi_t(\hat{\beta} + e_i\varepsilon_t - e_j\varepsilon_t).\\ &+ \psi_t(\hat{\beta} - e_i\varepsilon_t - e_j\varepsilon_t)]/4\varepsilon_t^2.\end{aligned}$$

Proof of Lemma 9. We provide a proof sketch following Theorem 7.4 from Newey & McFadden (1994) and Lemma 3.3 in Shapiro (1989). By Cond. 9.a and $t^{-1/2} = o(\varepsilon_t)$ we know for any vector $a \in \mathbb{R}^n$, it holds $\|\hat{\beta} + \varepsilon_t a - \beta_0\| = O_p(\varepsilon_t)$. Let $\beta = \hat{\beta} + a\varepsilon_t$. By a mean value theorem for locally Lipschitz functions (see Clarke (1990); the lemma is also used in the proof of Lemma 3.3 in Shapiro (1989)), there is a (sample-path dependent) β' on the segment joining β and β_0 such that

$$(\psi_t - \phi)(\beta) - (\psi_t - \phi)(\beta_0) = (\zeta_t^*)^\top(\beta - \beta_0).$$

for some $\zeta_t^* \in \partial\psi_t(\beta') - \nabla\phi(\beta')$. Then

$$\begin{aligned}|(\psi_t - \phi)(\beta) - (\psi_t - \phi)(\beta_0)| &\le \|\zeta_t\|\|\beta - \beta_0\| + \|\zeta_t^* - \zeta_t\|\|\beta - \beta_0\|\\ &= \|\zeta_t\|\|\beta - \beta_0\| + o_p(t^{-1/2} + \|\beta' - \beta_0\|)\|\beta - \beta_0\| && \text{(by Cond. 9.d)}\\ &= O_p(t^{-1/2})O_p(\varepsilon_t) + o_p(t^{-1/2} + O_p(\varepsilon_t))O_p(\varepsilon_t) && \text{(by Cond. 9.c)}\\ &= o_p(\varepsilon_t^2) && (22)\end{aligned}$$

Next by Cond. 9.b we have a quadratic expansion

$$\phi(\beta) - \phi(\beta_0) - \nabla\phi(\beta_0)^\top(\beta - \beta_0) - \frac{1}{2}(\beta - \beta_0)^\top H(\beta - \beta_0) = o_p(\varepsilon_t^2). \tag{23}$$

Let $a_{\pm\pm} = \pm e_i\varepsilon_t \pm e_j\varepsilon_t$, $\hat{\beta}_{\pm\pm} = \hat{\beta} + a_{\pm\pm}$ and $d_{\pm\pm} = \hat{\beta}_{\pm\pm} - \beta_0$. Then $d_{\pm\pm} = O_p(\varepsilon_t)$ and $d_{\pm\pm} = a_{\pm\pm} + o_p(\varepsilon_t)$. Applying the above bounds with $\beta \leftarrow \hat{\beta}_{\pm\pm}$, recalling the definition of $\hat{H}_{ij}$, we have

$$\begin{aligned}\hat{H}_{ij} &= [\psi_t(\hat{\beta}_{++}) - \psi_t(\hat{\beta}_{-+}) - \psi_t(\hat{\beta}_{+-}) + \psi_t(\hat{\beta}_{--})]/(4\varepsilon_t^2)\\ &= [\phi(\hat{\beta}_{++}) - \phi(\hat{\beta}_{-+}) - \phi(\hat{\beta}_{+-}) + \phi(\hat{\beta}_{--}) + o_p(\varepsilon_t^2)]/(4\varepsilon_t^2) && \text{(by Eq. (22))}\\ &= [\nabla\phi(\beta_0)^\top(d_{++} - d_{-+} - d_{+-} + d_{--})\\ &\quad + \frac{1}{2}(d_{++}^\top Hd_{++} - d_{+-}^\top Hd_{+-} - d_{-+}^\top Hd_{-+} + d_{--}^\top Hd_{--}) + o_p(\varepsilon_t^2)]/(4\varepsilon_t^2) && \text{(by Eq. (23))}\\ &= [0 + \frac{1}{2}(a_{++}^\top Ha_{++} - a_{-+}^\top Ha_{-+} - a_{+-}^\top Ha_{+-} + a_{--}^\top Ha_{--}) + o_p(\varepsilon_t^2)]/(4\varepsilon_t^2)\\ &= [4\varepsilon_t^2 H_{ij} + o_p(\varepsilon_t^2)]/(4\varepsilon_t^2)\\ &= H_{ij} + \frac{o_p(\varepsilon_t^2)}{\varepsilon_t^2} = H_{ij} + o_p(1).\end{aligned}$$

In the above we use $d_{++}^\top Hd_{++} = (d_{++} - a_{++})^\top Hd_{++} + (d_{++} - a_{++})^\top Ha_{++} + a_{++}^\top Ha_{++} = o_p(\varepsilon_t^2) + a_{++}^\top Ha_{++}$, and similarly for other terms. This completes the proof of Lemma 9. □

The original conditions for Lemma 9 in Newey & McFadden (1994) require the true parameter β_0 to lie in the interior of B. However, this condition is only used to derive the bound $\hat{\beta} - \beta_0 = O_p(t^{-1/2})$, which is assumed in our adapted version.

E.2. Showing stochastic equicontinuity via VC-subgraph function classes

Next we review classical results from the empirical process literature (Vaart & Wellner, 1996; Giné & Nickl, 2021).

We begin with the notions of Donsker function class and stochastic equicontinuity.

Let (Θ, P) be a probability space. Let $\mathcal{F}$ be a class of measurable functions of finite second moment. The class $\mathcal{F}$ is called P-Donsker if a certain central limit theorem holds for the class of random variables $\{\sqrt{t}(P_t - P)f : f \in \mathcal{F}\}$, where $P_t f = \frac{1}{t}\sum_{i=1}^{t} f(X_i)$ where X_i's are i.i.d. draws from P. Because Donskerness will be used as an intermediate step that we will not actually need to show directly or utilize directly, we refer the reader to Definition 3.7.29 from Giné & Nickl (2021) for a precise definition.

Lemma 10 (Donskerness $\Leftrightarrow$ stochastic equicontinuity, Theorem 3.7.31 from Giné & Nickl (2021))**.** *Let $d_P^2(f,g) = P(f-g)^2 - (P(f-g))^2$ and consider the pseudo-metric space $(\mathcal{F}, d_P)$. Assume $\mathcal{F}$ satisfies the condition $\sup_{f\in\mathcal{F}} |f(x) - Pf| < \infty$ for all $x \in \Theta$. Then the following are equivalent*

- *$\mathcal{F}$ is P-Donsker.*
- *$(\mathcal{F}, d_P)$ is totally bounded, and stochastic equicontinuity under the L2 function norm holds, i.e., for any $\delta_t = o(1)$,*

$$\sup_{(f,g)\in[\delta_t]_{L^2}} |\sqrt{t}\,(P_t - P)\,(f-g)| = o_p(1)$$

as $t \to \infty$, where $[\delta_t]_{L^2} = \{(f,g) : f, g \in \mathcal{F}, d_P(f,g) \le \delta_t\}$.

Lemma 10 reduces the problem of showing stochastic equicontinuity under the L2 function norm to showing Donskerness. In order to show Donskerness, we will show that our function class is *VC-subgraph*, which implies Donskerness. At the end, we will connect stochastic equicontinuity under the L2 function norm to the stochastic equicontinuity that we need (see Lemma 16).

Let $\mathcal{C}$ be a class of subsets of a set Θ. Let $A \subseteq \Theta$ be a finite set. We say that $\mathcal{C}$ *shatters* A if every subset of A is the intersection of A with some set $C \in \mathcal{C}$. The subgraph of a real function f on Θ is the set $G_f = \{(s,t) : s \in \Theta, t \in \mathbb{R}, t \le f(s)\}$.

Definition 4 (VC-subgraph function classes, Definition 3.6.1 and 3.6.8 from Giné & Nickl (2021))**.** *A collection of sets $\mathcal{C}$ is a Vapnik-Červonenkis class ($\mathcal{C}$ is VC) if there exists $k < \infty$ such that $\mathcal{C}$ does not shatter any subsets of Θ of cardinality k. A class of functions $\mathcal{F}$ is VC-subgraph if the class of sets $\mathcal{C} = \{G_f : f \in \mathcal{F}\}$ is VC.*

Lemma 11 (VC subgraph + envelop square integrability $\implies$ Donskerness, Theorem 3.7.37 from Giné & Nickl (2021))**.** *If $\mathcal{F}$ is VC-subgraph, and there exists a measurable F such that $f \le F$ for all $f \in \mathcal{F}$ with $PF^2 < \infty$, then $\mathcal{F}$ is P-Donsker.*

Since VC-subgraph implies Donskerness which is equivalent to stochastic equicontinuity, our problem reduces to showing the VC-subgraph property. The following lemmas show how to construct complex VC-subgraph function classes from simpler ones, and will be used in our proof.

Lemma 12 (Preservation of VC class of sets, Lemma 2.6.17 from Vaart & Wellner (1996))**.** *If $\mathcal{C}$ and $\mathcal{D}$ are VC classes of sets. Then $\mathcal{C} \cap \mathcal{D} = \{C \cap D : C \in \mathcal{C}, D \in \mathcal{D}\}$ is VC.*

Lemma 13 (Preservation of VC-subgraph function classes, Lemma 2.6.18 from Vaart & Wellner (1996))**.** *Let $\mathcal{F}$ and $\mathcal{G}$ be VC-subgraph classes of functions on a set Θ and $g : \Theta \mapsto \mathbb{R}$ be a fixed function. Then $\mathcal{F} \vee \mathcal{G} = \{f \vee g : f \in \mathcal{F}, g \in \mathcal{G}\}$, $\mathcal{F} + g = \{f + g : f \in \mathcal{F}\}$, $\mathcal{F} \circ \phi = \{f \circ \phi : f \in \mathcal{F}\}$ is VC-subgraph for fixed $\phi : \mathcal{X} \to \Theta$, and $\mathcal{F} \cdot g = \{fg : f \in \mathcal{F}\}$ are VC-subgraph;*

Lemma 14 (Problem 9 Section 2.6 from Vaart & Wellner (1996))**.** *If a collection of sets $\mathcal{C}$ is a VC-class, then the collection of indicators of sets in $\mathcal{C}$ is a VC-subgraph class of the same index.*

In general, the VC-subgraph property is not preserved by multiplication, whereas Donskerness is. Thus, our proof will use the VC-subgraph property up until a final step where we need to invoke multiplication, which will instead be applied on the Donskerness property.

Lemma 15 (Corollary 9.32 from Kosorok (2008))**.** *Let $\mathcal{F}$ and $\mathcal{G}$ be Donsker, then $\mathcal{F} \cdot \mathcal{G}$ is Donsker if both $\mathcal{F}$ and $\mathcal{G}$ are uniformly bounded.*

For parametric function classes, if the parametrization is continuous in a certain sense, then stochastic equicontinuity holds w.r.t. the norm in the parameter space.

