# Bootstrapping Fisher Market Equilibrium and First-Price Pacing Equilibrium

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

## Abstract

The linear Fisher market (LFM) is a basic equilibrium model from economics, which also has application in fair and efficient resource allocation. First-price pacing equilibrium (FPPE) is a model capturing budget-management mechanisms in first-price auctions. In certain practical settings such as advertising auctions, there is an interest in performing statistical inference over these models. A popular methodology for general statistical inference is the bootstrap procedure. Yet, for LFM and FPPE there is no existing theory for the valid application of bootstrap procedures. In this paper, we introduce and devise several statistically valid bootstrap inference procedures for LFM and FPPE. The most challenging part is to bootstrap general FPPE, which reduces to bootstrapping constrained M-estimators, a largely unexplored problem. We are able to devise a bootstrap procedure for FPPE under mild degeneracy conditions by using the powerful tool of epi-convergence theory. Experiments with synthetic and semi-real data verify our theory.

## 1. Introduction

The bootstrap (Efron & Tibshirani, 1994; Horowitz, 2001) is an automatic method for producing confidence intervals in statistical estimation. The theory of bootstrap has been extended to many areas of statistics, such as models with cube-root asymptotics (Cattaneo et al., 2020; Patra et al., 2018), semi-parametric models (Cheng & Huang, 2010; Ma & Kosorok, 2005) and so on. However, as far as we are concerned, there is no theory of bootstrap for competitive equilibrium settings.

In this paper, we study bootstrap inference in linear Fisher market (LFM) and first-price pacing equilibrium (FPPE). Fisher market equilibrium model has been used in the tech industry, such as the allocation of impressions to content in certain recommender systems (Murray et al., 2020), robust and fair work allocation in content review (Allouah et al., 2022); we refer readers to Kroer & Stier-Moses (2022) for a comprehensive review. Outside the tech industry, Fisher market equilibria also have applications to scheduling problems (Im et al., 2017), fair course seat allocation (Othman et al., 2010; Budish et al., 2016), allocating donations to food banks (Aleksandrov et al., 2015), sharing scarce compute resources (Ghodsi et al., 2011; Parkes et al., 2015; Kash et al., 2014; Devanur et al., 2018), and allocating blood donations to blood banks (McElfresh et al., 2020).

FPPE is a model for budget management in online advertising platforms. In these platforms, advertisers report advertising parameters, such as target audience, conversion locations, and budgets, and then the platform creates a proxy bidder to bid in individual auctions to maximize advertiser utilities and manage budgets. A common way to manage budgets is pacing, in which the platform modifies the advertiser's bids by applying a shading factor, referred to as *multiplicative pacing*. In the case where each auction is a first-price auction, FPPE captures 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).

Given the wide range of applications of LFM and FPPE, an inferential theory for LFM and FPPE is useful. Bootstrap, thanks to its convenience and conceptual simplicity, is a natural candidate as an inferential tool. However, due to the presence of an equilibrium structure in the dataset, the validity of bootstrap requires careful theoretical treatments, and practitioners should be cautious about the use of bootstrap when data arise from market equilibrium. For example, in Sec 4.2 we show that in the setting of first-price auction platforms, the traditional multinomial bootstrap may fail to consistently estimate the distribution of interest. Given the simplicity of resampling, it is fair to say bootstrap has been used in auction platforms as an infer-

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

*Proceedings of the* $41^{st}$ *International Conference on Machine Learning*, Vienna, Austria. PMLR 235, 2024. 

ential tool. It is thus urgent to develop a statistically valid bootstrap theory that accounts for the equilibrium effect in the data.

Our contributions are threefold.

**We characterize the full landscape of the asymptotics of FPPE.** The limit distribution of FPPE was studied in Liao & Kroer (2023) under a strict complementarity condition. Combining their results with a result of Shapiro (1989), we complete the characterization of the asymptotics of FPPE without strict complementarity, and show that it is captured by a quadratic program. We derive a new closed-form expression for this quadratic program, and use it to derive structural insights on the limit distribution in some special cases. Characterizing the general case of FPPE asymptotics is necessary in order to derive our bootstrap results, because we need to show that our bootstrapped distribution converges to the asymptotic distribution of FPPE.

**We develop bootstrap theory for LFM and FPPE.** A crucial fact for LFM and FPPE is that they both have an Eisenberg-Gale (EG) convex program characterization, and our bootstrap procedures rely on this program or its quadratic approximation. For LFM we study three types of bootstrap procedures: exchangeable bootstrap (Wellner & Zhan, 1996), numerical bootstrap (Hong & Li, 2020) and proximal bootstrap (Hong & Li, 2020). For FPPE the theory is a bit involved. We identify a bootstrap failure when some type of degenerate buyers are present in the market. Then different bootstrap procedures are proposed under certain assumptions on the market structures: full expenditure of budgets ($I_+ = \emptyset$), absence of degenerate buyers ($I_0 = \emptyset$), or fully general FPPE. We summarize the results in Tables 1 and 2.

**Numerical experiments demonstrate the validity of the theory.** We provide simulations and a semi-synthetic experiment based on a real-time bidding dataset from iPinYou (Liao et al., 2014).

| Exchangable BS | Numerical BS | Proximal BS |
|---|---|---|
| ✓ Thm 1 | ✓ Thm 7 | ✓ Thm 8 |

**Table 1:** Results for linear Fisher market.

**Notations.** The notation $\mathcal{N}(a, \Sigma)$ stands for a multivariate Gaussian distribution with mean $a$ and covariance $\Sigma$.

We use $W = (W_1, \ldots, W_t)$ to denote bootstrap weights in the paper. Different distributions imposed on $W$ correspond to different bootstrap resampling schemes. In the standard multinomial bootstrap $W = (W_1, \ldots, W_t)$ follows a multinomial with probabilities $(\frac{1}{t}, \ldots, \frac{1}{t})$. In exchangeable bootstrap $W$ is exchangeable: if for any permutation $\pi = (\pi_1, \ldots, \pi_t)$ of $(1, 2, \ldots, t)$, the joint distribution of $\pi(W) = (W_{\pi_1}, \ldots, W_{\pi_t})$ is the same as that of $W$. Given items $(\theta^\tau)_\tau$, we let $P_t$ be the expectation operator $P_t f = \frac{1}{t}\sum_{\tau=1}^t f(\theta^\tau)$. Given multinomial bootstrap weights $W$ and $(\theta^\tau)_\tau$, define the operator

| | Num. BS | Prox. BS | new methods |
|---|---|---|---|
| $I_+ = \emptyset$ (Sec 4.3) | ✓ Thm 3.1 | ✓ Thm 3.2 | |
| $I_0 = \emptyset$ (Sec 4.4) | ✗ NA | ✗ NA | ✓ Thm 4 |
| general (Sec 4.5) | ✗ NA | ✗ NA | ✓ Thm 6 |

**Table 2:** Results for first-price pacing equilibrium. NA means not applicable. $I_+ = \emptyset$ means full expenditure of budgets. $I_0 = \emptyset$ means absence of degenerate buyers.

$$P_t^b f = \frac{1}{t}\sum_{\tau=1}^t W_\tau f(\theta^\tau). \tag{1}$$

We write $P_t^{\text{ex},b} f = \frac{1}{t}\sum_{\tau=1}^t W_\tau f(\theta^\tau)$ for exchangeable bootstrap weights.

**Bootstrap Consistency** Most of our results will be concerned with the consistency of bootstrap procedures. To that end, we introduce the following definition of consistency. Given $t$ data points, a bootstrap estimate $X_t$ is a function of the data $(\theta^\tau)_{\tau=1}^t$ and bootstrap weights $W$, where the data and weights are assumed to be independent of each other. We say the conditional distribution of $(X_t)_t$ consistently estimates the distribution $L$, denoted $X_t \overset{p}{\rightsquigarrow} L$, if

$$\sup_{f\in \mathrm{BL}_1} \left| \mathbb{E}[f(X_t) \,|\, \{\theta^\tau\}_1^t] - \mathbb{E}_{X\sim L}[f(X)] \right| \overset{p}{\to} 0\,,$$

where $\mathrm{BL}_1$ is the space of functions $f : \mathbb{R}^n \to \mathbb{R}$ with $\sup_x |f(x)| \le 1$ and $|f(x) - f(y)| \le \|x - y\|_2$.

We survey related work in App A.2.

## 2. Review of Fisher Market and FPPE

Both LFM and FPPE have a set of buyers and a set of items being priced. Here we introduce some components that both models share. We have $n$ buyers and a possibly continuous set of items $\Theta$ with an integrating measure $\mathrm{d}\theta$. For example, $\Theta = [0, 1]$ with $\mathrm{d}\theta$ being the Lebesgue measure on $[0, 1]$. Both LFM and FPPE require 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 $s(\theta)$ units 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 $\mathrm{Var}[g] = \mathbb{E}[g^2] - (\mathbb{E}[g])^2$. Given $t$ i.i.d. draws $\{\theta^1, \ldots, \theta^t\}$ from $s$, let $P_t g(\cdot) = \frac{1}{t}\sum_{\tau=1}^t g(\theta^\tau)$.

Equilibria in both LFM and FPPE are characterized by an EG convex program. In both cases, the dual EG objective separates into per-item convex terms

$$F(\theta, \beta) = \max_{i \in [n]} \beta_i v_i(\theta) - \sum_{i=1}^n b_i \log \beta_i \ . \tag{2}$$

and the population and sample EG objectives are

$$H(\beta) = \mathbb{E}[F(\theta, \beta)] \ , \quad H_t(\beta) = P_t F(\cdot, \beta) \ . \tag{3}$$

We comment on the differential structure of $f(\theta, \beta) = \max_i \beta_i v_i(\theta)$ since it plays a role in later sections. Function $f(\beta, \theta)$ is a convex function of $\beta$ and its subdifferential $\partial_\beta f(\beta, \theta)$ is the convex hull of $\{v_i e_i$ : index $i$ such that $\beta_i v_i(\theta) = \max_k \beta_k v_k(\theta)\}$, with $e_i$ being the base vector in $\mathbb{R}^n$. When $\max_i \beta_i v_i(\theta)$ is attained by a unique $i^*$, the function $f$ is differentiable. In that case, the $i$-th entry of $\nabla_\beta f(\theta, \beta)$ is $v_i(\theta)$ for $i = i^*$ and zero otherwise.

### 2.1. Linear Fisher Markets (LFM)

In the LFM model, the goal is to divide items $\Theta$ in a fair and efficient way. It is well known that the competitive equilibrium from equal incomes (CEEI) mechanism produces an allocation that is Pareto efficient, envy-free and proportional (Nisan et al., 2007). LFM is also a useful tool for modeling competition in an economy.

We now describe the competitive equilibrium concept. Imagine there is a central policymaker that sets prices $p(\cdot)$ for the items $\Theta$. Upon observing the prices, the buyer $i$ maximizes their utility subject to the budget constraint. Their *demand set* is the set of bundles that are optimal under the prices:

$$D_i(p) := \operatorname*{arg\,max}_{x_i \in L_+^\infty(\Theta)} \left\{ \int v_i x_i s \, \mathrm{d}\theta : \int p x_i s \, \mathrm{d}\theta \le b_i \right\} .$$

Note that the demand set allows $x_i$ to take values greater than one.

Of course, due to the supply constraint, if prices are too low, there will be a supply shortage. On the other hand, if prices are too high, a surplus occurs. A competitive equilibrium is a set of prices and bundles such that all items are sold exactly at their supply (or have price zero). We call such an equilibrium the *limit LFM equilibrium* for the supply function $s$ (Gao & Kroer, 2023; Liao et al., 2023).

**Definition 1** (Limit Linear Fisher Market Equilibrium)**.** *The limit equilibrium, denoted* $\mathsf{LFM}(b, v, s, \Theta)$*, is an allocation-price tuple* $(x, p(\cdot))$ *such that the following holds.*

1. *Supply feasibility and market clearance:* $\sum_i x_i \le 1$ *and* $\int p(1 - \sum_i x_i)s \, \mathrm{d}\theta = 0$.
2. *Buyer optimality:* $x_i \in D_i(p)$ *all* $i$.

Given the equilibrium quantities $(x^*, p^*)$, let $u_i^* = \int v_i s x_i^* \, \mathrm{d}\theta$ be the buyer $i$ utility, and $\beta_i^* = b_i / u_i^*$ be the buyer $i$ inverse bang-per-buck. In an LFM, the equilibrium quantities $p^*, \beta^*, u^*$ are unique. Under twice differentiability (SMO; to be defined), the allocation $x^*$ is also unique.

Next we introduce the *finite* LFM, which models the data we observe in a market. Let $\gamma = \{\theta^1, \ldots, \theta^t\}$ be $t$ i.i.d. samples from the supply distribution $s$, each with supply $1/t$. See App A.4 for a full definition. For a finite LFM, let the equilibrium per-buyer inverse bang-per-bucks be denoted by $\beta^\gamma$.

**Definition 2** (Finite LFM, informal)**.** *The finite LFM equilibrium, denoted* $\widehat{\mathsf{LFM}}$*, is a limit LFM equilibrium where the item set* $\Theta$ *is the finite set of observed items* $\gamma$*.*

It is well-known (Eisenberg & Gale, 1959; Gao & Kroer, 2023) that the equilibrium inverse bang-per-buck $\beta^*$ in an limit (resp. finite) LFM uniquely solves the population (resp. sample) dual EG program

$$\beta^* = \operatorname*{arg\,min}_{\beta \in \mathbb{R}_+^n} H(\beta) \ , \ \beta^\gamma = \operatorname*{arg\,min}_{\beta \in \mathbb{R}_+^n} H_t(\beta) \ . \tag{4}$$

The asymptotics of LFM were studied in Liao et al. (2023) under twice differentiability (SMO; to be defined). Let $\mathcal{H} = \nabla^2 H(\beta^*)$. They show $\sqrt{t}(\beta^\gamma - \beta^*) \xrightarrow{d} \mathcal{J}_{\mathsf{LFM}}$, where

$$\mathcal{J}_{\mathsf{LFM}} = \mathcal{N}\big(0, \mathcal{H}^{-1}\mathbb{E}[\nabla F(\cdot, \beta^*) \nabla F(\cdot, \beta^*)^\top]\mathcal{H}^{-1}\big) \ . \tag{5}$$

### 2.2. First-Price Pacing Equilibrium (FPPE)

The FPPE setting (Conitzer et al., 2022a) models an economy that typically occurs on internet advertising platforms: the buyers (advertisers in the internet advertising setting) are subject to budget constraints, and must participate in a set of first-price auctions, each of which sells a single item. Each buyer chooses a *pacing multiplier* $\beta_i \in [0, 1]$ that scales down their bids in the auctions, and submits bids of the form $\beta_i v_i(\theta)$ for each item $\theta$, with the goal of choosing $\beta_i$ such that there is *no unnecessary pacing*, i.e. they spend their budget exactly, or they spend less than their budget but they do not scale down their bids. In the FPPE model, all auctions occur simultaneously, and thus the buyers choose a single $\beta_i$ that determines their bid in all auctions.

**Definition 3** (Limit FPPE, Gao & Kroer (2023); Conitzer et al. (2022a))**.** *A limit FPPE, denoted* $\mathsf{FPPE}(b, v, s, \Theta)$*, is the unique tuple* $(\beta, p(\cdot)) \in [0,1]^n \times L^1_+(\Theta)$ *such that there exist* $x_i : \Theta \to [0,1]$*,* $i \in [n]$ *satisfying*

1. *(First-price) Prices are determined by first-price auctions: for all items* $\theta \in \Theta$*,* $p(\theta) = \max_i \beta_i v_i(\theta)$*. Only the highest bidders win: for all* $i$ *and* $\theta$*,* $x_i(\theta) > 0$ *implies* $\beta_i v_i(\theta) = \max_k \beta_k v_k(\theta)$
2. *(Feasibility, market clearing) Let* $\mathsf{pay}_i = \int x_i(\theta)p(\theta)s(\theta)\,\mathrm{d}\theta$*. Buyers satisfy budgets: for all* $i$*,* $\mathsf{pay}_i \le b_i$*. There is no overselling: for all* $\theta$*,* $\sum_{i=1}^n x_i(\theta) \le 1$*. All items are fully allocated: for all* $\theta$*,* $p(\theta) > 0$ *implies* $\sum_{i=1}^n x_i(\theta) = 1$*.*
3. *(No unnecessary pacing) For all* $i$*,* $\mathsf{pay}_i < b_i$ *implies* $\beta_i = 1$*.*

FPPE is a hindsight and static solution concept for internet ad auctions. Suppose we know all the items that are going to show up on a platform. FPPE describes how we could configure the $\beta_i$'s in a way that ensures that all buyers satisfy their budgets, while maintaining their expressed valuation ratios between items. Typically, the $\beta_i$'s are chosen by a pacing algorithm that is run by the platform. FPPE has many nice properties, such as the fact that it is a competitive equilibrium, it is revenue-maximizing, revenue-monotone, shill-proof, has a unique set of prices, and so on (Conitzer et al., 2022a). We refer readers to Conitzer et al. (2022a); Liao & Kroer (2023) for more context about the use of FPPE in internet ad auctions.

Let $\beta^*$ and $p^*$ be the unique FPPE equilibrium multipliers and prices. Revenue in the limit FPPE is $\mathrm{REV}^* = \int p^*(\theta)s(\theta)\,\mathrm{d}\theta$ . We let the leftover budget be denoted by $\delta_i^* = b_i - \mathsf{pay}_i$. We say a buyer $i$ is *degenerate* if $\beta_i^* = 1$ and $\delta_i^* = 0$.

In FPPE the following regularity condition is important.

**Assumption 1** (SCS)**.** *There are no degenerate buyers, i.e.,* $\beta_i^* = 1$ *implies* $\delta_i^* > 0$*.*

This assumption is a strict complementary slackness condition since $\delta_i^*$ is the dual variable of $\beta_i^*$ in the EG program introduced below. We will study the asymptotics of FPPE without SCS. However, as we will see in Sec 4.4, condition SCS is helpful for bootstrap inference.

We let $\gamma = \{\theta^1, \dots, \theta^t\}$ be $t$ i.i.d. draws from $s$, each with supply $1/t$. They represent the items observed in an auction market. The definition of a finite FPPE is parallel to that of a limit FPPE, except that we change the supply function to be a discrete distribution supported on the finite set $\gamma$.

**Definition 4** (Finite FPPE, informal)**.** *A finite FPPE,* $\widehat{\mathsf{FPPE}}$*, is a limit FPPE where the item set is the finite set of observed items* $\gamma$*. See App A.4 for the full definition.*

It is well-known (Cole et al., 2017; Conitzer et al., 2022a; Gao & Kroer, 2023) that $\beta$ in a limit (resp. finite) FPPE uniquely solves the population (resp. sample) dual EG program

$$\beta^* = \operatorname*{arg\,min}_{\beta \in (0,1]^n} H(\beta) \ , \ \beta^\gamma = \operatorname*{arg\,min}_{\beta \in (0,1]^n} H_t(\beta) \ , \qquad (6)$$

where the objectives $H$ and $H_t$ is the same as Eq (4). The difference between the LFM and FPPE convex programs is that for FPPE we impose the constraint $(0,1]^n$.

The study of the asymptotics of FPPE was initiated by Liao & Kroer (2023). Let $\mathcal{J}_{\mathsf{FPPE}}$ be the limit distribution of $\sqrt{t}(\beta^\gamma - \beta^*)$, i.e.,

$$\sqrt{t}(\beta^\gamma - \beta^*) \xrightarrow{d} \mathcal{J}_{\mathsf{FPPE}} \, . \qquad (7)$$

They show that, with the strict complementary slackness assumption SCS, the distribution $\mathcal{J}_{\mathsf{FPPE}}$ simplifies to

$$\mathcal{N}\left(0, \, (P\mathcal{H}P)^\dagger \mathrm{Cov}[\nabla F(\cdot, \beta^*)](P\mathcal{H}P)^\dagger\right) \, , \qquad (8)$$

where $\mathcal{H} = \nabla^2 H(\beta^*)$ and $P = \mathrm{Diag}(1(\beta_i^* < 1))$. We will study the form of $\mathcal{J}_{\mathsf{FPPE}}$ assuming only twice differentiability (SMO) and not SCS. We will characterize $\mathcal{J}_{\mathsf{FPPE}}$ by a random quadratic program and provide several examples. Thus, a contribution of our paper is to remove the strict complementarity slackness assumption and characterize the full landscape of FPPE asymptotics.

### 2.3. Smoothness Assumptions

The following assumption will be made throughout the paper, for both LFM and FPPE.

**Assumption 2** (SMO)**.** *The EG population objective* $H(\cdot)$ *in Eq (3) is twice continuously differentiable in a neighborhood of* $\beta^*$*.*

Assumption 2 implies that the Hessian $\mathcal{H} = \nabla^2 H(\beta^*)$ is positive definite. Here $\beta^*$ is interpreted as the equilibrium inverse bang-per-buck in a limit LFM, and equilibrium pacing multipliers in a FPPE. See Liao & Kroer (2023); Liao et al. (2023) for discussions of implications and concrete examples of SMO holding.

Our research goal can now be stated as

> *Design bootstrap estimators of the distribution* $\mathcal{J}_{\mathsf{LFM}}$ *(resp.* $\mathcal{J}_{\mathsf{FPPE}}$*) given the observed market equilibrium* $\widehat{\mathsf{LFM}}$ *(resp.* $\widehat{\mathsf{FPPE}}$*).*

Inference on other quantities that are differentiable functions of $\beta^*$ can be achieved by the bootstrap delta method (Kosorok (2008, Theorem 12.1), Vaart & Wellner (2023, Theorem 3.10.11)). For example, utilities $u_i^* = b_i/\beta_i^*$ and the Nash social welfare $\sum_i b_i \log u_i^* = \sum_i b_i \log(b_i/\beta_i^*)$ are smooth functions of $\beta^*$. Revenue $\int \max_i\{\beta_i^* v_i(\theta)\} s(\theta)\, \mathrm{d}\theta$ is also a smooth function of $\beta^*$. For this reason, throughout the paper we will focus on inference of $\beta^*$, i.e., the utility prices in LFM and pacing multipliers in FPPE.

