

Experimenting in Equilibrium

Stefan Wager
 Graduate School of Business
 Stanford University
 swager@stanford.edu

Kuang Xu
 Graduate School of Business
 Stanford University
 kuangxu@stanford.edu

Draft Version August, 2019

Abstract

Classical approaches to experimental design assume that intervening on one unit does not affect other units. Recently, however, there has been considerable interest in settings where this non-interference assumption does not hold, e.g., when running experiments on supply-side incentives on a ride-sharing platform or subsidies in an energy marketplace. In this paper, we introduce a new approach to experimental design in large-scale stochastic systems with considerable cross-unit interference, under an assumption that the interference is structured enough that it can be captured using mean-field asymptotics. Our approach enables us to accurately estimate the effect of small changes to system parameters by combining unobtrusive randomization with light-weight modeling, all while remaining in equilibrium. We can then use these estimates to optimize the system by gradient descent. Concretely, we focus on the problem of a platform that seeks to optimize supply-side payments p in a centralized marketplace where different suppliers interact via their effects on the overall supply-demand equilibrium, and show that our approach enables the platform to optimize p based on perturbations whose magnitude can get vanishingly small in large systems. *Keywords:* experimental design, interference, mean-field model, stochastic system.

1 Introduction

Randomized controlled trials are widely used to guide decision making across several areas, ranging from classical industrial and agricultural settings [Fisher, 1935] to the modern tech sector [Athey and Luca, 2019, Kohavi et al., 2009, Tang et al., 2010]. A growing interest in randomized experiments, or A/B tests, has even led to the creation of specialized companies that help with rigorous statistical inference in dynamic designs [Johari et al., 2017b].

Much of the existing work on experimental design has focused on settings where we can intervene separately on different units, i.e., there is no cross-unit interference, and the lack of interference plays a key role in justifying standard analyses of randomized trials [Imbens and Rubin, 2015]. This non-interference assumption, however, is violated in many important applications. For example, Bottou et al. [2013] describe difficulties in using randomized experiments to study internet ad auctions: Advertisers participate in an auction to determine ad placements, and any intervention on one advertiser may change their behavior on

This work was partially supported by a seed grant from the Stanford Global Climate and Energy Project and a Facebook Faculty Award.

the auction and thus affect the opportunities available to other advertisers. In a different context, [Blake and Coey \[2014\]](#) document failures of the non-interference assumption due to an interaction between treated and control customers in an experiment run by an online marketplace.

The question of how to run experiments when simple A/B tests fail has proven to be challenging. Some approaches have been proposed that assume sparse interference patterns. For example, when studying Internet ad auctions, [Basse et al. \[2016\]](#), [Kohavi et al. \[2009\]](#) and [Ostrovsky and Schwarz \[2011\]](#) note that the auction type used for one ad keyword does not meaningfully affect how advertisers bid for other keywords, and then consider experiments that randomly assign keywords, rather than advertisers, to different conditions. Similarly, a social network might try to deploy different versions of a feature in different countries, and hope that the number of cross-border links is small enough to induce only negligible interference. The limitation of these approaches, however, is that the power of any experiment is limited by the number of non-interfering clusters available: For example, if a platform has 200 million customers in 100 countries, but chooses to randomize by country, then the largest effective sample size they can use for any experiment is 100, and not 200 million.

In this paper, we propose an alternative approach to experimentation in stochastic systems where a large number of, if not all, units interfere with one another. For concreteness, we focus on the problem of setting supply side payments in a centralized marketplace, where available demand is randomly allocated to a set of available suppliers. In these systems, different suppliers interact via their effects on the overall supply-demand equilibrium: The more suppliers choose to participate in the marketplace, the less demand on average an individual supplier would be able to serve in equilibrium. The objective of the system designer is to identify the optimal payment that maximizes the platform’s utility. Note that conventional randomized experimentation schemes that assume no interference fail in this system: For example, if we double the per-transaction payments made to a random half of suppliers, these suppliers will increase their production levels and reduce the amount of demand available to the remaining suppliers, and thus reduce their incentives to produce.

We consider a simple model of such a centralized marketplace, and design a class of “local” experimentation schemes that—by carefully leveraging the structure of the marketplace—enable us to optimize payments without disturbing the overall market equilibrium. To do so, we perturb the per-transaction payment p_i available to the i -th supplier by a small mean-zero shock, i.e., $p_i = p + \zeta \varepsilon_i$ where $0 < \zeta \ll 1$ and $\varepsilon_i = \pm 1$ independently and uniformly at random. A simple reduced form regression of per-supplier production on ε_i recovers a marginal response function which is not of direct relevance to policy making. However, in the limit where the number of suppliers is large, we can use mean-field modeling to translate the output of this reduced form regression into an estimate of the gradient of the platform’s utility with respect to p . We can then use these gradient estimates to optimize p via any stochastic first-order optimization method, such as stochastic gradient descent and its extensions.

The driving insight behind our result is that, although there is dependence across the behavior of a large number of units in the system, any such interference can only be channeled through a small number of key statistics: In our example, this corresponds to the total supply made available by all suppliers. Then, if we can intervene on individual units without meaningfully affecting the key statistics, we can obtain meaningful information about the system—at a cost that scales sub-linearly in the number of units. The type of interference that we consider, where the units experience global interference channeled through a small

number of key statistics, can manifest in a range of applications. We discuss some examples below.

Example 1 (Ride Sharing). Ride sharing platforms match customers who request a ride with nearby freelance drivers who are both active and not currently servicing another request. It is in the interest of the platform to have a reasonable amount of capacity available at all times to ensure a reliable customer experience. To this end, the platform may seek to increase capacity by increasing the rates paid to drivers for completing rides. And, when running experiments on the rates needed to achieve specific capacity levels, the platform needs to account for interference. If the platform in fact succeeds in increasing capacity by increasing rates—yet demand remains fixed—the expected utilization of each driver will go down and so the drivers’ expected revenue, i.e., the product of the rate and the expected utilization, will not increase linearly in the rate. Thus, if drivers respond to expected revenue when choosing whether to work for a platform, as empirical evidence suggests that they do [Hall et al., 2019], a platform that ignores interference effects will overestimate the power of rate hikes to increase capacity. However, as shown in our paper, we can accurately account for these interference effects via mean-field modeling because they are all channeled through a simple statistic, in this case total capacity.

Example 2 (Congestion Pricing). A policy maker may want to identify the optimal toll for congestion pricing [e.g., Goh, 2002]. We assume that drivers get positive utility from completing a trip, but get negative utility both from congestion delays and from paying tolls. Then, in studying the effect of a toll on congestion, the policy maker needs to address the fact that drivers interfere with one another through the overall state of congestion on the road: If we raise the tolls on a small subset of the drivers and hence discourage them from going on the road, those whose tolls remain unchanged may experience less congestion and hence be inclined to drive more. Therefore a policy maker that experiments with a small sub-population, without taking into account interference effects, may obtain an overly optimistic estimate of the true effect of a toll change when applied to all drivers. Again, however, all interference is channeled through a single statistic—congestion—and so mean-field modeling can capture its effect.

Example 3 (Renewable Energy Subsidies). In an electricity whole sale market, energy producers (e.g., generators) and consumers (e.g., utilities) make bids and offers in the day-ahead market, which is then cleared in a manner that balances the aggregate regional supply and demand. The operator of these markets, such as CAISO or ERCOT, may choose to provide subsidies or scheduling priorities to encourage renewable generation [see CAISO, 2009]. Suppose that the market operator would like to know the effect of increasing subsidies on energy generation. We expect that increased subsidies would increase both total and renewable energy production; the question is by how much, and what the effect of interference will be. It is plausible that the effect of subsidies on total supply will be mitigated by interference, because increased production from one supplier will decrease demand available to others. In contrast, interference may either mitigate or amplify the effect of subsidies on renewable energy production: Amplification effects may occur if subsidies affect profitability in a way that causes non-renewable producers to be replaced by new renewable entrants. In either case, all interference effects are channeled through global capacity, and so can be accounted for via mean-field modeling.

1.1 Related Work

The problem of experimental design under interference has received considerable attention in the statistics literature. The dominant paradigm has focused on robustness to interference, and on defining estimands in settings where some units may be exposed to spillovers from treating other units [Aronow and Samii, 2017, Athey et al., 2018, Basse et al., 2019, Eckles et al., 2017, Hudgens and Halloran, 2008, Manski, 2013, Sobel, 2006]. Depending on applications, the exposure patterns may be simple (e.g., the units are clustered such that exposure effects are contained within clusters) or more complicated (e.g., the units are connected in a network, and two units far from each other in graph distance are not exposed to each others’ treatments). Unlike this line of work that seeks robustness to interference driven by potentially complex and unknown mechanisms, the local randomization scheme proposed here crucially relies on having a stochastic model that lets us explain interference. Then, because all inference acts via a simple statistic, we can move beyond simply seeking robustness to interference and can in fact accurately predict interference effect using information gathered in equilibrium.

The idea that one can distill insights of a structural model down to the relationship between a small number of observable statistics has a long tradition in economics [e.g., Chetty, 2009, Harberger, 1964]. This approach can often be used for practical counterfactual analysis without needing to fit complicated structural models. Here, we use such an argument for experimental design rather than to guide methods for observational study analysis. At a high level, our paper also has a connection to results on learning in a setting where agents exhibit strategic behavior, including Feng et al. [2018], Iyer et al. [2014], and Kanoria and Nazerzadeh [2017], and in crowd-sourcing systems, including Johari et al. [2017a], Khetan and Oh [2016] and Massoulié and Xu [2018].

Our approach to optimizing p using gradients obtained from local experimentation intersects with the literature on continuous-arm bandits (or noisy zeroth-order optimization), which aims to optimize a function $f(x)$ by sequentially evaluating f at points x_1, x_2, \dots , and obtaining in return noisy versions of the function values $f(x_1), f(x_2), \dots$ [Bubeck et al., 2017, Spall, 2005]. A number of bandit methods first generate noisy gradient estimates of the function by comparing adjacent function values, and subsequently use these estimates in a first-order optimization method [Flaxman et al., 2005, Ghadimi and Lan, 2013, Jamieson et al., 2012, Kleinberg, 2005, Nesterov and Spokoiny, 2017]. In our model, this approach would amount to estimating utility gradients via global experimentation, by comparing the empirical utilities observed at two different payment levels. Compared to this literature, our paper exploits a cross-sectional structure not present in most existing zeroth-order models: We show that our local experimentation approach, which offers slightly different payments across a large number of units, is far more efficient at estimating the gradient than global experimentation, which offers all units the same payment on a given day. Such cross-sectional signals would be lost if we abstracted away the multiplicity of units, and only treated the average payment as a decision variable to be optimized. In Section 4.4, we provide a formal comparison for the regret of a platform deploying our approach versus a bandit-based algorithm, and establish sharp separation in terms of rates of convergence.

The limiting regime that we use, one in which the system size tends to infinity, is often known as the mean-field limit. It has a long history in the study of large-scale stochastic systems, such as the many-server regime in queueing networks [Bramson et al., 2012, Halfin and Whitt, 1981, Stolyar, 2015, Tsitsiklis and Xu, 2012, Vvedenskaya et al., 1996] and interacting particle systems [Graham and Méléard, 1994, Mézard et al., 1987, Sznitman,

1991]. Likewise, our proposed method leverages a key property of the mean-field limit: While changes to the behavior of a single unit may have significant impact on other units in a finite system, such interference diminishes as the system size grows and, in the limit, the behaviors among any finite set of units become asymptotically independent from one another, a phenomenon known as the propagation of chaos [Bramson et al., 2012, Graham and Méléard, 1994, Sznitman, 1991]. This asymptotic independence property underpins the effectiveness of our local experimentation scheme, and ensures that small, symmetric payment perturbations do not drastically alter the equilibrium demand-supply dynamics. Our work thus suggests that, just as mean-field models have been successful in the analysis of stochastic systems, they may be a useful paradigm for designing experiments in large stochastic systems.

2 Experimentation in Stochastic Systems

For concreteness, we focus our discussion on a simple setting inspired by a centralized marketplace for freelance labor that operates over a number of periods. In each period, the high-level objective of the decision maker (i.e., operator of the platform) is to match demand with a pool of potential suppliers in such a manner that maximizes the platform’s expected utility. To do so, the decision maker offers payments to each potential supplier individually, who in turn decides whether to become active/available based upon their belief of future revenue. Our main question is how the decision maker can use experimentation to efficiently discover their revenue-maximizing payment, despite not knowing the detailed parameterization of the model, and the presence of substantial stochastic uncertainty.

We formally describe a flexible stochastic model in Section 3; here, we briefly outline a simple variant of our model that lets us highlight some key properties of our approach. Each day $t = 1, \dots, T$ there are $i = 1, \dots, n$ potential suppliers, and demand for D_t identical tasks to be accomplished. A central platform chooses a distribution π_t , and then offers each supplier random payments $P_{it} \stackrel{\text{iid}}{\sim} \pi_t$ they commit to pay for each unit of demand served. The suppliers observe both π_t a state variable A_t that can be used to accurately anticipate demand D_t (e.g., A_t could capture local weather or events); however, the platform does not have access to A_t . Given their knowledge of P_{it} and A_t , each supplier independently chooses to become “active”; we write $Z_{it} = 1$ for active suppliers and $Z_{it} = 0$ else. Then, demand D_t is randomly allocated to active suppliers.

Our key assumption is that each supplier chooses to become active based on their expected revenue conditionally on being active: They first compute $q_{A_t}(\pi_t)$, their expected allocation rate (rate at which they will be matched with demand) conditionally on being active and given A_t and π_t . They then decide whether to become active by comparing the expected revenue $P_{it}q_{A_t}(\pi_t)$ with a random outside option.¹ The form of $q(\cdot)$ depends on both the amount of available supply and demand, and the efficiency with which supply can be matched with demand; see Section 3 for an example based on a queuing network. Finally, the platform’s utility U_t is given by the revenue from the demand served minus payments made to suppliers.

Figure 1 shows a simple example of an equilibrium resulting from this model in the limit as n gets large in a setting where all suppliers are offered the same payment p , for a specific

¹Here, we assume that suppliers don’t take into account the effect of their own decision to become active on their expected allocation rate. This is often taken to be a reasonable assumption in large stochastic systems [e.g., Chetty, 2009].

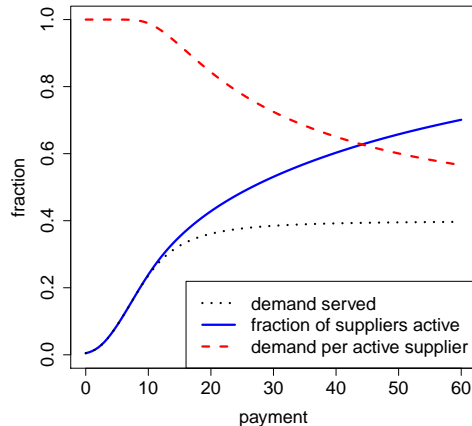


Figure 1: Example of large-sample behavior of market, conditionally on a realization of A . We show $\mu_A(p)$, the fraction of suppliers that choose to become active, $q_A(p)$ the expected amount of demand served per active supplier and $\mu_A(p)q_A(p)$ the expected amount of demand served (expressed as a multiple of the maximum capacity that would be available if all suppliers were active). The example is simulated in the mean-field limit, i.e., with number of potential suppliers n growing to infinity, such that $\mathbb{E}[D/n|A] = 0.4$. Individual supplier preferences are logistic (3.7) with $\alpha = 1$ with outside option $\log(B_i/20) \sim \mathcal{N}(0, 1)$. Supply-demand matching is characterized via the allocation function (3.5) with $L = 8$, visualized in Figure 4.

realization of demand D . We see that, as p gets larger, the active supply gets larger than demand and the utilization of active suppliers goes down.

Before presenting our proposed approach to learning p below, we first briefly review why standard approaches fall short. The core difficulty in our model comes from the interplay between network effects and market-wide demand fluctuations induced by the A_t .

The network effects break what one might call classical A/B-experimentation. Suppose that, on each day $t = 1, \dots, T$, the platform chose a small random fraction of suppliers and offer them an experimental payment p_{exp} , while everyone else gets offered the status quo payment p_{default} . We could then try to use the behavior of suppliers offered p_{exp} to estimate expected profit at p_{exp} , and then update the default payment. This approach allows for cheap experimentation because most of the suppliers get offered p_{default} . However, it will not consistently recover the optimal payment because it ignores feedback effects: When we raise payments, more suppliers opt to join the market and so the rate at which any given supplier is matched with demand goes down—and this attenuates the payment-sensitivity of supply relative to what is predicted by A/B testing.

Conversely, the market-wide demand fluctuations due to A_t degrade global optimization schemes that use payment variation across days for learning; such algorithms are equivalent to continuous-armed bandit algorithms considered in the optimization literature [Spall, 2005]. Suppose that, on each day $t = 1, \dots, T$, we randomly chose a payment p_t and made it available to all suppliers, and then observed realized profits U_t . We could then try estimate

profit gradients by comparing U_t to U_{t-1} . The problem is that, due to variation in daily context, the variation in per-supplier profit U_t/n given the chosen payment p_t is always of constant order, even in very large markets (i.e., in the limit $n \rightarrow \infty$); for example, in a ride-sharing setting, if day $t - 1$ is rainy and day t is sunny, then the effect of this weather change on profit may overwhelm the effect of any payment change deployed by the platform.² The upshot is that the platform cannot learn anything via global experimentation unless it considers large changes to the payments p_t that it offers to everyone. And such wide-spread payment changes are impractical for several reasons: They are expensive, and difficult to deploy.