Lemma 16 (From L^2-norm to parameter norm, Lemma 2.17 from Pakes & Pollard (1989); see also Lemma 1 from Chen et al. (2003))**.** *Suppose the function class $\mathcal{F} = \{f(\cdot, \beta), \beta \in B\}$, $B \subset \mathbb{R}^n$, is P-Donsker, with an envelope F such that $PF^2 < \infty$. Suppose $\int [f(\cdot, \beta) - f(\cdot, \beta_0)]^2 \, dP \to 0$ as $\beta \to \beta_0$. Then for any positive sequence $\delta_t = o(1)$, it holds*

$$\sup_{\beta : \|\beta - \beta_0\| < \delta_t} |\sqrt{t} (P_t - P) (f(\cdot, \beta) - f(\cdot, \beta_0))| = o_p(1). \tag{24}$$

Lemma 17 (Andrews (1994))**.** *If for any $\delta_t = o(1)$ Eq. (24) holds, then for any random elements β_t such that $\|\beta_t - \beta_0\|_2 = o_p(1)$, it holds $\sqrt{t}(P_t - P)(f(\cdot, \beta_t) - f(\cdot, \beta_0)) = o_p(1)$.*

Lemma 18. *Given any n fixed functions $v_i : \Theta \to \mathbb{R}$, $i \in [n]$, the following two function classes*

$$\mathcal{F}_1 = \{\theta \mapsto \max_i \{\beta_1 v_1(\theta), \ldots, \beta_n v_n(\theta)\} : \beta \in B\}$$
$$\mathcal{F}_2 = \{\theta \mapsto [g_1, \ldots, g_n](\theta, \beta) : \beta \in B\}$$

are VC-subgraph and Donsker. Here $g_i(\theta, \beta) = v_i(\theta) \prod_{k=1}^n \mathbb{1}(\beta_i v_i(\theta) \geq \beta_k v_k(\theta))$ and $B = [0, 1]^n$

Proof of Lemma 18. We show $\mathcal{F}_1$ is VC-subgraph. For each i, the class $\{\theta \mapsto v_i(\theta)\beta_i : \beta_i \in [0, 1]\}$ is VC-subgraph (Proposition 4.20 from Wainwright (2019), and Example 19.17 from Van der Vaart (2000)). By the fact the VC-subgraph function classes are preserved by pairwise maximum (Lemma 13), we know $\mathcal{F}_1$ is VC-subgraph. Moreover, the required envelop condition holds since $\operatorname{ess\,sup}_\theta f \leq \bar{v}$ for all $f \in \mathcal{F}_1$, so $\mathcal{F}_1$ is Donsker by Lemma 11.

We now show $\mathcal{F}_2$ is VC-subgraph. For a vector-valued function class, we say it is VC-subgraph if each coordinate is VC-subgraph. First, the class of sets $\{\{v \in \mathbb{R}^n : \beta_i v_i \geq \beta_k v_k\} \subset \mathbb{R}^n : \beta \in B\}$ is VC, for all $k \neq i$. By Lemma 12, we know the class of sets $\{\{v \in \mathbb{R}^n : \beta_i v_i \geq \beta_k v_k, \forall k \neq i\} \subset \mathbb{R}^n : \beta \in B\}$ is VC. By Lemma 14, we obtain that the class $\{\theta \mapsto \prod_{k=1}^n \mathbb{1}(\beta_i v_i(\theta) \geq \beta_k v_k(\theta)) : \beta \in B\}$ is VC-subgraph. Finally, multiplying all functions by a fixed function preserves VC-subgraph classes (Lemma 13), and so $\{\theta \mapsto v_i(\theta) \prod_{k=1}^n \mathbb{1}(\beta_i v_i(\theta) \geq \beta_k v_k(\theta)) : \beta \in B\}$ is VC-subgraph. Repeat the argument for each coordinate, and we obtain that $\mathcal{F}_2$ is VC-subgraph. Moreover, the required envelop condition holds since $\operatorname{ess\,sup}_\theta \|g(\theta, \beta)\|_2 \leq n\bar{v}$ for all $g \in \mathcal{F}_2$, and so $\mathcal{F}_2$ is Donsker by Lemma 11. We conclude the proof of Lemma 18. □

F. Proofs for Main Theorems

F.1. Proof of Thm. 1

Proof of Thm. 1. We verify all the conditions in Lemma 8. Recall $I = \{i : \beta_i^* = 1\}$ is the set of active constraints. The local geometry of B at β^* is described by the $|I|$ constraint functions $g_i(\beta) = e_i^\top \beta - 1$, $i \in I$.

First, we verify the conditions on the probability distribution and the sample function. A1 holds obviously for the map $\beta \mapsto \max_i \beta_i v_i(\theta)$. A2 holds by $f \leq \bar{v}$. A4 holds with Lipschitz constant $\bar{v}$. A5 holds since by SMO there is a neighborhood N_{diff} of β^* such that for all $\beta \in N_{\text{diff}}$, the set $\{\theta : f(\theta, \cdot) \text{ not differentiable at } \beta\}$ is measure zero (cf. Lemma 1). A6 holds by $\|\nabla f(\theta, \beta)\|_2 \leq n\bar{v}$. B4 holds by SMO.

Second, we verify the conditions on the optimality. B3 holds since the constraint functions are $g_i(\beta) = e_i^\top \beta - 1$, $i \in I$, whose gradient vectors are obviously linear independent. Moreover, the set $\{\beta : \beta_i > 0, i \in I\}$ is nonempty. B5 holds by $\nabla^2 H(\beta^*) = \nabla^2 \bar{f}(\beta^*) + \operatorname{diag}(b_i / {\beta_i^*}^2) \succcurlyeq \operatorname{diag}(b_i / {\beta_i^*}^2)$ being positive definite.

Finally, we verify the stochastic equicontinuity condition. Recall the definitions of the following two function classes from Lemma 18

$$\mathcal{F}_1 = \{\theta \mapsto \max_i \{\beta_1 v_1(\theta), \ldots, \beta_n v_n(\theta)\} : \beta \in B\},$$
$$\mathcal{F}_2 = \{\theta \mapsto g(\theta, \beta) = [g_1, \ldots, g_n](\theta, \beta) : \beta \in B\}.$$

Here $g_i(\theta, \beta) = v_i(\theta) \prod_{k=1}^n \mathbb{1}(\beta_i v_i(\theta) \geq \beta_k v_k(\theta))$ and $B = [0, 1]^n$. For any $\beta \in N_{\text{diff}}$ we have $\nabla f(\cdot, \beta) = g(\cdot, \beta) \in \mathcal{F}_2$. In Lemma 18 we show that $\mathcal{F}_2$ is VC-subgraph and Donsker. By Lemma 11 we know that a stochastic equicontinuity

condition w.r.t. the L^2 norm holds, i.e.,

$$\sup_{\beta \in [\delta_t]_{L^2}} \nu_t(g(\cdot,\beta) - g(\cdot,\beta^*)) = o_p(1) \tag{25}$$

where $[\delta_t]_{L^2} = \{\beta : \beta \in N_{\text{diff}}, \int \|g(\cdot,\beta) - g(\cdot,\beta^*)\|_2^2 \,\mathrm{d}S \le \delta_t\}$, $\int g \,\mathrm{d}S = \int gs \,\mathrm{d}\theta$, $\nu_t g = t^{-1/2}(P_t - P)g = t^{-1/2}\sum_{\tau=1}^t (g(\theta^\tau) - \int g \,\mathrm{d}S)$. Next, we note for (almost every) fixed θ, $\lim_{\beta\to\beta^*} \|g(\theta,\beta) - g(\theta,\beta^*)\|^2 = 0$ by $\Theta_{\text{tie}}(\beta^*)$ is measure zero (a condition implied by SMO). Moreover, note

$$\lim_{\beta\to\beta^*} \mathbb{E}[\|g(\theta,\beta) - g(\theta,\beta^*)\|_2^2] = \mathbb{E}\big[\lim_{\beta\to\beta^*} \|g(\theta,\beta) - g(\theta,\beta^*)\|_2^2\big] = 0$$

where the exchange of limit and expectation is justified by bounded convergence theorem, and by Lemma 16, we can replace $[\delta_t]_{L^2}$ with $[\delta_t] = \{\beta : \beta \in N_{\text{diff}}, \|\beta - \beta^*\|_2 \le \delta_t\}$ in Eq. (25). Finally, note $\nabla \bar{f}(\beta^*) = \mathbb{E}[g(\theta,\beta^*)]$, and if H_t is differentiable at $\beta \in N_{\text{diff}}$, then $\nabla f(\theta^\tau,\beta) = g(\theta^\tau,\beta)$ for all $\tau \in [t]$. Then

$$\begin{aligned}
&\sup_{[\delta_t]\cap\{\nabla H_t(\beta) \text{ exists}\}} \frac{\|(\nabla H_t - \nabla H)(\beta) - (\nabla H_t - \nabla H)(\beta^*)\|_2}{t^{-1/2} + \|\beta - \beta^*\|_2} \\
&= \sup_{[\delta_t]\cap\{\nabla H_t(\beta) \text{ exists}\}} \frac{\|(P_t - P)g(\cdot,\beta) - (P_t - P)g(\cdot,\beta^*)\|_2}{t^{-1/2} + \|\beta - \beta^*\|_2} \qquad (26)\\
&\le \sup_{[\delta_t]} \sqrt{t}\|(P_t - P)g(\cdot,\beta) - (P_t - P)g(\cdot,\beta^*)\|_2 = o_p(1) \qquad \text{(by Eq. (25))}
\end{aligned}$$

and thus the required stochastic equicontinuity condition holds.

Now we are ready to invoke Lemma 8. We need to find the three objects, C, q, ζ_t as in the lemma that characterize the limit distribution. The critical cone C is

$$\begin{aligned}
C &= \{w \in \mathbb{R}^n : w^\top e_i = 0 \text{ if } i \in I \text{ and } \delta_i^* > 0, \quad w^\top e_i \le 0 \text{ if } i \in I \text{ and } \delta_i^* = 0\} \\
&= \{w : Aw = 0\} \qquad \text{(SCS)}
\end{aligned}$$

where $A \in \mathbb{R}^{|I|\times n}$ whose rows are $\{e_i^\top, i \in I\}$. From here we can see the role of SCS is to ensure the critical cone is a hyperplane, which ensures asymptotic normality of β^γ.

If $|I| = 0$, i.e., β^* lies in the interior of B, then $\mathcal{P}$ is identity matrix, and the limit distribution is ture.

Now assume $|I| \ge 1$. Note $AA^\top$ is an identity matrix of size $|I|$ and $A^\top A = \mathrm{diag}(\mathbb{1}(i \in I)) = \mathrm{diag}(\mathbb{1}(\beta_i^* = 1))$. The optimal Lagrangian multiplier is unique and so the piecewise quadratic function q is $q(w) = w^\top \mathcal{H} w$. Finally, the gradient error term is

$$\zeta_t = \frac{1}{t}\sum_{\tau=1}^t \big(x^*(\theta^\tau) \odot v(\theta^\tau) - \bar{\mu}^*\big). \tag{27}$$

The unique minimizer of $w \mapsto \frac{1}{2}w^\top \mathcal{H} w + \zeta w$ over $\{w : Aw = 0\}$ is $-(\mathcal{P}\mathcal{H}\mathcal{P})^\dagger \zeta$ where

$$\mathcal{P} = I_n - A^\top (AA^\top)^\dagger A = \mathrm{diag}(\mathbb{1}(i \in I^c)) = \mathrm{diag}(\mathbb{1}(\beta_i^* < 1)).$$

For completeness, we provide details for solving this quadratic problem. By writing down the KKT conditions, the optimality condition is

$$\begin{bmatrix} \mathcal{H} & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} w \\ \lambda \end{bmatrix} = \begin{bmatrix} -\zeta \\ 0 \end{bmatrix} \implies \begin{bmatrix} w \\ \lambda \end{bmatrix} = \begin{bmatrix} -(\mathcal{H}^{-1} - \mathcal{H}^{-1}A^\top(A\mathcal{H}^{-1}A^\top)^{-1}A\mathcal{H}^{-1})\zeta \\ -((A\mathcal{H}^{-1}A^\top)^{-1}A\mathcal{H}^{-1})\zeta \end{bmatrix}$$

where $\lambda \in \mathbb{R}^{|I|}$ is the Lagrangian multiplier. By a matrix equality, for any symmetric positive definite $\mathcal{H}$ of size n and $A \in \mathbb{R}^{|I|\times n}$ of rank $|I|$, it holds

$$\mathcal{H}^{-1} - \mathcal{H}^{-1}A^\top(A\mathcal{H}^{-1}A^\top)^{-1}A\mathcal{H}^{-1} = P_A(P_A\mathcal{H}P_A)^\dagger P_A = (P_A\mathcal{H}P_A)^\dagger \tag{28}$$

with $P_A = I_n - A^\top(AA^\top)^\dagger A$. We conclude that the asymptotic expansion

$$\sqrt{t}(\beta^\gamma - \beta^*) = \frac{1}{\sqrt{t}}\sum_{\tau=1}^{t} D_\beta(\theta^\tau) + o_p(1) \tag{29}$$

holds, where

$$D_\beta(\theta) = -(\mathcal{P}\mathcal{H}\mathcal{P})^\dagger(x^*(\theta)\odot v(\theta) - \bar{\mu}^*),$$

and that the asymptotic distribution of $\sqrt{t}(\beta^\gamma - \beta^*)$ is $\mathcal{N}(0, \Sigma_\beta)$ with $\Sigma_\beta = \mathbb{E}[D_\beta D_\beta^\top]$. Note $\mathbb{E}[D_\beta] = 0$.

