Abstract

We study experimental design in large-scale stochastic systems with substantial uncertainty and structured cross-unit interference. We consider 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 propose a class of local experimentation schemes that can be used to optimize these payments without perturbing the overall market equilibrium. We show that, as the system size grows, our scheme can estimate the gradient of the platform’s utility with respect to $p$ while perturbing the overall market equilibrium by only a vanishingly small amount. We can then use these gradient estimates to optimize $p$ via any stochastic first-order optimization method. These results stem from the insight that, while the system involves a large number of interacting units, any interference can only be channeled through a small number of key statistics, and this structure allows us to accurately predict feedback effects that arise from global system changes using only information collected while remaining in equilibrium.

Keywords: interference, stochastic systems, mean-field limit.

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., 2017].

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 this lack of interference plays a key role in justifying standard analyses of randomized trials [Imbens and Rubin, 2015]. This interference-free 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

This work was partially supported by a seed grant from the Stanford Global Climate and Energy Project and a Facebook Faculty Award.
behavior on the auction and thus affect the opportunities available to other advertisers. In a
different context, Blake and Coey [2014] document failures of the no-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. 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 inter-
ference. 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 sys-
tems, 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 de-
signer 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 “lo-
cal” experimentation schemes that—by carefully leveraging the structure of the marketplace—
enable us to optimize payments without perturbing the overall market equilibrium. More
specifically, 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. Then, in the limit where the number of suppliers is large, we show
that we can estimate the gradient of the platform’s utility with respect to $p$ while perturbing
the overall market equilibrium by only a vanishingly small amount. 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 be-
havior 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, 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.
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 [Chetty, 2009], and 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.

Our approach to optimizing \( p \) using gradients obtained from local experimentation intersects with the literature on noisy zeroth-order optimization [e.g., Spall, 2005], which aims to optimize a function \( f(x) \) by sequentially evaluating \( f \) at points \( x_1, x_2, \ldots \), and obtaining in return noisy versions of the function values \( f(x_1), f(x_2), \ldots \). A number of zeroth-order optimization 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 [Jamieson et al., 2012, Duchi et al., 2012, Ghadimi and Lan, 2013, 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. Notably, local experimentation is beneficial even if the final objective is to identify a single optimal payment for all units. 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.

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 those arising in queueing networks [Halfin and Whitt, 1981, Vvedenskaya et al., 1996, Bramson et al., 2012, Tsitsiklis and Xu, 2012] and interacting particle systems [Mézard et al., 1987, Sznitman, 1991, Graham and Méléard, 1994]. 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 [Sznitman, 1991, Graham and Méléard, 1994, Bramson et al., 2012].
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 Local 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 efficiently to 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. For now, however, it is helpful to consider an elementary variant of our model that, despite its simplicity, lets us highlight some key properties of our approach. Each day \( k = 1, \ldots, K \) there are \( i = 1, \ldots, n \) potential suppliers, and demand for \( D_k \) identical tasks to be accomplished. A central platform offers a payment \( p_{ik} \) to each supplier that they can earn by servicing a unit of demand. The suppliers can accurately anticipate demand \( D_k \) (e.g., via their knowledge of local weather or events); however, the platform may not be able to. Given their knowledge of \( p_{ik} \) and \( D_k \), each supplier independently chooses to become “active”; we write \( Z_{ik} = 1 \) for active suppliers and \( Z_{ik} = 0 \) else. Then, demand \( D_k \) is randomly allocated to active suppliers (where each active supplier can service at most one unit of demand).