# 3. Bootstrapping Fisher Market Equilibrium

In this section we let $\beta^\gamma$ be the observed utility prices in $\widehat{\mathsf{LFM}}(b, v, 1/t, \gamma)$, where $\gamma$ consists of $t$ i.i.d. draws from supply $s$. As mentioned previously, $\beta^\gamma = \arg\min_{\mathbb{R}_+^n} H_t(\beta)$. The target distribution we want to estimate is $\mathcal{J}_{\mathsf{LFM}}$ in Eq (5).

## 3.1. Exchangeable Bootstrap

Define the exchangeable bootstrap by

$$\beta^b_{\mathsf{ex,LFM}} = \arg\min_{\beta \in \mathbb{R}_+^n} P_t^{\mathsf{ex},b} F(\cdot, \beta)\ . \tag{9}$$

Compared with the convex program for LFM in Eq (4), the exchangeable bootstrap replaces $P_t$ with $P_t^{\mathsf{ex},b}$. Exchangeable bootstrap is considered a smooth alternative to the traditional multinomial bootstrap (i.e. sampling with replacement) because it allows for a wider class of distributions of bootstrap weights (Præstgaard & Wellner, 1993). Concretely, we need the weights in the exchangeable bootstrap to satisfy the following conditions.

**Definition 5** (Exchangable bootstrap weights)**.** *(1) The random vector $W = (W_1, \dots, W_t)^\intercal$ is exchangeable. (2) $W_\tau \geq 0$, and $\sum_{\tau=1}^t W_\tau = t$. (3) $W_1$ has finite $(2+\epsilon)$ moment for some $\epsilon > 0$. (4) $\frac{1}{t}\sum_{\tau=1}^t (W_\tau - 1)^2 \xrightarrow{p} c^2 > 0$ as $t \to \infty$.*

Exchangeable bootstrap incorporates many popular forms of resampling as special cases such as the classical sampling with replacement, sampling without replacement, and normalized i.i.d. weights; see App A.3.

**Theorem 1.** $\sqrt{t}(\beta^b_{\mathsf{ex,LFM}} - \beta^\gamma) \overset{p}{\rightsquigarrow} c \cdot \mathcal{J}_{\mathsf{LFM}}$ *where the constant $c$ is defined in Def 5. Proof in App D.2.*

The proof of Thm 1 is complicated by the fact that the EG objective is nonsmooth due to the $\max$ operation in Eq (9). Establishing Thm 1 requires using the exchangeable bootstrap empirical process theory from Præstgaard & Wellner (1993) and Wellner & Zhan (1996) to establish a form of stochastic differentiability (Claim 1 in appendix), and applying the Taylor expansion-type analysis for nonsmooth objective functions from Pollard (1985).

In practice, approximate LMF equilibrium and bootstrap estimates suffice. Eq (9) need not be solved exactly; error in the objective up to order $o_p(1/n)$ suffices, i.e., $P_t^{\mathsf{ex},b} F(\cdot, \beta^b_{\mathsf{ex,LFM}}) \leq \min_\beta P_t^{\mathsf{ex},b} F(\cdot, \beta) + o_p(1/n)$. And $\beta^\gamma$ only needs to be an approximate Fisher market equilibrium: $P_t F(\cdot, \beta^\gamma) \leq \min_\beta P_t F(\cdot, \beta) + o_p(1/n)$. The proof of Thm 1 can be extended to account for the extra error from approximate optimization. In App A.5 we briefly review two other valid bootstrap procedures, proximal bootstrap, and numerical bootstrap, and the consistency theory based on Hong & Li (2020) and Li (2023). Proximal bootstrap has the advantage of solving quadratic programs only. However, those two methods converge at a rate slower than $1/\sqrt{t}$. In contrast, exchangeable bootstrap offers flexibility in choosing bootstrap weights, enjoys a $1/\sqrt{t}$ rate, and does not need parameter tuning.

# 4. Bootstrapping FPPE

In this section we let $\beta^\gamma$ be the pacing multiplier in $\widehat{\mathsf{FPPE}}(b, v, 1/t, \gamma)$, where $\gamma$ consists of $t$ i.i.d. draws from supply $s$. As mentioned previously, $\beta^\gamma = \arg\min_{(0,1]^n} H_t(\beta)$. The target distribution we want to estimate is $\mathcal{J}_{\mathsf{FPPE}}$ in Eq (7), the limit distribution of $\sqrt{t}(\beta^\gamma - \beta^*)$.

Bootstrapping FPPE is a significantly harder problem due to the presence of constraints in the EG program in Eq (6). We investigate the full landscape of FPPE asymptotics, i.e., $\mathcal{J}_{\mathsf{FPPE}}$, in Sec 4.1. In Sec 4.2, we show that the standard multinomial bootstrap fails to estimate $\mathcal{J}_{\mathsf{FPPE}}$ consistently. This also suggests that estimating $\mathcal{J}_{\mathsf{FPPE}}$ in full generality is difficult. Because of this, we divide our study into an easier case and the harder case. In the simpler case, we assume that all buyers exhaust their budget; for this case we show in Sec 4.3 that the bootstrap methods from Hong & Li (2020); Li (2023) are valid. A more realistic case is when some buyers do have leftover budgets. We design a bootstrap for this case in Sec 4.4, under an additional assumption of strong complementary slackness (SCS). Finally, to complete the picture, we present a bootstrap-based confidence region for fully general FPPE in Sec 4.5.

## 4.1. The Limit Distribution of General FPPE

The limit distribution of FPPE was studied in Liao & Kroer (2023) under Assumption SCS. In this section, we characterize the full landscape of the asymptotics of FPPE without strict complementarity. The convex program characterization in this section is a direct corollary of noticing the connection between the results of Shapiro (1989) and Liao & Kroer (2023). Concretely, Theorem 3.3 from Shapiro (1989) established asymptotic distribution results for general constrained programs under equicontinuity conditions, and the results of Liao & Kroer (2023) imply those equicontinuity conditions for the EG objective in Eq (2). This is how we derive the convex program characterization of the asymptotics below. We then derive a new closed-form expression for the convex program, which allows us to analyze the asymptotic structure for several example.

To describe $\mathcal{J}_{\mathsf{FPPE}}$ we need to introduce a quadratic program. Let $I = \{i : \beta_i^* = 1\}$ be the set of unpaced buyers and $I^c = [n] \setminus I$. We further partition $I$ into

$$I_+ = \{i : \beta_i^* = 1, \delta_i^* > 0\} \ , \ I_0 = \{i : \beta_i^* = 1, \delta_i^* = 0\} \ .$$

$I_+$ is the set of buyers with strictly positive leftover budgets, whereas $I_0$ are the degenerate buyers. From an optimization perspective, the set $I_+$ corresponds to the strongly active constraints in the program Eq (6), whose corresponding Lagrange multipliers are strictly positive, while the set $I_0$ are the weakly active constraints, whose Lagrange multipliers are zero. With these notations, we note SCS is the same as $I_0 = \emptyset$, and that the condition that all buyers exhaust their budgets is the same as $I_+ = \emptyset$. Define $h : \mathbb{R}^n \to \mathbb{R}^n$,

$$h(\xi) = \operatorname*{arg\,min}_{h \in \mathbb{R}^n; h_i = 0, i \in I_+; h_i \le 0, i \in I_0} \|h + \mathcal{H}^{-1}\xi\|_{\mathcal{H}}^2 \ , \quad (10)$$

where $\|a\|_{\mathcal{H}}^2 = a^\intercal \mathcal{H} a$. The program Eq (10) can be interpreted as projecting the vector $-\mathcal{H}^{-1}\xi$ onto the cone $\{h : h_i = 0, i \in I_+; h_i \le 0, j \in I_0\}$ w.r.t. the norm $\|\cdot\|_{\mathcal{H}}$. The function $h$ is continuous and positively homogeneous of degree 1, i.e., $h(t\xi) = th(\xi)$ for $t > 0$, but not necessarily linear. When $I_0 = \emptyset$, i.e., SCS holds, the function $h(\xi) = -(P\mathcal{H}P)^\dagger \xi$.

Combining Theorem 3.3 from Shapiro (1989) with the equicontinuity results of Liao & Kroer (2023), we have that under the SMO assumption,

$$\mathcal{J}_{\mathsf{FPPE}} = h\Big(\mathscr{N}\big(0, \mathrm{Cov}[\nabla F(\cdot, \beta^*)]\big)\Big) \ .$$

Below and in App A.6 we study the form of $\mathcal{J}_{\mathsf{FPPE}}$ under some special cases by deriving closed-form expression of the quadratic program Eq (10).

In the example below, we assume $I_+ = \emptyset$ for simplicity. Let $D = \mathrm{Diag}(\mathcal{H}^{-1})^{1/2}, \rho = D^{-1}\mathcal{H}^{-1}D^{-1}, Z = -D^{-1}\mathcal{H}^{-1}G$. where $G \sim \mathscr{N}(0, \mathrm{Cov}[\nabla F(\cdot, \beta^*)])$. Intuitively, $\rho$ is a normalized version of the inverse of the Hessian. Denote entries of $Z$ by $[Z_1, \dots, Z_n]^\intercal$.

**Example 1** (The case with $|I_0| = 1$.). *Let $I_+ = \emptyset$, $I_0 = \{1\}$ and $I^c = \{2, \dots, n\}$. Then $\mathcal{J}_{\mathsf{FPPE}} = -\mathcal{H}^{-1}G$ if $Z_1 < 0$, otherwise, if $Z_1 \ge 0$, then*

$$\mathcal{J}_{\mathsf{FPPE}} = D \begin{bmatrix} 0 \\ Z_2 - \rho_{12} Z_1 \\ \vdots \\ Z_n - \rho_{1n} Z_1 \end{bmatrix} \ . \quad (11)$$

Ex 1 and Ex 6 in appendix illustrate an interesting phenomenon that the limit marginal distribution of the degenerate buyers (those with $\beta_i^* = 1$ and $\delta_i^* = 0$) is a distribution with some probability weight at 0 and the rest on the negative reals. This makes sense intuitively since in a finite sample, $\beta_i^\gamma - \beta_i^* = \beta_i^\gamma - 1$ is always negative for $i \in I_0$. Another feature of $\mathcal{J}_{\mathsf{FPPE}}$ is that the limit distribution of $\sqrt{t}(\beta_i^\gamma - 1)$ is degenerate (a point mass at zero) if $i \in I_+$. This also implies $\beta_i^\gamma - 1 = o_p(\frac{1}{\sqrt{t}})$ if $i \in I_+$.

### 4.2. Failure of Multinomial Bootstrap for FPPE

As described in Andrews (2000), standard multinomial bootstrap might fail in constrained programs. In this section, we show that this is the case for FPPE.

Consider a one-buyer FPPE. Let $b_1 = 1$, $\mathbb{E}[v_1] = \int v_1 s \, d\theta = 1$ and $s$ is the supply (a probability density). Let $\gamma = \{\theta^\tau\}_\tau$ be i.i.d. draws from $s$. Let $\beta^\gamma$ be the pacing multiplier in $\widehat{\mathsf{FPPE}}(b, v, 1/t, \gamma)$ and $\beta^*$ be that in $\mathsf{FPPE}(b, v, s, \Theta)$.

Given the observed items, let $\{\theta^{\tau,b}\}_\tau$ be the resampled items (with replacement). For this instance, the bootstrapped FPPE with standard multinomial weights is

$$\beta^b = \operatorname*{arg\,min}_{\beta_1 \in (0,1]} \frac{1}{t} \sum_{\tau=1}^t \beta_1 v_1(\theta^{\tau,b}) - b_1 \log \beta_1 \ . \quad (12)$$

**Theorem 2** (Failure of Multinomial Bootstrap)**.** *The limit conditional distribution of $\sqrt{t}(\beta^b - \beta^\gamma)$ is not equal to the limit distribution of $\sqrt{t}(\beta^\gamma - \beta^*)$. Proof in App D.6.*

### 4.3. Bootstrapping FPPE with Poor Buyers

If FPPE has the additional structure that all buyers exhaust their budgets, i.e., $I_+ = \emptyset$, we can apply the numerical bootstrap (Hong & Li, 2020) and the proximal bootstrap (Li, 2023). Equivalently, it requires the population EG in Eq (6) does not have strongly active constraints, $\nabla H(\beta^*) = 0$, and the unconstrained optimum coincides with the constrained optimum.

Under this additional structure, Eq (10) becomes

$$\begin{aligned} \mathcal{J}_{\mathsf{FPPE}} &= \operatorname*{arg\,min}_{h : h_i \le 0, i \in I_0} \|h + \mathcal{H}^{-1} G\|_{\mathcal{H}}^2 \\ &= \operatorname*{arg\,min}_{h : h_i \le 0, i \in I_0} h^\intercal G + \tfrac{1}{2} h^\intercal \mathcal{H} h \end{aligned} \quad (13)$$

and $G \sim \mathscr{N}(0, \mathbb{E}[\nabla F(\cdot, \beta^*) \nabla F(\cdot, \beta^*)^\intercal])$.

To obtain numerical bootstrap and proximal bootstrap estimates, we require a smoothing parameter $\epsilon_t \downarrow 0$ such that $\epsilon_t \sqrt{t} \to \infty$. Then, to get $\beta^b_{\mathsf{nu,FPPE}}$ we solve

$$\operatorname*{arg\,min}_{\beta \in (0,1]^n} \frac{1}{t} \sum_{\tau=1}^t (1 + \epsilon_t \sqrt{t}(W_\tau - 1)) F(\theta^\tau, \beta) \ , \quad (14)$$

and to get $\beta^b_{\mathsf{pr,FPPE}}$ we solve

$$\arg\min_{\beta\in[0,1]^n} \epsilon_t (G^b)^\intercal(\beta-\beta^\gamma) + \tfrac{1}{2}\|\beta-\beta^\gamma\|^2_{\widehat{\mathcal{H}}} \,, \tag{15}$$

where

$$G^b = \sqrt{t}(P_t^b - P_t) D_F(\cdot,\beta^\gamma) \,,\; \widehat{\mathcal{H}} = (\widehat{\mathcal{H}}_{k,\ell})_{k,\ell} \,. \tag{16}$$

Here $D_F(\cdot,\beta^\gamma)$ is a deterministic element in the subdifferential $\partial_\beta F(\cdot,\beta^\gamma)$ [1]. The term $G^b$ estimates the Gaussian random variable $G$ in Eq (13). The numerical difference estimator is $\widehat{\mathcal{H}}_{k,\ell} = (\widehat{\nabla}^2_{k\ell,\eta_t} H_t)(\beta^\gamma)$, where

$$\begin{aligned}(\widehat{\nabla}^2_{k\ell,\eta} g)(\cdot) = [g(\cdot+\eta e_k+\eta e_\ell) - g(\cdot-\eta e_k+\eta e_\ell) \\ -g(\cdot+\eta e_k-\eta e_\ell) + g(\cdot-\eta e_k-\eta e_\ell)]/(4\eta^2),\end{aligned}$$

and $H_t$ is the finite-sample EG objective in Eq (3). In practice, both Eqs (14) and (15) only need to be solved approximately with error in the objective up to $o_p(\epsilon_t^2)$. The proximal bootstrap in Eq (15) is a bootstrap analogue of the distribution in Eq (13).

The following theorem shows that in the budget-exhaustion case, the numerical bootstrap and proximal bootstrap converge to the correct limit distribution. The proofs can be found in Appendices D.3 and D.5.

**Theorem 3.** *If all buyers exhaust their budgets ($I_+ = \emptyset$), then*

*3.1* $\epsilon_t^{-1}(\beta^b_{\mathsf{nu,FPPE}} - \beta^\gamma) \overset{p}{\rightsquigarrow} \mathcal{J}_{\mathsf{FPPE}}$ .

*3.2 If* $\widehat{\mathcal{H}} \overset{p}{\to} \mathcal{H}$, *then* $\epsilon_t^{-1}(\beta^b_{\mathsf{pr,FPPE}} - \beta^\gamma) \overset{p}{\rightsquigarrow} \mathcal{J}_{\mathsf{FPPE}}$ .

The proof proceeds by verifying conditions in Hong & Li (2020) and Li (2023). Stochastic equicontinuity of certain processes is verified using results from Liao & Kroer (2023).

### 4.4. Bootstrapping FPPE with SCS

In real-world auction markets such as those at internet companies, some fraction of buyers do have leftover budgets (Conitzer et al., 2022b). In this section, we give a bootstrap estimate of $\mathcal{J}_{\mathsf{FPPE}}$ under SCS, in which we allow users to have positive leftover budgets, but rule out degenerate buyers. Condition SCS is equivalent to requiring that in the population EG in Eq (6) there is no weakly active constraints (those whose Lagrangian multipliers are zero). Condition SCS, equivalent to $I_0 = \emptyset$, is realistic because degenerate buyers are a measure-zero edge case.

[1] We avoid writing $\nabla F(\cdot,\beta^\gamma)$ because in a finite FPPE there could be ties. And when ties happen for an item $\theta$, EG objective $\beta \mapsto F(\theta,\beta)$ is not differentiable at $\beta^\gamma$.

Choose two vanishing sequences $\delta_t$ and $\epsilon_t$. Define the *estimated* unpaced buyers $\widehat{I}_+ = \{i : \beta_i^\gamma > 1-\delta_t\}$ and the reduced feasible set $\widehat{B} = \{\beta\in[0,1]^n : \beta_i = 1 \text{ for } i\in\widehat{I}_+\}$. The proposed bootstrap estimator is

$$\beta^b = \arg\min_{\beta\in\widehat{B}} \epsilon_t (G^b)^\intercal(\beta-\beta^\gamma) + \tfrac{1}{2}\|\beta^\gamma-\beta\|^2_{\widehat{\mathcal{H}}} \,, \tag{17}$$

where $G^b$ and $\widehat{\mathcal{H}}$ are defined in Eq (16). The estimator has a nice geometric interpretation: we add certain appropriate noise to $\beta^\gamma$ and then project back to the reduced feasible set $\widehat{B}$. We call $\epsilon_t$ the bootstrap stepsize, whose effect is investigated in App B.

**Theorem 4.** *Let SCS hold in FPPE ($I_0 = \emptyset$). Let $\delta_t \asymp 1/\sqrt{t}$, $\epsilon_t = o(1)$ and $\epsilon_t\sqrt{t}\to\infty$. If $\widehat{\mathcal{H}} \overset{p}{\to} \mathcal{H}$, then $\epsilon_t^{-1}(\beta^b - \beta^\gamma) \overset{p}{\rightsquigarrow} \mathcal{J}_{\mathsf{FPPE}}$. Proof in App D.7.*

The estimator in Eq (17) is proposed following ideas from Li (2023); Cattaneo et al. (2020), where the bootstrap is in fact approximating the random quadratic program Eq (10). Many existing works (Geyer, 1994; Hong & Li, 2020; Li, 2023) require that strongly active constraints do not occur, and are thus not applicable for FPPE with buyers who have leftover budgets. As with proximal bootstrap, our approach requires solving quadratic programs only.

We briefly remark on the techniques used to prove Thm 4. We combine the theory of weak convergence (Vaart & Wellner, 2023) from statistics and epi-convergence theory (Rockafellar, 1970) from optimization. The reason is that weak convergence is a powerful tool to study asymptotics of statistical functionals, such as the $\arg\min$ function, and epi-convergence is designed for studying constrained programs. Such an approach dates back to Geyer (1994) and Molchanov (2005), and more recently was used by Parker (2019) for constrained quantile regression, and Hong & Li (2020) and Li (2023) in the context of bootstrap.

Both proximal bootstrap (Eq (15)) and our proposed bootstrap (Eq (17)) require a numerical difference estimate of the Hessian (Eq (16)). We provide a theorem to guide the choice of differencing stepsize.

**Theorem 5** (Hessian estimation, informal)**.** *Consider the finite difference estimate defined in Eq (16) with differencing stepsize $\eta_t = o(1)$ and $\eta_t\sqrt{t}\to\infty$. Under regularity conditions, $\widehat{\mathcal{H}}_{k\ell} - \mathcal{H}_{k\ell} \asymp \eta_t^2 + \frac{1}{\eta_t\sqrt{t}}$ + higher order terms. Proof in App D.4.*

By setting $\eta_t^2 = 1/(\eta_t\sqrt{t})$ we obtain the optimal choice $\eta_t \asymp t^{-1/6}$. The proof of Thm 5 uses empirical process theory to handle the nonsmoothness of the EG objective.

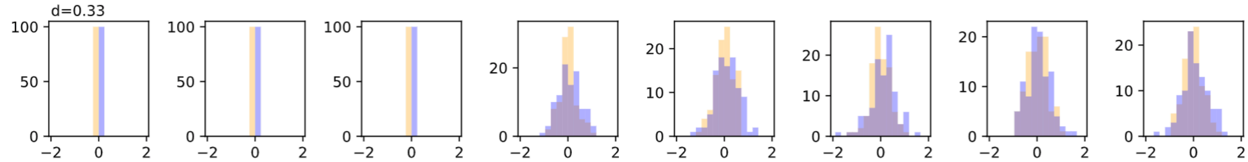


**Figure 1:** Bootstrap vs finite-sample distribution of an 8-buyer 1000-item FPPE. Values are i.i.d. uniformly distributed, and budgets are generated randomly in a way that the first three buyers have leftover budgets. Displayed are histograms of $\beta_1, \ldots, \beta_8$. Purple: 100 samples of $\epsilon_t^{-1}(\beta^b - \beta^\gamma)$ according to Eq (17) given one FPPE. Yellow: 100 samples of $\sqrt{t}(\beta^\gamma - \beta^*)$. Bootstrap distribution is very similar to FPPE distribution. The similarity is significant, because to obtain the distributions of FPPE, we need to observe multiple market equilibria, to which we usually do not have access. The bootstrap distribution, on the other hand, is generated based on just one finite FPPE.