One could further write out the expression. Assume $I = \{1, \dots, k\}$. Let $\Omega \equiv \mathrm{Cov}[\mu^*] = \mathrm{diag}(\mathrm{Var}[x_i^* v_i])$. Then the matrix $(\mathcal{P}\mathcal{H}\mathcal{P})^\dagger = \mathrm{diag}(0_{k\times k}, (\mathcal{H}_{I^cI^c})^{-1})$ and $\mathrm{Var}[x^*(\theta)\odot v(\theta) - \bar{\mu}^*] = \mathbb{E}[(x^*(\theta)\odot v(\theta) - \bar{\mu}^*)^{\otimes 2}] = \mathrm{diag}((\Omega_i^*)^2)$ where $(\Omega_i^*)^2 = \int (x_i^* v_i)^2 \,\mathrm{d}S - (\int x_i^* v_i \,\mathrm{d}S)^2$ is the variance of winned item value of buyer i.

Proof of $\beta^\gamma \xrightarrow{p} \beta^*$. This follows from Lemma 8.

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

Proof of CLT for revenue REV. We use a stochastic equicontinuity argument. Given the item sequence $\gamma = (\theta^1, \theta^2 \dots)$, define the (random) operator

$$\nu_t g = \sqrt{t}(P_t - P)g = \frac{1}{\sqrt{t}}\sum_{\tau=1}^{t}(g(\theta^\tau) - \mathbb{E}[g]) \ .$$

where $g : \Theta \to \mathbb{R}$, $\mathbb{E}[g] = \int g \,\mathrm{d}S$. Note $p^*(\theta) = \max_i \beta_i^* v_i(\theta) = f(\theta, \beta^*)$, $\mathrm{REV}^* = Pf(\cdot, \beta^*)$, $p^\gamma(\theta^\tau) = f(\theta^\tau, \beta^\gamma)$ and $\mathrm{REV}^\gamma = P_t f(\cdot, \beta^\gamma)$ we obtain the decomposition

$$\sqrt{t}(\mathrm{REV}^\gamma - \mathrm{REV}^*) = \underbrace{\frac{1}{\sqrt{t}}\sum_{\tau=1}^{t}\big(f(\theta^\tau, \beta^*) - \bar{f}(\beta^*)\big)}_{=:\mathrm{I}_t} + \underbrace{\nu_t(f(\cdot, \beta^\gamma) - f(\cdot, \beta^*))}_{=:\mathrm{II}_t} + \underbrace{\sqrt{t}(\bar{f}(\beta^\gamma) - \bar{f}(\beta^*))}_{=:\mathrm{III}_t}$$

For the term I_t, it can be written as $\mathrm{I}_t = \nu_t(p^*(\cdot) - \mathrm{REV}^*)$. By the linear representation for $\beta^\gamma - \beta^*$ in Eq. (29), applying the delta method, we get the linear representation result

$$\mathrm{III}_t = \frac{1}{\sqrt{t}}\sum_{\tau=1}^{t}\nabla\bar{f}(\beta^*)^\top D_\beta(\theta^\tau) + o_p(1) = \frac{1}{\sqrt{t}}\sum_{\tau=1}^{t}(\bar{\mu}^*)^\top D_\beta(\theta^\tau) + o_p(1)$$

We will show $\mathrm{II}_t = o_p(1)$. The difficulty lies in that the operator ν_t and the pacing multiplier β^γ depend on the same batch of items. This can be handled with the stochastic equicontinuity argument. The desired claim $\mathrm{II}_t = o_p(1)$ follows by verifying that the function class $\mathcal{F}_1 = \{\theta \mapsto f(\cdot, \beta) : \beta \in B\}$ (same as that defined in Lemma 18) is VC-subgraph and Donsker. This is true by Lemma 18. By Lemma 10 we know for any $\delta_t \downarrow 0$,

$$\sup_{w\in[\delta_t]_{L^2}} \nu_t(f(\cdot, w) - f(\cdot, \beta^*)) = o_p(1) \tag{30}$$

where $[\delta_t]_{L^2} = \{\beta : \beta \in B, \int (f(\cdot, \beta) - f(\cdot, \beta^*))^2 \,\mathrm{d}S \le \delta_t\}$. Noting that for all β, w, θ, it holds $|f(\theta, \beta) - f(\theta, w)| \le \bar{v}\|\beta - w\|_\infty$, we know that $\int [f(\cdot, \beta) - f(\cdot, \beta^*)]^2 \,\mathrm{d}S \to 0$ as $\beta \to \beta^*$. Then by Lemma 16, we know Eq. (30) holds with $[\delta_t]_{L^2}$ replaced with $[\delta_t] = \{\beta : \beta \in B, \|\beta - \beta^*\|_2 \le \delta_t\}$. Combined with the fact that $\beta^\gamma \xrightarrow{p} \beta^*$, by Lemma 17 we know $\mathrm{II}_t = o_p(1)$.

To summarize, we obtain the linear expansion

$$\sqrt{t}(\mathrm{REV}^\gamma - \mathrm{REV}^*) = \frac{1}{\sqrt{t}}\sum_{\tau=1}^{t}(p^*(\theta^\tau) - \mathrm{REV}^* + (\bar{\mu}^*)^\top D_\beta(\theta^\tau)) + o_p(1). \tag{31}$$

We complete the proof of Thm. 1.

□

F.2. Proof of Corollary 1

Full statement of Corollary 1. Under the same conditions as Thm. 1, $\sqrt{t}(u^\gamma - u^*)$, $\sqrt{t}(\delta^\gamma - \delta^*)$ and $\sqrt{t}(\mathrm{NSW}^\gamma - \mathrm{NSW}^*)$ are asymptotically normal with (co)variances $\Sigma_u \equiv \mathrm{diag}(b_i/(\beta_i^*)^2)\Sigma_\beta \mathrm{diag}(b_i/(\beta_i^*)^2)$, $\Sigma_\delta \equiv (I_n - \mathcal{H}(\mathcal{H}_B)^\dagger)\Omega(I_n - \mathcal{H}(\mathcal{H}_B)^\dagger)^\top$, and $\sigma^2_{\mathrm{NSW}} \equiv \mathrm{Vec}(b_i/\beta_i^*)^\top \Sigma_\beta \mathrm{Vec}(b_i/\beta_i^*)$, respectively.

Proof of Corollary 1. Proof of CLT for individual utility u. We use the delta method; see Theorem 3.1 from Van der Vaart (2000). Note $u^* = g(\beta^*)$ with $g : \mathbb{R}^n \to \mathbb{R}^n, g(\beta) = [b_1/\beta_1, \ldots, b_n/\beta_n]^\top$. By Eq. (29), it holds

$$\sqrt{t}(u^\gamma - u^*) = \frac{1}{\sqrt{t}}\sum_{\tau=1}^{t} \nabla g(\beta^*) D_\beta(\theta^\tau) + o_p(1). \tag{32}$$

Finally, noting $\nabla g(\beta^*) = \mathrm{diag}(-b_i/(\beta_i^*)^2)$ we complete the proof.

Proof of CLT for Nash social welfare NSW. We use the delta method. Note $\mathrm{NSW}^* = g(\beta^*)$ with $g : \mathbb{R}^n \to \mathbb{R}, g(\beta) = \sum_{i=1}^n b_i \log(b_i/\beta_i)$. By Eq. (29) it holds

$$\sqrt{t}(\mathrm{NSW}^\gamma - \mathrm{NSW}^*) = \frac{1}{\sqrt{t}}\sum_{\tau=1}^{t} \nabla g(\beta^*)^\top D_\beta(\theta^\tau) + o_p(1). \tag{33}$$

Finally, noting $\nabla g(\beta^*) = \mathrm{Vec}(b_i/\beta_i^*)$ we complete the proof of Corollary 1.

Proof of CLT for leftover budget δ. This is a direct consequence of Theorem 4.1 in Shapiro (1989). By that theorem, it holds that

$$\sqrt{t}\begin{bmatrix}\beta^\gamma - \beta^* \\ \delta_I^\gamma - \delta_I^*\end{bmatrix} \xrightarrow{d} \mathcal{N}(0, \Sigma_{\text{joint}})$$

with

$$\Sigma_{\text{joint}} = \begin{bmatrix}\mathcal{H} & A^\top \\ A & 0\end{bmatrix}^{-1}\begin{bmatrix}\Omega & 0 \\ 0 & 0\end{bmatrix}\begin{bmatrix}\mathcal{H} & A^\top \\ A & 0\end{bmatrix}^{-1} = \begin{bmatrix}(\mathcal{H}_B)^\dagger\Omega(\mathcal{H}_B)^\dagger & [Q\Omega(\mathcal{H}_B)^\dagger]^\top \\ Q\Omega(\mathcal{H}_B)^\dagger & Q\Omega Q^\top\end{bmatrix}$$

where $Q = (A\mathcal{H}^{-1}A^\top)^{-1}A\mathcal{H}^{-1} \in \mathbb{R}^{|I|\times n}$ and $\Omega = \mathrm{Cov}[x^* \odot v - \bar{\mu}^*]$. By a matrix equality, noting matrix A's rows are distinct basis vectors, it holds

$$(A\mathcal{H}^{-1}A^\top)^{-1}A\mathcal{H}^{-1} = A(I_n - \mathcal{H}(\mathcal{H}_B)^\dagger)$$

Moreover, for other entries of δ^γ, i.e., $\delta^\gamma_{I^c}$, their asymptotic variance will be zero. The matrix $(I_n - \mathcal{H}(\mathcal{H}_B)^\dagger)\Omega(I_n - \mathcal{H}(\mathcal{H}_B)^\dagger)^\top$ is zero at the (i,j)-th entry if i or $j \in I^c$. Summarizing, the asymptotic variance of $\sqrt{t}(\delta^\gamma - \delta^*)$ is $(I_n - \mathcal{H}(\mathcal{H}_B)^\dagger)\Omega(I_n - \mathcal{H}(\mathcal{H}_B)^\dagger)^\top$.

An alternative proof is by the delta method and a stochastic equicontinuity argument. It holds $\sqrt{t}(\delta^\gamma - \delta^*) \xrightarrow{d} \mathcal{N}(0, (I_n - \mathcal{H}(\mathcal{H}_B)^\dagger)\Omega(I_n - \mathcal{H}(\mathcal{H}_B)^\dagger)^\top)$ and the linear expansion

$$\sqrt{t}(\delta^\gamma - \delta^*) = t^{-1/2}\sum_{\tau=1}^{t}(I_n - \mathcal{H}(\mathcal{H}_B)^\dagger)(\mu^*(\theta^\tau) - \bar{\mu}^*) + o_p(1) \tag{34}$$

holds. In the case where $I = \emptyset$, i.e., $\delta_i^* = 0$ for all i, we have $\mathcal{H}_B = \mathcal{H}$ and so $I_n - \mathcal{H}(\mathcal{H}_B)^\dagger = 0$. □

F.3. Proof of Thm. 2

Proof of Thm. 2. Based on Le Cam's local asymptotic normality theory (Le Cam et al., 2000), to establish the local asymptotic minimax optimality of a statistical procedure, one needs to verify two things. First, the class of perturbed distributions (the class $\{s_{\alpha,g}\}_{\alpha,g}$ in our case) satisfies the locally asymptotically normal (LAN) condition (Vaart & Wellner, 1996; Le Cam et al., 2000). This part is completed by Lemma 8.3 from Duchi & Ruan (2021) since our construction of perturbed supply distributions follows theirs. Second, one should verify the asymptotic variance of the statistical procedure equals to the minimax optimal variance. To calculate this quantity, one needs to find the derivative of the map $\alpha \mapsto \mathrm{REV}^*_{\alpha,g}$ at $\alpha = 0$. For this part we present a formal derivation below.