2.1 Local Experimentation

Our goal is to use high-level information about the stochastic system described above to design a new experimental framework that lets us avoid the problems of both approaches described above: We want our experimental scheme to be consistent for the optimal payment (like global experimentation), but also to be cost-effective (like classical A/B testing) in that it only requires small perturbations to the status quo.

The driving insight behind our approach is that it is possible to learn about the relationship between profit and payment via unobtrusive randomization by randomly perturbing the payments P_{it} offered to supplier i in time period t . We propose setting

$$P_{it} = p_t + \zeta \varepsilon_{it}, \quad \varepsilon_{it} \stackrel{\text{iid}}{\sim} \{\pm 1\} \tag{2.1}$$

uniformly at random, where $\zeta > 0$ is a (small) constant that governs the magnitude of the perturbations, and regressing market participation Z_{it} on the payment perturbations ε_{it} . This regression lets us recover the *marginal response function*, i.e., the average payment sensitivity of a supplier in a situation where only they get different payments but others do not; see Section 3.2 for a formal definition.

This marginal response function is not directly of interest for optimizing p , as it ignores feedback effects. However, we find that—in our setting—this quantity captures relevant information for optimizing payments. More specifically we show in Section 3.2 that, provided we have good enough understanding of system dynamics to be able to anticipate match rates given the amount of supply and demand present in the market, in the mean-field limit where the market size grows, we can use consistent estimates of the marginal response function to derive consistent estimates of the actual payment-sensitivity of supply that accounts for network effects. Furthermore, we show in Section 4 that this approach enables us to optimize payments using vanishingly small-scale experimentation as the market gets large (i.e., we can take ζ in (2.1) to be very small when n is large).

Figure 2 shows results from our local experimentation approach on a simple simulation experiment in the setting of Figure 1, where the scaled demand $\mathbb{E}[D/n | A]$ follows a beta(15, 35) distribution. We initialize the system at $p_1 = 30$, and then each day run payment perturbations as in (2.1) to guide a payment update using an update rule described in Section 4.2. We see that the system quickly converges to a near-optimal payment of around 17.

²Of course, the platform may try to correct for contexts, e.g., by matching days with similar values of A_t with each other. One currently popular way of doing so in the technology industry is using synthetic controls [Abadie et al., 2010]. In practice, however, this approach may be difficult to implement, and will remain intractably noisy unless the platform can observe the full context A_t and use it to essentially perfectly predict demand. In this paper our goal is to develop methods for experimental design that do not require the platform to observe A_t .

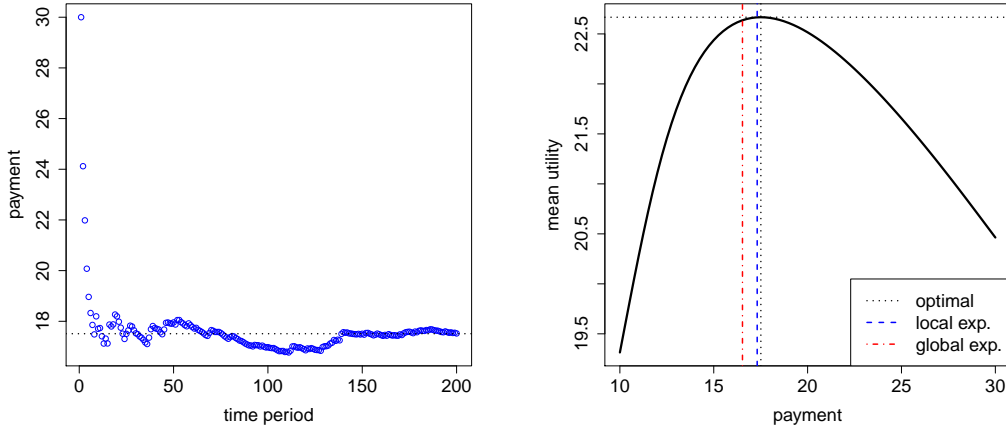


Figure 2: Results from learning p via local experimentation. The worker preference functions are as in Figure 1; the daily contexts are such that $\mathbb{E}[D/n | A] \sim \text{beta}(15, 35)$. The platform utility function is linear as in Lemma 3, with $\gamma = 100$. We learned gradients based on local randomization (2.1) with $\zeta = 0.5$, and then optimized payments via gradient descent as in (4.5) with a step size $\eta = 20$ and $I = (-\infty, \infty)$. The left panel shows the convergence of the p_t to the value p^* that optimizes mean utility. The right panel compares the average value of p_t over the last 100 steps of our algorithm to both a payment \hat{p} learned via global experimentation and the optimal payment p^* .

We also compare our results to what one could obtain using the baseline of global experimentation, where we randomize the payment $p_t \sim \text{Uniform}(10, 30)$ in each time period and measure resulting platform utility U_t , and then choose the final payment \hat{p} by maximizing a smooth estimate of the expectation of U_t given p_t . The left panel of Figure 3 shows the resulting (p_t, U_t) pairs, as well as the resulting \hat{p} . As seen in the right panel of Figure 2, the final \hat{p} obtained via this method is a reasonable estimate of the optimal p .

The major difference between the local and global randomization schemes is in the resulting cost of experimentation. In Section 4.3 we show that our local experimentation scheme pays a vanishing cost for randomization; the only regret relative to deploying the optimal p from the start is due to the rate of convergence of gradient descent. In contrast, the cost of experimentation incurred for finding \hat{p} via global experimentation is huge, because it needs to sometimes deploy very poor choices of p_t in order to learn anything. And, as shown in the right panel of Figure 3, after the first few days, the global experimentation approach in fact systematically achieves lower daily utilities U_t than local experimentation. We provide further numerical comparisons of local and global experimentation in Section 6.

3 Model: Stochastic Market with Centralized Pricing

We now present the general stochastic model we use to motivate our approach. All random variables are assumed to be independent across the periods and, within each period, are

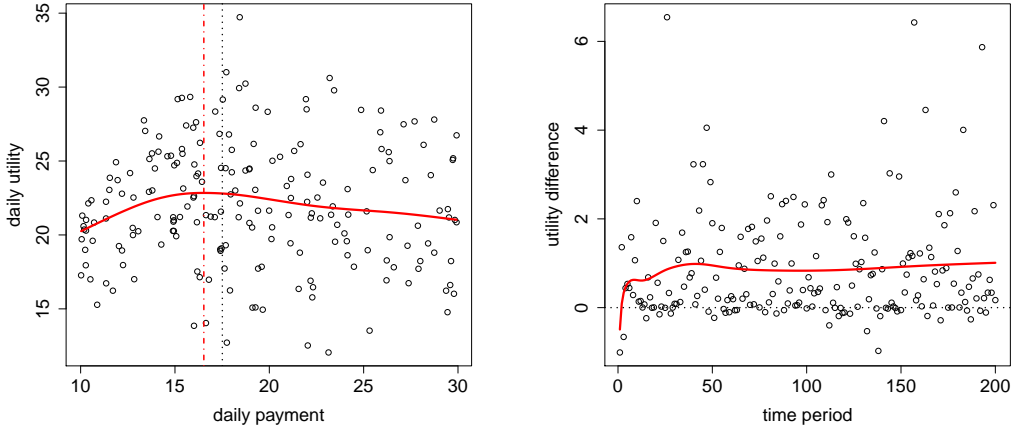


Figure 3: Results from learning p via global experimentation. The left panel shows pairs (p_t, U_t) resulting from daily experiments, along with both the resulting \hat{p} (dash-dotted line) and the optimal p^* (dotted line). The right panel shows the (scaled) difference in daily utility between our local experimentation approach and the global experimentation baseline (both approaches worked using the same demand sequence D_t).

independent from one another unless otherwise stated. We will consider a sequence of systems, indexed by $n \in \mathbb{N}$, where in the n th system there are n potential suppliers. We will refer to n as the market size. All variables in our model are thus implicitly dependent on the index, n , which we denote using the superscript (n) , e.g., $q^{(n)}$. We sometimes suppress this notation when the context is clear. In the rest of the section, we will focus on describing the model in a single time period.

Demand To reflect the reality that demand fluctuations may not concentrate with n , we allow for a random stochastic global state A drawn from a finite set \mathcal{A} . The global state affects demand, and is known to market participants (suppliers), but not to the platform (or the platform cannot react to it). For example, in a ride sharing example, A could capture the effect of weather (rain / shine) or major events (conference, sports game, etc.). Conditionally on the global state $A = a$, we assume that demand, D , is drawn from distribution $D \sim F_a$. We further assume that the demand scales proportionally with respect to the market size n , and that it concentrates after re-scaling by $1/n$. In particular, we assume that there exists $\{d_a\}_{a \in \mathcal{A}} \subset \mathbb{R}_+$, such that for all $a \in \mathcal{A}$, $\mathbb{E}[D/n | A = a] = d_a$ for all $n \in \mathbb{N}$,

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[(D/n - d_a)^2 \mid A = a \right] = 0, \quad (3.1)$$

and

$$\mathbb{P}(D/n \notin [d_a/2, 2d_a] \mid A = a) = o(1/n), \quad (3.2)$$

and as $n \rightarrow \infty$. In general, we will use the sub-script a to denote the conditioning that the global state $A = a$.

Matching Demand with Suppliers Depending on the realization of demand, all or a subset of the suppliers will be selected to serve the demand. In particular, the matching between the potential suppliers and demand occurs in three rounds:

Round 1: The platform chooses a payment distribution, π , and draws payments $P_i \stackrel{\text{iid}}{\sim} \pi$ for $i = 1, 2, \dots, n$. Then, for each supplier i , the platform announces both the payment P_i and the underlying distribution π , with the understanding that the supplier will be compensated with P_i for every unit of demand that they will be matched with eventually.

Round 2: Suppliers choose whether they want to be active. A supplier will not be matched with any demand if they choose to be inactive. We write $Z_i \in \{0, 1\}$ to denote whether the i -th participant chooses to participate in the marketplace, and write $T = \sum_{i=1}^n Z_i$ as the total number of active suppliers. The mechanism through which a supplier determines whether or not to become active will be described shortly.

Round 3: The platform employs some mechanism that randomly matches demand with active suppliers.

Denote by S_i the amount of demand that an active supplier i will be able to serve, and define

$$\Omega(d, t) \triangleq \mathbb{E}[S_i \mid D = d, T = t], \quad (3.3)$$

as the expected demand allocation to an active supplier under the payment distribution π , conditional on the total demand being d and total active suppliers being t . We allow for a range of possible matching mechanisms, but assume that in the limiting regime where t and d are large, $\Omega(d, t)$ converges to a “regular allocation function” that only depends on the ratio between the demand and active suppliers, d/t .

Definition 4 (Regular Allocation Function). A function $\omega : \mathbb{R}_+ \rightarrow [0, 1]$ is a regular allocation function if it satisfies the following:

1. $\omega(\cdot)$ is smooth, concave and non-decreasing.
2. $\lim_{x \rightarrow 0} \omega(x) = 0$ and $\lim_{x \rightarrow \infty} \omega(x) \leq 1$.
3. $\lim_{x \rightarrow 0} \omega'(x) \leq 1$.

The condition of ω being concave corresponds to the assumption that the marginal difficulty with which additional demand can be matched does not decrease as demand increases. The condition that $\lim_{x \rightarrow \infty} \omega(x) \leq 1$ asserts that the maximum capacity of all active suppliers be bounded after normalization.

Assumption 1. The function $\Omega : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$ satisfies the following:

1. $\Omega(d, t)$ is non-decreasing in d , and non-increasing in t .
2. There exists a bounded error function $l : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$ with

$$|l(d, t)| = o\left(1/\sqrt{t} + 1/\sqrt{d}\right), \quad (3.4)$$

such that $\Omega(d, t) = \omega(d/t) + l(d, t)$ for all $t, d \in \mathbb{R}_+$, where $\omega(\cdot)$ is a regular allocation function.

We provide below an example system in which the allocation rates are given by a regular allocation function (Definition 4).

Example 5 (Regular Allocation Function Example: Parallel Finite-Capacity Queues). Consider a service system where each active supplier operates as a single-server $M/M/1$ queue with a finite capacity, $L \in \mathbb{N}$, $L \geq 2$. A request that arrives at a queue is accepted if and only if the queue length is less than or equal to L , and is otherwise dropped. We assume that all servers operate at unit-rate, so that a request's service time is an independent exponential random variable with mean 1. Each unit demand generates an independent stream of requests which is modeled by a unit-rate Poisson process, so that the aggregate arrival process of requests is Poisson with rate D (by the merging property of independent Poisson processes). When a new request is generated within the system, the platform routes it to one of the T queues selected uniformly at random. The random routing corresponds, for instance, to a scenario where both the incoming requests and active suppliers are scattered across a geographical area, and as such, requests are assigned to the nearest server.

Within this model, each active supplier effectively functions an $M/M/1$ queue with service rate 1 and arrival rate D/T . Due to the capacity limit at L , some requests may be dropped if they are assigned to a queue currently at capacity. Using the theory of $M/M/1$ queues, it is not difficult to show that (cf. Eq. (5.6) of [Spencer et al., 2014]) if we denote D/T by x , then the rate at which requests are processed by a server, corresponding to the allocation rate, is given by

$$\omega(x) = \begin{cases} \frac{x-x^L}{1-x^L}, & x \neq 1, \\ 1 - \frac{1}{L}, & x = 1. \end{cases} \quad (3.5)$$

Numerical examples of $\omega(\cdot)$ are given in Figure 4. Note that $\omega(\cdot)$ satisfies all conditions in Definition 4 and is hence a regular allocation function. Finally, we may generalize the model to where the suppliers are partitioned into k equal-sized groups, so that each server operates at speed Tkm . The corresponding allocation function would have the same qualitative behavior.

Supplier Choice Behavior We assume that each supplier takes into account their expected revenue in equilibrium when making the decision of whether or not to become active. In particular, the of supplier i becoming active is given as follows, where T is the equilibrium number of active suppliers:

$$\mu_a^{(n)}(\pi) \triangleq \mathbb{P}_\pi [Z_i = 1 \mid A = a] = \mathbb{E}_\pi [f_{B_i}(P_i \mathbb{E}_\pi [\Omega(D, T) \mid A = a]) \mid A = a]. \quad (3.6)$$

Here, $\mathbb{E}_\pi [\Omega(D, T) \mid A = a]$ is the expected amount of demand served by each supplier given the platform's choice of π , and thus $P_i \mathbb{E}_\pi [\Omega(D, T) \mid A = a]$ is the expected revenue of the i -th supplier in mean-field equilibrium.³ B_i is a private feature that captures the heterogeneity across potential suppliers, such as a supplier's cost, or noise in their estimate of the expected revenue. We assume that the B_i 's are drawn i.i.d. from a set \mathcal{B} whose distribution may depend on A . The choice function $f_b(x)$ represents the of the supplier becoming active, when their private feature is b and expected equilibrium revenue is x . We assume the family of choice functions $\{f_b(\cdot)\}_{b \in \mathcal{B}}$ satisfies the following assumption.

³For now, assume that such equilibrium distribution is well defined, and we will justify its meaning rigorously in a moment.

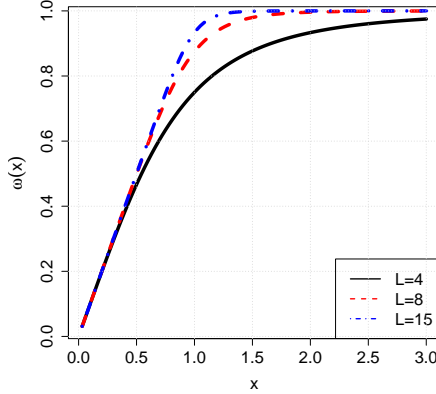


Figure 4: Examples of the regular allocation function $\omega(\cdot)$ in Example 5 under different values of capacity L .

Assumption 2. For all $b \in \mathcal{B}$, the choice function $f_b(\cdot)$ takes values in $[0, 1]$, is monotonically non-decreasing, and twice differentiable with a uniformly bounded second derivative.

Below is one example of a family of choice functions that satisfies Assumption 2:

Example 6 (Logistic Choice Function). A popular model in choice theory is the logit model (cf. Chapter 3 of Train [2009]), which, in our context, corresponds to the choice function being the logistic function:

$$\mathbb{P}[Z_i = 1 \mid P_i, \pi, A] = \frac{1}{1 + e^{-\alpha(P_i \mathbb{E}_\pi[\Omega(D, T) \mid A] - B_i)}}, \quad (3.7)$$

where $\alpha > 0$ is a parameter and the private feature B_i takes values in \mathbb{R}_+ and represent the break-even cost threshold of supplier i . In this example, the supplier's decision on whether to activate will depend on whether their expected revenue exceeds their break-even cost. The sensitivity of such dependence is modeled by the parameter α . Note that in the limit as $\alpha \rightarrow \infty$, the probability of the event $Z_i = 1$ conditionally on P_i, π and A is either 0 or 1. That is, a supplier will choose to be active if and only if they believe their expected revenue from Round 2 will exceed the break-even threshold B_i .

Platform Utility and Objective The platform's utility is defined to be the difference between revenue and total payment:

$$U = R(D, T) - \sum_{i=1}^n P_i Z_i S_i, \quad (3.8)$$

where S_i is the amount of demand that a supplier would serve if they become active, and $R(D, T)$ is the platform's expected revenue, with equilibrium active supply size T and total demand D . Analogous to the case of $\Omega(D, T)$, we will assume that the revenue function R is approximately linear in the sense that, for some function r , $R(D, T) \approx r(D/T)T$ when T and D are large. More precisely, assume the following:

Assumption 3. There exists a bounded error function $l : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$ with $|l(d, t)| = o(1/\sqrt{t} + 1/\sqrt{d})$ such that

$$R(d, t) = (r(d/t) - l(d, t))t, \quad \text{for all } t, d \in \mathbb{R}_+, \quad (3.9)$$

where $r : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is a smooth function with bounded derivatives.