### 4.5. Confidence Regions for General FPPE

In Sections 4.3 and 4.4 we assumed either $I_0$ or $I_+$ to be empty sets. Now we discuss bootstrap inference without such assumptions. We can construct a confidence region for $\beta^*$ using bootstrap test inversion. Suppose we have a scalar statistic $T(\beta^*, \delta^*, \theta^1, \ldots, \theta^t)$, and an upper bound estimate $c \in \mathbb{R}$ of the $(1-\alpha)$-quantile of its limit distribution. Then the region $\{(\beta, \delta) : T(\beta, \delta, \theta^1, \ldots, \theta^t) \le c\}$ is an asymptotically-valid confidence region for $(\beta^*, \delta^*)$.

First, we introduce a statistic based on the Lagrangian of the EG program. The idea of using the Lagrangian or Karush-Kuhn-Tucker (KKT) system for inference in constrained programs also appears in Li (2023); Hsieh et al. (2022). Consider the sample Lagrangian $L_t(\beta, \delta) = H_t(\beta) - \delta^\intercal(1_n - \beta)$ for $\beta \in (0,1]^n$ and $0 \le \delta \le b$. Define the statistic for some $\kappa \in (0, \infty]$:

$$T^\gamma(\beta, \delta) = -\inf_{h \in \mathbb{B}_\kappa} X^\gamma\big(\beta, \delta, \beta + \frac{h}{\sqrt{t}}\big) \ , \qquad (18)$$
$$X^\gamma(\beta, \delta, \beta') = t\big(L_t(\beta', \delta) - L_t(\beta, \delta)\big) \ ,$$

where $\mathbb{B}_\kappa = \{h \in \mathbb{R}^n : \|h\|_2 \le \kappa\}$. Given a threshold value $c$, the statistic $T^\gamma$ induces the region

$$C^\gamma(c) = \{(\beta, \delta) : T^\gamma(\beta, \delta) \le c, \beta \in [0,1]^n, 0 \le \delta \le b\} \ .$$

The region can be constructed as follows. Fix a $\delta$, and then collect all approximate local minimizers of the Lagrangian $L_t(\cdot, \delta)$, i.e., $\beta$ such that $L_t(\beta, \delta) \le \inf_{h \in \mathbb{B}_\kappa} L_t(\beta + \frac{h}{\sqrt{t}}, \delta) + \frac{c}{t}$. Next, we use bootstrap to estimate the distribution of $T^\gamma(\beta^*, \delta^*)$.

$$X^b(\beta) = (\epsilon_t (G^b)^\intercal(\beta - \beta^\gamma) + \tfrac{1}{2}\|\beta^\gamma - \beta\|^2_{\widehat{\mathcal{H}}})/(\epsilon_t)^2 \ ,$$
$$T^b = -\inf_{\beta \in \mathbb{R}^n_+} X^b(\beta) \ .$$

The function $X^b(\cdot)$ is in fact estimating a quadratic expansion of $X^\gamma(\beta^*, \delta^*, \cdot)$. Now we are ready to introduce the confidence region. Let $c^b_{1-\alpha}$ be the conditional $(1-\alpha)$-quantile of $T^b$. Then a confidence region for $(\beta^*, \delta^*)$ is $C^\gamma(c^b_{1-\alpha})$. Let $T^\infty$ be the limit distribution of $T^\gamma(\beta^*, \delta^*)$.

**Theorem 6.** *Suppose $\epsilon_t = o(1)$, $\epsilon_t\sqrt{t} \to \infty$, $\widehat{\mathcal{H}} \xrightarrow{p} \mathcal{H}$. If the CDF of $T^\infty$ is continuous at the $(1-\alpha)$-th quantile of $T^\infty$, then $\liminf_{t\to\infty} \mathbb{P}((\beta^*, \delta^*) \in C^\gamma(c^b_{1-\alpha})) \ge 1-\alpha$. Proof in App D.8.*

The condition on the continuity of the CDF is mild and commonly seen in the literature. The cost that comes with the general applicability of the confidence region $C^\gamma(c^b_{1-\alpha})$ is computational. To decide whether a point $(\beta, \delta)$ is in the region one solves the optimization problem in Eq (18).

## 5. Experiments

We now conduct experiments to investigate the performance of the bootstrap estimator Eq (17) in FPPE with SCS conditions. We aim to (1) verify that the bootstrap produces a consistent estimate of the FPPE asymptotic distribution, and (2) study the effect on the bootstrap of the stepsize parameter $\epsilon_t$ and market parameters, such as the number of items, number of buyers, proportion of budget-constrained buyers, and the value distributions.

**Synthetic experiments.** In App B.1 we consider an ideal scenario where buyers' values are i.i.d. draws from some distribution, i.e., $v_1, \ldots, v_n \sim_{iid} F_v$. To assess the effect of the tail of the value distributions, we take $F_v$ to be a uniform, exponential, or truncated normal distribution. We visualize and compare two setups: *1) true resampling,* where the finite-sample distribution of $\sqrt{t}(\beta^\gamma - \beta^*)$, obtained by repeatedly drawing independent FPPE instances, and *2) bootstrap:* $\epsilon_t^{-1}(\beta^b - \beta^\gamma)$ as defined in Eq (17), obtained by bootstrapping only one FPPE instance. We also vary the bootstrap stepsize $\epsilon_t$. Experiments confirm that our bootstrap Eq (17) is consistent, fairly robust under a wide range of market parameters when bootstrap stepsize is chosen appropriately.

**Semi-real experiments.** In App B.2 we construct realistic instances from real-world auction markets based on the iPinYou dataset (Liao et al., 2014). The dataset contains raw log data of the bid, impression, click, and conversion history on the iPinYou platform. From the dataset we estimate the click-through rate of impressions using logistic regression and simulate realistic advertisers' values by perturbing the regression coefficients. We treat the sum of pacing multipliers as the target parameter and use percentiles of the bootstrap estimates based on Eq (17) to construct confidence intervals. We assess the effect on the coverage rate of the number of items, number of advertisers, the bootstrap stepsize $\epsilon_t$, and the proportion of unpaced buyers. These experiments show that our bootstrap is suitable for realistic auction markets.

## 6. Future Directions

A bootstrap theory for FPPE without regularity conditions on either the buyers (e.g. Assumption 1) or the CDF assumption in Thm 6 would be desirable. However, we suspect that this will be a difficult task, since bootstrapping completely general constrained convex programs remains an open problem. Secondly, we saw in our experiments that Hessian estimation is important for the performance of our bootstrap methods. Thus, a better understanding of how to perform Hessian estimation for the best performance on real-world problems would be useful. In practice it would also be highly desirable to have a bootstrap theory that has some form of guarantees under nonstationary input data.

## Acknowledgements

This research was supported by the Office of Naval Research awards N00014-22-1-2530 and N00014-23-1-2374, and the National Science Foundation awards IIS-2147361 and IIS-2238960.

## Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

## References

Aleksandrov, M., Aziz, H., Gaspers, S., and Walsh, T. Online fair division: Analysing a food bank problem. *arXiv preprint arXiv:1502.07571*, 2015.

Allouah, A., Kroer, C., Zhang, X., Avadhanula, V., Dania, A., Gocmen, C., Pupyrev, S., Shah, P., and Stier, N. Robust and fair work allocation, 2022. URL https://arxiv.org/abs/2202.05194.

Andrews, D. W. Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica*, pp. 399–405, 2000.

Billingsley, P. *Convergence of probability measures*. John Wiley & Sons, 2013.

Bose, A. and Chatterjee, S. Generalised bootstrap in non-regular m-estimation problems. *Statistics & probability letters*, 55(3):319–328, 2001.

Bücher, A., Segers, J., and Volgushev, S. When uniform weak convergence fails: Empirical processes for dependence functions and residuals via epi-and hypographs. 2014.

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

Cattaneo, M. D., Jansson, M., and Nagasawa, K. Bootstrap-based inference for cube root asymptotics. *Econometrica*, 88(5):2203–2219, 2020.

Cheng, G. Moment consistency of the exchangeably weighted bootstrap for semiparametric m-estimation. *Scandinavian Journal of Statistics*, 42(3):665–684, 2015.

Cheng, G. and Huang, J. Z. Bootstrap consistency for general semiparametric m-estimation. 2010.

Chernozhukov, V. and Hong, H. Likelihood estimation and inference in a class of nonregular econometric models. *Econometrica*, 72(5):1445–1480, 2004.

Chernozhukov, V., Hong, H., and Tamer, E. Estimation and confidence regions for parameter sets in econometric models 1. *Econometrica*, 75(5):1243–1284, 2007.

Clarke, F. H. *Optimization and nonsmooth analysis*. SIAM, 1990.

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.

Devanur, N. R., Garg, J., Mehta, R., Vaziranb, V. V., and Yazdanbod, S. A new class of combinatorial markets with covering constraints: Algorithms and applications. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 2311–2325. SIAM, 2018.

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

Efron, B. and Tibshirani, R. J. *An introduction to the bootstrap*. CRC press, 1994.

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

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

Gao, Y. and Kroer, C. Infinite-dimensional fisher markets and tractable fair division. *Operations Research*, 71(2): 688–707, 2023.

Gao, Y., Peysakhovich, A., and Kroer, C. Online market equilibrium with application to fair division. *Advances in Neural Information Processing Systems*, 34:27305–27318, 2021.

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

Ghodsi, A., Zaharia, M., Hindman, B., Konwinski, A., Shenker, S., and Stoica, I. Dominant resource fairness: Fair allocation of multiple resource types. In *Nsdi*, volume 11, pp. 24–24, 2011.

Giné, E. On the bootstrap of M-estimators and other statistical functionals. *Exploring the limits of bootstrap*, 270: 13, 1992.

Giné, E. and Nickl, R. Uniform central limit theorems for kernel density estimators. *Probability Theory and Related Fields*, 141(3-4):333–387, 2008.

Hong, H. and Li, J. The numerical bootstrap. *The Annals of Statistics*, 48(1):397 – 412, 2020. doi: 10.1214/19-AOS1812. URL https://doi.org/10.1214/19-AOS1812.

Horowitz, J. L. The bootstrap. In *Handbook of econometrics*, volume 5, pp. 3159–3228. Elsevier, 2001.

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

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

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

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

Kim, J. and Pollard, D. Cube root asymptotics. *The Annals of Statistics*, pp. 191–219, 1990.

Knight, K. Epi-convergence in distribution and stochastic equi-semicontinuity. 1999.

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

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

Lahiri, S. On bootstrapping m-estimators. *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 157–170, 1992.

Lee, S. M. General m-estimation and its bootstrap. *Journal of the Korean Statistical Society*, 41(4):471–490, 2012.

Li, J. The proximal bootstrap for constrained estimators. May 2023. URL https://people.ucsc.edu/~jeqli/proximalbootstrap.pdf.

Liao, H., Peng, L., Liu, Z., and Shen, X. ipinyou global rtb bidding algorithm competition dataset. In *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising*, pp. 1–6, 2014.

Liao, L. and Kroer, C. Statistical Inference and A/B Testing for First-Price Pacing Equilibria. In *International Conference on Machine Learning*, 2023.

Liao, L., Gao, Y., and Kroer, C. Nonstationary dual averaging and online fair allocation. *Advances in Neural Information Processing Systems*, 35:37159–37172, 2022.

Liao, L., Gao, Y., and Kroer, C. Statistical inference for fisher market equilibrium. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=KemSBwOYJC.

Ma, S. and Kosorok, M. R. Robust semiparametric m-estimation and the weighted bootstrap. *Journal of Multivariate Analysis*, 96(1):190–217, 2005.

McElfresh, D. C., Kroer, C., Pupyrev, S., Sodomka, E., Sankararaman, K. A., Chauvin, Z., Dexter, N., and Dickerson, J. P. Matching algorithms for blood donation. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pp. 463–464, 2020.

Molchanov, I. S. *Theory of random sets*, volume 19. Springer, 2005.

Munro, E. Causal inference under interference through designed. 2023.

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

Murray, R., Kroer, C., Peysakhovich, A., and Shah, P. https://research.fb.com/blog/2020/09/robust-market-equilibria-how-to-model-uncertain-buyer-preferences/, Sep 2020.

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

Nisan, N., Roughgarden, T., Tardos, E., and Vazirani, V. V. *Algorithmic game theory*. Cambridge University Press, 2007.

Othman, A., Sandholm, T., and Budish, E. Finding approximate competitive equilibria: efficient and fair course allocation. In *AAMAS*, volume 10, pp. 873–880, 2010.

Parker, T. Asymptotic inference for the constrained quantile regression process. *Journal of Econometrics*, 213(1): 174–189, 2019.

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

Patra, R. K., Seijo, E., and Sen, B. A consistent bootstrap procedure for the maximum score estimator. *Journal of Econometrics*, 205(2):488–507, 2018.

Pflug, G. Asymptotic dominance and confidence for solutions of stochastic programs. 1991.

Pollard, D. New ways to prove central limit theorems. *Econometric Theory*, 1(3):295–313, 1985.

Pollard, D. *Convergence of stochastic processes*. Springer Science & Business Media, 2012.

Præstgaard, J. and Wellner, J. A. Exchangeably weighted bootstraps of the general empirical process. *The Annals of Probability*, pp. 2053–2086, 1993.

Resnick, S. I. *Extreme values, regular variation, and point processes*, volume 4. Springer Science & Business Media, 2008.

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

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

Römisch, W. Delta method, infinite dimensional. 2004.

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

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

Skorokhod, A. V. Limit theorems for stochastic processes. *Theory of Probability & Its Applications*, 1(3):261–290, 1956.

Vaart, A. v. d. and Wellner, J. A. Empirical processes. In *Weak Convergence and Empirical Processes: With Applications to Statistics*, pp. 127–384. Springer, 2023.

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.

Wellner, J. A. and Zhan, Y. Bootstrapping z-estimators. *University of Washington Department of Statistics Technical Report*, 308(5), 1996.

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

# A. Omitted Main Text

## A.1. Notations

Let $A = [A_1; \ldots; A_n]$ denote the matrix constructed by stacking $A_i$ from top to bottom. Vectors are column vectors by default. For a matrix $H$, a vector $G$, an index set $I \subset [n]$, we let $H_{II} = (H_{ij})_{i\in I, j\in I}$ to denote the $|I| \times |I|$ matrix consisting of entries in $H$, and $G_I$ be the subvector with entries indexed by $I$. We let $\mathbb{R}^n_+ = [0, \infty)^n$. Furthermore, we let $A^\dagger$ be the Moore-Penrose pseudo inverse of a matrix $A$. Denote by $e_j$ the $j$-th unit vector.

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 $\Theta$ w.r.t the integrating measure $\mathrm{d}\theta$ for any $p \in [1, \infty]$ (including $p = \infty$). We treat all functions that agree on all but a measure-zero set as the same. For a sequence of random variables $\{X_n\}$, we say $X_n = O_p(1)$ if for any $\epsilon > 0$ there exists a finite $M_\epsilon$ and a finite $N_\epsilon$ 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.

| Symbol | Meaning |
|---|---|
| $\xrightarrow{d}$ | convergence in distribution of random vectors |
| $\rightsquigarrow$ | weak convergence in metric space |
| $\overset{p}{\rightsquigarrow}$ | weak conditional convergence in metric space |
| $b, b_i$ | budgets |
| $\beta^*, \beta^\gamma, \beta^b$ | pacing multipliers |
| $e_i$ | the $i$-th basis vector |
| $\epsilon_t$ | stepsize parameter in numerical bootstrap and proximal bootstrap and the proposed bootstrap Eq (17) |
| $\eta_t$ | differencing stepsize in finite-difference estimator of Hessian |
| $\delta^*, \delta_i^*$ | leftover budget |
| $\delta_t$ | constraint slackness in the proposed bootstrap Eq (17) |
| $D_i(p)$ | demand set in a Fisher market |
| $F$ and $D_F$ | $F$ is the EG objective defined in Eq (2), and $D_F$ a deterministic selection of subgradients |
| $\mathrm{Cov}(\nabla F(\cdot, \beta^*))$ | $\mathbb{E}[(\nabla F(\cdot,\beta^*) - \mathbb{E}[\nabla F(\cdot,\beta^*)])(\nabla F(\cdot,\beta^*) - \mathbb{E}[\nabla F(\cdot,\beta^*)])^\top]$ |
| $\gamma$ | observed item set in LFM and FPPE |
| $G, G^b$ | a normal random variable $\mathscr{N}(0, \mathrm{Cov}(\nabla F(\cdot, \beta^*)))$ and its bootstrap estimate |
| $h$ | the quadratic program in Eq (10) |
| $\mathcal{H}, \widehat{\mathcal{H}}$ | the Hessian matrix of $H$ at $\beta^*$, and its finite-difference estimator |
| $H, H_t$ | population and sample dual EG objective |
| $I, I^c$ | The set of unpaced ($\beta_i^* = 1$) and paced buyers ($\beta_i^* < 1$) buyers, respectively |
| $I_0, I_+$ | The set of unpaced buyers with $\delta_i^* = 0$ and $\delta_i^* > 0$, respectively |
| $\mathcal{J}_{\mathsf{LFM}}, \mathcal{J}_{\mathsf{FPPE}}$ | limit distributions of interest, defined in Eqs (5) and (7) |
| $\ell^\infty(K)$ | the space of bounded functions $f : K \to \mathbb{R}$ |
| $p$ | price function in Fisher market and FPPE |
| $P$ | matrix whose diagonal is $\mathbb{1}(\beta_i^* < 1)$, $i \in [n]$ |
| $P_t, P_t^{\mathsf{ex},b}, P_t^b$ | expectation operators for the empirical distribution, exchangeable bootstrap distribution, and the classical multinomial bootstrap distribution |
| $u^*, u_i^*$ | equilibrium utility values in LFM and FPPE |
| $s(\cdot)$ | supply function (a probability density) |
| $v, v_i, v_i(\theta)$ | valuation functions |
| $x^*, x_i^*$ | equilibrium allocations in LFM and FPPE |

## A.2. Related Work

**Statistical Inference in Equilibrium Models.** Liao et al. (2023); Liao & Kroer (2023) study statistical properties of LFM and FPPE, respectively. 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. Johari et al. (2022) study a Markov chain model of two-sided platform and investigate the effect of bias under different market balance condition. Munro (2023) considers global treatment effects in a market where the allocation mechanism exhibits certain structures. Different from these work, this paper focuses on estimating the asymptotic distribution of the market equilibrium, uses bootstrap to conduct inference and develops its statistical theory in the specific models of LFM and FPPE.

**Bootstrapping $M$-estimators/mathematical programs.** There is a line of research on bootstrapping $M$-estimators (Lahiri, 1992; Giné, 1992; Wellner & Zhan, 1996; Bose & Chatterjee, 2001; Lee, 2012; Patra et al., 2018; Cattaneo et al., 2020). For constrained $M$-estimators, one needs to be cautious about bootstrap procedures since they could produce inconsistent estimates of the target distribution (Andrews, 2000). Bootstrapping constrained estimators is studied in Li (2023) and Hong & Li (2020); in these works it is assumed that there are no strongly active constraints. In the FPPE setting strongly active constraints do occur, and we use epi-convergence theory to remedy this.

### A.3. Examples of Exchangeable Bootstrap

**Example 2.** *The multinomial bootstrap corresponds to sampling with replacement. It satisfies Def 5 with the constant* $c^2 = 1$.

**Example 3.** *Sample without replacement. Let* $h = \lfloor \alpha t \rfloor$ *be the number of samples not chosen for some* $\alpha \in (0, 1)$. *Concretely, let* $w_\tau = \frac{t}{t-h}$ *for* $1 \le \tau \le t - h$ *and* $0$ *otherwise. Then* $W$ *is the vector of* $(w_1, \dots, w_t)$ *ordered at random independent of data. Def 5 is satisfied with the constant* $c^2 = \alpha/(1 - \alpha)$.

**Example 4.** *I.i.d. weights. Let* $w_1, \dots, w_t$ *be i.i.d. draws from some distribution with finite* $(2 + \epsilon)$ *moment, and* $\bar{w} = \frac{1}{t}\sum_{\tau=1}^t w_\tau$. *Define the bootstrap weights* $W_\tau = w_\tau/\bar{w}$. *Def 5 is satisfied with the constant* $c^2 = \mathrm{Var}(w_1)/(\mathbb{E}[w_1])^2$.

For more examples of exchangeable bootstrap weights we refer readers to Præstgaard & Wellner (1993) and Cheng (2015). The wide range of bootstrap weights allowed by Def 5 provides flexibility for practical application.

### A.4. Definition of Finite LFM and FPPE

Here we give a formal definition of finite LFM and FPPE. Let $v_i^\tau = v_i(\theta^\tau)$ be the valuation for the $\tau$'th sampled item.