For a given perturbation (α, g), we let $p^*_{\alpha,g}$ and $\mathrm{REV}^*_{\alpha,g}$ be the limit FPPE price and revenue under supply distribution $s_{\alpha,g}$. Let $S_{\alpha,g}(\theta) = \nabla_\alpha \log s_{\alpha,g}(\theta)$ be the score function. Obviously with our parametrization of $s_{\alpha,g}$ we have $S_{0,g}(\theta) = g(\theta)$ by Eq. (8). We next find the derivative of $\alpha \mapsto \mathrm{REV}^*_{\alpha,g}$ at $\alpha = 0$. Recall f is defined in Eq. (P-DualEG) and the price is produced by the highest bid, i.e., $p^*_{\alpha,g}(\theta) = \max_i \beta^*_{\alpha,g} v_i(\theta) = f(\theta, \beta^*_{\alpha,g})$.

$$\begin{aligned}\nabla_\alpha \mathrm{REV}^*_{\alpha,g} &= \nabla_\alpha \int f(\theta, \beta^*_{\alpha,g}) s_{\alpha,g}(\theta)\,\mathrm{d}\theta \\ &= \int [\nabla_\beta f(\theta, \beta^*_{\alpha,g}) \nabla_\alpha \beta^*_{\alpha,g} + f(\theta, \beta^*_{\alpha,g}) S_{\alpha,g}(\theta)] s_{\alpha,g}(\theta)\,\mathrm{d}\theta\ .\end{aligned}$$

Above we exchange the gradient and the expectation and then apply the chain rule. By a perturbation result by Lemma 8.1 and Prop. 1 from Duchi & Ruan (2021), under SMO and SCS,

$$\nabla_\alpha \beta^*_{\alpha,g}|_{\alpha=0} = -(\mathcal{H}_B)^\dagger \Sigma_{\mu^*,g}$$

with $\Sigma_{\mu^*,g} = \mathbb{E}_s[(\mu^*(\theta) - \bar{\mu}^*) g(\theta)^\top]$. Plugging in $\mathbb{E}_s[\nabla_\beta f(\theta, \beta^*_{0,g})] = \bar{\mu}^*$, $f(\theta, \beta^*_{0,g}) = p^*(\theta)$ (see App. C) and $S_{0,g} = g$, we obtain

$$\begin{aligned}\nabla_\alpha \mathrm{REV}^*_{\alpha,g}|_{\alpha=0} &= -(\bar{\mu}^*)^\top (\mathcal{H}_B)^\dagger \Sigma_{\mu^*,g} + \mathbb{E}_s[(p^*(\theta) - \mathrm{REV}^*) g(\theta)] \\ &= \mathbb{E}\big[\big(-(\bar{\mu}^*)^\top (\mathcal{H}_B)^\dagger (\mu^*(\theta) - \bar{\mu}^*) + (p^*(\theta) - \mathrm{REV}^*)\big) g(\theta)\big] \\ &= \mathbb{E}[D_{\mathrm{REV}}(\theta) g(\theta)].\end{aligned}$$

Now we have the two components required to invoke the local minimax result. Given a symmetric quasi-convex loss $L : \mathbb{R} \to \mathbb{R}$, we recall the local asymptotic risk for any procedure $\{\hat{r}_t : \Theta^t \to \mathbb{R}\}$ that aims to estimate the revenue:

$$\begin{aligned}&\mathrm{LAR}_{\mathrm{REV}}(\{\hat{r}_t\}) = \\ &\sup_{g \in G_d, d \in \mathbb{N}} \lim_{c \to \infty} \liminf_{t \to \infty} \sup_{\|\alpha\|_2 \le \frac{c}{\sqrt{t}}} \mathbb{E}_{s^{\otimes t}_{\alpha,g}}[L(\sqrt{t}(\hat{r}_t - \mathrm{REV}^*_{\alpha,g}))]\ .\end{aligned}$$

Following the argument in Duchi & Ruan (2021, Sec. 8.3) it holds

$$\inf_{\{\hat{r}_t\}} \mathrm{LAR}_{\mathrm{REV}}(\{\hat{r}_t\}) \ge \mathbb{E}[L(\mathcal{N}(0, \mathbb{E}[D^2_{\mathrm{REV}}(\theta)]))]\ .$$

We complete the proof of Thm. 2.

□

F.4. Proof of Thm. 3

Proof of Thm. 3. We first show $\hat{\mathcal{H}} \xrightarrow{p} \mathcal{H}$ by verifying conditions in Lemma 9. All conditions are easy to verify except the stochastic equicontinuity condition. By Lemma 18 we know the SE condition holds. We conclude $\hat{\mathcal{H}} \xrightarrow{p} \mathcal{H}$.

Next we show $\mathbb{P}(\hat{\mathcal{P}} = \mathcal{P}) \to 1$. Recall $\mathcal{P} = \mathrm{diag}(\mathbb{1}(\beta^*_i < 1))$ indicates the set of inactive constraints. For i such that $\beta^*_i = 1$, we know $\beta^\gamma_i - 1 = O_p(\frac{1}{\sqrt{t}})$ by the central limit theorem results Thm. 1 (actually the strong result $\beta^\gamma_i - 1 = o_p(\frac{1}{\sqrt{t}})$ holds). Then

$$\mathbb{P}(\beta^\gamma_i < 1 - \varepsilon_t) = \mathbb{P}(O_p(1) > \sqrt{t}\varepsilon_t) \to 0, \qquad \text{(by } \varepsilon_t\sqrt{t} \to \infty)$$

using the smoothing rate condition $\sqrt{t}\varepsilon_t \to \infty$. For i such that $\beta_i^* < 1$, we know $\beta^\gamma - \beta_i^* = o_p(1)$ by consistency of β^γ. Then

$$\mathbb{P}(\beta_i^\gamma < 1 - \varepsilon_t) = \mathbb{P}(o_p(1) < (1 - \beta_i^*) - \varepsilon_t) \to 1. \qquad \text{(by } \varepsilon_t = o(1) \text{ and } 1 - \beta_i^* > 0)$$

We conclude $\mathbb{P}(\hat{\mathcal{P}} = \mathcal{P}) \to 1$.

We now show $(\hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}})^\dagger \xrightarrow{p} (\mathcal{P}\mathcal{H}\mathcal{P})^\dagger$. For any $\epsilon > 0$,

$$\begin{aligned}
&\mathbb{P}(\|(\hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}})^\dagger - (\mathcal{P}\mathcal{H}\mathcal{P})^\dagger\|_F > \epsilon) \\
&\le \mathbb{P}(\|(\hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}})^\dagger - (\mathcal{P}\mathcal{H}\mathcal{P})^\dagger\|_F > \epsilon, \hat{\mathcal{P}} = \mathcal{P}) + \mathbb{P}(\hat{\mathcal{P}} \neq \mathcal{P}) \\
&= \mathbb{P}(\|[\hat{\mathcal{H}}_{I^cI^c}]^{-1} - [\mathcal{H}_{I^cI^c}]^{-1}\|_F > \epsilon) + \mathbb{P}(\hat{\mathcal{P}} \neq \mathcal{P}) \to 0 \qquad \text{(by } \hat{\mathcal{H}} \xrightarrow{p} \mathcal{H})
\end{aligned}$$

Next we show $\Omega = \mathbb{E}[(x^* \odot v - \bar{\mu}^*)^{\otimes 2}]$ can be consistently estimated by

$$\hat{\Omega} = \frac{1}{t}\sum_{\tau=1}^{t}(x^\gamma(\theta^\tau) \odot v(\theta^\tau) - \bar{\mu}^\gamma)^{\otimes 2}$$

Define for $\beta \in N_{\text{diff}}$

$$\begin{aligned}
\hat{\Omega}(\beta) &= \frac{1}{t}\sum_{\tau=1}^{t}\Big(d(\theta^\tau, \beta) - \frac{1}{t}\sum_{s=1}^{t} d(\theta^s, \beta)\Big)^{\otimes 2} \\
\Omega(\beta) &= \mathbb{E}[(\nabla f(\theta, \beta) - \nabla \bar{f}(\beta))^{\otimes 2}]
\end{aligned}$$

where

$$d(\theta, \beta) \in \partial f(\theta, \beta) = \text{convexhull}\{e_i v_i : i \text{ such that } \beta_i v_i(\theta) = \max_k \beta_k v_k(\theta)\}$$

is a fixed selection of an element from the subgradient set, N_{diff} is a neighborhood of β^* such that for all $\beta \in N$, the set $\{\theta : f(\theta, \cdot) \text{ not differentiable at } \beta\}$ is measure zero; see Lemma 1. [5] We also assume the subgradient selection coincides with the observed FPPE allocation selection, i.e.,

$$d(\theta^\tau, \beta^\gamma) = x^\gamma(\theta^\tau) \odot v(\theta^\tau) \in \partial f(\theta^\tau, \beta^\gamma).$$

Noting $\mathbb{E}[x^* \odot v] = \nabla \bar{f}(\beta^*) = \bar{\mu}^*$ and $\frac{1}{t}\sum_{\tau=1}^{t} d(\theta^\tau, \beta^\gamma) = \frac{1}{t}\sum_{\tau=1}^{t} x^\gamma(\theta^\tau) \odot v(\theta^\tau) = \bar{\mu}^\gamma$, we see that $\hat{\Omega}(\cdot)$ and $\Omega(\cdot)$ are constructed so that $\hat{\Omega}(\beta^\gamma) = \hat{\Omega}$ and $\Omega(\beta^*) = \Omega$. Moreover, for any fixed $\beta \in N_{\text{diff}}$, it holds $\mathbb{E}[d(\theta, \beta)] = \nabla \bar{f}(\beta)$ (Bertsekas, 1973, Prop. 2.2). Next we use a decomposition of $\hat{\Omega}(\beta)$

$$\hat{\Omega}(\beta) = \underbrace{\frac{1}{t}\sum_{\tau=1}^{t}\Big(d(\theta^\tau, \beta) - \nabla \bar{f}(\beta)\Big)^{\otimes 2}}_{\mathrm{I}(\beta)} - \underbrace{\Big(\frac{1}{t}\sum_{\tau=1}^{t} d(\theta^\tau, \beta) - \nabla \bar{f}(\beta)\Big)^{\otimes 2}}_{\mathrm{II}(\beta)}$$

We now argue that both terms converges in probability uniformly on N_{diff}. For the first term, consider a fixed $\beta \in N_{\text{diff}}$, we know

$$\{\theta : d(\theta, \cdot) \text{ not continuous at } \beta\} = \Theta_{\text{tie}}(\beta)$$

which is measure zero, and that $\nabla \bar{f}$ is continuous on N_{diff} by SMO. By Theorem 7.53 of Shapiro et al. (2021) (a uniform law of large number result for continuous random functions), it holds $\sup_{\beta \in N_{\text{diff}}} \|\frac{1}{t}\sum_{\tau=1}^{t}(d(\theta^\tau, \beta) - \nabla \bar{f}(\beta))^{\otimes 2} - \Omega(\beta)\| \xrightarrow{p} 0$ and $\sup_{\beta \in N_{\text{diff}}} \mathrm{II}(\beta) \xrightarrow{p} 0$. To summarize we have

$$\sup_{\beta \in N_{\text{diff}}} \|\hat{\Omega}(\beta) - \Omega(\beta)\| \xrightarrow{p} 0.$$

[5] Note this is different from the condition there is a neighborhood N such that the set $\{\theta : f(\theta, \cdot) \text{ is differentiable on } N\}$ is measure one. The map $(\theta, \beta) \mapsto \nabla f(\theta, \beta)$ is defined on $(\bigcap_{\beta \in N} \Theta_{\text{tie}}(\beta)) \times N$, and yet the set $\bigcap_{\beta \in N} \Theta_{\text{tie}}(\beta)$ might not be measure one. This is the reason we use subgradients in the definition of $\hat{\Omega}(\cdot)$.