As an example, the platform could receive a fixed amount γ from each unit of demand served, in which case we have $R(D, T) = \gamma(T\Omega(D, T))$. Given this notation, we write the platform's expected utility in the n -th system as

$$u_a^{(n)}(\pi) = \frac{1}{n} \mathbb{E}_n [U | A = a], \quad \text{and} \quad u^{(n)}(\pi) = \mathbb{E}_n [u_A^{(n)}(\pi)]. \quad (3.10)$$

Denote by δ_x the Dirac measure with unit mass on x . We consider two different objectives for the decision maker (i.e., platform operator). First, they may want to control *regret*, and deploy a sequence of payment distributions π whose utility nearly matches that of the optimal *fixed payment*, p^* . Second, they may want to estimate p^* . In Section 4, we provide results with guarantees along both objectives.

Symmetric Payment Perturbation An important family of payment distributions that will be used repeatedly throughout the paper is that of symmetric payment perturbation. Let $\{\varepsilon_i\}_{i \in \mathbb{N}}$ be a sequence of i.i.d. Bernoulli random variables with $\mathbb{P}(\varepsilon_i = -1) = \mathbb{P}(\varepsilon_i = +1) = \frac{1}{2}$. Fix $p > \zeta > 0$. We say the payments are ζ -perturbed from p , if

$$P_i = p + \zeta \varepsilon_i, \quad i \in \mathbb{N}. \quad (3.11)$$

In what follows, we will use $\pi_{p, \zeta}$ to denote the payment distribution when payments are ζ -perturbed from p , $\mu_a^{(n)}(p, \zeta)$ to denote $\mu_a^{(n)}(\pi_{p, \zeta})$. The meanings of $\mu_a^{(n)}(p, \zeta)$, $u_a^{(n)}(p, \zeta)$, etc., are to be understood analogously. When $\zeta = 0$, we may omit the dependence on ζ and write, for instance, $\mu_a^{(n)}(p)$ in place of $\mu_a^{(n)}(p, 0)$ or $\mu_a^{(n)}(\pi_{p, 0})$.

3.1 Mean-Field Limits

The model described above is framed in terms of an equilibrium active supply size, T . We now provide a formal definition, and verify existence and uniqueness.

Definition 7 (Active Supply Size in Equilibrium). We say that a random variable T is an *equilibrium supply size*, if, when all suppliers make activation choices according to (3.6), the resulting distribution for the number of active suppliers equals that of T .

Lemma 1. *Suppose that the conditions in Assumptions 1, 2 and 3 hold. Fix $p > 0$, $\zeta \in [0, p)$, and $a \in \mathcal{A}$. Let the payment distribution π be defined on \mathbb{R}_+ . Then, conditional on $A = a$, the equilibrium active supply size exists, is unique, and follows a Binomial distribution.*

Next, we define some quantities that will play a key role in our analysis, and verify that they converge to tractable mean-field limits. The first quantity we consider is the equilibrium number of active suppliers $\mu_a^{(n)}(p)$, as defined in (3.6). Second, we define the function $q(\cdot)$, which captures the expected amount of demand matched to each supplier if the total number of suppliers were exogenously drawn as a binomial (n, μ) random variable rather than determined by the equilibrium:

$$q_a^{(n)}(\mu) = \mathbb{E} [\Omega(D, X) | A = a], \quad X \sim \text{Binomial}(n, \mu). \quad (3.12)$$

Lemma 2. Under the conditions of Lemma 1, for all $a \in \mathcal{A}$, and $p, \mu \in \mathbb{R}_+$, the following hold:

$$\lim_{n \rightarrow \infty} \mu_a^{(n)}(p) = \mu_a(p), \quad (3.13)$$

$$\lim_{n \rightarrow \infty} q_a^{(n)}(\mu) = \omega(d_a/\mu), \quad (3.14)$$

$$\lim_{n \rightarrow \infty} u_a^{(n)}(p) = u_a(p) = (r(d_a/\mu_a(p)) - p\omega(d_a/\mu_a(p))) \mu_a(p), \quad (3.15)$$

$$\lim_{n \rightarrow \infty} (q_a^{(n)})'(\mu) = -\omega'(d_a/\mu) \frac{d_a}{\mu^2}, \quad (3.16)$$

where $\omega(\cdot)$ and $r(\cdot)$ are described in Definition 4 and Assumption 3, respectively. In (3.13), the limit $\mu_a(p)$ is the only solution to $\mu = \mathbb{E} [f_{B_1}(p\omega(d_a/\mu)) \mid A = a]$.

Finally, the following result establishes conditions under which the limiting utility functions $u_a(p)$ are concave, thus enabling us to globally optimize utility via first-order methods.

Lemma 3. Let $f_a(\cdot)$ be the average choice function: $f_a(x) = \mathbb{E} [f_{B_1}(x) \mid A = a]$. Fix $\gamma > 0$, $c_0 \in (0, \gamma)$ and $a \in \mathcal{A}$. Suppose the following holds:

1. We have a linear revenue function, $r(x) = \gamma\omega(x)$.
2. Let $\underline{x} = \inf_{p \in (c_0, \gamma)} pq_a(\mu_a(p))$ and $\bar{x} = \sup_{p \in (c_0, \gamma)} pq_a(\mu_a(p))$. The average choice function $f_a(\cdot)$ satisfies
 - (a) $f_a(\cdot)$ is strongly concave in the domain (\underline{x}, \bar{x}) .
 - (b) $f_a(\underline{x}) - f'_a(\underline{x})\underline{x} \geq 0$, or, equivalently, that there exists a differentiable, non-negative concave function $\tilde{f}(\cdot)$, such that $\tilde{f}(\underline{x}) = f_a(\underline{x})$ and $\tilde{f}'(\underline{x}) \leq f'_a(\underline{x})$.
3. The allocation function $\omega(\cdot)$ is strongly concave in the domain $(d_a/\mu_a(c_0), d_a/\mu_a(\gamma))$.

Then, under the conditions of Lemma 1, the limiting platform utility $u_a(\cdot)$ is strongly concave in the domain (c_0, γ) .

3.2 The Marginal Response Function

Finally, as discussed in Section 2, a key quantity that motivates our approach to experimentation is the marginal response function, $\Delta(p)$, which captures the average payment sensitivity of a supplier in a situation where only they get different payments but others do not (meaning that there are no network effects).

Definition 8 (Marginal Response Function). Fix $n \in \mathbb{N}$, $a \in \mathcal{A}$ and $p > 0$. The marginal response function is defined by

$$\Delta_a^{(n)}(p) = q_a^{(n)}(\mu_a^{(n)}(p)) \mathbb{E} \left[f'_{B_1} \left(pq_a^{(n)}(\mu_a^{(n)}(p)) \right) \mid A = a \right]. \quad (3.17)$$

This marginal response function Δ plays a key role in our analysis for the following reasons. First, as shown in the following section, in the mean-field limit as $n \rightarrow \infty$, Δ is easy to estimate using small random payment perturbations that do not meaningfully affect the overall equilibrium. Second, provided we have a good enough understanding of the underlying system dynamics to know the appropriate allocation function $\omega(\cdot)$, we can use

consistent estimates of Δ to estimate the true payment sensitivity of supply that accounts for feedback effects, $d\mu(p)/dp$. This fact is formalized in the following result. We note that, other than Δ , all terms on the right-hand side of (3.20) are readily estimated from observed data by taking averages.

Lemma 4. *Under the conditions of Lemma 1, for any $a \in \mathcal{A}$ and $p \in \mathbb{R}_+$, we have that*

$$\frac{d}{dp} \mu_a^{(n)}(p) = \frac{\Delta_a^{(n)}(p)}{1 - p\Delta_a^{(n)}(p) q_a^{(n)'} \left(\mu_a^{(n)}(p) \right) / q_a^{(n)} \left(\mu_a^{(n)}(p) \right)} \text{ for any } n \geq 1. \quad (3.18)$$

Furthermore, this relationship carries through in the mean-field limit,

$$\lim_{n \rightarrow \infty} \Delta_a^{(n)}(p) = \Delta_a(p) \triangleq \omega(d_a/\mu_a(p)) \mathbb{E} [f'_{B_1}(p\omega(d_a/\mu_a(p))) \mid A = a], \quad (3.19)$$

$$\lim_{n \rightarrow \infty} \frac{d}{dp} \mu_a^{(n)}(p) = \mu'_a(p) = \Delta_a(p) / \left(1 + \frac{pd_a\Delta_a(p)\omega'(d_a/\mu_a(p))}{\mu_a(p)^2\omega(d_a/\mu_a(p))} \right). \quad (3.20)$$

4 Learning via Local Experimentation

We present our main results in this section. The main framework we adopt for learning payments is based on first-order optimization. We first show that our local experimentation approach enables us to construct a useful estimate of the utility gradient at the current value of p (Section 4.1); then, we use these gradient estimates to update the payment using a form of gradient ascent (Section 4.2). Throughout this section, we focus on optimizing utility in the $(n \rightarrow \infty)$ mean-field limit, while verifying that finite- n errors have an asymptotically vanishing effect on learning. In Section 4.3, we study the cost of the local experimentation needed to estimate utility gradients, and verify that it scales sub-linearly in n . Finally in Section 4.4, we compare our results to those available to classical continuous-armed bandits.

4.1 Estimating Utility Gradients

Recall that, in our model, there are two sources of randomness. First, there is the stochastic global context $A \in \mathcal{A}$, which affects overall demand. In the context of ride-sharing, A could capture multiplicative demand fluctuations due to weather or holidays. Second, there is randomness due to decisions of individual market participants. This second source of error decays with market size size n . Our goal here is to verify that local experimentation allows us to eliminate errors of the second type via concentration as the market size n gets large. Conversely, because the context A affects everyone in the same way, there is no way to average out the effect of A without collecting data across many days.

Define $\bar{Z} = \frac{1}{n} \sum_{i=1}^n Z_i$ and $\bar{D} = D/n$. As discussed in Section 2 our proposal starts for perturbing individual payments as in (3.11), and then estimating the regression coefficient $\hat{\Delta}$ of market participation Z_i on the perturbation $\zeta_n \varepsilon_i$, i.e.,

$$\hat{\Delta} = \zeta_n^{-1} \sum_{i=1}^n (Z_i - \bar{Z})(\varepsilon_i - \bar{\varepsilon}) / \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2. \quad (4.1)$$

Our first result below relates this quantity $\hat{\Delta}$ we can estimate via local randomization to a quantity that is more directly relevant to estimating payments, namely the payments derivative of u conditionally on the global state A .

Theorem 5. *Suppose the conditions of Lemma 1 hold. Let*

$$\hat{\Upsilon} = \hat{\Delta} / \left(1 + \frac{p\bar{D}\hat{\Delta}\omega'(\bar{D}/\bar{Z})}{\bar{Z}^2\omega(\bar{D}/\bar{Z})} \right), \quad (4.2)$$

and

$$\hat{\Gamma} = \hat{\Upsilon} [r(\bar{D}/\bar{Z}) - p\omega(\bar{D}/\bar{Z}) - (r'(\bar{D}/\bar{Z}) - p\omega'(\bar{D}/\bar{Z}))\bar{D}/\bar{Z}] - \omega(\bar{D}/\bar{Z})\bar{Z}. \quad (4.3)$$

Then, assuming that the perturbations scale as $\zeta_n = \zeta n^{-\alpha}$ for some $0 < \alpha < 0.5$,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\left| \hat{\Gamma} - \frac{d}{dp} u_A(p) \right| > \varepsilon \right] = 0, \quad (4.4)$$

for any $\varepsilon > 0$.

4.2 A First-Order Algorithm

Our key use of Theorem 5 involves optimizing for a utility-maximizing p . At every time period t , $\hat{\Gamma}_t$ is a consistent estimate of the gradient of $u_{A_t}(\cdot)$ at p_{t-1} , and we can plug it into any first-order optimization method that allows for noisy gradients. The proposal below is a variant of mirror descent that allows us to constraint the p_t to an interval I [e.g., Beck and Teboulle, 2003]. We need to specify a step size η , an interval $I = [c_-, c_+]$, and an initial payment p_1 . Then, at time period $t = 1, 2, \dots$, we do the following:

1. Deploy randomized payment perturbations (3.11) around p_t to estimate $\hat{\Gamma}_t$ as in (4.3),
2. Perform a gradient update⁴

$$p_{t+1} = \operatorname{argmin}_p \left\{ \frac{1}{2\eta} \sum_{s=1}^t s(p - p_s)^2 - \theta_t p : p \in I \right\}, \quad \theta_t = \sum_{s=1}^t s \hat{\Gamma}_s. \quad (4.5)$$

The following result shows that if we run our method for T time periods in a large market-place and the reward functions $u_a(\cdot)$ are strongly concave, then the utility derived by our first-order optimization scheme is competitive with any fixed payment level p , up to regret that decays as $1/t$.⁵

Theorem 6. *Under the conditions of Theorem 5, suppose we run the above learning algorithm for T time periods and that $u_a(\cdot)$ is σ -strongly concave over the interval $p \in I$ for all a . Suppose, moreover, that we run (4.5) with step size $\eta > \sigma^{-1}$ and that the gradients of u are bounded, i.e., $|u'_a(p)| < M$ for all $p \in I$ and $a \in \mathcal{A}$. Then*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\frac{1}{T} \sum_{t=1}^T t (u_{A_t}(p) - u_{A_t}(p_t)) \leq \frac{\eta M^2}{2} \right] = 1, \quad (4.6)$$

for any $p \in I$ and $T \geq 1$.

⁴Note that, without the constraint to the interval I , this update is equivalent to basic gradient descent with $p_{t+1} = p_t + 2\eta\hat{\Gamma}_t/(t+1)$.

⁵In (4.6), we up-weight the regret terms $u_{A_t}(p) - u_{A_t}(p_t)$ in later time periods to emphasize their $1/t$ rate of decay. One could also use an analogous proof to verify that the unweighted average regret is bounded on the order of $T^{-1} \sum_{t=1}^T (u_{A_t}(p) - u_{A_t}(p_t)) = \mathcal{O}_P(\log(T))$.

The above result doesn't make any distributional assumptions on the contexts A_t ; rather, (4.6) bounds the regret of our payment sequence p_1, p_2, \dots along the realized sample path of A_t relative to any fixed oracle. We believe this aspect of our result to be valuable in many situations: For example, if A_t needs to capture weather phenomena that have a big effect on demand, it is helpful not to need to model the distribution of A_t , as the weather may have complex dependence in time as well as long-term patterns. However, if we are willing to assume that the A_t are independent and identically distributed, Theorem 6 also implies that an appropriate average of our learned payments is consistent for the optimal payment via online-to-batch conversion [Cesa-Bianchi et al., 2004].

Corollary 7. *Under the conditions of Theorem 6, suppose moreover that the A_t are independent and identically distributed and let $u(p) = \mathbb{E} [u_{A_t}(p)]$. Then, for any $\delta > 0$,*

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left[(p^* - \bar{p}_T)^2 \leq \frac{\eta M^2}{\sigma T} (16 \log(\delta^{-1}) + 4) \right] \geq 1 - \delta, \quad (4.7)$$

where $p^* = \operatorname{argmax} \{u(p) : p \in I\}$ and $\bar{p}_T = \frac{2}{T(T+1)} \sum_{t=1}^T t p_t$.

4.3 The Cost of Experimentation

Our argument so far has proceeded in two parts. In Section 4.1 we showed we could consistently use local experimentation to estimate gradients of the utility function $u_a(p)$. Then, in Section 4.2, we gave bounds on the regret that updates payments p_t via gradient descent—as though the platform could observe gradients $u'_a(p)$ at no additional cost. Here, we complete the picture, and show that local experimentation in fact induces negligible excess cost as we approach the mean-field limit. In general, a platform that randomizes payments around p_t will make lower profits than one that just pays everyone p_t ;⁶ the result below, however, shows that this excess cost decays quadratically in the magnitude of payment perturbations ζ .

Theorem 8. *Under the conditions of Theorem 5 there are constants $C, \alpha > 0$ such that*

$$\frac{1}{T} \sum_{t=1}^T (u_{A_t}(p_t) - u_{A_t}(p_t, \zeta_t)) \leq C \zeta^2 \quad \text{for all } 0 \leq \zeta < \alpha. \quad (4.8)$$

Recall that, as the market size gets large, Theorem 5 enables us to estimate gradients of $u_a(p)$ in large- n markets using an amount of randomization that scales as $n^{-\alpha}$ for some $0 < \alpha < 0.5$. Combined with Theorem 8, this result implies that we can in fact estimate gradients of $u_a(p)$ “for free” via local experimentation when n is large, and that the regret of a platform deploying our platform matches to first order the regret of an oracle who was able to run first-order optimization on the mean-field limit.