**Definition 6** (Finite LFM). *The finite observed LFM, denoted* $\widehat{\mathsf{LFM}}(b, v, \sigma, \gamma)$, *is a allocation-price tuple* $(x, p) \in \mathbb{R}_+^{t\times n} \times \mathbb{R}_+^n$ *such that the following hold:*

1. *Supply feasibility and market clearance:* $\sum_i x_i^\tau \le 1$ *and* $\sum_\tau p^\tau(1 - \sum_i x_i^\tau) = 0$.
2. *Buyer optimality:* $x_i \in D_i(p) = \arg\max_{x_i}\{\sum_\tau x_i^\tau v_i^\tau : \sigma \sum_\tau x_i^\tau p^\tau \le b_i, 0 \le x_i^\tau \le 1\}$, *the demand set given the prices.*

Suppose we have a finite LFM equilibrium $(x, p) = \widehat{\mathsf{LFM}}(b, v, \sigma = 1/t, \gamma)$. Then $u_i^\gamma = \sigma \sum_{\tau=1}^t x_i^\tau v_i^\tau$ is the utility of buyer $i$ in equilibrium, and $\beta_i^\gamma = b_i/u_i^\gamma$ is the utility price of buyer $i$.

**Definition 7** (Finite FPPE, Conitzer et al. (2022a)). *The finite observed FPPE,* $\widehat{\mathsf{FPPE}}(b, v, \sigma, \gamma)$, *is the unique tuple* $(\beta, p) \in [0, 1]^n \times \mathbb{R}_+^t$ *such that there exists* $x_i^\tau \in [0, 1]$ *satisfying:*

1. *(First-price) For all* $\tau$, $p^\tau = \max_i \beta_i v_i^\tau$. *For all* $i$ *and* $\tau$, $x_i^\tau > 0$ *implies* $\beta_i v_i^\tau = \max_k \beta_k v_k^\tau$.
2. *(Supply and budget feasible) For all* $i$, $\sigma \sum_\tau x_i^\tau p^\tau \le b_i$. *For all* $\tau$, $\sum_i x_i^\tau \le 1$.
3. *(Market clearing) For all* $\tau$, $p^\tau > 0$ *implies* $\sum_i x_i^\tau = 1$.
4. *(No unnecessary pacing) For all* $i$, $\sigma \sum_\tau x_i^\tau p^\tau < b_i$ *implies* $\beta_i = 1$.

### A.5. Numerical Bootstrap and Proximal Bootstrap for LFM

We briefly review two valid bootstrap procedures and the consistency theory based on Hong & Li (2020); Li (2023). In this section we only consider multinomial bootstrap weights.

Given a sequence $\epsilon_t$ of positive numbers converging zero, the numerical bootstrap estimator is defined as

$$\beta^b_{\mathsf{nu,LFM}} = \arg\min_{\beta \in \mathbb{R}_+^n} (P_t + \epsilon_t \sqrt{t}(P_t^b - P_t))F(\cdot, \beta)$$
$$= \arg\min_{\beta \in \mathbb{R}_+^n} \frac{1}{t}\sum_{\tau=1}^t (1 + \epsilon_t \sqrt{t}(W_\tau - 1))F(\theta^\tau, \beta)$$

**Theorem 7.** *Let $\epsilon_t = o(1)$ and $\epsilon_t\sqrt{t} \to \infty$. Then $\epsilon_t^{-1}(\beta^b_{\mathsf{nu,LFM}} - \beta^\gamma) \overset{p}{\rightsquigarrow} \mathcal{J}_{\mathsf{LFM}}$.*

The proof is in App D.3.

Numerical bootstrap does not offer computational benefits, since it requires solving EG programs. However, as we see in Sec 4.3 the idea of proximal bootstrap extends to a special case of FPPE where all buyers spend their budgets. The regular multinomial bootstrap is recovered by setting $\epsilon_t = 1/\sqrt{t}$.

To describe proximal bootstrap, we define

$$G^b = \sqrt{t}(P_t^b - P_t)D_F(\cdot, \beta^\gamma)$$

and the numerical difference estimator of the Hessian matrix $\widehat{\mathcal{H}}$, whose $(k,\ell)$-th entry is $\widehat{\mathcal{H}}_{k,\ell} = (\widehat{\nabla}^2_{k\ell,\eta_t} H_t)(\beta^\gamma)$, where $(\widehat{\nabla}^2_{k\ell,\epsilon} g)(\cdot) = [g(\cdot + \epsilon e_k + \epsilon e_\ell) - g(\cdot - \epsilon e_k + \epsilon e_\ell) - g(\cdot + \epsilon e_k - \epsilon e_\ell) + g(\cdot - \epsilon e_k - \epsilon e_\ell)]/(4\epsilon^2)$. And $D_F(\cdot, \beta^\gamma)$ is a deterministic element in $\partial F(\cdot, \beta^\gamma)$.

The proximal bootstrap estimator is defined as

$$\beta^b_{\mathsf{pr,LFM}} = \arg\min_{\beta \in \mathbb{R}^n_+} \{\epsilon_t (G^b)^\top (\beta - \beta^\gamma) + \frac{1}{2}\|\beta^\gamma - \beta\|^2_{\widehat{\mathcal{H}}}\} \tag{19}$$

**Theorem 8.** *Let $\epsilon_t\sqrt{t} \to \infty$ and $\epsilon_t \downarrow 0$. Then $\epsilon_t^{-1}(\beta^b_{\mathsf{pr,LFM}} - \beta^\gamma) \overset{p}{\rightsquigarrow} \mathcal{J}_{\mathsf{LFM}}$.*

The proof is in App D.5.

Proximal bootstrap is clearly computationally cheap since it only requires solving an unconstrained convex quadratic program (as opposed to the exponential cone program for EG). On the other hand, the numerical bootstrap requires estimation of the Hessian matrix. See Thm 5 for a discussion of stepsize selection when using finite difference methods to estimate the Hessian.

### A.6. Examples of FPPE limit distributions

**Example 5** (The case with $I_0 = \emptyset$, SCS holds)**.** *Suppose $|I| = k$. Assume $I_0 = \emptyset$, $I_+ = \{1, \ldots, k\}$. Let $\tilde{\mathcal{H}} = \mathcal{H}_{I^c I^c}$, a square matrix of size $(n-k)$ and $\tilde{G} = G_{I^c}$. Then*

$$\mathcal{J}_{\mathsf{FPPE}} = [0_{k\times 1}; -\tilde{\mathcal{H}}^{-1}\tilde{G}] \tag{20}$$

*which is the same as Eq (8). This agrees with the result from Liao & Kroer (2023).*

**Example 6** (The case with $|I_0| = 2$.)**.** *Let $I_+ = \emptyset$, $I_0 = \{1, 2\}$ and $I^c = \{3, \ldots, n\}$. Then*

$$\mathcal{J}_{\mathsf{FPPE}} = \begin{cases} DZ = -\mathcal{H}^{-1}G & \text{if } Z_1 < 0, Z_2 < 0 \\ D\begin{bmatrix} 0 \\ Z_2 - \rho_{12}Z_1 \\ \vdots \\ Z_n - \rho_{1n}Z_1 \end{bmatrix} & \text{if } Z_1 \geq 0, Z_2 - \rho_{12}Z_1 < 0 \\ D\begin{bmatrix} Z_1 - \rho_{21}Z_2 \\ 0 \\ Z_3 - \rho_{23}Z_2 \\ \vdots \\ Z_n - \rho_{2n}Z_2 \end{bmatrix} & \text{if } Z_2 \geq 0, Z_1 - \rho_{21}Z_2 < 0 \\ [0_{2\times 1}; -\tilde{\mathcal{H}}^{-1}\tilde{G}] & \text{o.w.} \end{cases}, \tag{21}$$

*where $\tilde{\mathcal{H}} = \mathcal{H}_{I^c I^c}$ and $\tilde{G} = G_{I^c}$. We present the derivation in App A.6.*

DERIVING CLOSED-FORM EXPRESSION FOR $\mathcal{J}_{\mathsf{FPPE}}$

We recall a few definitions regarding the constraints. Let $I = \{i : \beta_i^* = 1\}$, $I^c = [n] \setminus I$. We further partition $I$ into

$$I_+ = \{i : \beta_i^* = 1, \delta_i^* > 0\}, I_0 = \{i : \beta_i^* = 1, \delta_i^* = 0\}.$$

Let $A$ (resp. $B$) be a matrix whose rows are $e_i^\intercal, i \in I_+$ (resp. $i \in I_0$). So $A$ is a $|I_+| \times n$ matrix and $B$ is $|I_0| \times n$.

A combinatorial expression of $\mathcal{J}_{\mathsf{FPPE}}$ is available. One can solve the quadratic program Eq (10), which contains linear inequality constraints, by solving at most $2^{|I_0|}$ linearly constrained programs. First, one create a candidate linearly constrained program by turning some inequality constraints to be equality ones, and then record the optimal objective value. Then the smallest value out of all $2^{|I_0|}$ candidate programs must be the same as the original program.

Given $G$, let $Q_j$ and $h_j$ be the optimal value and the optimal solution to the program

$$\min_{h \in \mathbb{R}^n} (h + \mathcal{H}^{-1}G)^\intercal \mathcal{H}(h + \mathcal{H}^{-1}G) \text{ s.t. } [A; B_j]h = 0. \tag{22}$$

Here $B_j$ consists of some (possibly zero) rows of $B$, $j = 1, \dots, 2^{|I_0|}$.

The program Eq (22) is just projecting the vector $-\mathcal{H}^{-1}G$ onto the linear subspace spanned by $\Gamma_j = [A; B_j]$ w.r.t. the norm $\|\cdot\|_{\mathcal{H}}$. With this geometric interpretation, it is easy to write down the solution. Define the projection matrix $P_j = I - \mathcal{H}^{-1}\Gamma_j^\intercal(\Gamma_j\mathcal{H}^{-1}\Gamma_j^\intercal)^{-1}\Gamma_j$. Then the closed-form expressions for $Q_j$ and $h_j$ are

$$Q_j = \|(I - P_j)\mathcal{H}^{-1}G\|_{\mathcal{H}}^2 = (\mathcal{H}^{-1}G)^\intercal\Gamma_j^\intercal(\Gamma_j\mathcal{H}^{-1}\Gamma_j^\intercal)^{-1}\Gamma_j(\mathcal{H}^{-1}G)$$
$$h_j = -P_j\mathcal{H}^{-1}G,$$

Then it is obvious that

$$\mathcal{J}_{\mathsf{FPPE}} = h_{j(G)} = -P_{j(G)}\mathcal{H}^{-1}G \tag{23}$$

where $j(G) = \arg\min_j\{Q_j : Bh_j \le 0\}$. Equivalently,

$$\mathcal{J}_{\mathsf{FPPE}} = \tag{24}$$
$$\sum_{j=1}^{2^{|I_0|}} -\left(\mathbb{1}(Bh_j \le 0) \prod_{\ell=1}^{2^{|I_0|}} \mathbb{1}(Q_j \le Q_\ell \text{ or } Bh_\ell \not\le 0)\right)P_j\mathcal{H}^{-1}G \tag{25}$$

a random vector of length $n$. Only one of the term will be selected for each realization of $G$. The representation allows us to derive the exact distribution in some cases.

*Omitted details in Ex 6.* We show $\mathcal{J}$ in Eq (24) reduces to the claimed expression Eq (21).

Consider the programs

$$\min_{h \in \mathbb{R}^n} (h + \mathcal{H}^{-1}G)^\intercal \mathcal{H}(h + \mathcal{H}^{-1}G)$$
$$\text{subject to no constraints} \tag{$Q_1$}$$
$$\text{or subject to } h^\intercal e_1 = 0 \tag{$Q_2$}$$
$$\text{or subject to } h^\intercal e_2 = 0 \tag{$Q_3$}$$
$$\text{or subject to } h^\intercal e_1 = 0, h^\intercal e_2 = 0 \tag{$Q_4$}$$

For $Q_1$ the optimal solution is $h_1 = -\mathcal{H}^{-1}G$.

For $Q_2$, the optimal value is $Q_2 = (\mathcal{H}^{-1}G)^\intercal e_1(e_1^\intercal\mathcal{H}^{-1}e_1)^{-1}e_1^\intercal(\mathcal{H}^{-1}G) = (G^\intercal\mathcal{H}^{-1}e_1)^2/(\mathcal{H}^{-1})_{11}^2 = Z_1^2$ and the optimal solution $h_2$ is

$$
\begin{aligned}
h_2 &= -[I - \mathcal{H}^{-1}e_1(e_1^\intercal \mathcal{H}^{-1}e_1)^{-1}e_t^\intercal]\mathcal{H}^{-1}G \\
&= \begin{bmatrix} 0 & & & \\ \frac{(\mathcal{H}^{-1})_{21}}{(\mathcal{H}^{-1})_{11}} & -1 & & \\ \vdots & & \ddots & \\ \frac{(\mathcal{H}^{-1})_{n1}}{(\mathcal{H}^{-1})_{11}} & & & -1 \end{bmatrix} \mathcal{H}^{-1}G \\
&= D \begin{bmatrix} 0 \\ Z_2 - \rho_{12}Z_1 \\ \vdots \\ Z_n - \rho_{1n}Z_1 \end{bmatrix}
\end{aligned}
$$

where we recall $Z = -D^{-1}\mathcal{H}^{-1}G = [Z_1, \ldots, Z_n]^\intercal$. For $Q_3$, the optimal value is $Z_2^2$ and the optimal solution $h_3$ is the third display in Eq (21).

For $Q_4$, the optimal solution $h_4$ is the fourth display in Eq (21).

We consider the indicator part for $j = 2$ in Eq (24), i.e., the expression

$$
\mathbb{1}(Bh_2 \le 0) \prod_{\ell=1}^{4} \mathbb{1}(Q_2 \le Q_\ell \text{ or } Bh_\ell \not\le 0)
$$

Then $B = [e_1^\intercal; e_2^\intercal]$. Note $Bh_1 = [Z_1/\sqrt{(\mathcal{H}^{-1})_{11}}, Z_2/\sqrt{(\mathcal{H}^{-1})_{22}}]^\intercal$, $Bh_2 = [0, Z_2 - \rho_{12}Z_1]^\intercal$, $Bh_3 = [Z_1 - \rho_{12}Z_2, 0]^\intercal$. Obviously, both $Q_2, Q_3 \ge Q_1$, and both $Q_2, Q_3 \le Q_4$. It can be shown

$$
\begin{aligned}
&\mathbb{1}(Bh_2 \le 0)\mathbb{1}(Q_1 \ge Q_2 \text{ or } Bh_1 \not\le 0)\mathbb{1}(Q_3 \ge Q_2 \text{ or } Bh_3 \not\le 0)\mathbb{1}(Q_4 \ge Q_2 \text{ or } Bh_4 \not\le 0) \\
&= \mathbb{1}(Bh_2 \le 0)\mathbb{1}(Bh_1 \not\le 0)\mathbb{1}(Q_3 \ge Q_2 \text{ or } Bh_3 \not\le 0) \times 1 \\
&= \mathbb{1}(Z_2 - \rho_{12}Z_1 \le 0)\mathbb{1}(Z_1 \ge 0)\mathbb{1}(Z_2^2 \ge Z_1^2 \text{ or } Z_1 - \rho_{12}Z_2 > 0) \quad \text{almost surely} \\
&= \mathbb{1}(Z_2 - \rho_{12}Z_1 \le 0)\mathbb{1}(Z_1 \ge 0) \quad \text{almost surely,}
\end{aligned}
$$

where the last equality follows by a case-by-case analysis. The indicator parts for $j = 1, 3, 4$ are analyzed similarly. □

## B. Experiments

### B.1. Simulation: Verify Bootstrap Consistency for FPPE

In this section we verify the consistency of our bootstrap estimators, and investigate the effect of the bootstrap stepsize $\epsilon_t$ (in Eq (17)) on the quality of bootstrap approximation in FPPE on fully synthetic data.

We consider an 8-buyer FPPE instance with 100 items sampled with i.i.d. values. Budgets of buyers are selected so that the first three buyers are unpaced ($\beta = 1$). This is to model the fact that in reality there could be buyers with leftover budgets. We use dual averaging (Xiao, 2010; Gao & Kroer, 2020; Liao et al., 2022) to compute the limit FPPE pacing multiplier $\beta^*$. Finite FPPEs are computed with MOSEK. We draw 100 finite FPPEs and obtain the finite FPPE distribution by plotting the histogram of $\sqrt{t}(\beta^\gamma - \beta^*)$. We call this *true resampling*, which would not be possible in practice. Finally, we then generate *a single* FPPE and resample 100 bootstrapped $\beta$'s according to Eq (17), obtaining the bootstrap distribution estimate. To experiment with different tail behaviors for values, we run three sets of experiments: uniform, exponential and truncated normal values. We also vary the choice of bootstrap stepsize $\epsilon_t = t^{-d}$.

*Results* In Figures 2 to 4 we present the finite-sample distribution of $\beta^\gamma$ and $\beta^b$. Each column corresponds to the pacing multiplier of a buyer, and each row corresponds to a choice of $d$. First, we observe that with a suitable choice of $d$, the bootstrap distribution is a good approximation to that of finite FPPE with true resampling. For buyers with $\beta_i^* = 1$ the proposed bootstrap is able to correctly identify them. For buyers with $\beta_i^* < 1$, bootstrap correctly captures the range and the shape of the distribution. This result is significant, because to obtain the distributions of FPPE, we need to observe multiple market equilibria, to which we usually do not have access. The bootstrap distribution, on the other hand, is generated based on just one finite FPPE. Second, we also observe that if $\epsilon_t$ is too large ($d$ too small), the quality of approximation degrades. In particular, in the bottom rows of plots for uniform and normal values, bootstrap tends to ignore the right part of the distribution of finite FPPEs.

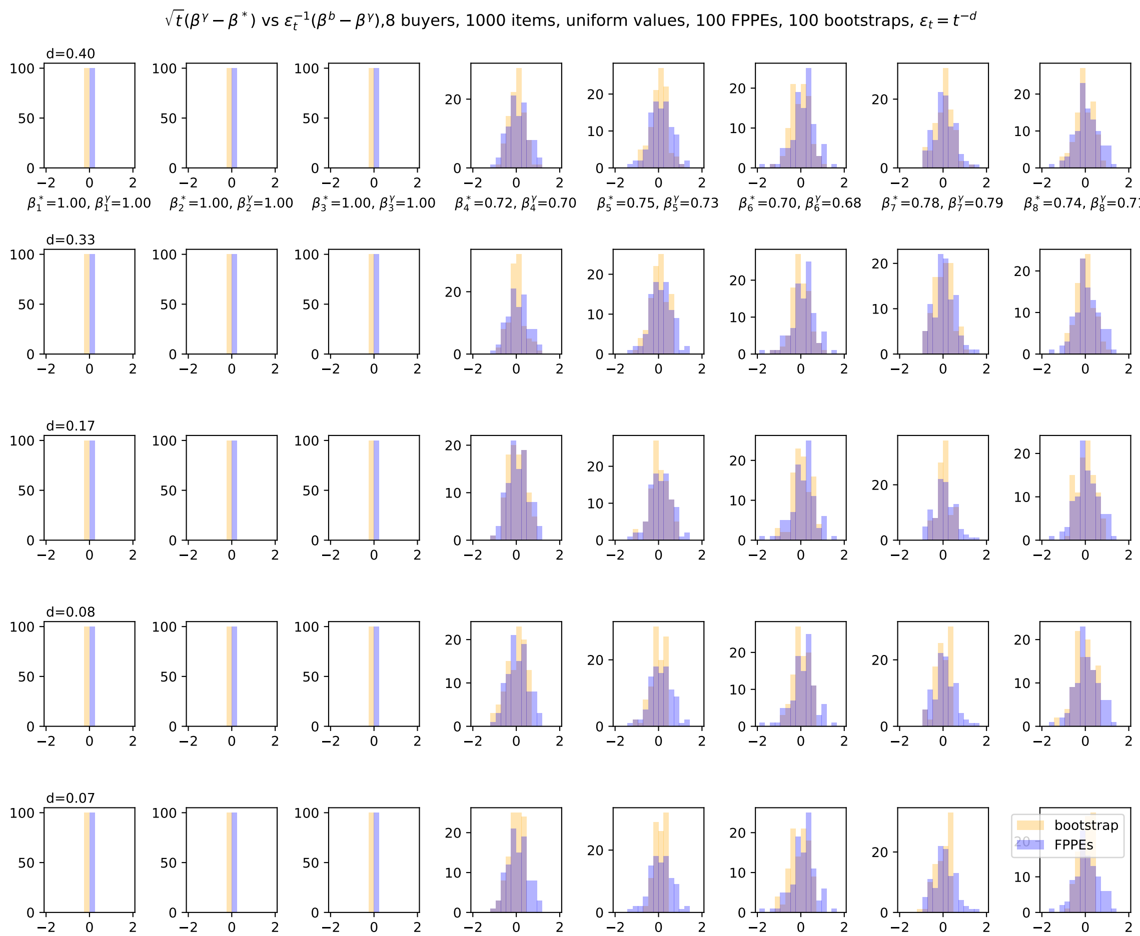


**Figure 2:** Comparison of Bootstrap and FPPE finite item distribution.

### B.2. Semi-Real Data

In this section we apply our bootstrap estimator to a real-world dataset, the iPinYou dataset (Liao et al., 2014).