Combined with the fact that $\beta^\gamma \xrightarrow{p} \beta^*$ we have $\hat{\Omega}(\beta^\gamma) \xrightarrow{p} \Omega(\beta^*)$. From this we conclude $\hat{\Omega} \xrightarrow{p} \Omega$.

Proof of $\hat{\Sigma}_\beta \xrightarrow{p} \Sigma_\beta$. We rewrite $\hat{\Sigma}_\beta$ as

$$\begin{aligned}\hat{\Sigma}_\beta &= (\hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}})^\dagger \left(\frac{1}{t}\sum_{\tau=1}^{t}(x^\gamma(\theta^\tau)\odot v(\theta^\tau) - \bar{\mu}^\gamma)^{\otimes 2}\right)(\hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}})^\dagger \\ &= (\hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}})^\dagger \hat{\Omega}(\hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}})^\dagger \\ &\xrightarrow{p} (\mathcal{P}\mathcal{H}\mathcal{P})^\dagger \Omega(\mathcal{P}\mathcal{H}\mathcal{P})^\dagger = \Sigma_\beta\end{aligned}$$

Proof of $\hat{\sigma}^2_{\text{REV}} \xrightarrow{p} \sigma^2_{\text{REV}}$. In Lemma 18 we have shown both $\mathcal{F}_1$ and $\mathcal{F}_2$ are VC-subgraph, and thus a uniform law of large number holds. We rewrite

$$\begin{aligned}\sigma^2_{\text{REV}} = &\underbrace{\mathbb{E}[(p^* - \text{REV}^*)^2]}_{\text{I}_t} + \underbrace{(\bar{\mu}^*)^\top(\mathcal{P}\mathcal{H}\mathcal{P})^\dagger\Omega(\mathcal{P}\mathcal{H}\mathcal{P})^\dagger\bar{\mu}^*}_{\text{II}_t} \\ &+ \underbrace{2\mathbb{E}[(p^* - \text{REV}^*)(x^*\odot v - \bar{\mu}^*)]^\top(\mathcal{P}\mathcal{H}\mathcal{P})^\dagger\bar{\mu}^*}_{\text{III}_t}\end{aligned}$$

and

$$\begin{aligned}\hat{\sigma}^2_{\text{REV}} = &\underbrace{\frac{1}{t}\sum_{\tau=1}^{t}(p^\gamma(\theta^\tau) - \text{REV}^\gamma)^2}_{\hat{\text{I}}_t} + \underbrace{(\bar{\mu}^\gamma)^\top(\hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}})^\dagger\hat{\Omega}(\hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}})^\dagger\bar{\mu}^\gamma}_{\hat{\text{II}}_t} \\ &+ \underbrace{2\left(\frac{1}{t}\sum_{\tau=1}^{t}(p^\gamma(\theta^\tau) - \text{REV}^\gamma)(x^\gamma(\theta^\tau)\odot v(\theta^\tau) - \bar{\mu}^\gamma)\right)^\top(\hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}})^\dagger\bar{\mu}^\gamma}_{\hat{\text{III}}_t}\end{aligned}$$

We have $\hat{\text{I}}_t \xrightarrow{p} \text{I}_t$ by invoking Lemma 18, applying a uniform LLN and using the fact that $\beta^\gamma \xrightarrow{p} \beta^*$. And $\hat{\text{II}}_t \xrightarrow{p} \text{II}$ holds by $\bar{\mu}^\gamma \xrightarrow{p} \bar{\mu}^*$, $(\hat{\mathcal{P}}\hat{\mathcal{H}}\hat{\mathcal{P}})^\dagger \xrightarrow{p} (\mathcal{P}\mathcal{H}\mathcal{P})^\dagger$ and $\hat{\Omega} \xrightarrow{p} \Omega$, and applying Slutsky's theorem. Finally, $\hat{\text{III}}_t \xrightarrow{p} \text{III}$ by $\mathcal{F}_1 \cdot \mathcal{F}_2$ is Donsker by Lemma 15 and thus a uniform law of large number holds, and that $\beta^\gamma \xrightarrow{p} \beta^*$.

We complete the proof of Thm. 3. □

F.5. Proof of Thm. 4

Proof of Thm. 4. By the EG characterization of FPPE, we know that $\beta^\gamma(1)$, the pacing multiplier of the observed FPPE $\widehat{\mathsf{FPPE}}\big(\pi b, v(1), \frac{\pi}{t_1}, \gamma(1)\big)$, solves the following dual EG program

$$\min_B \frac{1}{t_1}\sum_{\tau=1}^{t_1}\max_i v_i(\theta^\tau)\beta_i - \sum_{i=1}^{n} b_i \log(\beta_i) \tag{35}$$

The major technical challenge is that the number of summands in the first summation is also random. Given a fixed integer k and a sequence of items $(\theta^{1,1}, \ldots, \theta^{1,k})$, define

$$\begin{aligned}&\beta^{\text{lin},k}(1) = \beta^*(1) + \frac{1}{k}\sum_{\tau=1}^{k} D_\beta(1, \theta^{1,\tau}), \\ &\beta^k(1) = \text{the unique pacing multiplier in } \widehat{\mathsf{FPPE}}(b, v(1), k^{-1}, (\theta^{1,1}, \ldots, \theta^{1,k}))\end{aligned}$$

Here $D_\beta(1,\cdot) = -(\mathcal{H}_B(1))^\dagger(\mu^*(1,\cdot) - \bar{\mu}^*(1))$ where $\mathcal{H}_B(1), \mu^*(1,\cdot)$ and $\bar{\mu}^*(1)$ are the projected Hessian in Eq. (5), item utility function in Eq. (4), and total item utility vector in Eq. (1) in the limit market $\mathsf{FPPE}(b, v(1), s)$. Note $\mathbb{E}[D_\beta(1,\cdot)] = 0$. Note $\beta^\gamma(1) = \beta^{t_1}$ since scaling the supply and the budget at the same time does not change the equilibrium pacing multiplier. We introduce the following asymptotic equivalence results:

Lemma 19. *Recall $t_1 \sim Bin(\pi, t)$. If SCS and SMO hold for the limit market FPPE$(b, v(1), s)$, then*

- $\sqrt{t}(\beta^{\gamma}(1) - \beta^{lin,t_1}) = o_p(1)$ *as* $t \to \infty$.
- $\sqrt{t}(\beta^{lin,t_1} - \beta^{lin,\lfloor \pi t \rfloor}) = o_p(1)$ *as* $t \to \infty$.

Here $\lfloor a \rfloor$ is the greatest integer less than or equal to $a \in \mathbb{R}$. A similar result holds for the market limit FPPE$(b, v(0), s)$ and the influence function $D_\beta(0, \cdot)$ is defined similarly.

With Lemma 19, we write

$$\begin{aligned}
&\sqrt{t}(\hat{\tau}_\beta - \tau_\beta) \\
&= \sqrt{t}(\beta^{\gamma}(1) - \beta^*(1)) - \sqrt{t}(\beta^{\gamma}(0) - \beta^*(0)) \\
&= \sqrt{t}\left(\frac{1}{\sqrt{\lfloor \pi t \rfloor}} \sum_{\tau=1}^{\lfloor \pi t \rfloor} D_\beta(1, \theta^{1,\tau}) - \frac{1}{\sqrt{\lfloor (1-\pi)t \rfloor}} \sum_{\tau=1}^{\lfloor (1-\pi)t \rfloor} D_\beta(0, \theta^{0,\tau}) \right) + o_p(1) && \text{(Lemma 19)} \\
&\xrightarrow{d} \mathcal{N}\left(0, \frac{1}{\pi}\mathrm{Var}[D_\beta(1,\cdot)] + \frac{1}{(1-\pi)}\mathrm{Var}[D_\beta(0,\cdot)]\right). && \text{(independence between } \{\theta^{1,\tau}\}_\tau \text{ and } \{\theta^{0,\tau}\}_\tau)
\end{aligned}$$

Proof of CLT for τ_β. It follows from the above discussion.

Proof of CLT for τ_u. We begin with the linear expansion Eq. (32) and repeat the same argument.

Proof of CLT for τ_{REV}. We begin with the linear expansion Eq. (31) and repeat the same argument.

Proof of CLT for τ_{NSW}. We begin with the linear expansion Eq. (33) and repeat the same argument.

We complete the proof of Thm. 4. □

In order to prove Lemma 19, we will need the following lemma.

Lemma 20. *If $X_t = o_p(1)$ and $T \sim Bin(\pi, t)$ and T and the sequence are independent, then $X_T = o_p(1)$.*

Proof of Lemma 20. By $X_t = o_p(1)$ we know for all $\epsilon > 0$ it holds $\mathbb{P}(|X_t| > \epsilon) \to 0$, or equivalently $\sup_{k \geq t} \mathbb{P}(|X_k| > \epsilon) \to 0$ as $t \to \infty$. By a concentration for binomial distribution, we know for all $\delta > 0$, it holds $\mathbb{P}(|T - \pi t| > \delta \pi t) \leq 2\exp(-\delta^2 \pi t/3)$. Now write

$$\begin{aligned}
\mathbb{P}(|X_T| > \epsilon) &\leq \mathbb{P}(|X_T| > \epsilon, T \in (1 \pm \delta)\pi t) + \mathbb{P}(T \notin (1 \pm \delta)\pi t) \\
&\leq \sum_{k \in (1 \pm \delta)\pi t} \mathbb{P}(|X_k| > \epsilon)\mathbb{P}(T = k) + 2\exp(-\delta^2 \pi t/3) \\
&\leq \sup_{k \geq (1-\delta)\pi t} \mathbb{P}(|X_k| > \epsilon) + 2\exp(-\delta^2 \pi t/3) \to 0 \text{ as } t \to \infty
\end{aligned}$$

where in the second inequality we use the independence between T and the sequence. We conclude $X_T = o_p(1)$, completing proof of Lemma 20. □

Proof of Lemma 19. The first statement uses the independence between t_1 and the items $(\theta^{1,1}, \theta^{1,2}, \dots)$. Define $R(k) = \sqrt{t}(\beta^k(1) - \beta^{\text{lin},k}(1))$. By Eq. (29), we have $R(k) = o_p(1)$ as $k \to \infty$. With this notation, the first statement is equivalent to $R(t_1) = o_p(1)$ where $t_1 \sim \text{Bin}(\pi, t)$, which holds true by Lemma 20.