4.4 Comparison with Bandit Algorithms

The above results present a marked contrast to the case of continuous-armed bandits (or zeroth-order optimization), where the platform only gets to observe realized utility $u_{A_t}(p_t)$,

⁶This is because randomization will not affect active supply size to first order, but suppliers randomized to higher payments are more likely to be active. Randomization thus increases the average per-unit payment the platform needs to suppliers without increasing the amount of demand the platform is able to serve.

and does not get any additional information about $u'_{A_t}(p_t)$. Unlike our local experimentation approach, continuous-armed bandit-based policies cannot effectively exploit the strong concavity of $u_{A_t}(\cdot)$. Thus, whereas our regret bound (4.6) corresponds to a $1/T$ rate of convergence, bandit-based policies can only attain average regret scaling as $1/\sqrt{T}$ with the number of time periods T ; see Bubeck et al. [2017] and Shamir [2013] for algorithms and matching lower bounds.⁷ Furthermore, even getting a $1/\sqrt{T}$ regret with a bandit-based policy requires some care. For example, Flaxman et al. [2005] and Kleinberg [2005] proposed continuous-armed bandit algorithms that estimate derivatives by noisy function evaluations and—while desirable due to their transparency and ease of implementation—these methods only attain a $1/\sqrt[4]{T}$ rate of convergence in average regret.

In other words, our ability to use mean-field modeling to leverage small-scale payment variation within (rather than across) time periods enables us to fundamentally alter the difficulty of the problem of learning the optimal p , and to improve our rate of convergence in T . Here, we only considered using local experimentation to drive a first-order optimization algorithm, but our conclusions extend more broadly than that. For example, the lower-bounds of Shamir [2013] imply that even Bayesian zeroth-order optimization as considered in, e.g., Letham et al. [2018] cannot attain average regret bounds that decay faster than $1/\sqrt{T}$. However, a Bayesian learning method that incorporated our estimates of $u'_{A_t}(p_t)$ derived via local experimentation could potentially do better.

Finally, we note that the well-known slow rates of convergence for continuous-armed bandits have led some authors to studying a query model where we can evaluate the unknown functions $u_{A_t}(\cdot)$ twice rather than once; for example, Duchi et al. [2015] show that two function evaluations can result in substantially faster rates of convergence than one. The reason for this gain is that, given two function evaluations, the analyst directly cancel out the main effect of the global noise term A_t . In our setting, it is implausible that a platform could carry out such paired function evaluations in practice unless, e.g., they simultaneously run experiments across two identical twin cities. But in this paper, we found that—by leveraging structural information and mean-field modeling—local experimentation can be used to obtain similar gains over zeroth-order optimization as one could get via twin evaluation.

5 Generalizations: Risk Aversion and Surge Pricing

So far, we have focused our discussion on a specific a model of a centralized market for freelance labor: The platform chooses a distribution π and then, for each supplier i , draws $P_i \sim \pi$ and promises to pay the supplier P_i per unit of demand served; the supplier computes $q_A(\pi)$, the expected number of units of demand they will get to serve if they join the market; finally, each supplier compares their expected revenue $P_i q_A(\pi)$ to their outside option and chooses whether or not to join the marketplace. Our main results were that: 1) In large markets, we can unobtrusively estimate a marginal response function via local experimentation; 2) The behavior of this marketplace can be characterized by a mean-field limit; 3) In the mean-field limit, we can transform estimates of the marginal response function into predictions of the effect of policy-relevant interventions. Thus, in large markets, we can use local experimentation for optimizing platform choices.

As outlined in the introduction, however, we expect the general principles outlined here

⁷As discussed in Shamir [2013], the $1/\sqrt{T}$ lower bound on regret holds even if A_t are independent and identically distributed and the objective function $u(\cdot)$ is quadratic.

to be more broadly applicable, beyond the model given above. While a full theory of experimental design powered by mean-field equilibria is beyond the scope of this paper, in this section we discuss two extensions of our model that are of considerable practical interest: risk-averse suppliers and surge pricing. We define models for both problems below, and write down balance conditions generalizing (3.6). Afterwards, we conjecture the existence and form of a mean-field equilibrium, and show that the conjectured equilibrium model lets us again map from consistent estimates of a marginal response function to relevant counterfactual predictions—using the same recipe as deployed in the rest of this paper.

Example 9 (Risk Aversion). Under risk aversion, supplier utility functions may not scale linearly with their revenue, and instead there is a concave function β such that the relevant quantity for understanding the suppliers’ choices is the expectation of $\beta(\text{revenue})$ [Holt and Laury, 2002, Pratt, 1978]. Suppose that $\beta(0) = 0$, and that each worker can serve 0 or 1 units of demand.⁸ Then our balance condition (3.6) becomes

$$\mu_a^{(n)}(\pi) = \mathbb{P}_\pi [Z_i = 1 \mid A = a] = \mathbb{E}_\pi \left[f_{B_i} \left(\beta(P_i) q_a^{(n)}(\mu_a^{(n)}(\pi)) \right) \mid A = a \right]. \quad (5.1)$$

The curvature of the function $\beta(\cdot)$ thus corresponds to the degree of a supplier’s risk aversion, and setting $\beta(p) = p$ recovers our original risk-neutral model.

Example 10 (Supply-Side Surge Pricing). Several prominent ride sharing platforms deploy surge pricing where, in case of heavy demand, the platform applies a multiplier (generally greater than 1) to the original payment in order to encourage higher supplier participation [Cachon et al., 2017, Hall et al., 2015]. As a simple model, suppose that surge is triggered automatically based on the supply-demand ratio, i.e., there is a function $s : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that, in each period, the i -th supplier gets paid $s(D/T)P_i$ per unit of demand served. Suppliers can anticipate surge and, as in the rest of the paper, they make decisions based on limiting values of all random variables. Thus, suppliers anticipate payments $s(d_a/\mu_a(\pi))P_i$, resulting in a balance condition

$$\mu_a^{(n)}(\pi) = \mathbb{P}_\pi [Z_i = 1 \mid A = a] = \mathbb{E}_\pi \left[f_{B_i} \left(s \left(\frac{d_a}{\mu_a^{(n)}(\pi)} \right) P_i q_a^{(n)}(\mu_a^{(n)}(\pi)) \right) \mid A = a \right], \quad (5.2)$$

where again $s(x) = 1$ recovers our original model.

In both examples above, we conjecture that—in analogy to Lemma 4—a mean-field limit exists and that it can be characterized by analogues of (5.1) and (5.2) but without the n -superscripts. In this case, we can write both mean-field limits in a unified form via *generalized earning functions*, $\theta : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$, so that the asymptotic balance condition is

$$\mu_a(\pi) = \mathbb{P}_\pi [Z_i = 1 \mid A = a] = \mathbb{E}_\pi \left[f_{B_i} (\theta(P_i, q_a(\mu_a(\pi)))) \mid A = a \right]. \quad (5.3)$$

In the case of (5.1), we have $\theta_{risk}(p, q) = \beta(p)q$. Meanwhile, for (5.2), recall that in the mean-field limit the matching of supply and demand is characterized by the identity $q_a(\mu_a(\pi)) = \omega(d_a/\mu_a(\pi))$. Thus, our conjecture means that (5.2) converges to (5.3) with generalized earning function $\theta_{surge}(p, q) = pqs(\omega^{-1}(q))$.

We close this section by carrying out “step 3” of the analysis outlined in the first paragraph of this section, i.e., by showing how (5.3) lets us map from a marginal response

⁸Generalizations to workers who can serve many units of demand are immediate, at the expense of more involved notation.

function to utility gradients with respect to surge; we leave verification of the conjectured convergence to (5.3) for further work. To this end, fix $a \in \mathcal{A}$. First, it is not difficult to show that the changes caused by the introduction of $\theta(\cdot)$ affects the computation of utility derivative $u'_a(p)$ only through the expression for $\mu'_a(p)$ (cf. the proof of Proposition 14). Hence, we here only focus on expressions for $\mu'_a(p)$.

Now, we can directly check that a reduced form expression as in (2.1) allows us to estimate the following marginal response function via local randomization,

$$\Delta_a(p) = (\nabla\theta)_1(p, q_a(\mu_a(p)))\mathbb{E} [f'_{B_1}(\theta(p, q_a(\mu_a(p))) | A = a]. \quad (5.4)$$

where $(\nabla\theta)_i(\cdot, \cdot)$ denotes the i th coordinate of the gradient of θ . Meanwhile, an argument based on the chain rule similar to that in the proof of Lemma 4 shows that

$$\mu'_a(p) = \Delta_a(p) / \left(1 + \frac{(\nabla\theta)_2(p, q_a(\mu_a(p)))}{(\nabla\theta)_1(p, q_a(\mu_a(p)))} \omega' \left(\frac{d_a}{\mu_a(p)} \right) \frac{d_a}{\mu_a^2(p)} \Delta_a(p) \right). \quad (5.5)$$

Note, furthermore, that all quantities in (5.5) except $\Delta_a(p)$ are either known a-priori or can be estimated via observed averages. The upshot is that the mean-field equilibrium characterized by (5.3) enables us to map an easy-to-estimate marginal response function to $\mu'_a(p)$ via (5.5). These estimates of $\mu'_a(p)$ can then be directly used to compute utility gradients $u'_a(p)$ that can be used for first-order optimization.

6 Simulation Results

We now consider a more comprehensive empirical evaluation of the performance of local versus global experimentation, building on in the simulation results of Section 2, and compare mean performance of local experimentation and global experimentation across 1,000 simulation replications. Local experimentation is run for 200 steps, exactly as described in Section 2, with a random initialization $p_1 \sim \text{Unif}(10, 30)$. Meanwhile, for global experimentation, we consider a collection of strategies that first randomly draw payments $p_t \sim \text{Unif}(10, 30)$ for the first $1 \leq t \leq T$ time periods, fit a spline to the data (as in the left panel of Figure 3), and then deploy the learned policy for the remaining $200 - T$ time periods. We consider the choices $T \in \{40, 60, 80, \dots, 200\}$. For both methods, we report both in-sample regret, i.e., the mean utility shortfall relative to deploying the population-optimal p^* for the 200 learning periods, as well as future expected regret, i.e., the expected utility shortfall from deploying the learned policy \hat{p} after the 200 learning periods. For local experimentation, we set $\hat{p} = 2 \sum_{t=1}^T t p_t / (T(T+1))$ following Corollary 7, whereas for global experimentation we set \hat{p} to be the output of spline optimization discussed above.

As seen in the left panel of Figure 5, local experimentation outperforms global experimentation by an order of magnitude along both metrics. Quantitatively, local experimentation achieved mean in-sample regret of 0.025 and mean future regret of 0.0045. In contrast, the best numbers achieved by global experimentation for these metrics were 0.57 and 0.12 respectively—and there was not a single choice of tuning parameters that achieved both. In general, we see that a larger choice of T always improves future regret, whereas for in-sample regret there is an optimal middle ground that balances exploration and exploitation (here, $T = 80$).

Next, we consider an analogous simulation design, but with supply-side surge pricing. As discussed in Section 5, we assume that the platform makes a public commitment to mechanically increase supply side payments by a multiplicative factor $s(D/T)$ once the

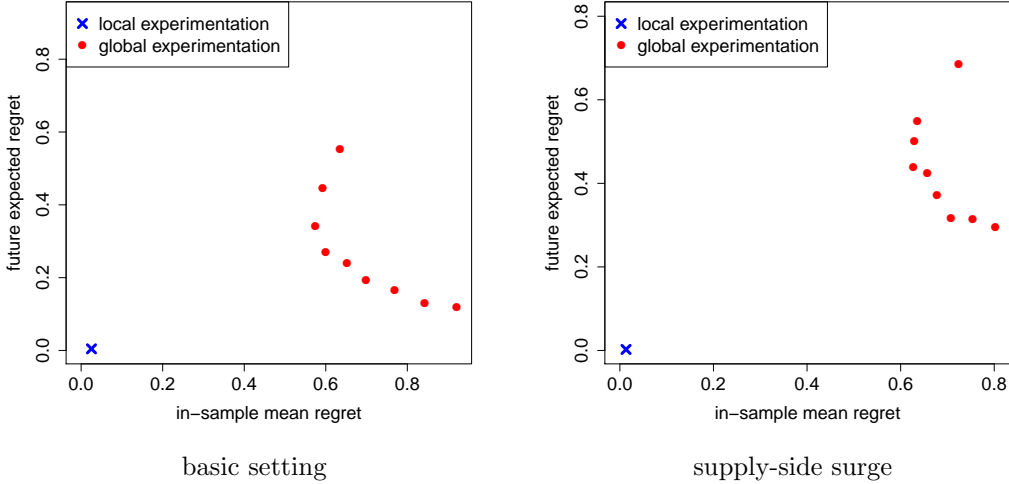


Figure 5: Comparison of the regret of local and global experimentation in the setting of Section 2, averaged across 1,000 simulation replications. The global experimentation path is detailed in Section 6. In the right panel, the platform makes a public commitment to multiply supply-side payments by a surge factor (6.1).

demand D and supply T are realized, and suppliers take this commitment into account when choosing whether or not to join the marketplace. Here, we use

$$s(D/T) = \frac{D}{T} / \omega\left(\frac{D}{T}\right), \quad (6.1)$$

meaning that, by the properties of $\omega(\cdot)$ as outlined in Definition 4, the surge multiplier is 1 when D is small relative to T , but eventually climbs up to the ratio D/T as demand outpaces supply. This choice of $s(\cdot)$ is by no means optimal; it is simply an example.

As discussed in Section 5, our analysis of surge relies on a conjecture that relevant properties of mean-field limits as discussed in Section 3.1 still hold with surge. We work with a limiting platform utility function that depends linearly on revenue minus costs as in Lemma 3,

$$u_a(p) = \left(\gamma - ps\left(\frac{d_a}{\mu_a(p)}\right)\right) \omega\left(\frac{d_a}{\mu_a(p)}\right) \mu_a(p). \quad (6.2)$$

As discussed above, we can estimate the p -derivative of the expected scaled active supply size, $\mu'_a(p)$, by local experimentation via (5.4) and (5.5). Moreover, following the argument of Theorem 5, we obtain p -derivatives of $u_a(p)$ via

$$\begin{aligned} u_a(p) &= \left(\gamma - ps\left(\frac{d_a}{\mu_a(p)}\right)\right) \left(-\omega'\left(\frac{d_a}{\mu_a(p)}\right) \frac{d_a}{\mu_a(p)} + \omega\left(\frac{d_a}{\mu_a(p)}\right)\right) \mu'_a(p) \\ &\quad - s\left(\frac{d_a}{\mu_a(p)}\right) \omega\left(\frac{d_a}{\mu_a(p)}\right) \mu_a(p) + ps'\left(\frac{d_a}{\mu_a(p)}\right) \frac{d_a}{\mu_a(p)} \omega\left(\frac{d_a}{\mu_a(p)}\right) \mu'_a(p). \end{aligned} \quad (6.3)$$

We turn this into a feasible estimator plugging in our local experimentation estimates of $\hat{\mu}'_a(p)$ for $\mu'_a(p)$, and estimating the ratio $d_a/\mu_a(p)$ via its sample analogue D/T .

Results for learning p are given in the right panel of Figure 5. Qualitatively, the results match those obtained without surge, and local experimentation still outperforms global experimentation by an order of magnitude. Local experimentation achieved mean in-sample regret of 0.013 and mean future regret of 0.0024, while the best corresponding numbers achieved by global experimentation for these metrics were 0.63 and 0.29 respectively. We also note that adding the automatic surge multiplier as in (6.1) decreased the optimal base payment from 17.6 to 15.7, while increasing optimal mean platform utility by 0.06 (the median utility difference is 0.04). Thus, in this example, the regret of global experimentation is much larger than the utility gain from using surge as in (6.1) relative to not using surge—whereas the regret of local experimentation is less than the effect of adopting surge.

Finally, as discussed in Section 2, the global experimentation baseline used here is fairly simple, and it is plausible that a more sophisticated implementation of global experimentation using, e.g., Bayesian principles, could do better. We note, however, that other approaches to global experimentation that we tried here (e.g., following Flaxman et al. [2005] or Kleinberg [2005]) performed worse than the baseline we report results for, and that formal results presented in Section 4.4 imply a separation in rates of convergence for in-sample regret that can be achieved via local versus global experimentation. Thus, we expect the order-of-magnitude advantage our method enjoys relative to global experimentation to withstand improvements that could be obtained via specialized variants of global experimentation.

7 Discussion

We introduced in this paper a new framework for experimental design in stochastic systems with significant cross-unit interference. The key insight is that, in certain families of models, the inference is structured enough to be captured by a small number of key statistics, such as the global demand-supply equilibrium, and the impact of interference can be subsequently accounted for using mean-field asymptotics. We then proposed an approach based on local experimentation that would allow us to accurately and efficiently estimate the utility gradient in the large-system limit, and use these gradient estimates to perform first-order optimization.

We believe that the general approach proposed in this paper, one that leverages stochastic modeling and mean-field asymptotics in experimental design, has the potential to be applicable in a wider range of problems. As one example, we may consider models in which the key statistics that capture the interference patterns are multi-dimensional. This could occur in a marketplace which, instead of being full centralized, consists of a small number of inter-connected sub-markets. For instance, in a ride-sharing platform, the sub-markets may correspond to neighboring cities connected by highways and bridges. In these systems, suppliers' behaviors remain to be primarily influenced by the local supply-demand equilibrium in their respective sub-markets. These local equilibria in turn interact with one another due to network effects. Nevertheless, in a large-market regime where the numbers of market participants are relatively large in all sub-markets, while the total number of sub-markets remains the same, we may still use the type of mean-field asymptotics in this paper to account for the interference across both individual units and sub-markets to efficiently estimate the effect of payment adjustments. In another direction, we may extend the one-shot equilibrium model adopted in this paper to dynamic settings where the equilibrium emerges

gradually according to a stochastic process (e.g., suppliers may adapt to payment variations only over time), and study whether a dynamic version of our mean-field model can be used to analysis the effects of local experimentation in these systems.