*The data.* The iPinYou dataset (Liao et al., 2014) contains raw log data of the bid, impression, click, and conversion history on the iPinYou platform in the weeks of March 11–17, June 8–15 and October 19–27. We use the impression and click data of 5 advertisers on June 6, 2013, containing a total of 1.8 million impressions and 1,200 clicks. As in the main text, let $i \in \{1, 2, 3, 4, 5\}$ index advertisers (buyers) and let $\tau$ index impressions/users (items in FPPE terminology). The five advertiser are labeled by number and their categories are revealed: 1459 (Chinese e-commerce), 3358 (software), 3386 (international e-commerce) and 3476 (tire). From the raw log data, the following dataset can be extracted. The response variable is a binary variable $\text{CLICK}_i^\tau \in \{0, 1\}$ that indicates whether the user clicked the ad or not. The relevant predictors include a categorical variable ADEXCHANGE of three levels that records from which ad-exchange the impression was generated, a categorical variable REGION of 35 levels indicating provinces of user IPs, and finally 44 boolean variables, USERTAG's, indicating whether a user belongs to certain user groups defined based on demographic, geographic and other information. We select the top-10 most frequent user tags and denote them by $\text{USERTAG}_1, \ldots, \text{USERTAG}_{10} \in \{0, 1\}$. Both ADEXCHANGE and USERTAG are masked in the dataset, and we do not know their real-world meanings.

*Simulate advertisers with logistic regression.* The raw data contains only five advertisers. In order to simulate more realistic

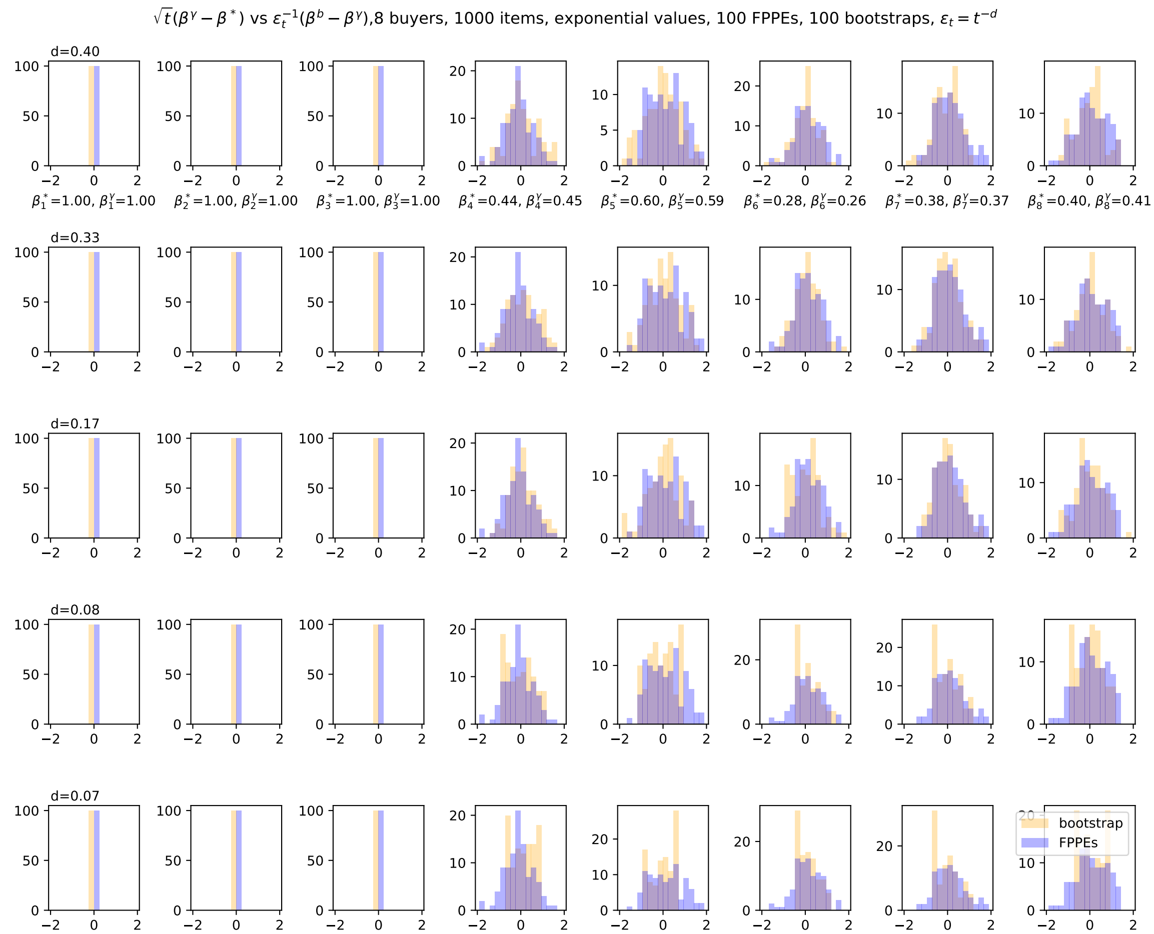


**Figure 3:** Comparison of Bootstrap and FPPE finite item distribution.

advertiser values, we fit a logistic regression and then perturb the fitted coefficients to generate more advertisers. We posit the following logistic regression model for click-through rates (CTRs). For a user $\tau$ that saw the ad of advertiser $i$, the click process is governed by

$$\mathrm{CTR}_i^\tau = \mathbb{P}(\textsc{click}_i^\tau = 1 \mid \theta^\tau) = \frac{1}{1+\exp(w_i^\top \theta^\tau)}$$

$$\theta^\tau = [1, \textsc{AdExchange}_2, \textsc{AdExchange}_3, \textsc{Region}_2, \ldots, \textsc{Region}_{35}, \textsc{UserTag}_1, \ldots, \textsc{UserTag}_{10}] \in \{1\} \times \{0,1\}^{46}$$

where the weight vectors $w_i \in \mathbb{R}^{47}$ are the coefficients to be estimated from the data. Note that $\textsc{AdExchange}_1$ and $\textsc{Region}_1$ are absorbed in the intercept. By running 5 logistic regressions, we obtain regression coefficients $w_1, w_2, \ldots, w_5$. To visualize the fitted regression, in Figure 5 we show the estimated click-through rate distributions of the five advertisers. The diagonal plots are the histogram of CTRs, and the off-diagonal panels are the pair-wise scatter plots of CTRs. To generate more advertisers, we take a convex combination of the coefficients $w_i$'s, add uniform noise, and obtain a new parameter, say $w'$. Given an item, the CTR of the newly generated advertisers will be $\frac{1}{1+\exp(\theta^\top w')}$. The limit value distribution in Def 3 is the historical distribution of the simulated advertisers' predicted CTRs of the 1.8 million impressions.

*Experiment setup.* In this section we aim to produce confidence interval of the sum $\sum_i \beta_i^*$ with the bootstrap estimator Eq (17). Firstly, the sum equals $n$ times the average price-per-utility of advertisers, a measure of efficiency of the system. Secondly, since most quantities in FPPE, such as revenue and social welfare, are smooth functions of pacing multipliers, being able to perform inference about a linear combination of $\beta$'s indicates the ability to infer first-order estimates of those quantities.

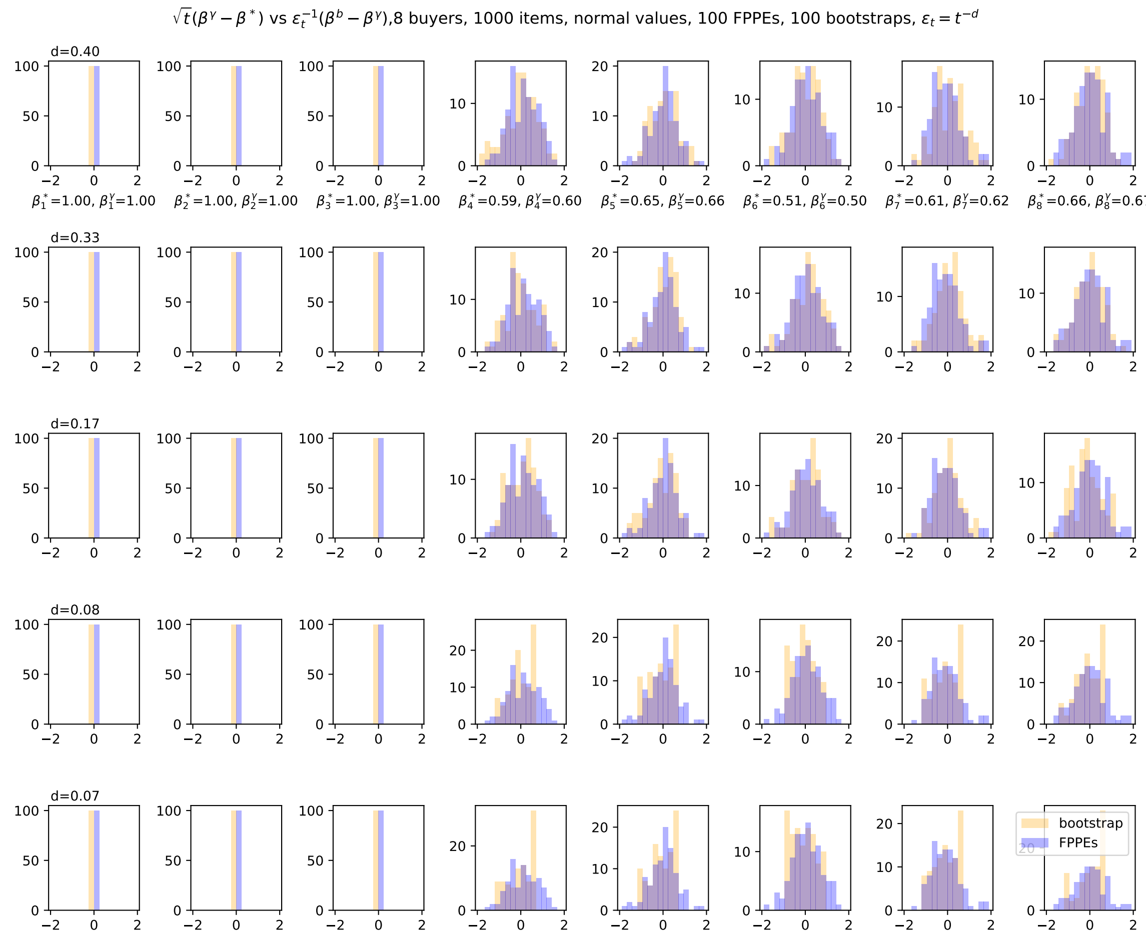


**Figure 4:** Comparison of Bootstrap and FPPE finite item distribution.

The estimator requires an initial consistent estimate of the Hessian matrix, which is implemented with finite difference in Eq (16) with differencing stepsize $\epsilon = t^{-0.4}$. The estimator also requires a bootstrap stepsize $\epsilon_t = t^{-d}$. We try $d$ over the grid $\{0.4, 0.3, 0.2, 0.1, 0.05\}$.

An experiment has parameters $(t, n, d, \alpha)$. Here $t \in \{100, 300, 500\}$ is the number of items and $n \in \{10, 20, 30, 50\}$ the number of advertisers. Parameter $d$ is the exponent of the bootstrap stepsize, and $\alpha \in \{0.1, 0.3, 0.5\}$ is the proportion of advertisers that are not budget-constrained (i.e., $\beta = 1$). To control $\alpha$ in the experiments, we select budgets as follows. Give infinite budgets to the first $\lfloor \alpha n \rfloor$ advertisers. Initialize the rest of the advertisers' budgets randomly, and keep decreasing their budgets until their pacing multipliers are strictly less than 1. For the experiment $(t, n, d, \alpha)$, we first compute the pacing multiplier in the limit market using dual averaging (Xiao, 2010; Gao et al., 2021; Liao et al., 2022). In one simulation of the experiment $(t, n, d, \alpha)$, we sample one FPPE by drawing values from the limit value distribution. Now given one FPPE, we generate bootstrapped pacing multipliers $\{\beta^{b,1}, \dots, \beta^{b,B}\}$ by Eq (17). We calculate the set of sums $S = \{s^{b,1}, \dots, s^{b,B}\}$ where $s^{b,1} = \sum_i \beta_i^{b,1}$ and so on. To obtain a confidence interval with nominal coverage 95%, we let $\ell, u$ be the 2.5% and 97.5% percentiles of $S$. We report the coverage rate and the width of $[\ell, u]$ in Table 3. We perform $B = 100$ bootstrap replications in each simulation. The reported coverage rate for an experiment with parameters $(t, n, d, \alpha)$ is averaged over 100 simulations.

*Results.* For an appropriate choice of $d \in [0.2, 0.3]$, the finite-sample coverage rate agrees with the nominal coverage 95%. Although our theory suggests that as long as $d < 1/2$, the bootstrapped distribution is asymptotically consistent, parameter $d$ does affect finite-sample coverage. Too small a $d$ (for example, 0.10 or 0.05) results in over-coverage and a large $d$ results in under-coverage. We also observe that for $d = 0.4$ and $n = 50$, the finite-sample coverage rate is undesirable for item size $t = 100$. Reassuringly, it increases to a nominal coverage of 95% as item size increases. We also see that the width of the confidence interval decreases as the number of items increases while maintaining nominal coverage. This is expected since the interval width decreases at a rate of $1/\sqrt{t}$. Finally, for appropriately chosen $d$ and item size $t$, the proportion of unpaced advertisers $\alpha$ does not affect finite-sample coverage rates, which demonstrates the robustness of the proposed bootstrap estimator.

**Table 3:** Coverage and width of CI in parentheses. $\alpha$ = proportion of $\beta_i = 1$, $d$ is the exponent in bootstrap stepsize $\epsilon_t = t^{-d}$.

| | $\alpha$ | 0.1 | | | | 0.3 | | | | 0.5 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | buyer | 10 | 20 | 30 | 50 | 10 | 20 | 30 | 50 | 10 | 20 | 30 | 50 |
| $d$ | item | | | | | | | | | | | | |
| 0.40 | 100 | 0.79 (1.07) | 0.82 (1.94) | 0.81 (3.02) | 0.88 (5.17) | 0.82 (1.20) | 0.84 (2.02) | 0.82 (3.47) | 0.3 (4.53) | 0.7 (1.11) | 0.91 (2.18) | 0.86 (2.67) | 0.36 (3.44) |
| | 300 | 0.73 (0.67) | 0.82 (1.14) | 0.85 (1.67) | 0.85 (2.74) | 0.89 (0.83) | 0.76 (1.21) | 0.9 (2.24) | 0.53 (3.17) | 0.82 (0.74) | 0.85 (1.51) | 0.95 (2.04) | 0.49 (2.65) |
| | 500 | 0.78 (0.55) | 0.83 (0.92) | 0.79 (1.34) | 0.89 (2.13) | 0.88 (0.68) | 0.83 (0.97) | 0.88 (1.87) | 0.83 (2.78) | 0.8 (0.60) | 0.85 (1.28) | 0.91 (1.69) | 0.81 (2.39) |
| 0.30 | 100 | 0.94 (1.68) | 0.94 (3.05) | 0.98 (4.53) | 0.97 (7.75) | 0.94 (1.81) | 0.93 (3.01) | 0.97 (4.95) | 0.8 (6.20) | 0.85 (1.67) | 0.95 (3.16) | 0.98 (4.00) | 0.77 (4.89) |
| | 300 | 0.96 (1.21) | 0.96 (2.01) | 0.97 (2.89) | 0.97 (4.69) | 0.96 (1.40) | 0.93 (2.08) | 0.99 (3.76) | 0.99 (5.08) | 0.91 (1.28) | 0.91 (2.39) | 0.99 (3.27) | 0.93 (4.32) |
| | 500 | 0.95 (1.03) | 0.97 (1.71) | 0.98 (2.40) | 1.0 (3.91) | 0.98 (1.24) | 0.98 (1.73) | 0.99 (3.24) | 1.0 (4.69) | 0.93 (1.10) | 0.96 (2.20) | 0.98 (2.90) | 0.99 (4.09) |
| 0.20 | 100 | 0.99 (2.39) | 0.99 (4.48) | 1.0 (6.45) | 0.98 (10.86) | 0.99 (2.63) | 1.0 (4.41) | 0.99 (6.51) | 0.96 (8.26) | 0.96 (2.41) | 1.0 (4.39) | 1.0 (4.94) | 0.96 (6.43) |
| | 300 | 1.0 (1.97) | 0.99 (3.33) | 0.99 (4.77) | 1.0 (7.63) | 0.99 (2.19) | 1.0 (3.40) | 1.0 (5.54) | 1.0 (7.36) | 0.96 (2.11) | 0.98 (3.65) | 0.99 (4.70) | 1.0 (6.01) |
| | 500 | 1.0 (1.80) | 0.99 (3.05) | 0.99 (4.13) | 1.0 (6.57) | 1.0 (2.09) | 0.99 (3.03) | 1.0 (5.14) | 1.0 (7.04) | 0.99 (1.90) | 0.95 (3.57) | 0.99 (4.55) | 1.0 (5.95) |
| 0.10 | 100 | 1.0 (3.36) | 1.0 (5.99) | 1.0 (8.56) | 1.0 (13.93) | 1.0 (3.44) | 1.0 (5.97) | 1.0 (8.04) | 1.0 (10.07) | 0.97 (3.24) | 0.99 (5.30) | 0.99 (6.18) | 0.98 (7.52) |
| | 300 | 1.0 (2.98) | 1.0 (5.01) | 1.0 (6.89) | 1.0 (10.81) | 1.0 (3.20) | 1.0 (4.83) | 1.0 (7.29) | 1.0 (9.09) | 1.0 (3.02) | 0.94 (4.79) | 0.92 (6.04) | 1.0 (7.67) |
| | 500 | 1.0 (2.87) | 1.0 (4.50) | 1.0 (6.41) | 1.0 (9.86) | 1.0 (3.09) | 1.0 (4.60) | 1.0 (6.94) | 1.0 (9.39) | 0.99 (2.85) | 0.98 (4.80) | 0.93 (6.02) | 0.94 (7.52) |
| 0.05 | 100 | 1.0 (3.70) | 1.0 (6.73) | 1.0 (9.61) | 1.0 (15.67) | 1.0 (3.82) | 0.99 (6.59) | 1.0 (8.46) | 1.0 (10.66) | 0.98 (3.67) | 0.95 (5.59) | 0.93 (6.47) | 1.0 (8.10) |
| | 300 | 1.0 (3.47) | 1.0 (5.86) | 1.0 (8.06) | 1.0 (12.21) | 1.0 (3.57) | 1.0 (5.72) | 1.0 (8.06) | 0.99 (10.36) | 1.0 (3.34) | 0.97 (5.48) | 0.9 (6.45) | 0.94 (8.22) |
| | 500 | 1.0 (3.35) | 1.0 (5.45) | 1.0 (7.38) | 1.0 (11.32) | 1.0 (3.61) | 1.0 (5.43) | 1.0 (7.74) | 1.0 (9.93) | 0.99 (3.33) | 0.97 (5.24) | 0.93 (6.50) | 0.98 (8.13) |

## C. Review of Weak Epi-Convergence

The Hoffman-Jørgensen weak convergence theory is a powerful tool to study the asymptotics convergence of statistical functional, especially the argmin functional. To apply the theory to specific applications, one needs to instantiate it with a metric space, usually a space of functions, verify continuity or some form of differentiability of the statistical functional and the weak convergence of certain processes, and then finally invoke continuous mapping theorem or functional delta method. Common choices of metric spaces include the space of bounded functions (on some metric space) with the uniform metric (Van der Vaart, 2000; Kosorok, 2008; Billingsley, 2013), the space of locally bounded functions on $\mathbb{R}^n$ with the topology of uniform convergence on compacta (Kim & Pollard, 1990) or the topology of induced by the hypi-semimetric (Bücher et al., 2014), the space of positive signed measures or point measures on the real line with an appropriate metric (Resnick, 2008), the space of bounded real-valued continuous functions on $\mathbb{R}^n$ with the usual sup-norm (Giné & Nickl, 2008), or the space of cadlag (left limit right continuous) functions on $\mathbb{R}$ with the Skorohod $J_1$ metric (Skorokhod, 1956; Pollard, 2012). To study asymptotics of constrained minimizers, a suitable choice of metric space is the space of extended real-valued lower semi-continuous functions with the metric that induces the topology of epi-convergence. Such an approach dates back to Geyer (1994) and Molchanov (2005), and used in Chernozhukov & Hong (2004) for nonregular models and Chernozhukov et al. (2007) for set estimation, and more recently is used by Parker (2019) for constrained quantile regression, and Hong & Li (2020) and Li (2023) in the context of bootstrap.

To begin with, we introduce the concept of epi-convergence and its probabilistic extensions. Consider

$$\mathscr{L}_n = \{f : \mathbb{R}^n \to \bar{\mathbb{R}} : f \text{ is proper lower semi-continuous (lsc)}\},$$
$$\mathscr{C}\mathscr{S}_n = \{A : A \text{ is a nonempty closed set in } \mathbb{R}^n\}.$$

For $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, we let $\operatorname{epi} f = \{(x, v) \in \mathbb{R}^{n+1} : f(x) \le v\}$ be its epi-graph. Also for $C$ a nonempty closed subset of $\mathbb{R}^{n+1}$, and a point $v \in \mathbb{R}^{n+1}$, let $d_C(v) = \inf\{\|u - v\|_2 : u \in C\}$ be the distance of $v$ to the set $C$, and for nonempty sets $A$ and $B$, define $d_\rho(A, B) = \max\{|d_A(v) - d_B(v)| : \|v\|_2 \le \rho\}$. Define the Attouch-Wets metric on $\mathscr{C}\mathscr{S}_{n+1}$ by

$$d_{\text{AW}}(A, B) = \int_0^\infty d_\rho(A, B) \exp(-\rho) \, \mathrm{d}\rho \tag{26}$$

And for $f, g \in \mathscr{L}_n$, define the metric $d_{\text{epi}}(f, g) = d_{\text{AW}}(\operatorname{epi} f, \operatorname{epi} g)$. In $\mathscr{C}\mathscr{S}_n$, the topology induced by $d_{\text{AW}}$ is equivalent to Wijsman topology and the topology of Painlevé-Kuratowski set convergence (Römisch, 2004). The metric space $(\mathscr{C}\mathscr{S}_n, d_{\text{AW}})$ is complete and separable (Rockafellar & Wets, 2009, Theorem 4.42, Proposition 4.45). Also, the metric space $(\mathscr{L}_n, d_{\text{epi}})$ is complete and separable (Rockafellar & Wets, 2009, Theorem 7.58). We say a sequence of functions $f_t \in \mathscr{L}_n$ epi-converges to $f \in \mathscr{L}_n$ if $d_{\text{epi}}(f_t, f) \to 0$.