The second statement is equivalent to $\sqrt{\lfloor \pi t \rfloor}\big(\beta^{\text{lin},t_1}(1) - \beta^{\text{lin},\lfloor \pi t \rfloor}(1)\big) = o_p(1)$. To prove this we use a Komogorov's inequality. By Theorem 2.5.5 from Durrett (2019), for any $\epsilon > 0$, (let $\sigma_{D_\beta} = \mathbb{E}[\|D_\beta(1, \theta)\|_2^2]^{1/2}$)

$$\mathbb{P}\left(\sqrt{\lfloor \pi t \rfloor} \sup_{(1-\epsilon)\lfloor \pi t \rfloor \leq m \leq (1+\epsilon)\lfloor \pi t \rfloor} \|\beta^{\text{lin},m}(1) - \beta^{\text{lin},(1-\epsilon)\lfloor \pi t \rfloor}(1)\|_2 \geq \delta \sigma_{D_\beta} \right) \leq \frac{2\epsilon}{\delta^2}.$$

Then

$$\begin{aligned}
&\mathbb{P}\big(\sqrt{\lfloor \pi t \rfloor}\big\|\beta^{\text{lin},t_1}(1) - \beta^{\text{lin},\lfloor \pi t \rfloor}(1)\big\|_2 \geq \delta\big) \\
&\leq \mathbb{P}\big(\sqrt{\lfloor \pi t \rfloor}\big\|\beta^{\text{lin},t_1}(1) - \beta^{\text{lin},\lfloor \pi t \rfloor}(1)\big\|_2 \geq \delta,\ (1-\epsilon)\lfloor \pi t \rfloor \leq t_1 \leq (1+\epsilon)\lfloor \pi t \rfloor\big) \\
&\quad + \mathbb{P}\big(t_1 \notin \big[(1-\epsilon)\lfloor \pi t \rfloor, (1+\epsilon)\lfloor \pi t \rfloor\big]\big) \\
&\leq \frac{2\epsilon\sigma^2_{D_\beta}}{\delta^2} + \mathbb{P}\big(t_1 \notin \big[(1-\epsilon)\lfloor \pi t \rfloor, (1+\epsilon)\lfloor \pi t \rfloor\big]\big) \to \frac{2\epsilon\sigma^2_{D_\beta}}{\delta^2}
\end{aligned}$$

Finally, since the above holds for all $\epsilon > 0$, we obtain $\sqrt{\lfloor \pi t \rfloor}(\beta^{\text{lin},t_1} - \beta^{\text{lin},\lfloor \pi t \rfloor}) = o_p(1)$. We complete the proof of Lemma 19. □

G. Experiment Details

G.1. Constraint identification and fast convergence rate; see Figures 1 to 3

We verify that β_i^γ converges to 1 at a faster speed than the usual $O_p(t^{-1/2})$ if $\beta_i^* = 1$, i.e., the constraint is active. We choose the FPPE instances as follows.

- $n = 25$ buyers and $t = 1000$ items.
- budget: $b_i = U_i + 1$ for $i = 1, \ldots, 5$ and $b_i = U_i$ for $i = 6, \ldots, 25$, U_i's are i.i.d. uniforms. The extra budgets are to ensure we observe $\beta_i^* = 1$ for the first few buyers.
- value and supply: we let $\{v_1, \cdots, v_n\}$ be identically and independently distributed as either uniforms, exponential, or truncated standard normal.

Under each configuration we form 100 observed FPPEs (100 trials), and plot the histogram of each $\sqrt{t}(\beta_i^\gamma - \beta_i^*)$. The population EG Eq. (P-DualEG) is a constrained stochastic program and can be solved with stochastic gradient based method. In particular, the true value β^* is computed by the dual averaging algorithm (Xiao, 2010). The mean square error decays as $\mathbb{E}[\|\beta^{\text{da},t} - \beta^*\|^2] = O(t^{-1})$ with t being the number of iteration, and so if we choose t to be large enough, it will be accurate to replace β^*, and we still observe CLT for the quantities $\sqrt{t}(\beta^{\text{da},t} - \beta^\gamma)$.

We clearly see that (i) if $\beta_i^* < 1$ then the finite sample distribution is close to a normal distribution, and (ii) if $\beta_i^* = 1$ (or very close to 1, such as $\beta_{14,21}$ in the uniform value plots, $\beta_{20,23}$ in exponential), the finite sample distribution puts most of the probability mass at 1. For cases where β_i^* is close to 1, we need to futher increase number of items to observe normality.

G.2. Choice of smoothing parameter ε_t; see Figures 3 to 5.

Recall that a key component in the variance estimator is the Hessian estimation, during which we choose a smoothing parameter ε_t. In particular, the smoothing ε_t is used to (1) estimate the active constraints and (2) construct the numerical difference estimator $\hat{\mathcal{H}}$. Thm. 3 suggests a choice of $\varepsilon_t = t^{-d}$ for some $0 < d < \frac{1}{2}$. We investigate the effect of d in this experiment.

We look at the following configuration of H_t defined in Eq. (S-DualEG) and the smoothing parameter d. Note we will be evaluating Hessian at a prespecified point and do not need to form any market equilibria in this experiment.

- $n = 9$ buyers. The item size t ranges from 200 to 5000, at a log scale. Concretely, we choose $t \in$ [199, 223, 249, 279, 311, 348, 389, 434, 486, 543, 606, 678, 757, 846, 946, 1057, 1181, 1319, 1474, 1647, 1841, 2057, 2298, 2568, 2870, 3207, 3583, 4004, 4474, 5000]
- budget: it does not play a role in Hessian estimation.
- value and supply: uniform, exponential, or truncated standard normal
- the smoothing parameter $d \in$ [0.10, 0.17, 0.25, 0.32, 0.40, 0.47, 0.55, 0.62, 0.70].

We evaluate the Hessian $\nabla^2 H$ at a pre-specified point $\beta = [0.200, 0.333, 0.467, 0.600, 0.733, 0.867, 1.000]$, and plot the estimated diagonal values, $\hat{\mathcal{H}}_{ii}$ for $i \in [7]$, against the number of items t. Under each configuration we repeat for 10 trials. We see that d represents a bias-variance trade-off. For a small d (0.10, 0.17, 0.25), the variance of the estimated value $\hat{\mathcal{H}}_{ii}$

is small and yet bias is large (since the plots seem to be trending to some point as number of item increases). For a large d (0.55, 0.62, 0.70) variance is large and yet the bias is small (the estimates are stationary around some point). It is suggested to use $d \in (0.32, 0.47)$.

G.3. Coverage rate of revenue confidence interval Eq. (11); see Tables 1 to 3.

- $n \in$ [10, 20, 30, 40, 50, 60, 100, 200, 300] buyers. Number of items $t \in$ [40, 60, 80, 100, 200, 400].
- budget and proportion of buyer with leftover budgets: $b_i = U_i + 1$ for $i = 1, \ldots, [\alpha n]$ and $b_i = U_i$ for $i = [\alpha n] + 1, \ldots, n$, U_i's are i.i.d. uniforms. Here $\alpha \in$ [0.4, 0.6, 0.8] is the proportion of buyers with leftover budgets.
- value and supply: uniform, exponential, or truncated standard normal
- the smoothing parameter in Hessian estimation $d = 0.4$.
- Coverage rate $\alpha = 0.05$, so we construct 90% CIs.

Under each configuration, we calculate the true revenue in the limit FPPE with dual averaging and sampling, and then form 100 (or 50 in the larger size experiment) observed FPPEs. For each observed FPPE, we construct the revenue confidence interval as in Eq. (11). Then we see if the interval covers the true revenue. After 100 trials (or 50 in larger size experiments), we report the coverage rate for that configuration.

We observe that the empirical coverage rate agrees with the nominal coverage rate 90%.

G.4. Effect of treatment assignment probability π in A/B testing; see Table 4.

In this experiment we study the effect of treatment assignment probability π on the coverage rate of treatment effect confidence interval Eq. (17).

- $n \in$ [30, 60, 100] buyers. The item size $t \in$ [100, 200, 400].
- budget and proportion of buyer with leftover budgets: $b_i = U_i + 1$ for $i = 1, \ldots, [\alpha n]$ and $b_i = U_i$ for $i = [\alpha n] + 1, \ldots, n$, U_i's are i.i.d. uniforms. Here $\alpha = .3$ is the proportion of buyers with leftover budgets.
- treatment effects on values: suppose before treatment values are uniformly distributed, and after treatment values become exponentially distributed.
- the smoothing parameter in Hessian estimation $d = .4$.
- the treatment probability $\pi \in$ [.1, .3, .5, .7, .9]

We do observe that for markets with fewer buyers (say 30), the treatment probability has an effect on coverage rate if one type of treatment is applied to a small proportion of items; the empirical coverage rates are slightly lower than the nominal 90%; see for example the entries corresponding to "treatment prob = 0.9" and "number of buyers = 30". However, when we increase the number of buyers, the empirical coverage rate agrees with the nominal 90%.

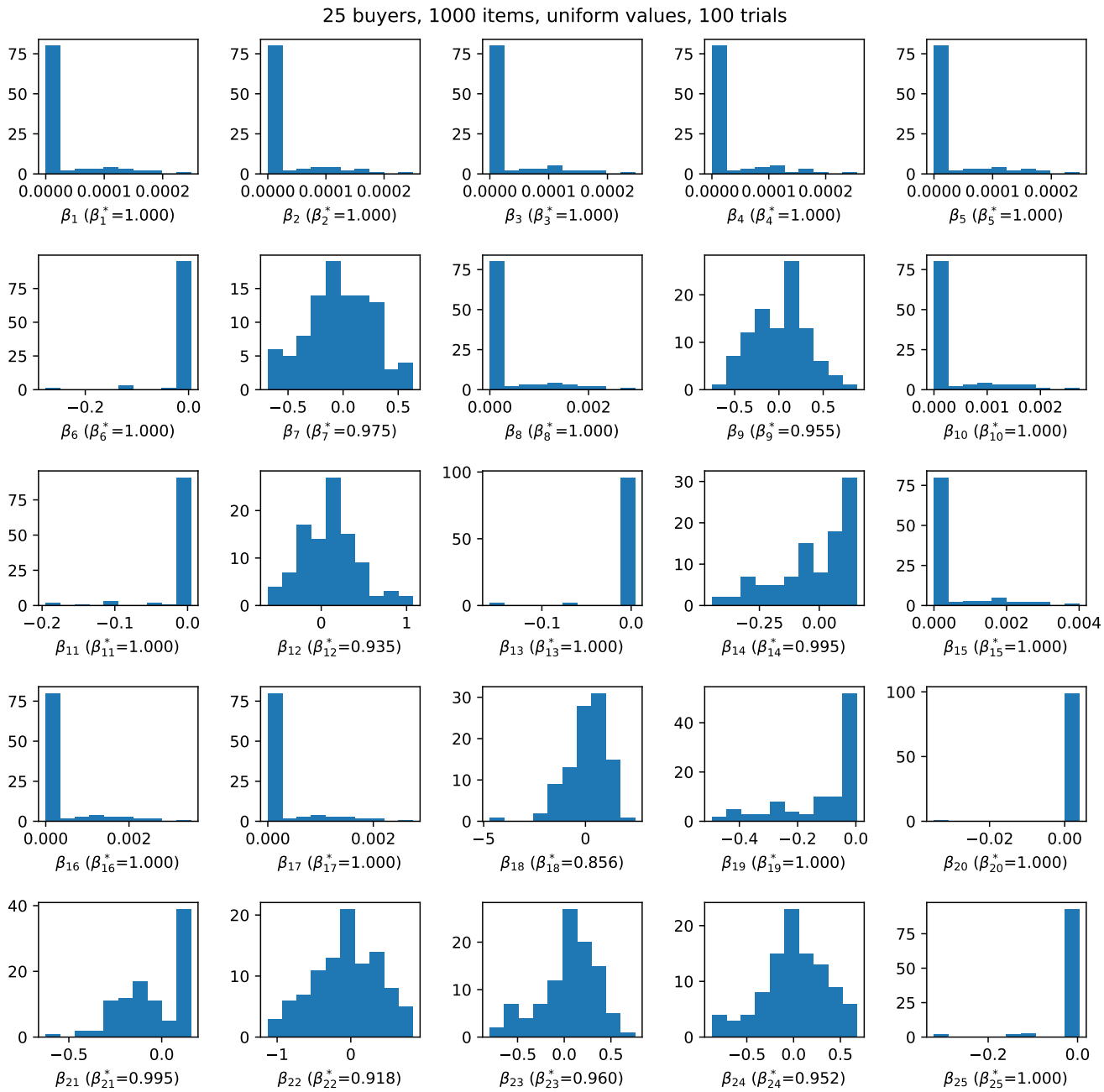

Figure 1: Distribution of $\sqrt{t}(\beta_i^\gamma - \beta_i^*)$ for all $i \in [n]$. Nonnormality and fast convergence for buyers with $\beta_i^* = 1$. Uniform values.