8 Proof of Main Results

8.1 Proof of Lemma 1

Fix $a \in \mathcal{A}$. Recall that all suppliers know the realization of the global state, a , and the probability that a given supplier will choose to become active is given by (3.6). Define

$$\psi_a^{(n)}(\mu, \pi) \triangleq \mathbb{E} [f_{B_1} (P_1 \mathbb{E} [\Omega(D, X) | A = a]) | A = a], \quad X \sim \text{Binomial}(n, \mu). \quad (8.1)$$

That is, $\psi_a^{(n)}(\mu, \pi)$ is the probability of a supplier becoming active under the payment distribution π , if they believe that the active supply size is Binomial (n, μ) . Note that, since the same payment distribution applies uniformly across all suppliers, so are the probabilities of the suppliers becoming active. As a result, for any π , the actual active supply size T will follow a Binomial distribution. In particular, this implies that T is an equilibrium active supply size if and only if it is binomial with mean μ that satisfies the following fixed-point equation:

$$\psi_a^{(n)}(\mu, \pi) = \mu. \quad (8.2)$$

It suffices to show that (8.2) admits a unique solution in the domain $\mu \in [0, 1]$. Because $f_b(\cdot)$ is by construction non-decreasing, it follows that $\psi_a^{(n)}(\mu, \pi)$ is a continuous function and non-increasing in μ : A supplier is more discouraged from becoming active, if they believe there will be more active suppliers in the market eventually. In particular, the left-hand side of (8.2), $\psi_a^{(n)}(\mu, \pi)$, is a non-negative, continuous and non-increasing function over $\mu \in [0, 1]$, which implies that (8.2) admits a unique solution in $[0, 1]$.

8.2 Proof of Lemma 2

Our argument will leverage the following simple expression for the limiting derivative of $q(\cdot)$, the proof of which follows immediately from a generalization of a classical result of Stein [1981] to exponential families; The proof is given in Appendix A.1.

Proposition 9. *Fix $a \in \mathcal{A}$ and $\mu > 0$. Then, $\frac{d}{d\mu} q_a^{(n)}(\mu)$ is non-positive, and*

$$\lim_{n \rightarrow \infty} \frac{d}{d\mu} q_a^{(n)}(\mu) = \frac{d}{d\mu} \omega(d_a/\mu) = -\omega'(d_a/\mu) \frac{d_a}{\mu^2}. \quad (8.3)$$

The claim in (3.14) follows directly from the definition of Ω and Assumption 1, i.e., that $\Omega(d, t)$ converges to $\omega(d/t)$ as $t \rightarrow \infty$, and the fact that conditional on $A = a$, D/n concentrates on d_a as $n \rightarrow \infty$. For (3.13), recall that $\mu_a^{(n)}(p)$ is the solution to the balance equation in (8.2): $\mu = \mathbb{E} [f_{B_1}(p q_a^{(n)}(\mu)) | A = a]$. By (3.14), and the monotonicity of the functions $f_b(\cdot)$ and $\omega(\cdot)$, we have that $\mu_a^{(n)}(p)$ converges to $\mu_a(p)$ as $n \rightarrow \infty$, where $\mu_a(p)$ is the solution to the limiting balance equation given in the statement of Lemma 2. The claim in (3.15) follows from (3.13), (3.14) and Assumption 3. The convergence of $q_a^{(n)'}(\mu)$ in (3.16) follows from Proposition 9.

8.3 Proof of Lemma 3

Fix $a \in \mathcal{A}$. It follows from Lemma 2 that

$$u_a(p) = \gamma \mu_a(p) q_a(\mu_a(p)) - p \mu_a(p) \omega(d_a/p) = (\gamma - p) \mu_a(p) q_a(\mu_a(p)), \quad (8.4)$$

where

$$q_a(\mu) = \omega(d_a/\mu). \quad (8.5)$$

The following basic fact will be useful; The proof is given in Appendix A.2.

Proposition 10. *Let $g : \mathbb{R}_+ \rightarrow \mathbb{R}$ be a concave function such that $g(0) \geq 0$. Then*

$$g(x) \geq g'(x)x, \quad \text{for all } x > 0. \quad (8.6)$$

We have that

$$u_a''(p) = (\gamma - p) \frac{d^2}{dp^2} (q_a(\mu_a(p)) \mu_a(p)) - 2 \frac{d}{dp} (q_a(\mu_a(p)) \mu_a(p)). \quad (8.7)$$

Note that $q_a(\mu_a(p)) \mu_a(p)$ is the normalized amount of demand that ends up being served. The next result shows that $q_a(\mu_a(p)) \mu_a(p)$ is non-decreasing in the payment p ; The proof is given in Appendix A.3.

Proposition 11. *There exists $c > 0$, such that*

$$\frac{d}{dp} (q_a(\mu_a(p)) \mu_a(p)) \geq c, \quad \text{for all } p \in (c_0, \gamma). \quad (8.8)$$

Because $p < \gamma$, in light of Proposition 11, in order to show that $u_a(\cdot)$ is strictly concave, it suffices to demonstrate that

$$\frac{d^2}{dp^2} (q_a(\mu_a(p)) \mu_a(p)) \leq 0. \quad (8.9)$$

To this end, we have that

$$\begin{aligned} & \frac{d^2}{dp^2} (q_a(\mu_a(p)) \mu_a(p)) \\ &= \mu_a''(p) (\mu_a(p) q'(\mu_a(p)) + q_a(\mu_a(p))) + \mu_a'(p)^2 (\mu_a(p) q''(\mu_a(p)) + 2q'_a(\mu_a(p))) \\ &\stackrel{(a)}{=} \mu_a''(p) (\mu_a(p) q'(\mu_a(p)) + q_a(\mu_a(p))) + \mu_a'(p)^2 \frac{d_a^2 \omega''(d_a/\mu_a(p))}{\mu_a(p)^3} \\ &\stackrel{(b)}{=} \mu_a''(p) \left(\omega(d_a/\mu_a(p)) - \omega'(d_a/\mu_a(p)) \frac{d_a}{\mu_a(p)} \right) + \mu_a'(p)^2 \frac{d_a^2 \omega''(d_a/\mu_a(p))}{\mu_a(p)^3} \end{aligned} \quad (8.10)$$

where steps (a) and (b) follow from the fact that $q_a(\mu) = \omega(d_a/\mu)$. Because $\omega(\cdot)$ is concave, it follows that the second term in (8.10) is non-positive. Furthermore, recall from Definition 4 that $\omega(\cdot)$ is concave and $\omega(0) = 0$. By Proposition 10, we have that

$$\omega(d_a/\mu_a(p)) - \omega'(d_a/\mu_a(p)) \frac{d_a}{\mu_a(p)} \geq 0. \quad (8.11)$$

In the remainder of the proof, we will focus on showing that

$$\mu_a''(p) \leq 0, \quad (8.12)$$

which would imply the strong concavity of $u_a(\cdot)$.

Recall that, by construction, the average choice function $f_a(\cdot)$ is non-decreasing and concave. Recall from (8.2) that $\mu_a(p)$ satisfies the fixed-point equation:

$$\mu_a(p) = \psi_a(\mu_a(p), p) \quad (8.13)$$

where

$$\psi_a(\mu, p) = f_a(pq_a(\mu)). \quad (8.14)$$

Twice-differentiating (8.13) with respect to p , we obtain that

$$\mu_a''(p) = \left\{ \frac{\partial}{\partial \mu} \psi_a(\mu, p) \right\}_{\mu=\mu_a(p)} \mu_a''(p) + (\mu_a'(p), 1) \mathbf{H}_{\psi_a}(\mu_a(p), p) (\mu_a'(p), 1)^\top, \quad (8.15)$$

where $\mathbf{H}_{\psi_a}(\cdot, \cdot)$ denotes the Hessian of $\psi_a(\cdot)$. This leads to

$$\mu_a''(p) = \left(1 - \left\{ \frac{\partial}{\partial \mu} \psi_a(\mu, p) \right\}_{\mu=\mu_a(p)} \right)^{-1} (\mu_a'(p), 1) \mathbf{H}_{\psi_a}(\mu_a(p), p) (\mu_a'(p), 1)^\top. \quad (8.16)$$

Note that since f and q are non-increasing, we have that $\left\{ \frac{\partial}{\partial \mu} \psi_a(\mu, p) \right\}_{\mu=\mu_a(p)} \leq 0$. It remains to verify that

$$(\mu_a'(p), 1) \mathbf{H}_{\psi_a}(\mu_a(p), p) (\mu_a'(p), 1)^\top \leq 0. \quad (8.17)$$

Recall that $\psi_a(\mu, p) = f_a(pq_a(\mu))$. We have that

$$\begin{aligned} \frac{\partial^2}{\partial \mu^2} \psi_a(\mu, p) &= f_a''(pq_a(\mu)) (pq_a'(\mu))^2 + f_a'(pq_a(\mu)) pq_a''(\mu), \\ \frac{\partial^2}{\partial p^2} \psi_a(\mu, p) &= f_a''(pq_a(\mu)) q_a(\mu)^2 \\ \frac{\partial^2}{\partial p \partial \mu} \psi_a(\mu, p) &= f_a''(pq_a(\mu)) pq_a'(\mu) q_a(\mu) + f_a'(pq_a(\mu)) q_a'(\mu). \end{aligned} \quad (8.18)$$

Rearranging terms, and using the fact that $q_a(\mu) = \omega(d_a/\mu)$, we can decompose $\mathbf{H}_{\psi_a}(\mu, p)$ as follows:

$$\mathbf{H}_{\psi_a}(\mu, p) = f_a''(pq_a(\mu)) \mathbf{A} + \frac{f_a'(pq_a(\mu)) \omega''(d_a/\mu) d_a^2}{\mu^4} \mathbf{B} + \frac{f_a'(pq_a(\mu)) \omega'(d_a/\mu) d_a}{\mu^2} \mathbf{C} \quad (8.19)$$

where

$$\begin{aligned} \mathbf{A} &= (pq_a'(\mu), q_a(\mu)) \otimes (pq_a'(\mu), q_a(\mu)), \\ \mathbf{B} &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} \frac{2p}{\mu} & -1 \\ -1 & 0 \end{bmatrix}. \end{aligned}$$

We make the following observations concerning the three terms in (8.19). For the first term, \mathbf{A} is the outer product of $(pq'_a(\mu), q_a(\mu))$ with itself and is hence positive semi-definite. Since $f_a(\cdot)$ is concave and hence $f''_a(\cdot) < 0$, we have that $f''_a(pq_a(\mu))\mathbf{A}$ is negative semi-definite, i.e.,

$$f''_a(pq_a(\mu))\mathbf{A} \leq 0. \quad (8.20)$$

For the second term, note that $f'_a(pq_a(\mu)) > 0$ and $\omega''(\cdot) < 0$ due to the concavity of $\omega(\cdot)$. Therefore, we have that

$$\frac{f'_a(pq_a(\mu))\omega''(d_a/\mu)d_a^2}{\mu^4}\mathbf{B} \leq 0. \quad (8.21)$$

For the third term, since we are only interested in the properties of \mathbf{C} along the specific direction $(\mu'_a(p), 1)$, it suffices to show that when $\mu = \mu_a(p)$, $(\mu'_a(p), 1)\mathbf{C}(\mu'_a(p), 1)^\top$ is non-positive. This claim is isolated in the form of the following proposition; The proof is given in Appendix A.4.

Proposition 12.

$$(\mu'_a(p), 1)\mathbf{C}(\mu'_a(p), 1)^\top \leq 0. \quad (8.22)$$

By combining (8.20), (8.21) and Proposition 12, we have proven (8.17), i.e.,

$$(\mu'_a(p), 1)\mathbf{H}_{\psi_a}(\mu_a(p), p)(\mu'_a(p), 1)^\top \leq 0. \quad (8.23)$$

This in turn proves Lemma 3. □

8.4 Proof of Lemma 4

We start by verifying (3.18). By (8.2) and the chain rule, we have that

$$\begin{aligned} \frac{d}{dp}\mu_a^{(n)}(p) &= \frac{d}{dp}\psi(\mu_a^{(n)}, \delta_p) \\ &= \frac{d}{dp}\mathbb{E}\left[f_{B_1}\left(pq_a^{(n)}(\mu_a^{(n)}(p))\right) \mid A = a\right] \\ &= \mathbb{E}\left[f'_{B_1}\left(pq_a^{(n)}(\mu_a^{(n)}(p))\right)\left(q_a^{(n)}(\mu_a^{(n)}(p)) + p(q_a^{(n)})'(\mu_a^{(n)}(p))(\mu_a^{(n)})'(p)\right) \mid A = a\right] \\ &= q_a^{(n)}(\mu_a^{(n)}(p))\mathbb{E}\left[f'_{B_1}\left(pq_a^{(n)}(\mu_a^{(n)}(p))\right) \mid A = a\right] \\ &\quad + p(q_a^{(n)})'(\mu_a(p))(\mu_a^{(n)})'(p)\mathbb{E}\left[f'_{B_1}\left(pq_a^{(n)}(\mu_a^{(n)}(p))\right) \mid A = a\right] \\ &= \Delta_a^{(n)}(p) + p\Delta_a^{(n)}(p)(q_a^{(n)})'(\mu_a^{(n)}(p))(\mu_a^{(n)})'(p)/q_a^{(n)}(\mu_a^{(n)}(p)). \end{aligned}$$

The last expression above is linear in $(\mu_a^{(n)})'(p)$. Re-arranging the equation and solving for $(\mu_a^{(n)})'(p)$ leads to the desired result. Finally, (3.19) is a direct consequence of Lemma 2, while (3.20) follows by combining from (3.18) with (3.19) and Lemma 2.

8.5 Proof of Theorem 5

The proof will make use of the following two technical results. The first concerns the sensitivity of the system dynamics with respect to small perturbations ζ , and the second extends calculations from Section 3.1 to the utility function $u_a^{(n)}(p)$. The proofs of these

results are given in Appendices A.5 and A.6, respectively. Recall that $\mu_a^{(n)}(p, \zeta)$ is the expected fraction of active suppliers in equilibrium when the payments are ζ -perturbed from p , i.e., $\mu_a^{(n)}(p, \zeta) \triangleq \mathbb{E}[T(p, \zeta)/n \mid A = a]$.

Proposition 13. *Fix $p > 0$, $a \in \mathcal{A}$ and $n \in \mathbb{N}$. $\mu_a^{(n)}(p, \zeta)$ and $q_a^{(n)}(p, \zeta)$ are twice differentiable functions with respect to ζ , and satisfy:*

1. $\left\{ \frac{\partial}{\partial \zeta} \mu_a^{(n)}(p, \zeta) \right\}_{\zeta=0} = \left\{ \frac{\partial}{\partial \zeta} q_a^{(n)}(\mu_a^{(n)}(p, \zeta)) \right\}_{\zeta=0} = 0$ for all $n \in \mathbb{N}$.
2. There exists $\alpha > 0$ such that $\left\{ \frac{\partial^2}{\partial \zeta^2} \mu_a^{(n)}(p, \zeta) \right\}_{\zeta=\zeta_0}$ and $\left\{ \frac{\partial^2}{\partial \zeta^2} q_a^{(n)}(\mu_a^{(n)}(p, \zeta)) \right\}_{\zeta=\zeta_0}$ are bounded uniformly over all $\zeta_0 \in (0, \alpha)$ and $n \in \mathbb{N}$.

Proposition 14. *Fix $p > 0$ and $a \in \mathcal{A}$. We have that*

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{d}{dp} u_a^{(n)}(p) &= u'_a(p) \\ &= \mu'_a(p) \left[r \left(\frac{d_a}{\mu_a(p)} \right) - p\omega \left(\frac{d_a}{\mu_a(p)} \right) - \left(r' \left(\frac{d_a}{\mu_a(p)} \right) - p\omega' \left(\frac{d_a}{\mu_a(p)} \right) \right) \frac{d_a}{\mu_a(p)} \right] \\ &\quad - \omega \left(\frac{d_a}{\mu_a(p)} \right) \mu_a(p), \end{aligned}$$

where $u_a(\cdot)$ is defined in (3.15).

Now, recall that we are considering the case where the platform employs an η -perturbed payment distribution $\pi_{p,\eta}$, with $P_i = p + \eta\varepsilon_i$ ((3.11)). Define the estimators

$$\bar{D} = D/n, \quad \bar{Z} = T/n = \frac{1}{n} \sum_{i=1}^n Z_i, \quad (8.24)$$

$$\hat{\Delta} = \zeta_n^{-1} \widehat{\text{Cov}}[Z_i, \varepsilon_i] / \widehat{\text{Var}}[\varepsilon_i], \quad (8.25)$$

so that \bar{D} and \bar{Z} correspond to the scaled demand and active suppliers, respectively, and $\hat{\Delta}$ is the scaled regression coefficient of Z_i on ε_i . Finally, define the estimator

$$\hat{\Upsilon} = \hat{\Delta} / \left(1 + \frac{p\bar{D}\hat{\Delta}\omega'(\bar{D}/\bar{Z})}{\bar{Z}^2\omega(\bar{D}/\bar{Z})} \right). \quad (8.26)$$

Our main remaining task is to show that, under the stated conditions,

$$\bar{D} \rightarrow d_a, \quad \bar{Z} \rightarrow \mu_a(p), \quad \hat{\Delta} \rightarrow \Delta_a(p), \quad \text{and} \quad \hat{\Upsilon} \rightarrow \mu'_a(p), \quad (8.27)$$

in L_2 as $n \rightarrow \infty$, for any $a \in \mathcal{A}$. In light of Proposition 14, the desired conclusion (4.4) then follows immediately by combining (4.3) with (8.27) with (8.24) and invoking Slutsky's lemma.