**Definition 8** (Epi-convergence in probability). *Let $Z_t : \Omega \times \mathbb{R}^n \to \bar{\mathbb{R}}$ and $Z : \Omega \times \mathbb{R}^n \to \bar{\mathbb{R}}$ be random lsc, extended real-valued functions. We write $Z_t \xrightarrow{\text{epi}} Z$ in probability if for any $\epsilon > 0$ it holds $\mathbb{P}(\omega : d_{\text{epi}}(Z_t(\omega, \cdot), Z(\omega, \cdot)) > \epsilon) \to 0$ as $t \to \infty$.*

**Definition 9** (Epi-convergence in distribution, Knight (1999)). *We say $Z_t \overset{\text{epi}}{\rightsquigarrow} Z$ if for any closed rectangles $R_1, \dots, R_k$ with open interiors $R_1^o, \dots, R_k^o$, it holds the random vector $(\inf_{R_j} Z_t(\cdot), j = 1, \dots, k) \xrightarrow{d} (\inf_{R_j} Z(\cdot), j = 1, \dots, k)$ and $(\inf_{R_j^o} Z_t(\cdot), j = 1, \dots, k) \xrightarrow{d} (\inf_{R_j^o} Z(\cdot), j = 1, \dots, k)$, or equivalently, for any real numbers $a_1, \dots, a_k$,*

$$\begin{aligned}
&\mathbb{P}\left(\inf_{u \in R_1} Z(u) > a_1, \cdots, \inf_{u \in R_k} Z(u) > a_k\right) \\
&\le \liminf_{t \to \infty} \mathbb{P}\left(\inf_{u \in R_1} Z_t(u) > a_1, \cdots, \inf_{u \in R_k} Z_t(u) > a_k\right) \\
&\le \limsup_{t \to \infty} \mathbb{P}\left(\inf_{u \in R_1^o} Z_t(u) \ge a_1, \cdots, \inf_{u \in R_k^o} Z_t(u) \ge a_k\right) \\
&\le \mathbb{P}\left(\inf_{u \in R_1^o} Z(u) \ge a_1, \cdots, \inf_{u \in R_k^o} Z(u) \ge a_k\right)
\end{aligned}$$

*By Corollary 2.4 from Pflug (1991), $Z_t \overset{\text{epi}}{\rightsquigarrow} Z$ is equivalent to the weak convergence of $\operatorname{epi} Z_t$ to $\operatorname{epi} Z$ in the metric space $(\mathscr{C}\mathscr{S}_{n+1}, d_{AW})$.*

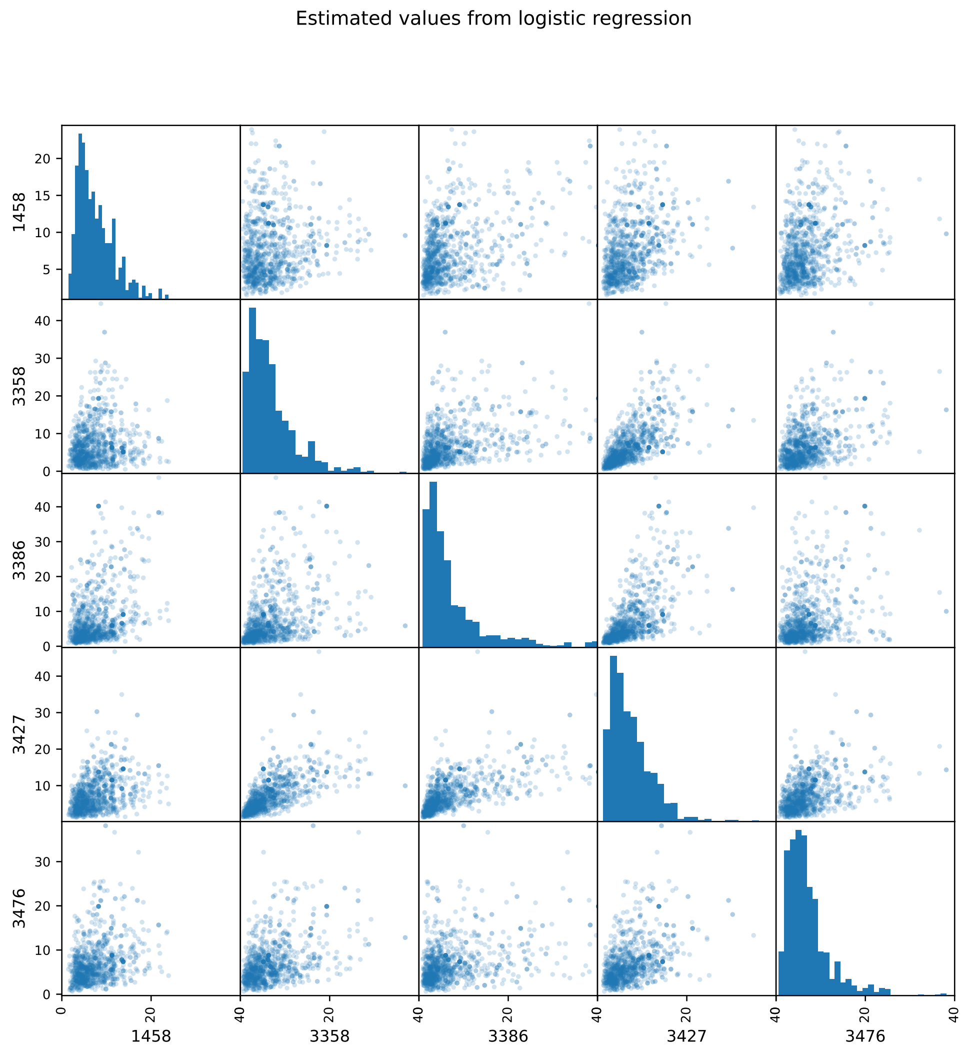


**Figure 5:** Click-through rate (in 0.01%) distributions from logistic regression.

The following lemma is Slutsky's theorem (Kosorok, 2008, Theorem 7.15) specialized to the space $(\mathscr{L}_n, d_{\text{epi}})$.

**Lemma 1.** *Let $(Z_t)_t, (Y_t)_t$ and $Z$ be random lsc, extended real-valued functions and $Y$ be a deterministic element in $\mathscr{L}_n$. If $Y_t \xrightarrow{\text{epi}} Y$ in probability and $Z_t \overset{\text{epi}}{\rightsquigarrow} Z$. Then $Z_t + Y_t \overset{\text{epi}}{\rightsquigarrow} Z + Y$.*

We also introduce a bootstrap version of weak epi-convergence following Section 2.2.3 in Kosorok (2008). A bootstrap version of continuous mapping theorem can also be stated (Hong & Li, 2020).

**Definition 10** (Conditional weak epi-convergence in probability)**.** *Let $BL$ denote the space of Lipschitz functions $f : (\mathscr{L}_n, d_{AW}) \to \mathbb{R}$ with Lipschitz parameter equal 1, i.e., $\sup|f| \le 1$ and $|f(x) - f(y)| \le d_{AW}(x, y)$. And suppose $Z_t = Z_t(X, W)$ is defined on a product probability space, where $W$ represents bootstrap weights, and $X$ represents data. The process $Z_t$ converges to $Z$ in the sense of weak epi-convergence conditionally in probability if $\sup_{f\in BL} |\mathbb{E}_W[f(Z_t)|X] - \mathbb{E}[f(Z)]| \to 0$ in probability, along with certain measurability conditions.*

## D. Proofs

### D.1. Stochastic Equicontinuity Results for the EG Objective

Let $\epsilon_t = o(1)$ and $K$ be a compact set. Let $D_F(\theta, \beta) \in \partial F(\theta, \beta)$ be a deterministic element of the subgradient. Note by SMO $D^*_F(\cdot) = D_F(\cdot, \beta^*) = \nabla F(\cdot, \beta^*)$. We also let $F(\beta) = F(\cdot, \beta)$. Note that in the following claims, we do not need $\nabla H(\beta^*) = 0$. They work for any $\beta^*$ at which $H$ is continuously differentiable in a neighborhood.

**Claim 1.**

$$\sup_{h\in K} (P_t - P)(F(\beta^* + \epsilon_t h) - F(\beta^*) - \epsilon_t h^\top D^*_F(\cdot)) = o_p(\epsilon_t/\sqrt{t})$$
$$\sup_{h\in K} (P^{\text{ex},b}_t - P_t)(F(\beta^* + \epsilon_t h) - F(\beta^*) - \epsilon_t h^\top D^*_F(\cdot)) = o_p(\epsilon_t/\sqrt{t})$$

*Proof of Claim 1.* Let $r_{1,F}(\cdot, \beta) = F(\cdot, \beta) - F(\cdot, \beta^*) - D^*_F(\cdot)^\top(\beta - \beta^*)$.

By SMO there is a neighborhood of $\beta^*$, say $N$, on which $H$ is differentiable. Then for any $\beta \in N$, the set $\{\theta : \beta \mapsto f(\theta, \beta)$ differentiable at $\beta\}$ is measure one. Choose $t$ large enough so that the ball $\{\beta : \|\beta - \beta^*\|_2 \le \delta_t\}$ is contained in $N$. By a mean value theorem for locally Lipschitz functions, (Clarke, 1990, Theorem 2.3.7), it holds $(P_t - P)(F(\beta) - F(\beta^*)) = \zeta^\top(\beta - \beta^*)$ where $\zeta \in \partial(P_t - P)F(\tilde\beta)$ and $\tilde\beta$ lies on the segment joining $\beta$ and $\beta^*$. By $\tilde\beta \in N$, it holds $\zeta = (P_t - P)D_F(\cdot, \tilde\beta)$. Then the desired claim is equivalent to

$$\begin{aligned}
&\sup_{\|\beta-\beta^*\|_2\le\delta_t} \frac{(P_t - P)r_{1,F}(\cdot, \beta)}{\frac{1}{\sqrt{t}}\|\beta - \beta^*\|_2} \\
=& \sup_{\|\beta-\beta^*\|_2\le\delta_t} \frac{(P_t - P)(D_F(\cdot, \tilde\beta) - D_F(\cdot, \beta^*))^\top(\tilde\beta - \beta^*)}{\frac{1}{\sqrt{t}}\|\beta - \beta^*\|_2} \\
\le& \sup_{\|\beta-\beta^*\|_2\le\delta_t} \|\sqrt{t}(P_t - P)(D_F(\cdot, \tilde\beta) - D_F(\cdot, \beta^*))\|_2 = o_p(1)
\end{aligned} \tag{27}$$

where the last equality is due to Liao & Kroer (2023).

The assumption on the bootstrap weights (Def 5) implies that a bootstrap version of Eq (27) holds, i.e., $\sup_{\|\beta-\beta^*\|_2\le\delta_t} \|\sqrt{t}(P^{\text{ex},b}_t - P_t)(D_F(\cdot, \tilde\beta) - D_F(\cdot, \beta^*))\|_2 = o_p(1)$ (Wellner & Zhan, 1996, Lemma 4.1). The same argument goes through for the proof of bootstrap differentiability. We finish the proof of the lemma. ☐

**Claim 2.**

$$\sup_{\|h-s\|_2=o(1),s,h\in K} (P_t - P)(F(\beta^* + \epsilon_t h) - F(\beta^* + \epsilon_t s)) = o_p(\epsilon_t/\sqrt{t})$$
$$\sup_{\|h-s\|_2=o(1),s,h\in K} (P^{\text{ex},b}_t - P_t)(F(\beta^* + \epsilon_t h) - F(\beta^* + \epsilon_t s)) = o_p(\epsilon_t/\sqrt{t})$$

*Proof.* This is implied by Claim 1. ☐

**Claim 3.**

$$\sup_{h\in K} P_t(F(\beta^* + \epsilon_t h) - F(\beta^*) - \epsilon_t h^\intercal D^*(\cdot) - \epsilon_t^2 h^\intercal \mathcal{H} h) = o_p(\epsilon_t/\sqrt{t} + \epsilon_t^2)$$

*Proof.* Let $r_{2,F}(h) = F(\cdot, \beta^* + h) - F(\cdot, \beta^*) - D_F(\cdot, \beta^*)^\intercal h - \frac{1}{2} h^\intercal \mathcal{H} h$.

Split the LHS by $\sup_h |P_t r_{2,F}(\epsilon_t h)| \le \sup_h |P r_{2,F}(\epsilon_t h)| + \sup_h |(P_t - P) r_{2,F}(\epsilon_t h)|$. The first term is $o(\epsilon_t^2)$ by twice differentiability. The second term is bounded by $o_p(\epsilon_t/\sqrt{t})$ as in Claim 1. □

**Notations in the proof sections**. Define the demeaned and the centered function $\tilde{F}(\cdot, \beta) = F(\cdot, \beta) - \mathbb{E}[F(\theta, \beta)]$ and $\bar{F}(\cdot, \beta) = F(\cdot, \beta) - F(\cdot, \beta^*)$. Let $H^b(\beta) = P_t^b F(\cdot, \beta)$ be the bootstrapped EG objective.

### D.2. Proof of Thm 1

*Proof of Thm 1.* In the proof, we use $\beta^b$ to denote the exchangeable bootstrap estimator Eq (9) and use $P_t^b$ to denote the bootstrap empirical operator with exchangeable weights (Def 5).

*Step 1*. Show $\beta^b \xrightarrow{p} \beta^*$. The consistency of the bootstrap estimator is implied by uniform convergence of $H^b(\cdot)$ to $H(\cdot)$ and uniqueness of $\beta^*$. For proof, we refer readers to the proof of Theorem 3.5 from Giné (1992).

*Step 2*. Show $\beta^b - \beta^* = O_p(1/\sqrt{t})$.

Define

$$\Delta^\gamma = \mathcal{H}^{-1}(P_t - P) D_F(\cdot, \beta^*),$$
$$\Delta^b = \mathcal{H}^{-1}(P_t^b - P_t) D_F(\cdot, \beta^*).$$

Let $r_{1,F}(\cdot, \beta) = F(\cdot, \beta) - F(\cdot, \beta^*) - D_F^*(\cdot)^\intercal(\beta - \beta^*)$, $D_F^*(\cdot) = D_F(\cdot, \beta^*) = \nabla F(\cdot, \beta^*)$.

We begin with the optimality of $\beta^b$ and then apply the definition of $r$. For ease of notation, we let $F(\beta) = F(\cdot, \beta)$. We have

$$\begin{aligned}
0 &\ge P_t^{\mathsf{ex},b}(F(\beta^b) - F(\beta^*)) \\
&= (P_t^{\mathsf{ex},b} - P_t + P_t - P)(F(\beta^b) - F(\beta^*)) + P(F(\beta^b) - F(\beta^*)) \\
&= (\Delta^b + \Delta^\gamma)^\intercal \mathcal{H}(\beta^b - \beta^*) + P(F(\beta^b) - F(\beta^*)) \qquad (28)\\
&\quad + (P_t^{\mathsf{ex},b} - P_t + P_t - P) r_{1,F}(\cdot, \beta^b) \\
&\ge (\Delta^b + \Delta^\gamma + o_p(1))^\intercal(\beta^b - \beta^*) + c \cdot \|\beta^b - \beta^*\|_2^2 \qquad (29)
\end{aligned}$$

where in the last inequality we used (i) $(P_t^{\mathsf{ex},b} - P_t) r_{1,F}(\cdot, \beta^b) = o_p(\frac{1}{\sqrt{t}} + \|\beta^b - \beta^*\|_2) \cdot \|\beta^b - \beta^*\|_2 = o_p(1)\|\beta^b - \beta^*\|_2$ by Claim 1, (ii) $(P_t - P) r_{1,F}(\cdot, \beta^b) = o_p(1)\|\beta^b - \beta^*\|_2$ by Claim 1, and (iii) $\beta \mapsto PF(\cdot, \beta)$ is locally strongly convex at $\beta^*$, and so there is a neighborhood of $\beta^*$ and a constant $c > 0$ such that $P(F(\beta) - F(\beta^*)) \ge c\|\beta - \beta^*\|_2^2$ for all $\beta$ in this neighborhood. The expression Eq (29) now becomes $0 \ge O_p(t^{-1/2})\|\beta^b - \beta^*\|_2 + c\|\beta^b - \beta^*\|_2^2$. Since the case $\beta^b - \beta^* = 0$ can be easily excluded, we divide both sides by $\|\beta^b - \beta^*\|_2$ and conclude that $(\beta^b - \beta^*) = O_p(1/\sqrt{t})$.

*Step 3*. Find the asymptotic distribution. Since $\beta^b$ is the minimizer of $P_t^{\mathsf{ex},b} F$ over $\mathbb{R}_+^n$, defining $\bar{F}(\beta) = F(\cdot, \beta) - F(\cdot, \beta^*)$, we have

$$\begin{aligned}
0 &\ge P_t^{\mathsf{ex},b}(\bar{F}(\beta^b) - \bar{F}(\Delta^b + \Delta^\gamma + \beta^*)) \\
&= [(P_t^{\mathsf{ex},b} - P_t) + (P_t - P)](\bar{F}(\beta^b) - \bar{F}(\Delta^b + \Delta^\gamma + \beta^*)) + P(\bar{F}(\beta^b) - \bar{F}(\Delta^b + \Delta^\gamma + \beta^*)) \\
&= (\Delta^b + \Delta^\gamma)^\intercal \mathcal{H}(\beta^b - \beta^* - (\Delta^b + \Delta^\gamma)) + \frac{1}{2}\|\beta^b - \beta^*\|_{\mathcal{H}}^2 - \frac{1}{2}\|\Delta^b + \Delta^\gamma\|_{\mathcal{H}}^2 \\
&\quad + [(P_t^{\mathsf{ex},b} - P_t) + (P_t - P)](r_{1,F}(\cdot, \beta^b) - r_{1,F}(\cdot, \Delta^b + \Delta^\gamma + \beta^*)) \\
&= \frac{1}{2}\|\Delta^b + \Delta^\gamma + (\beta^b - \beta^*)\|_{\mathcal{H}}^2 + o_p(1/t)
\end{aligned}$$

where in the last line we used (i) $(P_t^{\mathsf{ex},b} - P_t) r_{1,F}(\cdot, \beta) = o_p(1/t)$ for any random $\beta$ such that $\beta = \beta^* + O_p(1/\sqrt{t})$ by Claim 1, (ii) $(P_t - P) r_{1,F}(\cdot, \beta) = o_p(1/t)$ for any random $\beta$ such that $\beta = \beta^* + O_p(1/\sqrt{t})$ by Claim 1, and (iii) $P\bar{F}(\beta) = \frac{1}{2}\|\beta - \beta^*\|^2_{\mathcal{H}} + o(\|\beta - \beta^*\|_2^2)$ for $\beta \to \beta^*$ due to $\nabla H(\beta^*) = 0$.

Rearranging gives $t\|\Delta^b + \Delta^\gamma + (\beta^b - \beta^*)\|_2^2 = o_p(1)$. Next , using $\beta^\gamma - \beta^* = -\Delta^\gamma + o_p(1/\sqrt{t})$ (Liao & Kroer, 2023) we have

$$\sqrt{t}(\beta^b - \beta^\gamma) = -\sqrt{t}\Delta^b + o_p(1)$$

By an exchangeable bootstrap CLT (Præstgaard & Wellner, 1993), we know $\sqrt{t}\Delta^b \xrightarrow{d} c \cdot \mathcal{H}^{-1}\mathcal{N}(0, \mathbb{E}[\nabla F(\cdot, \beta^*)\nabla F(\cdot, \beta^*)^\intercal])$ conditional on almost all data sequence $\gamma$, where $c$ is the constant defined in Def 5. This concludes the proof. □

### D.3. Proof of Numerical Bootstrap (Thm 7 and Thm 3.1)

**Theorem 7.** *Let $\epsilon_t = o(1)$ and $\epsilon_t\sqrt{t} \to \infty$. Then $\epsilon_t^{-1}(\beta^b_{\mathsf{nu,LFM}} - \beta^\gamma) \overset{p}{\rightsquigarrow} \mathcal{J}_{\mathsf{LFM}}$.*

*Proof of Thm 7 .* We verify the assumptions in Theorem 4.1 of the numerical bootstrap paper (Hong & Li, 2020).

For Thm 7, let $B = \mathbb{R}^n_{++}$, $\beta^b$ be $\beta^b_{\mathsf{nu,LFM}}$, and $\beta^*$ be the equilibrium pacing multiplier in Fisher market. We restate the assumptions in Theorem 4.1 of (Hong & Li, 2020) in our notations.

(i) $H_t(\beta^\gamma) \le \inf_B H_t + o_p(1/t)$, and $H^b_t(\beta^b) \le \inf_B H^b_t + o^*_p(\epsilon_t^2)$.

(ii) $\beta^\gamma \xrightarrow{p} \beta^*$, and $\beta^b - \beta^\gamma = o^*_p(1)$.

(iii) $\beta^*$ is an interior point of $B$.

(iv) The class $\{F(\cdot, \beta) - F(\cdot, \beta^*) : \|\beta - \beta^*\|_2 \le R\}$ is uniformly manageable.

(v) $H$ is twice differentiable at $\beta^*$ with positive definite Hessian $\mathcal{H}$.