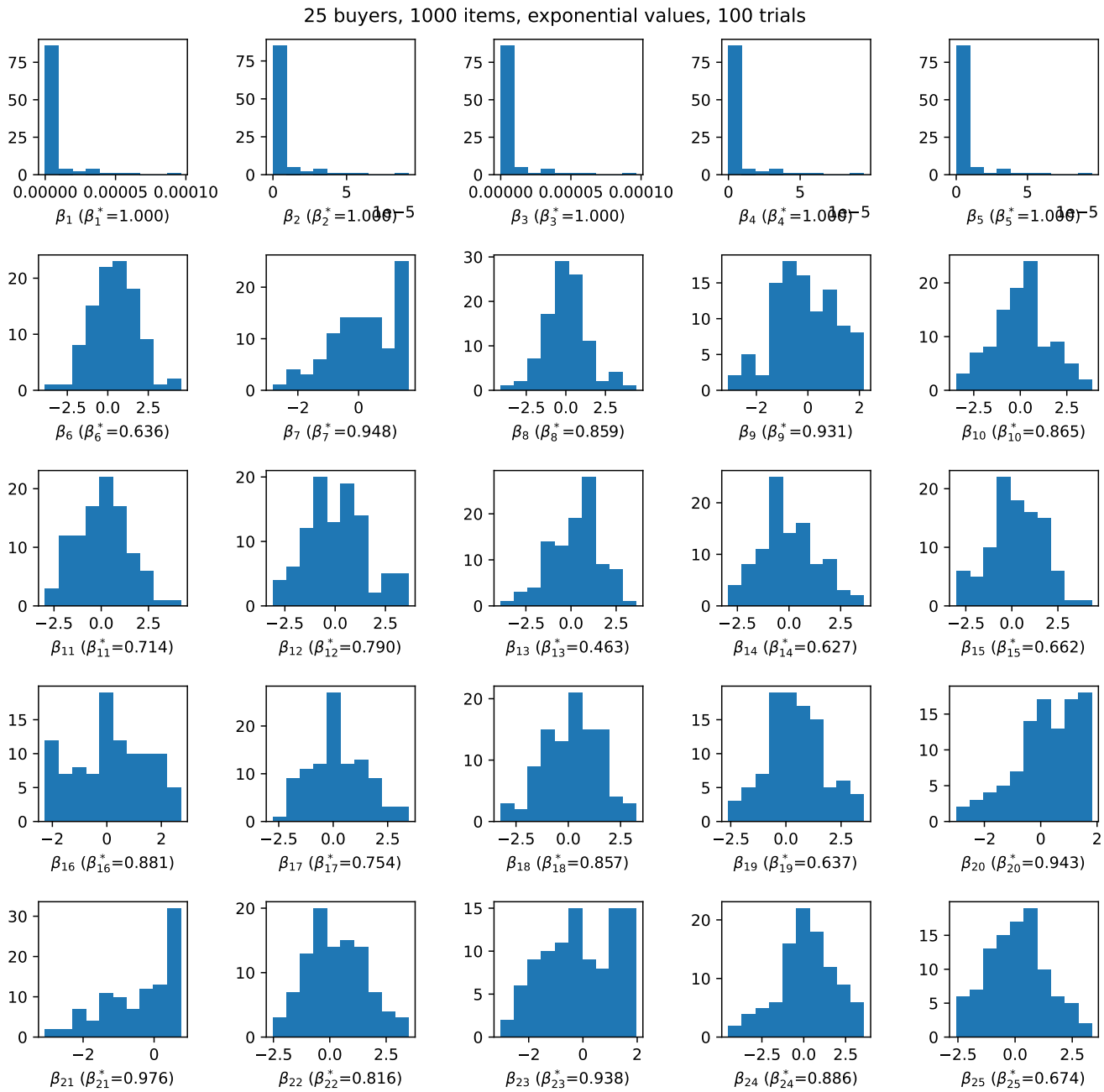

Figure 2: Distribution of $\sqrt{t}(\beta_i^\gamma - \beta_i^*)$ for all $i \in [n]$. Nonnormality and fast convergence for buyers with $\beta_i^* = 1$. Exponential values.

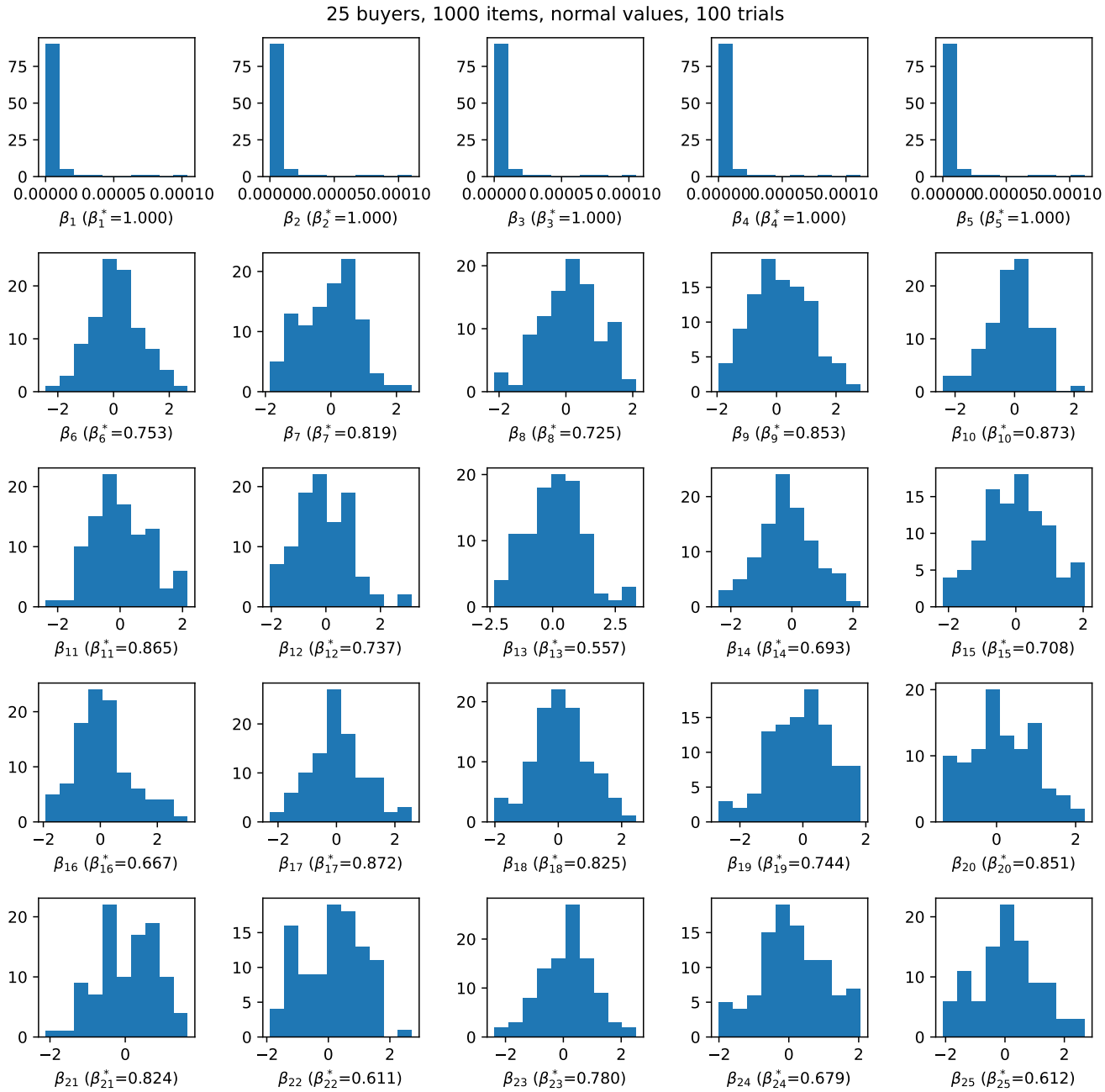

Figure 3: Distribution of $\sqrt{t}(\beta_i^\gamma - \beta_i^*)$ for all $i \in [n]$. Nonnormality and fast convergence for buyers with $\beta_i^* = 1$. Truncated normal values.

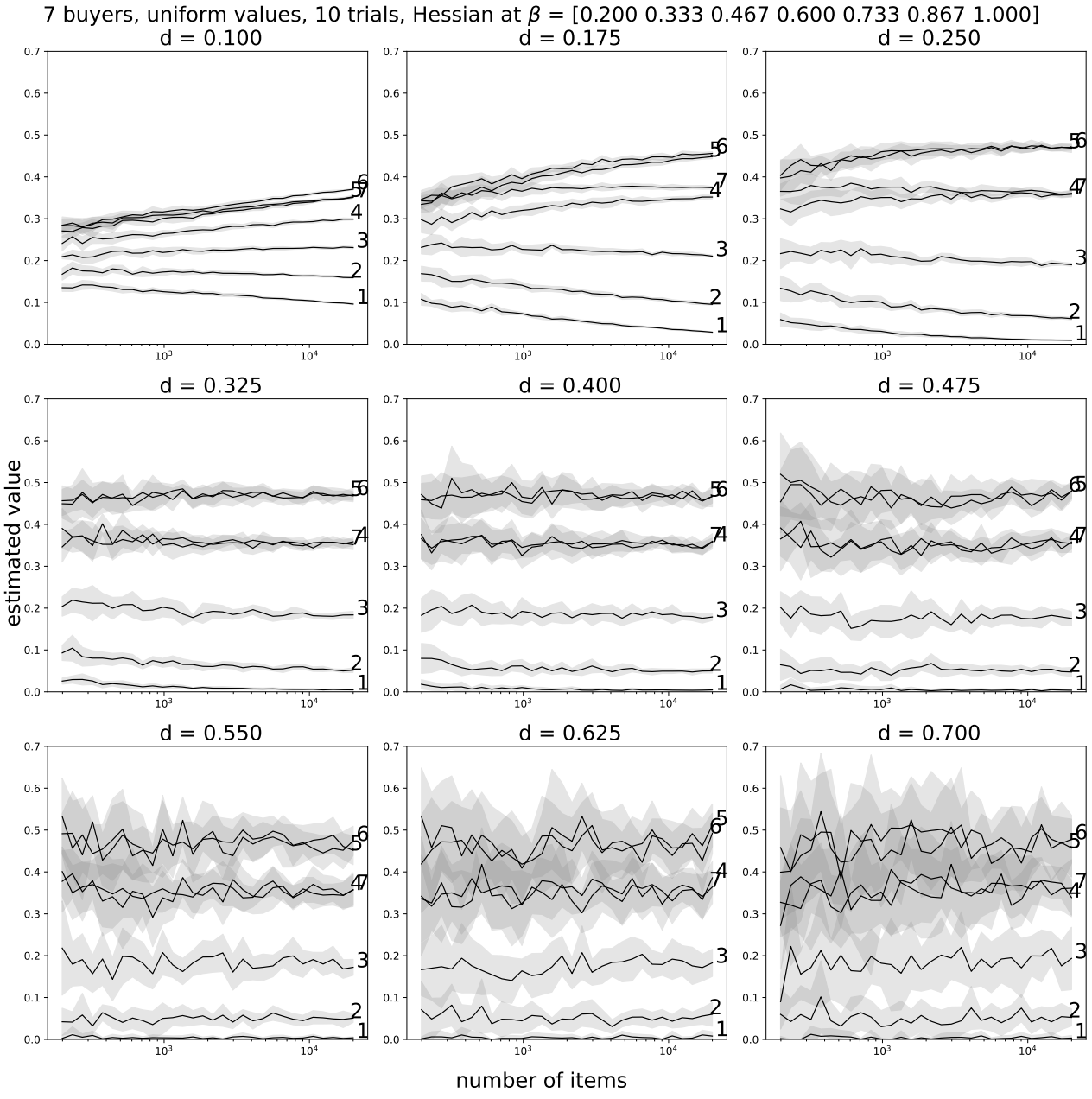

Figure 4: Effect of smoothing parameter on numerical difference estimation of Hessian. Each curve represents the estimated value of $\mathcal{H}_{ii}$. Uniform values.