We now turn to proving (8.27). First, we note that the fact that $\hat{\Upsilon} \rightarrow \mu'_a(p)$ follows directly by combining the first three convergence claims in (8.27) with Lemma 4. The fact that $\bar{Z} \rightarrow \mu_a(p)$ follows from our definition (3.1). For $\bar{Z} \rightarrow \mu_a(p)$, note that by Chernoff bound

we know that \bar{Z} concentrates on $\mu_a^{(n)}(p, \zeta_n)$ as $n \rightarrow \infty$. Furthermore, we have that

$$\begin{aligned} \lim_{n \rightarrow \infty} \mu_a^{(n)}(p, \zeta_n) &= \lim_{n \rightarrow \infty} \mu_a^{(n)}(p) + \lim_{n \rightarrow \infty} \left(\mu_a^{(n)}(p, \zeta_n) - \mu_a^{(n)}(p) \right) \\ &\stackrel{(a)}{=} \mu_a(p) + \lim_{n \rightarrow \infty} \left(\mu_a^{(n)}(p, \zeta_n) - \mu_a^{(n)}(p) \right) \\ &\stackrel{(b)}{=} \mu_a(p), \end{aligned} \tag{8.28}$$

where steps (a) and (b) follow from Lemma 2 and Proposition 13, respectively. Together, this shows that $\bar{Z} \rightarrow (p)$ in L_2 .

Finally, it remains to show that $\hat{\Delta} \rightarrow \Delta_a(p)$. To this end, we first observe the following fact: There exists a constant $C > 0$ such that, for every ζ and n

$$\left| \frac{1}{\zeta} \text{Cov}_n [Z_i, \varepsilon_i | A = a] - \Delta_a^{(n)}(p) \right| \leq \zeta C. \tag{8.29}$$

To prove (8.29), note that given ζ -perturbed payments $P_i = p + \zeta \varepsilon_i$, we have

$$\text{Cov}_n [Z_i, \varepsilon_i | A = a] = \mathbb{E}_n \left[\varepsilon_i f_{B_i} \left((p + \zeta \varepsilon_i) q_a^{(n)} \left(\mu_a^{(n)}(p, \zeta) \right) \right) | A = a \right].$$

We can then take the limit $\zeta \rightarrow 0$, and verify that there exists $C > 0$ such that

$$\begin{aligned} &\left| \mathbb{E}_n \left[\frac{1}{\zeta} \varepsilon_i f_{B_i} \left((p + \zeta \varepsilon_i) q_a^{(n)} \left(\mu_a^{(n)}(p, \zeta) \right) \right) | A = a \right] \right. \\ &\quad \left. - q_a^{(n)}(p, 0) \mathbb{E}_n \left[f'_{B_i} (p q_a^{(n)}(\mu_a^{(n)}(p, 0))) | A = a \right] \right| \leq \zeta C \end{aligned}$$

Here, we used Proposition 13, and specifically the fact that $\left\{ \frac{\partial}{\partial \zeta} q_a^{(n)} \left(\mu_a^{(n)}(p, \zeta) \right) \right\}_{\zeta=0} = 0$, both $f_{B_i}(\cdot)$ and $q_a^{(n)}(\mu_a^{(n)}(p, \cdot))$ are twice differentiable with bounded second derivatives uniformly over n , and ε_i has variance 1. Finally, by Slutsky's lemma and conditionally on $A = a$, we have

$$\hat{\Delta} - \frac{\text{Cov}_n [Z_i, \zeta_n \varepsilon_i | A = a]}{\text{Var}_n [\zeta_n \varepsilon_i | A = a]} \rightarrow_p 0.$$

As $\text{Var}_n [\varepsilon_i | A = a] = 1$, we conclude that $\hat{\Delta} \rightarrow_p \Delta_a(p)$ using (8.29) and Lemma 4.

8.6 Proof of Theorem 6

Given the form of (4.5), we can use Lemma 1 of Orabona et al. [2015] to check that

$$\sum_{t=1}^T t(p - p_t) \hat{\Gamma}_t \leq \frac{1}{2\eta} \sum_{t=1}^T t(p - p_t)^2 + \frac{\eta}{2} \sum_{t=1}^T \hat{\Gamma}_t^2. \tag{8.30}$$

We then can replace the gradient estimates $\hat{\Gamma}_t$ with their mean-field limits $u'_{A_t}(p_t)$ provided we add appropriate error terms as follows,

$$\begin{aligned} \sum_{t=1}^T t(p - p_t) u'_{A_t}(p_t) &\leq \frac{1}{2\eta} \sum_{t=1}^T t(p - p_t)^2 + \frac{\eta}{2} \sum_{t=1}^T u'_{A_t}(p_t)^2 \\ &\quad + \sum_{t=1}^T t(p - p_t) \left(u'_{A_t}(p_t) - \hat{\Gamma}_t \right) + \frac{\eta}{2} \sum_{t=1}^T \left(\hat{\Gamma}_t^2 - u'_{A_t}(p_t)^2 \right). \end{aligned}$$

Then, given the result in Theorem 5 we see that, for any $\varepsilon > 0$,

$$\sum_{t=1}^T t(p - p_t)u'_{A_t}(p_t) \leq \frac{1}{2\eta} \sum_{t=1}^T t(p - p_t)^2 + \frac{\eta}{2} \sum_{t=1}^T u'_{A_t}(p_t)^2 + \varepsilon \quad (8.31)$$

with probability tending to 1 as n gets large. Noting that $|u'_{A_t}(p_t)| < M$, this simplifies to

$$\sum_{t=1}^T t(p - p_t)u'_{A_t}(p_t) \leq \frac{1}{2\eta} \sum_{t=1}^T t(p - p_t)^2 + \frac{\eta M^2 T}{2} \quad (8.32)$$

with probability tending to 1. The desired statement (4.6) follows by leveraging the remaining assumptions from the theorem statement: σ -strong concavity of $u_{A_t}(\cdot)$ implies that

$$u_{A_t}(p) \leq u_{A_t}(p_t) + (p - p_t)u'_{A_t}(p_t) - \frac{\sigma}{2}(p - p_t)^2, \quad (8.33)$$

and we use the above to replace the left-hand side expression of (8.32) while noting that $\sigma > \eta^{-1}$.

8.7 Proof of Corollary 7

Let p^* be the maximizer of $u(\cdot)$ over $I = [c_-, c_+]$. By (4.6) we have

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\frac{1}{T} \sum_{t=1}^T t(u(p^*) - u(p_t)) \leq \frac{Z_T}{T} + \frac{\eta M^2}{2} \right] = 1, \quad (8.34)$$

$$Z_T = \sum_{t=1}^T t(u(p^*) - u(p_t) + u_{A_t}(p^*) - u_{A_t}(p_t)).$$

Paired with strong concavity of $u(p)$ around p^* and the fact that $u'(p^*) = 0$, this implies

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\frac{\sigma}{2} \frac{1}{T} \sum_{t=1}^T t(p^* - p_t)^2 \leq \frac{Z_T}{T} + \frac{\eta M^2}{2} \right] = 1. \quad (8.35)$$

In order to verify the desired result, our next step is to bound Z_T . First, because p_t is chosen before we get to learn about A_t , Z_t is a martingale. Second, because the derivative of $u_a(p)$ is uniformly bounded by M , we have $|Z_t - Z_{t-1}| \leq 2Mt|p_t - p^*|$ for all t . Thus, using Hoeffding's lemma to bound the moment-generating function of a bounded random variable, these two facts together imply that

$$\mathbb{E} \left[\exp(c(Z_t - Z_{t-1})) \mid p_t, \{Z_s, p_s\}_{s=1}^{t-1} \right] \leq \exp \left(\frac{1}{2} c^2 M^2 t^2 (p_t - p^*)^2 \right),$$

and so

$$Y_t = \exp \left(cZ_t - \frac{1}{2} c^2 M^2 \sum_{s=1}^t s^2 (p_s - p^*)^2 \right)$$

is a super-martingale for any $c > 0$. Thus, by Markov's inequality,

$$\mathbb{P} \left[Z_T \geq \frac{cM^2}{2} \sum_{t=1}^T t^2 (p_t - p^*)^2 + \frac{-\log(\delta)}{c} \right] \leq \delta. \quad (8.36)$$

for any $0 < \delta < 1$. Pairing (8.35) and (8.36) with $c = \sigma/(2M^2T)$ then yields (recall that $\eta > \sigma^{-1}$)

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left[\frac{\sigma}{4} \frac{1}{T} \sum_{t=1}^T t(p^* - p_t)^2 \leq \eta M^2 (2 \log(\delta^{-1}) + 1/2) \right] \geq 1 - \delta. \quad (8.37)$$

Finally, the desired result follows by noting that

$$\frac{T^2}{2} (p^* - \bar{p}_T)^2 \leq \sum_{t=1}^T t(p^* - \bar{p}_T)^2 \leq \sum_{t=1}^T t(p^* - p_t)^2.$$

8.8 Proof of Theorem 8

Using Proposition 13 and a first-order Taylor expansion with a Lagrange-form remainder, we immediately see that there is a $C > 0$ such that, for all $n \geq n_0$ and $0 \leq \zeta < \alpha$,

$$\sum_{t=1}^T \left(u_{A_t}^{(n)}(p_t) - u_{A_t}^{(n)}(p_t, \zeta) \right) \leq CT\zeta^2.$$

Because this bound holds uniformly over all large enough n , it also holds in the limit, and so (4.8) holds.

References

- Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of californias tobacco control program. *Journal of the American Statistical Association*, 105(490), 2010.
- Peter M Aronow and Cyrus Samii. Estimating average causal effects under general interference, with application to a social network experiment. *The Annals of Applied Statistics*, 11(4):1912–1947, 2017.
- Susan Athey and Michael Luca. Economists (and economics) in tech companies. *Journal of Economic Perspectives*, 33(1):209–30, 2019.
- Susan Athey, Dean Eckles, and Guido W Imbens. Exact p-values for network interference. *Journal of the American Statistical Association*, 113(521):230–240, 2018.
- Guillaume W Basse, Hossein Azari Soufiani, and Diane Lambert. Randomization and the pernicious effects of limited budgets on auction experiments. In *Artificial Intelligence and Statistics*, pages 1412–1420, 2016.
- Guillaume W Basse, Avi Feller, and Panos Toulis. Randomization tests of causal effects under interference. *Biometrika*, 2019.
- Amir Beck and Marc Teboulle. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters*, 31(3):167–175, 2003.
- Thomas Blake and Dominic Coey. Why marketplace experimentation is harder than it seems: The role of test-control interference. In *Proceedings of the fifteenth ACM conference on Economics and computation*, pages 567–582. ACM, 2014.

- Léon Bottou, Jonas Peters, Joaquin Quinonero Candela, Denis Xavier Charles, Max Chickering, Elon Portugaly, Dipankar Ray, Patrice Y Simard, and Ed Snelson. Counterfactual reasoning and learning systems: The example of computational advertising. *Journal of Machine Learning Research*, 14(1):3207–3260, 2013.
- Maury Bramson, Yi Lu, and Balaji Prabhakar. Asymptotic independence of queues under randomized load balancing. *Queueing Systems*, 71(3):247–292, 2012.
- Sébastien Bubeck, Yin Tat Lee, and Ronen Eldan. Kernel-based methods for bandit convex optimization. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 72–85. ACM, 2017.
- Gerard P Cachon, Kaitlin M Daniels, and Ruben Lobel. The role of surge pricing on a service platform with self-scheduling capacity. *Manufacturing & Service Operations Management*, 19(3):368–384, 2017.
- CAISO. Renewable resources and the California electric power industry: System operations, wholesale markets and grid planning. *California ISO Report*, 2009.
- Nicolo Cesa-Bianchi, Alex Conconi, and Claudio Gentile. On the generalization ability of on-line learning algorithms. *IEEE Transactions on Information Theory*, 50(9):2050–2057, 2004.
- Raj Chetty. Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. *Annual Review of Economics*, 1(1):451–488, 2009.
- John C Duchi, Michael I Jordan, Martin J Wainwright, and Andre Wibisono. Optimal rates for zero-order convex optimization: The power of two function evaluations. *IEEE Transactions on Information Theory*, 61(5):2788–2806, 2015.
- Dean Eckles, Brian Karrer, and Johan Ugander. Design and analysis of experiments in networks: Reducing bias from interference. *Journal of Causal Inference*, 5(1), 2017.
- Zhe Feng, Chara Podimata, and Vasilis Syrgkanis. Learning to bid without knowing your value. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 505–522. ACM, 2018.
- Ronald A Fisher. *The Design of Experiments*. Oliver and Boyd, Edinburgh, 1935.
- Abraham D Flaxman, Adam Tauman Kalai, Adam Tauman Kalai, and H Brendan McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 385–394. Society for Industrial and Applied Mathematics, 2005.
- Saeed Ghadimi and Guanghui Lan. Stochastic first- and zeroth-order methods for nonconvex stochastic programming. *SIAM Journal on Optimization*, 23(4):2341–2368, 2013.
- Mark Goh. Congestion management and electronic road pricing in singapore. *Journal of transport geography*, 10(1):29–38, 2002.
- Carl Graham and Sylvie Méléard. Chaos hypothesis for a system interacting through shared resources. *Probability Theory and Related Fields*, 100(2):157–174, 1994.

- Shlomo Halfin and Ward Whitt. Heavy-traffic limits for queues with many exponential servers. *Operations research*, 29(3):567–588, 1981.
- Jonathan Hall, Cory Kendrick, and Chris Nosko. The effects of ubers surge pricing: A case study. *The University of Chicago Booth School of Business*, 2015.
- Jonathan V Hall, John J Horton, and Daniel T Knoepfle. Pricing efficiently in designed markets: The case of ride-sharing. 2019.
- Arnold C Harberger. The measurement of waste. *The American Economic Review*, pages 58–76, 1964.
- Charles A Holt and Susan K Laury. Risk aversion and incentive effects. *American economic review*, 92(5):1644–1655, 2002.
- Michael G Hudgens and M Elizabeth Halloran. Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482):832–842, 2008.
- Guido W Imbens and Donald B Rubin. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press, 2015.
- Krishnamurthy Iyer, Ramesh Johari, and Mukund Sundararajan. Mean field equilibria of dynamic auctions with learning. *Management Science*, 60(12):2949–2970, 2014.
- Kevin G Jamieson, Robert Nowak, and Ben Recht. Query complexity of derivative-free optimization. In *Advances in Neural Information Processing Systems*, pages 2672–2680, 2012.
- Ramesh Johari, Vijay Kamble, and Yash Kanoria. Matching while learning. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pages 119–119. ACM, 2017a.
- Ramesh Johari, Pete Koomen, Leonid Pekelis, and David Walsh. Peeking at a/b tests: Why it matters, and what to do about it. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1517–1525. ACM, 2017b.
- Yash Kanoria and Hamid Nazerzadeh. Dynamic reserve prices for repeated auctions: Learning from bids. *Available at SSRN 2444495*, 2017.
- Ashish Khetan and Sewoong Oh. Achieving budget-optimality with adaptive schemes in crowdsourcing. In *Advances in Neural Information Processing Systems*, pages 4844–4852, 2016.
- Robert D Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. In *Advances in Neural Information Processing Systems*, pages 697–704, 2005.
- Ron Kohavi, Roger Longbotham, Dan Sommerfield, and Randal M Henne. Controlled experiments on the web: survey and practical guide. *Data mining and knowledge discovery*, 18(1):140–181, 2009.
- Benjamin Letham, Brian Karrer, Guilherme Ottoni, and Eytan Bakshy. Constrained Bayesian optimization with noisy experiments. *Bayesian Analysis*, 2018.

- Charles F Manski. Identification of treatment response with social interactions. *The Econometrics Journal*, 16(1):S1–S23, 2013.
- Laurent Massoulié and Kuang Xu. On the capacity of information processing systems. *Operations Research*, 66(2):568–586, 2018.
- Marc Mézard, Giorgio Parisi, and Miguel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Company, 1987.
- Yurii Nesterov and Vladimir Spokoiny. Random gradient-free minimization of convex functions. *Foundations of Computational Mathematics*, 17(2):527–566, 2017.
- Francesco Orabona, Koby Crammer, and Nicolo Cesa-Bianchi. A generalized online mirror descent with applications to classification and regression. *Machine Learning*, 99(3):411–435, 2015.
- Michael Ostrovsky and Michael Schwarz. Reserve prices in internet advertising auctions: a field experiment. *EC*, 11:59–60, 2011.
- John W Pratt. Risk aversion in the small and in the large. In *Uncertainty in Economics*, pages 59–79. Elsevier, 1978.
- Ohad Shamir. On the complexity of bandit and derivative-free stochastic convex optimization. In *Conference on Learning Theory*, pages 3–24, 2013.
- Michael E Sobel. What do randomized studies of housing mobility demonstrate? causal inference in the face of interference. *Journal of the American Statistical Association*, 101(476):1398–1407, 2006.
- James C Spall. *Introduction to stochastic search and optimization: estimation, simulation, and control*, volume 65. John Wiley & Sons, 2005.
- Joel Spencer, Madhu Sudan, and Kuang Xu. Queuing with future information. *Annals of Applied Probability*, 24(5):2091–2142, 2014.
- Charles M Stein. Estimation of the mean of a multivariate normal distribution. *The annals of Statistics*, pages 1135–1151, 1981.
- Alexander L Stolyar. Pull-based load distribution in large-scale heterogeneous service systems. *Queueing Systems*, 80(4):341–361, 2015.
- Alain-Sol Sznitman. Topics in propagation of chaos. In *Ecole d’été de probabilités de Saint-Flour XIX1989*, pages 165–251. Springer, 1991.
- Diane Tang, Ashish Agarwal, Deirdre O’Brien, and Mike Meyer. Overlapping experiment infrastructure: More, better, faster experimentation. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 17–26. ACM, 2010.
- Kenneth E Train. *Discrete choice methods with simulation*. Cambridge university press, 2009.