(vi) The limit $\Sigma(s,t) = \lim_{\epsilon\downarrow 0} \frac{1}{\epsilon^2}\mathbb{E}\big[\big(\tilde{F}(\cdot, \beta^* + \epsilon s) - \tilde{F}(\cdot, \beta^*)\big)\big(\tilde{F}(\cdot, \beta^* + \epsilon t) - \tilde{F}(\cdot, \beta^*)\big)\big]$ exists for each $s, t$.

(vii) For all $\delta > 0$ it holds $\lim_{\epsilon\downarrow 0} \frac{1}{\epsilon}\mathbb{E}[(F(\cdot, \beta^* + \epsilon s) - F(\cdot, \beta^*))\mathbb{1}(|F(\cdot, \beta^* + \epsilon s) - F(\cdot, \beta^*)| > \delta)] = 0$, where $\tilde{F}(\cdot, \beta) = F(\cdot, \beta) - \mathbb{E}[F(\theta, \beta)]$.

(viii) Let $G_R(\cdot) = \sup_{\|\beta - \beta^*\|_2 \le R} |F(\cdot, \beta) - F(\cdot, \beta^*)|$. As $R \to 0$, $\mathbb{E}[G_R^2] = O(R^2)$.

(ix) $\sqrt{t}\epsilon_t \to \infty$ and $\epsilon_t \downarrow 0$.

(x) $\mathbb{E}[G_R^2 \mathbb{1}(RG_R > \eta)] = o(R^2)$ for all $\eta > 0$.

(xi) There is a neighborhood of $\beta^*$ such that $\mathbb{E}[|F(\cdot, \beta) - F(\cdot, \beta')|] = O(\|\beta - \beta'\|_2^2)$ for $\beta, \beta'$ in this neighborhood.

Implicit in the paper, it is also required that the gradient of the population objective is zero at optimum. This is true for Fisher market.

We now verify these conditions. Condition (i) holds because we consider exact minimizers. Condition (ii): The $\beta^\gamma - \beta^* \xrightarrow{p} 0$ part has been verified in Liao & Kroer (2023). It remains to show the $\beta^b - \beta^\gamma \xrightarrow{p} 0$ part. Since $\beta \mapsto P^b_t F(\cdot, \beta)$ converges to $H$ in probability pointwise and that $\beta \mapsto P^b_t F(\cdot, \beta)$ is convex, it holds that the convergence is uniform over compact sets. By uniqueness of $\beta^*$ it holds $\beta^b \xrightarrow{p} \beta^*$. Condition (iii) naturally holds for the linear Fisher market. Condition (iv) requires that there is an $R_0 > 0$ such that the function class $\{F(\cdot, \beta) - F(\cdot, \beta^*) : \|\beta - \beta^*\|_2 \le R\}$ is uniformly manageable for all $R \le R_0$. It holds because the class $\{\theta \mapsto f(\theta, \beta) - f(\theta, \beta^*) : \beta \in B\}$ is uniformly manageable, which is verified in Liao & Kroer (2023). Condition (v) is assumed in SMO. In condition (vi) the function $\Sigma$ is the covariance kernel for some

Gaussian process. We show $\Sigma(s,t) = s^\top \mathbb{E}[\nabla\tilde F(\cdot,\beta^*)\nabla\tilde F(\cdot,\beta^*)^\top]t = s^\top\mathbb{E}[\nabla F(\cdot,\beta^*)\nabla F(\cdot,\beta^*)^\top]t$. By the dominated convergence theorem, we can pass the limit inside the expectation and obtain

$$\begin{aligned}&\lim_{\epsilon\downarrow 0}\mathbb{E}[(\tilde F(\cdot,\beta^*+\epsilon t)-\tilde F(\cdot,\beta^*))(\tilde F(\cdot,\beta^*+\epsilon s)-\tilde F(\cdot,\beta^*))]/\epsilon^2\\ &= t^\top\mathbb{E}[\nabla\tilde F(\cdot,\beta^*)\nabla\tilde F(\cdot,\beta^*)^\top]s\end{aligned}$$

Condition (vii): by dominated convergence theorem we can move the limit inside expectation.

$$\begin{aligned}&\lim_{\epsilon\downarrow 0}\left|\frac{1}{\epsilon}\mathbb{E}[(F(\cdot,\beta^*+\epsilon s)-F(\cdot,\beta^*))\mathbb{1}(|F(\cdot,\beta^*+\epsilon s)-F(\cdot,\beta^*)|>\delta)]\right|\\ &\le \mathbb{E}\left[\lim_{\epsilon\downarrow 0}\left|\frac{1}{\epsilon}(F(\cdot,\beta^*+\epsilon s)-F(\cdot,\beta^*))\right|\mathbb{1}(|F(\cdot,\beta^*+\epsilon s)-F(\cdot,\beta^*)|>\delta)\right]\\ &\le L\|s\|_2\mathbb{E}\Big[\lim_{\epsilon\downarrow 0}\mathbb{1}(|F(\cdot,\beta^*+\epsilon s)-F(\cdot,\beta^*)|>\delta)\Big]=0\end{aligned}$$

where the last equality holds because $\beta\mapsto F(\theta,\beta)$ is continuous for all $\theta$. Now we show conditions (viii) and (x). Consider the set $\prod_{i=1}^n[\beta_i^*/2,1]$. On this set, for all $\theta$, the function $\beta\mapsto F(\theta,\beta)$ is Lipschitz with parameter $L=\bar v+2\sqrt{n}$ w.r.t. $\ell_2$ norm. So for $R$ small enough the ball $\{\beta:\|\beta-\beta^*\|_2\le R\}$ is contained in the set $\prod_{i=1}^n[\beta_i^*/2,1]$. The first requirement follows by noting $\mathbb{E}[G_R^2]\le L^2R^2=O(R^2)$. The second requirement follows from the arguments $\lim_{R\to 0}\mathbb{E}[\frac{G_R^2}{R^2}\mathbb{1}(RG_R>\eta)]\le\mathbb{E}[L^2\cdot\lim_{R\to 0}\mathbb{1}(G_R>\eta/R)]=0$ where the last equality holds due to $G_R$ being bounded for $R$ small enough. Condition (ix) requires that $\epsilon_t\sqrt{t}\to\infty$, which is assumed. Condition (xi): Inspecting the proof this condition is used to show Claim 2, which we proved separately.

Now we can invoke Theorem 4.1 from Hong & Li (2020) and conclude $\epsilon_t^{-1}(\beta^b_{\mathsf{nu,LFM}}-\beta^\gamma)\overset{p}{\rightsquigarrow}\mathcal{J}_{\mathsf{LFM}}$. □

*Proof of Thm 3.1.* For Thm 3.1, let $B=[0,1]^n$. Let $\beta^b$ refer to $\beta^b_{\mathsf{nu,FPPE}}$, and $\beta^*$ be the equilibrium pacing multiplier in FPPE. As before, it is implicitly assumed in Hong & Li (2020) that the gradient of $H$ equals zero at $\beta^*$. This is true for FPPE if and only if all buyers spend their budgets. Theorem 4.2 from Hong & Li (2020) requires that all conditions stated above except (i) and (iii) hold, and that $\beta^*$ uniquely minimizes $H$ over $B$, which is true in FPPE. Now we can invoke Theorem 4.2 from Hong & Li (2020) and conclude $\epsilon_t^{-1}(\beta^b_{\mathsf{nu,FPPE}}-\beta^\gamma)\overset{p}{\rightsquigarrow}\mathcal{J}_{\mathsf{FPPE}}$. Note that under the assumption that all buyers spend their budgets, the limit distribution simplifies to $\mathcal{J}_{\mathsf{FPPE}}=\arg\min_{h\in C}G^\top h+\frac{1}{2}h^\top\mathcal{H}h$ where $G\sim\mathcal{N}(0,\mathbb{E}[\nabla F(\cdot,\beta^*)\nabla F(\cdot,\beta^*)^\top])$ and $C=\{h\in\mathbb{R}^n: h_i\le 0, i\in I_0\}$ □

### D.4. Proof of Thm 5 and formal statement

Below we start by giving a more formal version of Thm 5 and then prove it.

**Theorem 9** (Hessian estimation)**.** *Assume $H(\cdot)$ is four times continuously differentiable in a neighborhood of $\beta^*$. Consider the finite difference estimate defined in Eq (16) with differencing stepsize $\eta_t=o(1)$ and $\eta_t\sqrt{t}\to\infty$. For some intermediate quantity $\check H_{k\ell}$, it holds $\widehat{\mathcal{H}}_{k\ell}-\check{\mathcal{H}}_{k\ell}=o_p(\eta_t^2+\frac{1}{\sqrt{t\eta_t^2}})+O_p(\frac{1}{\sqrt t})$, $\mathbb{E}[(\check{\mathcal{H}}_{k\ell}-\mathcal{H}_{k\ell})^2]=\Theta(\eta_t^4+\frac{1}{t\eta_t^2})+o(\eta_t^4+\frac{1}{t\eta_t^2})$ where the $O_p(1/\sqrt{t})$ part does not depend on $\eta_t$. Proof in App D.4.*

*Proof of Thm 5 and Thm 9.* The proof follows the idea in Lemma 2 from Cattaneo et al. (2020). The main difference is their result is for cube-root asymptotics, while our setting is the usual square-root asymptotics. Define

$$\check{\mathcal{H}}_{k\ell}=(\widehat\nabla^2_{k\ell,\eta_t}H_t)(\beta^*),\quad \bar{\mathcal{H}}_{k\ell}(\beta)=(\widehat\nabla^2_{k\ell,\eta_t}H)(\beta)$$

Then $\widehat{\mathcal{H}}_{k\ell}-\check{\mathcal{H}}_{k\ell}=R+S$ where

$$R=\widehat{\mathcal{H}}_{k\ell}-\check{\mathcal{H}}_{k\ell}-\bar{\mathcal{H}}_{k\ell}(\beta^\gamma)+\bar{\mathcal{H}}_{k\ell}(\beta^*)=\widehat\nabla^2_{k\ell,\eta_t}[(H_t-H)(\beta^\gamma)-(H_t-H)(\beta^*)],\quad S=\bar{\mathcal{H}}_{k\ell}(\beta^\gamma)-\bar{\mathcal{H}}_{k\ell}(\beta^*).$$

It will be shown that (i) $R=o_p(1/(\sqrt{t}\eta_t))$, (ii) $S=O_p(\frac{1}{\sqrt t})+o_p(\eta_t^2)$, and (iii) $\mathbb{E}[(\mathcal{H}_{k\ell}-\check{\mathcal{H}}_{k\ell})^2]=\Theta(\eta_t^4+\frac{1}{t\eta_t^2})+o(\eta_t^4+\frac{1}{t\eta_t^2})$.

To show (i), by Claim 2,

$$\eta_t^{-1}\sqrt{t}\sup_{T_t}(P_t-P)(F(\cdot,\beta^*+t_1\eta_t)-F(\cdot,\beta^*+t_2\eta_t))=o_p(1) \tag{30}$$

where the supremum runs over the set $T_t=\{(t_1,t_2):\|t_1\|,\|t_2\|\le M,\|t_1-t_2\|\le\delta_t\}$ for some $M>0$ and $\delta_t\downarrow 0$. Alternatively, the claim Eq (30) can be proved following the proof of Theorem 4.1 from Hong & Li (2020) or Lemma 4.6 from Kim & Pollard (1990). With this bound, letting $\delta=(\beta^\gamma-\beta^*)/\eta_t=o_p(1)$, terms such as

$$\begin{aligned}&(P_t-P)(F(\cdot,\beta^\gamma+\eta_t(e_k+e_\ell))-F(\cdot,\beta^*+\eta_t(e_k+e_\ell)))\\&=(P_t-P)(F(\cdot,\beta^*+\eta_t(\underbrace{\delta+e_k+e_\ell}_{t_1}))-F(\cdot,\beta^*+\eta_t(\underbrace{e_k+e_\ell}_{t_2})))\\&\le\sup_{T_t}(P_t-P)(F(\cdot,\beta^*+t_1\eta_t)-F(\cdot,\beta^*+t_2\eta_t))\\&=o_p(\eta_t/\sqrt{t})\end{aligned}$$

can be upper bounded as above. And so $R=\frac{1}{\eta_t^2}o_p(\eta_t/\sqrt{t})=o_p(1/\sqrt{t\eta_t^2})$.

To show (ii), by Taylor's theorem, $S=(\nabla_\beta\nabla^2_{k\ell}H(\beta)|_{\beta=\beta^*})^\intercal(\beta^\gamma-\beta^*)+o_p(\eta_t^2)=O_p(\frac{1}{\sqrt{t}})+o_p(\eta_t)$, and the $O_p(1/\sqrt{t})$ term does not involve $\eta_t$.

Finally, to show (iii), we calculate the bias and variance of $\check{\mathcal{H}}_{k\ell}$. Let $d(\cdot)=F(\cdot,\beta^*+\eta_t(e_k+e_\ell))-F(\cdot,\beta^*+\eta_t(-e_k+e_\ell))-F(\cdot,\beta^*+\eta_t(e_k-e_\ell))+F(\cdot,\beta^*+\eta_t(-e_k-e_\ell))$. Then $\check{\mathcal{H}}_{k\ell}=\frac{1}{4\eta_t^2}P_td(\cdot)$. For the bias, following the proof of Lemma 2 from Cattaneo et al. (2020), by Taylor's theorem,

$$\mathbb{E}[\check{\mathcal{H}}_{k\ell}-\mathcal{H}_{k\ell}]=\frac{1}{6}(\nabla_k^2\nabla_{k\ell}^2H(\beta^*)+\nabla_\ell^2\nabla_{k\ell}^2H(\beta^*))\eta_t^2+o(\eta_t^2)$$

To see this, let $g(\epsilon)=H(\beta^*+\epsilon h)-H(\beta^*)$. Then $g(0)=0$. Also let $\mathcal{H}(\beta)=\nabla^2H(\beta)$. Now $g^{(1)}(\epsilon)=\nabla H(\beta^*+\epsilon h)^\intercal h$, $g^{(2)}(\epsilon)=h^\intercal\mathcal{H}(\beta^*+\epsilon h)h$, $g^{(3)}(\epsilon)=\sum_{i=1}^n h_i^2h^\intercal\nabla_\beta\mathcal{H}_{ii}(\beta^*+\epsilon h)+2\sum_{i<j}h_ih_jh^\intercal\nabla_\beta\mathcal{H}(\beta^*+\epsilon h)$, and $g^{(4)}=\sum_{i=1}^n h_i^2h^\intercal\nabla_\beta^2\mathcal{H}_{ii}(\beta^*+\epsilon h)h+2\sum_{i<j}h_ih_jh^\intercal\nabla^2_\beta\mathcal{H}(\beta^*+\epsilon h)h$.

For the variance, note $\mathrm{Var}(\check{\mathcal{H}}_{k\ell})=\frac{1}{16t\eta_t^4}\mathrm{Var}(d(\cdot))=\frac{1}{16t\eta_t^4}\mathbb{E}[d(\cdot)^2]+O(1/t)$. Next,

$$\eta_t^{-2}\mathbb{E}[(F(\cdot,\beta^*+\eta_t(e_k+e_\ell))-F(\cdot,\beta^*))^2]\to(e_k+e_\ell)^\intercal\mathbb{E}[\nabla F(\cdot,\beta^*)^{\otimes 2}](e_k+e_\ell)$$

and so $\mathbb{E}[d(\cdot)^2]=\Theta(\eta_t^2)$. Conclude that

$$\mathrm{Var}(\check{\mathcal{H}}_{k\ell})=\Theta(\frac{1}{t\eta_t^2})+O(1/t).$$

□

### D.5. Proof of Proximal Bootstrap (Thm 8 and Thm 3.2)

*Proof of Thm 8.* As $t\to\infty$, $\epsilon_t^{-1}(\beta^b-\beta^\gamma)=-\widehat{\mathcal{H}}^{-1}G^b$ with probability approaching 1 due to $\epsilon_t=o(1)$ and $\epsilon_t\sqrt{t}\to\infty$. Next, $G^b=\sqrt{t}(P_t^b-P_t)D_F(\cdot,\beta^\gamma)=\sqrt{t}(P_t^b-P_t)(D_F(\cdot,\beta^\gamma)-D_F(\cdot,\beta^*))+\sqrt{t}(P_t^b-P_t)D_F(\cdot,\beta^*)=\sqrt{t}(P_t^b-P_t)D_F(\cdot,\beta^*)+o_p(1)\xrightarrow{d}\mathcal{N}(0,\mathbb{E}[\nabla F(\cdot,\beta^*)\nabla F(\cdot,\beta^*)^\intercal])$ conditional on data, by Wellner & Zhan (1996, Lemma 4.1). Also $\widehat{\mathcal{H}}\xrightarrow{p}\mathcal{H}$ and by continuity of matrix inverse, $\widehat{\mathcal{H}}^{-1}\xrightarrow{p}\mathcal{H}^{-1}$. We conclude $\epsilon_t^{-1}(\beta^b-\beta^\gamma)$ converges in distribution to $\mathcal{J}_{\mathsf{LFM}}$ conditionally in probability. □

*Proof of Thm 3.2.* We present proof following the idea in the working paper of Li (2023). In fact, most of the conditions required in that paper have been established in Liao & Kroer (2023), such as stochastic equicontinuity of certain processes.

Let $B=[0,1]^n$. Let $\beta^b$ refer to $\beta^b_{\mathsf{nu,FPPE}}$, and $\beta^*$ be the equilibrium pacing multiplier in FPPE. It is assumed in Li (2023) that the gradient of $H$ equals zero at $\beta^*$. This is true for FPPE if and only if all buyers spend their budgets, i.e., $I_+=\emptyset$.

Step 1. Show $\beta^b \xrightarrow{p} \beta^*$. Since $\epsilon_t \to 0$ and $G^b = O_p(1)$, we have $\epsilon_t G^b = o_p(1)$. Then for each $\beta \in \mathbb{R}^n$

$$\begin{aligned}
&\epsilon_t (G^b)^\top \beta + \frac{1}{2}\|\beta^\gamma - \beta\|^2_{\widehat{\mathcal{H}}} + \chi(\beta \in [0,1]^n) \\
&= \frac{1}{2}\|\beta^* - \beta\|^2_{\widehat{\mathcal{H}}} + \chi(\beta \in [0,1]^n) + \left(\epsilon_t (G^b)^\top \beta + (\beta^* - \beta)^\top \widehat{\mathcal{H}}(\beta^\gamma - \beta^*) + \frac{1}{2}\|\beta^\gamma - \beta^*\|^2_{\widehat{\mathcal{H}}}\right) \\
&= \frac{1}{2}\|\beta^* - \beta\|^2_{\widehat{\mathcal{H}}} + \chi(\beta \in [0,1]^n) + o_p(1) \\
&\xrightarrow{p} \frac{1}{2}\|\beta^* - \beta\|^2_{\mathcal{H}} + \chi(\beta \in [0,1]^n)
\end{aligned}$$

By convexity, it implies uniform convergence on compact sets in probability. Since $\beta^*$ uniquely minimize $\beta \mapsto \|\beta^* - \beta\|^2_{\mathcal{H}} + \chi(\beta \in [0,1]^n)$, we conclude $\beta^b \xrightarrow{p} \beta^*$ (Newey & McFadden, 1994, Theorem 2.7).

Step 2. Identify the limit distribution of $\beta^b$. Note that when $I_+ = \emptyset$, $\mathbb{E}[\nabla F(\cdot, \beta^*)] = 0$. Define $X_t(h) = (G^b)^\top (h + \frac{\beta^* - \beta^\gamma}{\epsilon_t}) + \frac{1}{2}\|h + \frac{\beta^* - \beta^\gamma}{\epsilon_t}\|^2_{\widehat{\mathcal{H}}}$. We first show $X_t(h) \rightsquigarrow G^\top h + \frac{1}{2} h^\top \mathcal{H} h$ in $\ell^\infty(K)$ for any compact $K \subset \mathbb{R}^n$, where $G \sim \mathcal{N}(0, \mathbb{E}[\nabla F(\cdot, \beta^*) \nabla F(\cdot, \beta^*)^\top])$. The proof is identical to the proof of Claim 4 and is omitted here. Next, by a change of variable $h = \epsilon_t^{-1}(\beta - \beta^*)$, the inclusion $\beta \in [0,1]$ becomes $h \in ([0,1]^n - \beta^*)/\epsilon_t$, and

$$\epsilon_t^{-1}(\beta^b - \beta^*) = \operatorname*{arg\,min}_{h \in ([0,1]^n - \beta^*)/\epsilon_t} X_t(h) \xrightarrow{d} \operatorname*{arg\,min}_{h \in \mathbb{R}^n : h_i \le 0, i \in I_0} G^\top h + \frac{1}{2} h^\top \mathcal{H} h = \mathcal{J}_{\mathsf{FPPE}}$$

where the last equality uses $I_+ = \emptyset$ and thus $([0,1]^n - \beta^*)/\epsilon_t \xrightarrow{\text{epi}} \{h \in \mathbb{R}^n : h_i \le 0, i \in I_0\}$. We conclude the proof of Thm 3.2. □

### D.6. Proof of Thm 2

The limit FPPE is $\beta_1^* = 1$ and $\delta_1^* = 0$. The observed FPPE is $\beta_1^\gamma = \min\{1, 1/\bar{v}^t\}$ where $\bar{v}^t = \frac{1}{t}\sum_{\tau=1}^t v_1(\theta^\tau)$. The bootstrapped FPPE Eq (12) is $\beta_1^b = \min\{1, 1/\bar{v}^{t,b}\}$ where $\bar{v}^{t,b} = \frac{1}{t}\sum_{\tau=1}^t v_1(\theta^{\tau,b})$.