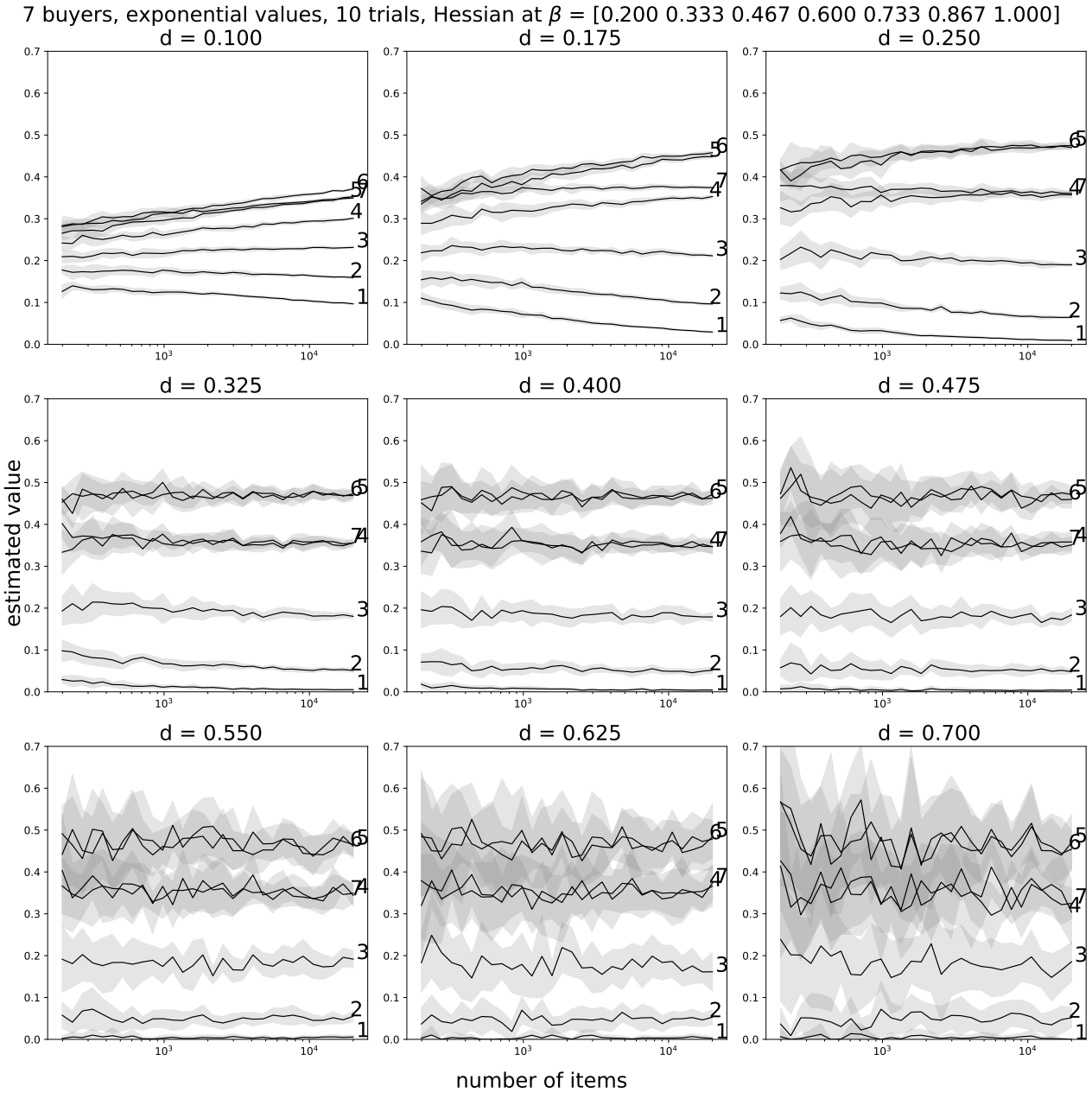

Figure 5: Effect of smoothing parameter on numerical difference estimation of Hessian. Each curve represents the estimated value of $\mathcal{H}_{ii}$. Exponential values.

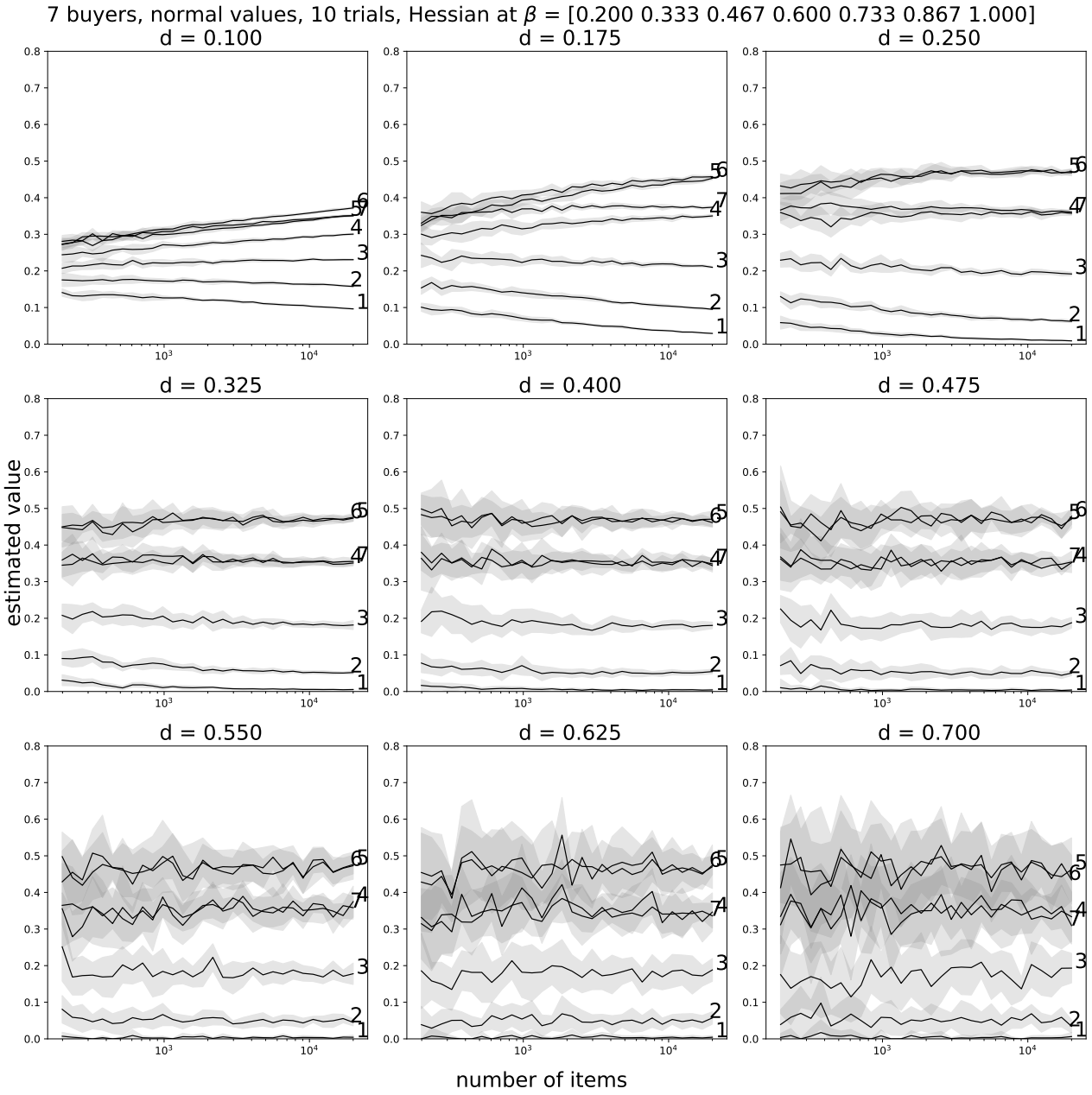

Figure 6: Effect of smoothing parameter on numerical difference estimation of Hessian. Each curve represents the estimated value of $\mathcal{H}_{ii}$. Truncated normal values.

	number of buyers	10			20			30		
	beta = 1 proportion	0.4	0.6	0.8	0.4	0.6	0.8	0.4	0.6	0.8
number of items	value distribution									
	uniform	0.91	0.88	0.84	0.92	0.88	0.86	0.91	0.93	0.90
40	normal	0.92	0.87	0.85	0.88	0.89	0.92	0.87	0.86	0.90
	exponential	0.89	0.90	0.92	0.88	0.85	0.91	0.84	0.93	0.79
	uniform	0.88	0.90	0.89	0.92	0.86	0.89	0.92	0.90	0.90
60	normal	0.89	0.94	0.85	0.90	0.90	0.89	0.95	0.83	0.87
	exponential	0.84	0.87	0.93	0.91	0.88	0.81	0.87	0.88	0.93
	uniform	0.84	0.92	0.92	0.87	0.88	0.92	0.92	0.95	0.87
80	normal	0.86	0.86	0.92	0.85	0.86	0.89	0.89	0.87	0.87
	exponential	0.90	0.90	0.86	0.88	0.87	0.93	0.90	0.88	0.90

Table 1: Coverage rate of the 90% revenue CI defined in Eq. (11). Number of trials for each entry = 100.

	number of buyers	40			50			60		
	beta = 1 proportion	0.4	0.6	0.8	0.4	0.6	0.8	0.4	0.6	0.8
number of items	value distribution									
	uniform	0.93	0.89	0.92	0.85	0.87	0.87	0.85	0.85	0.91
100	normal	0.89	0.87	0.84	0.91	0.88	0.84	0.83	0.82	0.90
	exponential	0.93	0.90	0.95	0.83	0.80	0.85	0.89	0.90	0.87
	uniform	0.82	0.88	0.86	0.85	0.88	0.96	0.88	0.89	0.81
200	normal	0.91	0.87	0.87	0.86	0.86	0.88	0.89	0.92	0.86
	exponential	0.92	0.85	0.93	0.94	0.87	0.92	0.86	0.85	0.97
	uniform	0.80	0.92	0.90	0.91	0.89	0.92	0.85	0.89	0.89
400	normal	0.87	0.86	0.88	0.83	0.89	0.91	0.86	0.91	0.86
	exponential	0.90	0.86	0.92	0.84	0.84	0.81	0.86	0.93	0.92

Table 2: Coverage rate of the 90% revenue CI defined in Eq. (11). Number of trials for each entry = 100.

	number of buyers	100			200			300		
	beta = 1 proportion	0.4	0.6	0.8	0.4	0.6	0.8	0.4	0.6	0.8
number of items	value distribution									
	uniform	0.92	0.88	0.98	0.90	0.80	0.92	0.94	0.92	0.84
100	normal	0.90	0.92	0.86	0.84	0.94	0.86	0.90	0.92	0.88
	exponential	0.80	0.80	0.74	0.88	0.82	0.72	0.90	0.80	0.92
	uniform	0.82	0.92	0.90	0.90	0.90	0.86	0.92	0.82	0.90
200	normal	0.88	0.88	0.88	0.92	0.84	0.90	0.92	0.72	0.96
	exponential	0.84	0.80	0.92	0.90	0.78	0.86	0.94	0.80	0.84
	uniform	0.86	0.90	0.88	0.98	0.88	0.94	0.86	0.86	0.76
400	normal	0.78	0.86	0.90	0.88	0.90	0.88	0.80	0.82	0.82
	exponential	0.88	0.80	0.88	0.80	0.80	0.70	0.86	0.92	0.88

Table 3: Coverage rate of the 90% revenue CI defined in Eq. (11). Number of trials for each entry = 50.

	treatment prob	0.1	0.3	0.5	0.7	0.9
number of buyers	number of items					
	100	0.84	0.82	0.82	0.87	0.79
30	200	0.89	0.87	0.76	0.85	0.86
	500	0.88	0.89	0.95	0.81	0.68
	100	0.87	0.88	0.81	0.87	0.75
60	200	0.89	0.87	0.85	0.83	0.85
	500	0.88	0.91	0.86	0.85	0.90
	100	0.89	0.84	0.95	0.85	0.82
100	200	0.92	0.90	0.83	0.85	0.86
	500	0.92	0.94	0.89	0.93	0.89

Table 4: Coverage rate of the 90% CI for revenue treatment effect defined in Eq. (17). Number of trials for each entry = 100.