John N Tsitsiklis and Kuang Xu. On the power of (even a little) resource pooling. *Stochastic Systems*, 2(1):1–66, 2012.

Nikita Dmitrievna Vvedenskaya, Roland L’vovich Dobrushin, and Fridrikh Izrailevich Karpelevich. Queueing system with selection of the shortest of two queues: An asymptotic approach. *Problemy Peredachi Informatsii*, 32(1):20–34, 1996.

A Proofs

A.1 Proof of Proposition 9

We start by verifying a useful property that applies to any exponential family with discrete support.

Definition 11. Let $\{X\}$ a family of discrete random variables and parameterized by $\theta \in \Theta \subset \mathbb{R}$. We say that $\{X\}$ is an exponential family, if the probability mass function (PMF) f_θ for X can be expressed as

$$f_\theta(x) = h(x) \exp(\eta(\theta)T(x) - A(\theta)), \quad x \in \mathbb{Z}. \quad (\text{A.1})$$

where $T(\cdot)$ is referred to as the sufficient statistic, and $\eta(\theta)$ the natural parameter.

We have the following identity, which is a simple generalization of a result proved by [Stein \[1981\]](#) for Gaussian random variables.

Lemma 15. Fix an exponential family of random variables X with discrete support \mathcal{X} parametrized by $\theta \in \Theta$ defined over a finite subset of \mathbb{R} , with sufficient statistic $T(X)$ and natural parameter $\eta(\theta)$. Then, for any function $g : \mathcal{X} \rightarrow \mathbb{R}$, we have that

$$\frac{d}{d\theta} \mathbb{E}_\theta [g(X)] = \text{Cov}_\theta [g(X), T(X)], \quad (\text{A.2})$$

for all θ in the interior of Θ .

Proof. First, observe that $\sum_x f_\theta(x) = h(x) \exp(\eta(\theta)T(x) - A(\theta)) = 1$. Taking derivatives on both sides with respect to θ , we obtain that

$$\begin{aligned} 0 &= \sum_x h(x) (T(x)\eta'(\theta) - A'(\theta)) (\exp(\eta(\theta)T(x) - A(\theta))) \\ &= \mathbb{E}_\theta [T(X)] \eta'(\theta) - A'(\theta) \end{aligned} \quad (\text{A.3})$$

which implies that

$$\mathbb{E}_\theta [T(X)] = A'(\theta)/\eta'(\theta). \quad (\text{A.4})$$

We have that

$$\begin{aligned}
\frac{d}{d\theta} \mathbb{E}_\theta [g(X)] &= \sum_{x \in \mathcal{X}} g(x) \frac{\partial}{\partial \theta} f_\theta(x) \\
&= \sum_{x \in \mathcal{X}} g(x) (T(x)\eta'(\theta) - A'(\theta)) h(x) \exp(\eta(\theta)T(x) - A(\theta)) \\
&= \eta'(\theta) \mathbb{E}_\theta [g(X) (T(X) - A'(\theta)/\eta'(\theta))] \\
&\stackrel{(a)}{=} \eta'(\theta) \mathbb{E}_\theta [g(X) (T(X) - \mathbb{E}_\theta [T(X)])] \\
&\stackrel{(b)}{=} \eta'(\theta) \mathbb{E}_\theta [(g(X) - \mathbb{E} [g(X)]) (T(X) - \mathbb{E} [T(X)])] \\
&= \text{Cov}_\theta [g(X), T(X)], \tag{A.5}
\end{aligned}$$

where (a) follows from (A.4), and (b) from the fact that $\mathbb{E}_\theta [T(X) - \mathbb{E}_\theta [T(X)]] = 0$. This proves Lemma 15. \square

We are now ready to prove the stated result. We will assume that all probabilities are calculated by conditioning on $A = a$, and thus omit it from our notation. The fact that $\frac{d}{d\mu} q_a^{(n)}(\mu)$ is non-positive follows directly from the fact that $\Omega(d, t)$ is non-increasing in t (Assumption 1). The PMF of an (n, μ) Binomial random variable X can be written as

$$f_\mu(x) = \binom{n}{x} \exp\left(x \log\left(\frac{\mu}{1-\mu}\right) + n \log(\mu(1-\mu))\right). \tag{A.6}$$

In particular, the set of Binomial random variables forms an exponential family, with natural parameter $\eta(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$ and sufficient statistic $T(X) = X$. We now employ Lemma 15 above. Define

$$H \triangleq X - \mathbb{E}_\mu [X] = X - n\mu. \tag{A.7}$$

For a fixed d , we have that⁹

$$\begin{aligned}
\frac{d}{d\mu} \mathbb{E}_\mu [\Omega(d, X)] (\eta'(\mu))^{-1} &\stackrel{(a)}{=} \text{Cov}_\mu [\Omega(d, X), X] \\
&= \mathbb{E}_\mu [\Omega(d, X)H] \\
&\stackrel{(b)}{=} \mathbb{E}_\mu [\omega(d/X)H] + \mathbb{E}_\mu [l(d, X)H] \\
&\stackrel{(c)}{\in} \mathbb{E}_\mu [\omega(d/X)H] \pm \sqrt{\mathbb{E}_\mu [l(d, X)^2] \mathbb{E} [H^2]} \\
&= \text{Cov}_\mu [\omega(d/X), X] \pm \sqrt{\mathbb{E}_\mu [l(d, X)^2] \mathbb{E} [H^2]} \tag{A.8}
\end{aligned}$$

where (a) follows from Lemma 15, (b) from Assumption 1, and (c) from the CauchySchwarz inequality. Taking expectation with respect to $d \sim D$ on both sides, we obtain

$$\frac{d}{d\mu} \mathbb{E}_\mu [\Omega(D, X)] (\eta'(\mu))^{-1} \in \mathbb{E}_\mu [\omega(D/X)] \pm \sqrt{\mathbb{E}_\mu [l(D, X)^2] \mathbb{E} [H^2]}. \tag{A.9}$$

We next bound each of the two terms in (A.9). For the second term, recall that $l(\cdot)$ is bounded and $|l(d, t)| = o(1/\sqrt{d} + 1/\sqrt{t})$ (Assumption 1). Furthermore, it follows from the Chernoff bound and (3.2), respectively, that

$$\mathbb{P}_\mu [X \geq n\mu/2], \quad \mathbb{P} [D \geq nd_a/2] = o(1/n) \text{ as } n \rightarrow \infty. \tag{A.10}$$

⁹The notation $x \in a \pm b$ denotes $x \in [a - b, a + b]$.

This implies that

$$\mathbb{E}_\mu [l(D, X)^2] = o(1/n), \text{ as } n \rightarrow \infty. \quad (\text{A.11})$$

Furthermore, note that

$$\mathbb{E}_\mu [H^2] = \text{Var}_\mu [X] = \mathcal{O}(n). \quad (\text{A.12})$$

Combining the above two equations, we conclude that

$$\sqrt{\mathbb{E}_\mu [l(D, X)^2] \mathbb{E} [H^2]} = o(1), \text{ as } n \rightarrow \infty. \quad (\text{A.13})$$

Next, we turn to the first term in (A.9), which will follow from the following result. Fix $\delta \in (0, \mu)$, and define the event

$$\mathcal{E} = \{|H/n| < \delta\}. \quad (\text{A.14})$$

Using Taylor expansion on the function

$$h(x) \triangleq \omega \left(\frac{d/n}{\mu + x} \right) \quad (\text{A.15})$$

and the smoothness of ω , we have that there exists a constant $c_1 > 0$ such that, for all n and d ,

$$h(x) \in h(0) + h'(0)x \pm c_1 x^2, \quad \forall x \in [-\delta, \delta]. \quad (\text{A.16})$$

Fix $d \in \mathbb{R}_+$. We have that

$$\begin{aligned} \mathbb{E}_\mu [\omega(d/X) H] &= \mathbb{E}_\mu \left[\omega \left(\frac{d/n}{(n\mu + H)/n} \right) H \right] \\ &= \mathbb{E}_\mu [h(H/n) H] \\ &= \mathbb{E}_\mu [\mathbf{1}_{\mathcal{E}} h(H/n) H] + \mathbb{E}_\mu [\mathbf{1}_{\bar{\mathcal{E}}} h(H/n) H] \\ &\stackrel{(a)}{\in} \mathbb{E}_\mu [\mathbf{1}_{\mathcal{E}} h(H/n) H] \pm c_2 \sqrt{\mathbb{P}(\bar{\mathcal{E}}) \mathbb{E}_\mu [H^2]} \\ &\stackrel{(b)}{\in} \mathbb{E}_\mu [\mathbf{1}_{\mathcal{E}} h(H/n) H] \pm o(1) \\ &\stackrel{(c)}{\in} \mathbb{E}_\mu \left[\left(h(0) + h'(0) \frac{H}{n} \right) H \right] \pm c_1 \mathbb{E}_\mu \left[\frac{H^2}{n^2} H \right] \pm c_2 \mathbb{P}(\bar{\mathcal{E}}) \pm o(1) \\ &\in \mathbb{E}_\mu \left[\left(h(0) + h'(0) \frac{H}{n} \right) H \right] \pm c_1 \mathbb{E}_\mu \left[\frac{H^2}{n^2} H \right] \pm o(1) \\ &\stackrel{(d)}{\in} \mathbb{E}_\mu \left[\left(h(0) + h'(0) \frac{H}{n} \right) H \right] \pm \mathcal{O}(1/n) \pm o(1) \\ &\stackrel{(e)}{\in} \frac{d}{d\mu} \omega \left(\frac{d/n}{\mu} \right) \mathbb{E}_\mu [H^2] \pm o(1) \\ &= \frac{d}{d\mu} \omega \left(\frac{d/n}{\mu} \right) \mu_a^{(n)} (1 - \mu) \pm o(1) \\ &= \frac{d}{d\mu} \omega \left(\frac{d/n}{\mu} \right) (\eta'(\mu))^{-1} \pm o(1) \end{aligned} \quad (\text{A.17})$$

where $c_2 = \max_{d, x \in \mathbb{R}_+} \Omega(d, x)$, and $c_3 = \max_{x \in [-\delta, \delta]} h(x)$, and the $o(1)$ term does not depend on d . Step (a) is based on the CauchySchwarz inequality, (b) from the fact that $\mathbb{P}(\bar{\mathcal{E}})$

converges to 0 exponentially fast in n by the Chernoff bound and that $\mathbb{E}[H^2] = \mathcal{O}(n)$, (c) from the Taylor expansion in (A.16), and (d) from the fact that $|\mathbb{E}[H^3]| = \mathcal{O}(n)$ as a result of X being a Binomial random variable. Finally, step (e) follows from the definition of h in (A.15).

Recall from (3.1) that, conditional on $A = a$, D/n concentrates on d_a as $n \rightarrow \infty$. (A.17) thus implies that

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E}_\mu [\omega(D/X) H] &= \lim_{n \rightarrow \infty} \sum_{d \in \mathbb{Z}_+} \mathbb{E}_\mu [\omega(d/X) H] \mathbb{P}[D = d | A = a] \\ &= \frac{d}{d\mu} \omega(d_a/\mu) (\eta'(\mu))^{-1} \end{aligned} \quad (\text{A.18})$$

Substituting (A.13) and (A.18) into (A.9), we obtain that

$$\lim_{n \rightarrow \infty} \frac{d}{d\mu} \mathbb{E}_\mu [\Omega(D, X) | A = a] = \frac{d}{d\mu} \omega(d_a/\mu). \quad (\text{A.19})$$

This proves Proposition 9. \square

A.2 Proof of Proposition 10

We have that

$$g(x) = g(0) + \int_0^x g'(s) ds \stackrel{(a)}{\geq} \int_0^x g'(s) ds \stackrel{(b)}{\geq} \omega'(x)x, \quad (\text{A.20})$$

where (a) follows from the assumption that $g(0) \geq 0$, and (b) from the concavity of $g(\cdot)$, which implies that $g'(\cdot)$ is non-increasing over $x > 0$. \square

A.3 Proof of Proposition 11

By the chain rule, and the fact that $q_a(\mu) = \omega(d_a/\mu)$, we have that

$$\begin{aligned} \frac{d}{dp} (q_a(\mu_a(p))\mu_a(p)) &= q'(\mu_a(p))\mu'_a(p)\mu_a(p) + q_a(\mu_a(p))\mu'_a(p) \\ &= (\omega(d_a/\mu_a(p)) - \omega'(d_a/\mu_a(p))d_a/\mu_a(p))\mu'_a(p) \end{aligned} \quad (\text{A.21})$$

Using the expression for $\mu'_a(p)$ (cf. (3.20)), it is not difficult to show that, as a result of the strong concavity of $f_a(\cdot)$ in the interval (\underline{x}, \bar{x}) , we have that $\inf_{p \in (c_0, p)} \mu'_a(p) > 0$. Furthermore, using the same argument as in the proof of Proposition 10 and the fact that $\omega(\cdot)$ is strongly concave with $\omega(0) \geq 0$, we have that $\inf_{p \in (c_0, \gamma)} (\omega(d_a/\mu_a(p)) - \omega'(d_a/\mu_a(p))d_a/\mu_a(p)) > 0$. Together, this implies that $\inf_{p \in (c_0, \gamma)} \frac{d}{dp} (q_a(\mu_a(p))\mu_a(p)) > 0$, thus proving our claim. \square

A.4 Proof of Proposition 12

Note that

$$(\mu'_a(p), 1)\mathbf{C}(\mu'_a(p), 1)^\top = \omega(d_a/\mu_a(p)) \frac{2d_a\mu'_a(p)}{\mu_a(p)^2} (p\mu'_a(p) - \mu_a(p)). \quad (\text{A.22})$$

It therefore suffices to show that

$$p\mu'_a(p) - \mu_a(p) \leq 0. \quad (\text{A.23})$$

From Lemma 4, we have that

$$\mu'_a(p) = \frac{\Delta_a(p)}{1 - p\Delta_a(p)q'_a(\mu_a(p))/q_a(\mu_a(p))}, \quad (\text{A.24})$$

where

$$\Delta_a(p) = q_a(\mu_a(p))f'_a(pq_a(\mu_a(p))), \quad (\text{A.25})$$

and

$$\mu_a(p) = f_a(pq_a(\mu_a(p))). \quad (\text{A.26})$$

Multiplying the left-hand side of (A.24) by $p/\mu_a(p)$, we obtain

$$\begin{aligned} \frac{\mu'_a(p)p}{\mu_a(p)} &= \frac{p\Delta_a(p)/\mu_a(p)}{1 - p\Delta_a(p)q'_a(\mu_a(p))/q_a(\mu_a(p))} \\ &= \frac{f'_a(pq_a(\mu_a(p)))(pq_a(\mu_a(p)))}{f(pq_a(\mu_a(p)))} \cdot \frac{1}{1 - p\Delta_a(p)q'_a(\mu_a(p))/q_a(\mu_a(p))} \\ &\leq \frac{\tilde{f}'(pq_a(\mu_a(p)))(pq_a(\mu_a(p)))}{\tilde{f}(pq_a(\mu_a(p)))} \cdot \frac{1}{1 - p\Delta_a(p)q'_a(\mu_a(p))/q_a(\mu_a(p))} \\ &\stackrel{(a)}{\leq} \frac{1}{1 - p\Delta_a(p)q'_a(\mu_a(p))/q_a(\mu_a(p))} \\ &\stackrel{(b)}{\leq} 1, \end{aligned}$$

where (a) follows from Proposition 10 combined with the non-negativity and concavity of $\tilde{f}(\cdot)$, and (b) from the fact that $q'_a(\mu) = -\omega(d_a/\mu)d_a/\mu^2 \leq 0$. This proves (A.23) and hence the proposition. \square

A.5 Proof of Proposition 13

Fix $a \in \mathcal{A}$ and $n \in \mathbb{N}$. Denote by $\pi_{p,\zeta}$ the ζ -perturbed payment distribution centered at p (3.11). We first prove that $\left\{ \frac{\partial}{\partial \zeta} \mu_a^{(n)}(p, \zeta) \right\}_{\zeta=0} = 0$. By (8.2), $\mu_a^{(n)}(p, \zeta)$ satisfies

$$\frac{\partial^k}{\partial^k \zeta} \mu_a^{(n)}(p, \zeta) = \frac{\partial^k}{\partial^k \zeta} \psi_a^{(n)}(\mu_a^{(n)}(p, \zeta), \pi_{p,\zeta}), \quad k \in \mathbb{N}. \quad (\text{A.27})$$