First, we derive the limit distribution of the observed FPPE. We have

$$\sqrt{t}(\beta_1^\gamma - 1) = \frac{1}{\bar{v}^t}\min\{\sqrt{t}(1 - \bar{v}^t), 0\} \xrightarrow{d} \min\{Z, 0\} = \mathcal{J}_{\mathsf{FPPE}}.$$

where $Z \sim \mathcal{N}(0, \mathrm{Var}(v_1))$. In the above we used $\bar{v}^t \xrightarrow{p} \mathbb{E}[v_1] = 1$, Slutsky's Theorem, $\sqrt{t}(1 - \bar{v}^t) \xrightarrow{d} Z$, and the continuous mapping theorem.

Now we analyze the limit distribution of the bootstrapped FPPE. Define the set $A_c = \{\limsup_t \sqrt{t}(1 - \bar{v}^t) > c\}$ for any $c > 0$. By the law of the iterated logarithm,

$$\mathbb{P}\left(\limsup_{t \to \infty} \frac{\sqrt{t}(1 - \bar{v}^t)}{\sqrt{2 \log \log t}} = 1\right) = 1,$$

and thus $\mathbb{P}(A_c) = 1$ for all $0 < c < \infty$. Note that it holds $\bar{v}^{t,b} - 1 \xrightarrow{p} 0$ and $\sqrt{t}(\bar{v}^{t,b} - \bar{v}^t) \xrightarrow{d} \mathcal{N}(0, \mathrm{Var}(v_1))$ conditional on observed items (by triangular-array versions of the law of large numbers and the central limit theorem, see Theorem 2.2.6 and Theorem 3.4.10 from Durrett (2019)). On the event $A_c$, we can choose a subsequence $\{t_k\}_k$, such that $\sqrt{t_k}(1 - \bar{v}^{t_k}) \ge c$ for all $k$. Now let $t$ be an element of this subsequence. Then we have

$$\begin{aligned}
\sqrt{t}(\beta_1^b - \beta_1^\gamma) &= \sqrt{t}(\min\{1, 1/\bar{v}^{t,b}\} - \min\{1, 1/\bar{v}^t\}) \\
&\ge \sqrt{t}(\min\{1, 1/\bar{v}^{t,b}\} - 1) \\
&= \frac{1}{\bar{v}^{t,b}}\min\{0, \sqrt{t}(\bar{v}^t - \bar{v}^{t,b} + 1 - \bar{v}^t)\} \\
&\ge \frac{1}{\bar{v}^{t,b}}\min\{0, \sqrt{t}(\bar{v}^t - \bar{v}^{t,b}) + c\} \\
&\xrightarrow{d} \min\{0, Z + c\} \ge \min\{0, Z\},
\end{aligned}$$

where we used Slutsky's theorem for the convergence in probability. The last inequality is strict with strictly positive probability. We conclude that the standard multinomial bootstrap $\sqrt{t}(\beta^b - \beta^\gamma)$ fails to converge to the desired distribution $\mathcal{J}_{\mathsf{FPPE}}$.

### D.7. Proof of Thm 4

*Proof of Thm 4.* Define the estimated critical cone

$$\widehat{C} = \{h : h_i = \frac{1-\beta_i^*}{\delta_t} \text{ for } i \in \widehat{I}_+\}$$

Recall under SCS the critical cone is $C = \{h \in \mathbb{R}^n : h_i = 0, i \in I_+\}$.

*Step 1.* We show that the critical cone is correctly estimated in the sense that

$$\chi(\cdot \in \widehat{C}) \xrightarrow{\text{epi}} \chi(\cdot \in C) \quad \text{in probability} \tag{31}$$

The claim is equivalent to $d_{AW}(\widehat{C}, C) \xrightarrow{p} 0$, where $d_{AW}$ is defined in Eq (26). For any $\epsilon > 0$, the event $\{d_{AW}(\widehat{C}, C) > \epsilon\}$ is equivalent to $\{\widehat{I}_+ \neq I_+\}$. First we bound $\mathbb{P}(\widehat{I}_+ \neq I_+)$.

$$\begin{aligned}\mathbb{P}(\widehat{I}_+ \neq I_+) &= \mathbb{P}(\exists i \in I_+, 1-\beta_i^\gamma > \delta_t, \text{ or } \exists i \in I^c, 1-\beta_i^\gamma < \delta_t) \\ &\leq \sum_{i \in I_+} \mathbb{P}(1-\beta_i^\gamma > \delta_t) + \sum_{i \in I^c} \mathbb{P}(1-\beta_i^\gamma < \delta_t)\end{aligned}$$

For $i \in I_+$, by $\beta_i^\gamma - 1 = o_p(\frac{1}{\sqrt{t}})$, we have $\mathbb{P}(1-\beta_i^\gamma > \delta_t) = \mathbb{P}(o_p(1) > \delta_t\sqrt{t}) \to 0$ since $\delta_t\sqrt{t} \to c > 0$. For $i \in I^c$, since $\beta_i^* - 1 < 0$, $\mathbb{P}(1-\beta_i^\gamma < \delta_t) = \mathbb{P}(\beta_i^* - \beta_i^\gamma < \beta_i^* - 1 + \delta_t) = \mathbb{P}(o_p(1) < \beta_i^* - 1 + \delta_t) \to 0$ by $\delta_t \downarrow 0$. We conclude $\mathbb{P}(\widehat{I}_+ \neq I_+) \to 0$ and so $\chi(\cdot \in \widehat{C}) \xrightarrow{\text{epi}} \chi(\cdot \in C)$ in probability.

*Step 2.* Next, we show $\beta^b \xrightarrow{p} \beta^*$. Since $\epsilon_t \to 0$ and $G^b = O_p(1)$, we have $\epsilon_t G^b = o_p(1)$. Then for each $\beta \in \mathbb{R}^n$

$$\begin{aligned}&\epsilon_t G^b + \frac{1}{2}\|\beta^\gamma - \beta\|_{\widehat{\mathcal{H}}}^2 + \chi(\beta \in \widehat{B}) \\ &= \frac{1}{2}\|\beta^* - \beta\|_{\widehat{\mathcal{H}}}^2 + \chi(\beta \in \widehat{B}) + \left(\epsilon_t G^b + (\beta^* - \beta)^\intercal \widehat{\mathcal{H}}(\beta^\gamma - \beta^*) + \frac{1}{2}\|\beta^\gamma - \beta^*\|_{\widehat{\mathcal{H}}}^2\right) \\ &= \frac{1}{2}\|\beta^* - \beta\|_{\widehat{\mathcal{H}}}^2 + \chi(\beta \in \widehat{B}) + o_p(1) \\ &\xrightarrow{p} \frac{1}{2}\|\beta^* - \beta\|_{\mathcal{H}}^2 + \chi(\beta \in B')\end{aligned}$$

where $B' = \{\beta \in [0,1]^n : \beta_i = 1, i \in I_+\}$. By convexity, it implies uniform convergence on compact sets in probability. Since $\beta^*$ uniquely minimize $\|\beta^* - \beta\|_{\mathcal{H}}^2 + \chi(\beta \in B)$, we conclude $\beta^b \xrightarrow{p} \beta^*$ (Newey & McFadden, 1994, Theorem 2.7).

*Step 3.* Identify the limit distribution of $\beta^b$.

Note $\epsilon_t^{-1}(\beta^b - \beta^\gamma) = \epsilon_t^{-1}(\beta^b - \beta^*) + \epsilon_t^{-1}(\beta^* - \beta^\gamma) = \epsilon_t^{-1}(\beta^b - \beta^*) + o_p(1)$, it suffices to show $\epsilon_t^{-1}(\beta^b - \beta^*) \overset{p}{\rightsquigarrow} \mathcal{J}_{\mathsf{FPPE}}$. First with a change of variable $h = \epsilon_t^{-1}(\beta - \beta^*)$ and so $\epsilon_t^{-1}(\beta - \beta^\gamma) = h + \epsilon_t^{-1}(\beta^* - \beta^\gamma)$, dividing the objective function by $\epsilon_t^2$, the bootstrap estimator in Eq (17), with probability approaching one, can be written as

$$\frac{\beta^b - \beta^*}{\epsilon_t} = \arg\min_{h \in \mathbb{R}^n} \Big\{ \underbrace{(G^b)^\intercal(h + \frac{\beta^* - \beta^\gamma}{\epsilon_t}) + \frac{1}{2}\|h + \frac{\beta^* - \beta^\gamma}{\epsilon_t}\|_{\widehat{\mathcal{H}}}^2}_{:=X_t(h)} + \chi(h \in \widehat{C}) \Big\}$$

**Claim 4.** *$X_t(h) \rightsquigarrow (\overset{p}{\rightsquigarrow}) G^\intercal h + \frac{1}{2}h^\intercal \mathcal{H} h$ in $\ell^\infty(K)$ for any compact $K \subset \mathbb{R}^n$.*

By Knight (1999), we know convergence in distribution with respect to the topology of uniform convergence on compact sets implies weak epi-convergence, and so $X_t \overset{\text{epi}}{\rightsquigarrow} G^\intercal h + \frac{1}{2}h^\intercal \mathcal{H} h$. Combining $X_t \overset{\text{epi}}{\rightsquigarrow} G^\intercal h + \frac{1}{2}h^\intercal \mathcal{H} h$, Eq (31) and Lemma 1, we have

$$X_t(h) + \chi(h \in \widehat{C}) \overset{\text{epi}}{\rightsquigarrow} G^\intercal h + \frac{1}{2}\|h\|_{\mathcal{H}}^2 + \chi(h \in C)$$

The above implies conditional weak epi-convergence in probability of the processes $h \mapsto X_t(h) + \chi(h \in \widehat{C})$. A bootstrap version of the continuous mapping theorem can be stated for conditional weak epi-convergence in probability (Hong & Li, 2020). Then $\epsilon_t^{-1}(\beta^b - \beta^*) \overset{p}{\rightsquigarrow} \arg\min_{h\in C} G^\top h + \frac{1}{2} h^\top \mathcal{H} h$, which is exactly $\mathcal{J}_{\mathsf{FPPE}}$. □

*Proof of Claim 4.* First, we show for any compact set $K \in \mathbb{R}^n$,

$$\sup_{h\in K} |X_t(h) - ((G^b)^\top h + \frac{1}{2}\|h\|^2_{\mathcal{H}})| = o_p(1) \tag{32}$$

Note the LHS can be upper bounded by

$$(G^b)^\top(\beta^\gamma - \beta^*)\epsilon_t^{-1} + \frac{1}{2}\|\beta^\gamma - \beta^*\|^2_{\mathcal{H}}\epsilon_t^{-2} + \sup_K h^\top \widehat{\mathcal{H}}(\beta^\gamma - \beta^*)\epsilon_t^{-1} + \sup_K h^\top(\widehat{\mathcal{H}} - \mathcal{H})h$$

We just need to show $(G^b)^\top(\beta^\gamma - \beta^*) = o_p(\epsilon_t)$, $\|\beta^\gamma - \beta^*\|_{\widehat{\mathcal{H}}} = o_p(\epsilon_t^2)$ and $\sup_K h^\top\widehat{\mathcal{H}}(\beta^\gamma - \beta^*) = o_p(\epsilon_t)$ and $\sup_K |h^\top\mathcal{H}h - h^\top\widehat{\mathcal{H}}h| = o_p(1)$ (Thm 5). This holds by $G^b = O_p(1)$, $\beta^\gamma - \beta^* = O_p(1/\sqrt{t})$, $\|\mathcal{H} - \widehat{\mathcal{H}}\|_2 = o_p(1)$ and $1/\sqrt{t} = o(\epsilon_t)$.

Next we show for any compact $K \subset \mathbb{R}^n$,

$$(G^b)^\top h + \frac{1}{2}\|h\|^2_{\widehat{\mathcal{H}}} \rightsquigarrow G^\top h + \frac{1}{2}\|h\|^2_{\mathcal{H}} \quad \text{in } \ell^\infty(K) \tag{33}$$

where $G \sim \mathcal{N}(0, \mathrm{Cov}(\nabla F(\cdot, \beta^*)))$. It suffices to show $G^b \xrightarrow{d} G$ and $\widehat{\mathcal{H}} \xrightarrow{p} \mathcal{H}$. In Thm 5 is has been shown that $\widehat{\mathcal{H}} - \mathcal{H} = o_p(1)$. Note $G^b = \sqrt{t}(P_t^b - P_t)D_F(\cdot, \beta^\gamma) = \sqrt{t}(P_t^b - P_t)(D_F(\cdot,\beta^\gamma) - D_F(\cdot,\beta^*)) + \sqrt{t}(P_t^b - P_t)D_F(\cdot,\beta^*) = \sqrt{t}(P_t^b - P_t)D_F(\cdot,\beta^*) + o_p(1)$ by Wellner & Zhan (1996, Lemma 4.1). We conclude $G^b \xrightarrow{d} G$ and $X_t \rightsquigarrow h^\top G + \frac{1}{2}h^\top\mathcal{H}h$. To show conditional convergence $\overset{p}{\rightsquigarrow}$, one can use arguments analogous to Theorem 2.9.6 in Van der Vaart (2000). □

### D.8. Proof of Thm 6

*Proof of Thm 6.* Let $T^{b,\infty}$ be the conditional limit distribution of $T^b$, and $T^\infty$ be the limit distribution of $T^\gamma(\beta^*, \delta^*)$. The proof relies on the following result. For two real-valued random variables $X$ and $Y$, we say $X$ is stochastically dominated by $Y$, denoted $X \le_{st} Y$ if $\mathbb{P}(X > x) \le \mathbb{P}(Y > x)$ for all $x \in \mathbb{R}$.

**Theorem 10.** *For all $\kappa \in (0, \infty)$, $T^\infty \le_{st} T^{b,\infty}$. When $\kappa = \infty$, $T^\infty = T^{b,\infty}$.*

We recall a result regarding quantiles.

**Lemma 2** (Lemma 21.1 from Van der Vaart (2000))**.** *Let $F(x) = \mathbb{P}(X \le x)$ be the CDF of a real-valued random variable $X$. And let $F^{-1}(p) = \inf(x \in \mathbb{R} : F(x) \ge p)$ for $p \in (0,1)$ be the quantile. Then $F(F^{-1}(p)) \ge p$ for $p \in (0,1)$. Equality holds if $F$ is continuous at $F^{-1}(q)$.*

Recall the following condition

$$\text{the CDF of } T^\infty \text{ is continuous at the } (1-\alpha)\text{-th quantile of } T^\infty. \tag{34}$$

Let $c_{1-\alpha}$ be the $(1-\alpha)$-th quantile of $T^\infty$. Then Lemma 2 implies that $\mathbb{P}(T^\infty \le c_{1-\alpha-\epsilon}) \ge 1 - \alpha - \epsilon$ for all $\epsilon \ge 0$ small enough. The assumption in Eq (34) implies $\mathbb{P}(T^\infty \le c_{1-\alpha}) = 1 - \alpha$.

For any $\epsilon > 0$, let $A$ be the event $\{\gamma : \mathbb{P}(T^b \le c_{1-\alpha} \,|\, \gamma) - (1-\alpha) \le \epsilon\}$. Let $c^b_{1-\alpha}$ be that the $(1-\alpha)$-th quantile of $T^b$ conditional on $\gamma$.

First we show $\mathbb{P}(A) \to 1$. By $T^{b,\infty} \ge_{st} T^\infty$ we know $\mathbb{P}(T^{b,\infty} \le c_{1-\alpha}) \le \mathbb{P}(T^\infty \le c_{1-\alpha}) = 1 - \alpha$. Then

$$\begin{aligned}\mathbb{P}(A^c) &= \mathbb{P}(\mathbb{P}(T^b \le c_{1-\alpha} \,|\, \gamma) - (1-\alpha) > \epsilon) \\ &\le \mathbb{P}(\mathbb{P}(T^b \le c_{1-\alpha} \,|\, \gamma) - \mathbb{P}(T^{b,\infty} \le c_{1-\alpha}) > \epsilon) \to 0\end{aligned}$$

due to $T^b \overset{p}{\rightsquigarrow} T^{b,\infty}$. Under event $A$, we have $c_{1-\alpha-\epsilon} \leq c^b_{1-\alpha}$. To see this note $\mathbb{P}(T^\infty \leq c^b_{1-\alpha} \,|\, \gamma) \geq \mathbb{P}(T^b \leq c^b_{1-\alpha}) - \epsilon \geq 1 - \alpha - \epsilon$. Then

$$
\begin{aligned}
&\mathbb{P}\big((\beta^*, \delta^*) \in C^\gamma(c^b_{1-\alpha})\big) \\
&= \mathbb{P}(T^\gamma(\beta^*, \delta^*) \leq c^b_{1-\alpha}) \\
&\geq \mathbb{P}(T^\gamma(\beta^*, \delta^*) \leq c^b_{1-\alpha}, A) \\
&\geq \mathbb{P}(T^\gamma(\beta^*, \delta^*) \leq c_{1-\alpha-\epsilon}, A) \\
&\geq \mathbb{P}(T^\gamma(\beta^*, \delta^*) \leq c_{1-\alpha-\epsilon}) - \mathbb{P}(A^c) \geq 1 - \alpha - \epsilon + o(1)
\end{aligned}
$$

where the last line follows from $\mathbb{P}(T^\gamma(\beta^*, \delta^*) \leq c_{1-\alpha-\epsilon}) \to \mathbb{P}(T^\infty \leq c_{1-\alpha-\epsilon}) \geq 1 - \alpha - \epsilon$ and $\mathbb{P}(A^c) \to 0$. Since $\epsilon > 0$ is arbitrary, we conclude $\liminf \mathbb{P}((\beta^*, \delta^*) \in C^\gamma(c^b_{1-\alpha})) \geq 1 - \alpha$. □

*Proof of Thm 10.* We study the asymptotic distributions of $T^\gamma(\beta^*, \delta^*)$ and $T^b$. Recall $\nabla H(\beta^*) = -\delta^*$.

Step 1. We will show

$$
t(L_t(\beta^* + \tfrac{h}{\sqrt{t}}, \delta^*) - L_t(\beta^*, \delta^*)) \rightsquigarrow h^\top G + \frac{1}{2} h^\top \mathcal{H} h
$$

in $\ell^\infty(K)$ for any compact $K \subset \mathbb{R}^n$.

$$
\begin{aligned}
&t(L_t(\beta^* + \tfrac{h}{\sqrt{t}}, \delta^*) - L_t(\beta^*, \delta^*)) \\
&= t(H_t(\beta^* + \tfrac{h}{\sqrt{t}}) - H_t(\beta^*) + (\delta^*)^\top(\tfrac{h}{\sqrt{t}})) \\
&= \sqrt{t}(P_t - P)(\nabla F(\cdot, \beta^*))^\top h + \frac{1}{2} h^\top \mathcal{H} h + o_p(1) \\
&\rightsquigarrow h^\top G + \frac{1}{2} h^\top \mathcal{H} h \text{ in } \ell^\infty(K)
\end{aligned}
$$

where $G \sim \mathcal{N}(0, \mathrm{Cov}(\nabla F(\cdot, \beta^*)))$, $o_p(1)$ term is uniform over $h \in K$ by Claim 3. Applying a continuous mapping theorem

$$
\begin{aligned}
T^\gamma(\beta^*, \delta^*) &= -\inf_{h \in \mathbb{B}_\kappa} t(L_t(\beta^* + \tfrac{h}{\sqrt{t}}) - L_t(\beta^*)) \\
&\overset{d}{\to} -\inf_{h \in \mathbb{B}_\kappa} (G^\top h + \frac{1}{2} h^\top \mathcal{H} h) =: T
\end{aligned}
$$

Step 2. In Claim 4, we have shown

$$
X^b(\beta^* + \epsilon_t h) \overset{p}{\rightsquigarrow} G^\top h + \frac{1}{2} h^\top \mathcal{H} h
$$

in $\ell^\infty(K)$ for any compact $K \in \mathbb{R}^n$. Next we study the asymptotic distribution of $T^b$.

$$
\begin{aligned}
T^b &= -\inf_{\beta \in \mathbb{R}^n_+} X^b(\beta) \\
&= -\inf_{h \in (\mathbb{R}^n_+ - \beta^*)/\epsilon_t} X^b(\beta^* + \epsilon_t h) \\
&\overset{p}{\rightsquigarrow} -\inf_{h \in \mathbb{R}^n} (G^\top h + \frac{1}{2} h^\top \mathcal{H} h) =: T^{b,\infty}
\end{aligned}
$$

where the last line follows due to $\beta^*$ lying in the interior of $\mathbb{R}^n_+$, and a bootstrap version of continuous mapping theorem; see Theorem 10.8 in Kosorok (2008). We can see for each draw of $G$, we have the dominance relationship $T^{b,\infty} \geq T$. We conclude the $(1-\alpha)$-quantile of $T^{b,\infty}$ is greater than or equal to that of $T$.

The claim that when $\kappa = \infty$, $T^b \overset{p}{\rightsquigarrow} T^\infty$ is obvious.

**Remark 1.** *One could also use the statistic*

$$T^{\gamma}(\beta) = \inf_{0 \le \delta \le b, \delta^{\mathsf{T}}(1_n - \beta) = 0} \left( L_t(\beta, \delta) - \inf_{h \in \mathbb{B}_\kappa} L_t(\beta + h/\sqrt{t}, \delta) \right)$$

and the region $\{\beta \in (0,1]^n : T^{\gamma}(\beta) \le \iota\}$ to do inference on just $\beta$. Noting $T^{\gamma}(\beta^*) \le L_t(\beta^*, \delta^*) - \inf_{h \in \mathbb{B}_\kappa} L_t(\beta^* + h/\sqrt{t}, \delta^*)$, we can estimate an upper bound of the quantile of its limit distribution by bootstrap. □