It therefore suffices to evaluate the right-hand side of the above equation. To this end:

$$\begin{aligned} &\psi_a^{(n)}(\mu_a^{(n)}(p, \zeta), \pi_{p,\zeta}) - \psi_a^{(n)}(\mu_a^{(n)}(p), \delta_p) \\ &\stackrel{(a)}{=} \frac{1}{2} \left(\mathbb{E} \left[f_{B_1} \left((p + \zeta) q_a^{(n)}(\mu_a^{(n)}(p, \zeta)) \right) \mid A = a \right] + \mathbb{E} \left[f_{B_1} \left((p - \zeta) q_a^{(n)}(\mu_a^{(n)}(p, \zeta)) \right) \mid A = a \right] \right) \\ &\quad - \mathbb{E} \left[f_{B_1} \left(p q_a^{(n)}(\mu_a^{(n)}(p)) \right) \mid A = a \right] \\ &= \frac{1}{2} \left(\mathbb{E} \left[f_{B_1} \left((p + \zeta) q_a^{(n)}(\mu_a^{(n)}(p, \zeta)) \right) \mid A = a \right] - \mathbb{E} \left[f_{B_1} \left(p q_a^{(n)}(\mu_a^{(n)}(p)) \right) \mid A = a \right] \right) \\ &\quad + \frac{1}{2} \left(\mathbb{E} \left[f_{B_1} \left((p - \zeta) q_a^{(n)}(\mu_a^{(n)}(p, \zeta)) \right) \mid A = a \right] - \mathbb{E} \left[f_{B_1} \left(p q_a^{(n)}(\mu_a^{(n)}(p)) \right) \mid A = a \right] \right), \end{aligned} \quad (\text{A.28})$$

where (a) follows from the definition of ζ -perturbation ((3.11)) and the independence of perturbations $\{\varepsilon_i\}_{i \in \mathbb{N}}$ from the rest of the system. Since both $f_{B_1}(\cdot)$ and $q_a^{(n)}(\cdot)$ are bounded, for the first term on the right-hand side of (A.28), it is not difficult to show using the dominated convergence theorem that there exists $c > 0$ such that¹⁰

$$\begin{aligned} & \mathbb{E} \left[f_{B_1} \left((p + \zeta) q_a^{(n)}(\mu_a^{(n)}(p, \zeta)) \right) \mid A = a \right] - \mathbb{E} \left[f_{B_1} \left(p q_a^{(n)}(\mu_a^{(n)}(p)) \right) \mid A = a \right] \\ & \in v\zeta \pm c\zeta^2, \end{aligned} \quad (\text{A.29})$$

for all sufficiently small ζ , where $v \triangleq \left\{ \frac{\partial}{\partial \zeta} \mathbb{E} \left[f_{B_1} \left((p + \zeta) q_a^{(n)}(\mu_a^{(n)}(p, \zeta)) \right) \mid A = a \right] \right\}_{\zeta=0}$. Applying the same argument to the second term in (A.28), we have that there exists c , such that for all sufficiently small ζ

$$\frac{\psi_a^{(n)}(\mu_a^{(n)}(p, \zeta), \zeta) - \psi_a^{(n)}(\mu_a^{(n)}(p, 0), 0)}{\zeta} \in \pm \frac{c\zeta^2}{\zeta} = \pm c\zeta, \quad (\text{A.30})$$

which further implies that

$$\left\{ \frac{\partial}{\partial \zeta} \mu_a^{(n)}(p, \zeta) \right\}_{\zeta=0} = 0. \quad (\text{A.31})$$

For the derivative of $q_a^{(n)}(\mu_a^{(n)}(p, \cdot))$, note that by chain rule, we have

$$\frac{\partial}{\partial \zeta} q_a^{(n)}(\mu_a^{(n)}(p, \zeta)) = (q_a^{(n)})'(\mu_a^{(n)}(p, \zeta)) \frac{\partial}{\partial \zeta} \mu_a^{(n)}(p, \zeta). \quad (\text{A.32})$$

Since $\left\{ \frac{\partial}{\partial \zeta} \mu_a^{(n)}(p, \zeta) \right\}_{\zeta=0} = 0$ by (A.31), and $(q_a^{(n)})'(\mu_a^{(n)}(p, \zeta))$ is finite by Proposition 9, we have that $\left\{ \frac{\partial}{\partial \zeta} q_a^{(n)}(\mu_a^{(n)}(p, \zeta)) \right\}_{\zeta=0} = 0$. This proves the first claim of Proposition 13.

For the second claim, define

$$g_{\varepsilon_1}(\zeta) \triangleq (p + \varepsilon_1 \zeta) q_a^{(n)}(\mu_a^{(n)}(p, \zeta)). \quad (\text{A.33})$$

Applying the chain rule to (A.27), we have that

$$\begin{aligned} \frac{\partial^2}{\partial^2 \zeta} \mu_a^{(n)}(p, \zeta) &= \mathbb{E} \left[\frac{\partial^2}{\partial^2 \zeta} f_{B_1}(g_{\varepsilon_1}(\zeta)) \mid A = a \right] \\ &= \mathbb{E} \left[f_{B_1}''(g_{\varepsilon_1}(\zeta)) g_{\varepsilon_1}'(\zeta)^2 + f_{B_1}'(g_{\varepsilon_1}(\zeta)) g_{\varepsilon_1}''(\zeta) \mid A = a \right] \end{aligned} \quad (\text{A.34})$$

Note that

$$\begin{aligned} g_{\varepsilon_1}'(\zeta) &= \frac{\partial}{\partial \zeta} \left[(p + \varepsilon_1 \zeta) q_a^{(n)}(\mu_a^{(n)}(p, \zeta)) \right] \\ &= p \frac{\partial}{\partial \zeta} q_a^{(n)}(\mu_a^{(n)}(p, \zeta)) + \varepsilon_1 q_a^{(n)}(\mu_a^{(n)}(p, \zeta)) + \varepsilon_1 \zeta \frac{\partial}{\partial \zeta} q_a^{(n)}(\mu_a^{(n)}(p, \zeta)), \end{aligned} \quad (\text{A.35})$$

and

$$g_{\varepsilon_1}''(\zeta) = p \frac{\partial^2}{\partial^2 \zeta} q_a^{(n)}(\mu_a^{(n)}(p, \zeta)) + 2\varepsilon_1 \frac{\partial}{\partial \zeta} q_a^{(n)}(\mu_a^{(n)}(p, \zeta)) + \varepsilon_1 \zeta \frac{\partial^2}{\partial^2 \zeta} q_a^{(n)}(\mu_a^{(n)}(p, \zeta)). \quad (\text{A.36})$$

¹⁰Notation: $x \in y \pm z \leftrightarrow x \in [y - z, y + z]$.

By chain rule, we have

$$\begin{aligned}
& \left\{ \frac{\partial^2}{\partial^2 \zeta} q_a^{(n)}(\mu_a^{(n)}(p, \zeta)) \right\}_{\zeta=0} \\
&= (q_a^{(n)})''(\mu_a^{(n)}(p, 0)) \left\{ \frac{\partial}{\partial \zeta} \mu_a^{(n)}(p, \zeta) \right\}_{\zeta=0}^2 + (q_a^{(n)})'(\mu_a^{(n)}(p, 0)) \left\{ \frac{\partial^2}{\partial^2 \zeta} \mu_a^{(n)}(p, \zeta) \right\}_{\zeta=0} \\
&= (q_a^{(n)})'(\mu_a^{(n)}(p)) \left\{ \frac{\partial^2}{\partial^2 \zeta} \mu_a^{(n)}(p, \zeta) \right\}_{\zeta=0}, \tag{A.37}
\end{aligned}$$

where the last step follows from the fact that $\left\{ \frac{\partial}{\partial \zeta} \mu_a^{(n)}(p, \zeta) \right\}_{\zeta=0} = 0$. Applying (A.32) and (A.37) to (A.35) and (A.36), we have

$$g'_{\varepsilon_1}(0) = p \left\{ \frac{\partial}{\partial \zeta} q_a^{(n)}(\mu_a^{(n)}(p, \zeta)) \right\}_{\zeta=0} + \varepsilon_1 q_a^{(n)}(\mu_a^{(n)}(p, 0)) + 0 = \varepsilon_1 q_a^{(n)}(\mu_a^{(n)}(p)), \tag{A.38}$$

and

$$\begin{aligned}
g''_{\varepsilon_1}(0) &= p \left\{ \frac{\partial^2}{\partial^2 \zeta} q_a^{(n)}(\mu_a^{(n)}(p, \zeta)) \right\}_{\zeta=0} + 2\varepsilon_1 \left\{ \frac{\partial}{\partial \zeta} q_a^{(n)}(\mu_a^{(n)}(p, \zeta)) \right\}_{\zeta=0} + 0 \\
&= p (q_a^{(n)})'(\mu_a^{(n)}(p)) \left\{ \frac{\partial^2}{\partial^2 \zeta} \mu_a^{(n)}(p, \zeta) \right\}_{\zeta=0}. \tag{A.39}
\end{aligned}$$

Substituting the expressions for $g'_{\varepsilon_1}(0)$ and $g''_{\varepsilon_1}(0)$ into (A.34), we obtain:

$$\begin{aligned}
& \left\{ \frac{\partial^2}{\partial^2 \zeta} \mu_a^{(n)}(p, \zeta) \right\}_{\zeta=0} \\
&= \mathbb{E} \left[f''_{B_1}(g_{\varepsilon_1}(0)) \varepsilon_1^2 q_a^{(n)}(\mu_a^{(n)}(p))^2 + f'_{B_1}(g_{\varepsilon_1}(0)) p (q_a^{(n)})'(\mu_a^{(n)}(p)) \left\{ \frac{\partial^2}{\partial^2 \zeta} \mu_a^{(n)}(p, \zeta) \right\}_{\zeta=0} \mid A = a \right] \\
&= \mathbb{E} [f''_{B_1}(g_{\varepsilon_1}(0)) \mid A = a] q_a^{(n)}(\mu_a^{(n)}(p))^2 \\
&+ \mathbb{E} [f'_{B_1}(g_{\varepsilon_1}(0)) \mid A = a] p (q_a^{(n)})'(\mu_a^{(n)}(p)) \left\{ \frac{\partial^2}{\partial^2 \zeta} \mu_a^{(n)}(p, \zeta) \right\}_{\zeta=0}, \tag{A.40}
\end{aligned}$$

where the last step follows from the fact that $\varepsilon_1 \in \{-1, 1\}$ and hence $\varepsilon_1^2 = 1$. After rearrangement, the above equation yields

$$\left\{ \frac{\partial^2}{\partial^2 \zeta} \mu_a^{(n)}(p, \zeta) \right\}_{\zeta=0} = \frac{\mathbb{E} [f''_{B_1}(g_{\varepsilon_1}(0)) \mid A = a] q_a^{(n)}(\mu_a^{(n)}(p))^2}{1 - \mathbb{E} [f'_{B_1}(g_{\varepsilon_1}(0)) \mid A = a] (q_a^{(n)})'(\mu_a^{(n)}(p)) p}, \tag{A.41}$$

and by (A.37), we have

$$\begin{aligned}
& \left\{ \frac{\partial^2}{\partial^2 \zeta} q_a^{(n)}(\mu_a^{(n)}(p, \zeta), \zeta) \right\}_{\zeta=0} \\
&= (q_a^{(n)})'(\mu_a^{(n)}(p, 0)) \left\{ \frac{\partial^2}{\partial^2 \zeta} \mu_a^{(n)}(p, 0) \right\}_{\zeta=0} \\
&= (q_a^{(n)})'(\mu_a^{(n)}(p)) \left(\frac{\mathbb{E} [f''_{B_1}(g_{\varepsilon_1}(0)) \mid A = a] q_a^{(n)}(\mu_a^{(n)}(p))^2}{1 - \mathbb{E} [f'_{B_1}(g_{\varepsilon_1}(0)) \mid A = a] (q_a^{(n)})'(\mu_a^{(n)}(p)) p} \right). \tag{A.42}
\end{aligned}$$

Finally, we check the uniform boundedness of the second derivatives with respect to all n and all sufficiently small ζ . To show that $\left\{ \frac{\partial^2}{\partial^2 \zeta} \mu_a^{(n)}(p, \zeta) \right\}_{\zeta=0}$, note that $f'_{B_1}(\cdot)$ is non-negative and $(q_a^{(n)})'(\cdot)$ non-positive (Proposition 9). Therefore, the term $\mathbb{E} [f'_{B_1}(g_{\varepsilon_1}(0)) | A = a] (q_a^{(n)})'(\mu_a^{(n)}(p))p$ is non-positive. By (A.41), this implies the uniform boundedness of $\frac{\partial^2}{\partial^2 \zeta} \mu_a^{(n)}(p, \zeta)$. Note that by Proposition 9, $(q_a^{(n)})'(\mu_a^{(n)}(p))$ is non-positive and bounded, and with (A.42) this shows that $\frac{\partial^2}{\partial^2 \zeta} q_a^{(n)}(\mu_a^{(n)}(p, \zeta))$ is bounded for all n and all sufficiently small ζ . This proves the second claim and thus completes the proof of Proposition 13.

A.6 Proof of Proposition 14

Fix $a \in \mathcal{A}$ and $n \in \mathbb{N}$. Consider the case where the payment distributions where all potential suppliers are offered a fixed payment, p , i.e., $\pi = \delta_p$. Recall from (3.8) and (3.10) that

$$\begin{aligned} \frac{d}{dp} u_a^{(n)}(p) &= \frac{d}{dp} \frac{1}{n} \mathbb{E} \left[R(D, T) - \sum_{i=1}^n P_i Z_i S_i \mid A = a \right] \\ &= \frac{d}{dp} \mathbb{E} \left[\frac{1}{n} R(D, T) \mid A = a \right] - \frac{d}{dp} \left(\frac{1}{n} p \mathbb{E} \left[\sum_{i=1}^n Z_i S_i \mid A = a \right] \right) \\ &= \frac{d}{dp} \mathbb{E} \left[\frac{1}{n} R(D, T) \mid A = a \right] - \frac{d}{dp} \left(p \mathbb{E} \left[\frac{1}{n} \Omega(D, T) T \mid A = a \right] \right), \end{aligned} \quad (\text{A.43})$$

where $T \sim \text{Binomial}(\mu_a^{(n)}(p), n)$. We have by the chain rule:

$$\frac{d}{dp} \mathbb{E} [R(D, T)] = \mathbb{E} \left[(\mu_a^{(n)})'(p) \left\{ \frac{d}{d\mu} \mathbb{E}_\mu [R(D, X) \mid A = a] \right\}_{\mu=\mu_a^{(n)}(p)} \right] \quad (\text{A.44})$$

where $X \sim \text{Binomial}(\mu, n)$, and similarly

$$\frac{d}{dp} \mathbb{E} [\Omega(D, T) T] = \mathbb{E} \left[(\mu_a^{(n)})'(p) \left\{ \frac{d}{d\mu} \mathbb{E}_\mu [\Omega(D, X) T \mid A = a] \right\}_{\mu=\mu_a^{(n)}(p)} \right]. \quad (\text{A.45})$$

Using arguments essentially identical to that of Proposition 9, we can show that for all $a \in \mathcal{A}$ and $\mu > 0$

$$\lim_{n \rightarrow \infty} \frac{d}{d\mu} \mathbb{E}_\mu \left[\frac{1}{n} R(D, X) \mid A = a \right] = \frac{d}{d\mu} (r(d_a/\mu)\mu) \quad (\text{A.46})$$

$$\lim_{n \rightarrow \infty} \frac{d}{d\mu} \mathbb{E}_\mu \left[\frac{1}{n} \Omega(D, X) X \mid A = a \right] = \frac{d}{d\mu} (\omega(d_a/\mu)\mu), \quad (\text{A.47})$$

where the limiting functions ω and r are defined in Assumptions 3 and 1, respectively. Substituting (A.46) and (A.47) into (A.44) and (A.45), respectively, and observing that

$$\lim_{n \rightarrow \infty} \mathbb{E} [\Omega(D, T) T / n \mid A = a] = \omega(d_a/\mu_a(p)) \mu_a(p), \quad (\text{A.48})$$

we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{d}{dp} \mathbb{E} \left[\frac{1}{n} R(D, T) \mid A = a \right] &= \mu'_a(p) \left\{ \frac{d}{d\mu} (r(d_a/\mu)\mu) \right\}_{\mu=\mu_a(p)}, \\ \lim_{n \rightarrow \infty} \frac{d}{dp} \left(p \mathbb{E} \left[\frac{1}{n} \Omega(D, T) T \mid A = a \right] \right) &= p \mu'_a(p) \left\{ \frac{d}{d\mu} (\omega(d_a/\mu)\mu) \right\}_{\mu=\mu_a(p)} + \omega(d_a/\mu_a(p)) \mu_a(p), \end{aligned}$$

where $\mu_a(p) = \lim_{n \rightarrow \infty} \mu_a^{(n)}(p)$ is defined in Lemma 4.

$$\begin{aligned}
& \lim_{n \rightarrow \infty} \frac{d}{dp} u_a^{(n)}(p) \\
&= \mu_a'(p) \left\{ \frac{d}{d\mu} (r(d_a/\mu)\mu) \right\}_{\mu=\mu_a(p)} - p \mu_a'(p) \left\{ \frac{d}{d\mu} (\omega(d_a/\mu)\mu) \right\}_{\mu=\mu_a(p)} + \omega(d_a/\mu_a(p)) \mu_a(p) \\
&= \mu_a'(p) \left(\left\{ \frac{d}{d\mu} (r(d_a/\mu)\mu) \right\}_{\mu=\mu_a(p)} - p \left\{ \frac{d}{d\mu} (\omega(d_a/\mu)\mu) \right\}_{\mu=\mu_a(p)} \right) \\
&\quad - \omega(d_a/\mu_a(p)) \mu_a(p) \\
&= \mu_a'(p) \left[r \left(\frac{d_a}{\mu_a(p)} \right) - p \omega \left(\frac{d_a}{\mu_a(p)} \right) - \left(r' \left(\frac{d_a}{\mu_a(p)} \right) - p \omega' \left(\frac{d_a}{\mu_a(p)} \right) \right) \frac{d_a}{\mu_a(p)} \right] \\
&\quad - \omega \left(\frac{d_a}{\mu_a(p)} \right) \mu_a(p) \\
&= u_a'(p),
\end{aligned}$$

where $u_a(\cdot)$ is defined in (3.15). This recovers the desired result. \square