Our key assumption is that each supplier chooses to become active based on their expected revenue conditionally on being active: They first compute \( q_{D_k}(p_{ik}) \), the probability that they will be matched with a unit of demand conditionally on being active (which also depends on the payments offered to other supplies), and then take expected revenue to be \( p_{ik}q_{D_k}(p_{ik}) \). For now, we assume the following simple functional form

\[
P[Z_{ik} \mid p_{ik}, D_k] = \frac{1}{1 + e^{-\beta(p_{ik}q_{D_k}(p_{ik}) - B_i)}},
\]

where \( B_i \) measures the value of a random supplier-specific outside option; as discussed in Section 3, this specific functional form does not matter for our results—the important point is suppliers only interact via \( q_D(p) \). Finally, assume that the utility of the platform is as follows (later, we will relax the functional form of \( R \)),

\[
U = R(T, D) - \bar{p} \min \{T, D\}, \quad R(T, D) = CD \left(1 - e^{-D/T}\right),
\]

where \( T = \sum_{i=1}^{n} Z_i \) is the total number of active suppliers, \( \bar{p} \) is the average payment made to suppliers who service a unit of demand, and \( C \) is a scaling constant.

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 price \( p \), for a specific realization of demand \( D \); here \( D = 0.3n \), \( C = 100 \), and \( \log(B_i/20) \) has a standard
Figure 1: Example of large-sample behavior of market, conditionally on a realization of demand $D$. Here $\mu_D(p)$ represents the probability that a random supplier decides to become active at $p$, whereas $q_D(p)$ is the expected fraction of active suppliers who will be matched with a unit of demand.

Gaussian distribution. Overall, we see that the expected fraction of active suppliers $\mu_D(p) = \mathbb{E}_p[T/n \mid D]$ increases with $p$, whereas $q_D(p)$ decreases in $p$. There is a noticeable kink in the slope of $\mu_D$ once $q_D$ falls below 1 due to equilibrium effects.

In order to optimize payments $p$, the platform needs to understand the expected utility function $\chi(p) = \mathbb{E}_p[U/n]$ well enough to find its maximum. Before presenting our proposal, we briefly outline two baselines that are often discussed in practice.

- **Global experimentation:** On each day $k = 1, \ldots, K$, we randomly choose a payment $p_k$ that is made available to all suppliers. We then observe utility $U_k$, which acts as a noisy estimate of $\chi(p_k)$. Given sufficient time periods, this approach will consistently find the maximizer $p^*$ of $\chi(\cdot)$; however, the cost of experimentation is substantial, as we may end up repeatedly offering poorly chosen payments $p_k$ to all suppliers in the marketplace.

- **Classical A/B experimentation:** On each day $k = 1, \ldots, K$, we choose a small random fraction of suppliers and offer them a payment $p_k$, while everyone else gets offered the status quo payment $p_0$. We could then try to use the behavior of suppliers offered $p_k$ to estimate $\chi(p_k)$. This approach allows for cheap experimentation because most of the suppliers get offered $p_0$. However, it will not consistently recover the optimal payment $p^*$ because it ignores feedback effects: When we raise payments, more suppliers opt to join the market and so the probability of any given supplier being matched with demand goes down—and this attenuates the payment-sensitivity of supply relative to what is predicted by A/B testing.

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 \( p^* \) 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.

### 2.1 Estimating Utility Gradients in Equilibrium

The driving insight behind our approach is that it is possible to learn about \( \chi(p) \) via unobstrusive randomization by randomly perturbing the prices \( p_{ik} \) offered to supplier \( i \) in time period \( k \): We propose setting

\[
p_{ik} = p_k + \zeta \varepsilon_{ik}, \quad \varepsilon_{ik} \overset{iid}{\sim} \{ \pm 1 \}
\]  

(2.3)

uniformly at random, where \( \zeta > 0 \) is a (small) constant that governs the magnitude of the perturbations. Then, by regressing market participation \( Z_{ik} \) on the price perturbations \( \varepsilon_{ik} \), we can estimate the average local payment sensitivity of suppliers ignoring feedback effects [Imbens and Angrist, 1994]

\[
\Delta_D(p) = \left\{ \frac{d}{dp} \mathbb{E} \left[ \frac{1}{1 + \exp(-\beta (p' q_D(p) - B_0))} \right] \right\}_{p'}
\]

(2.4)

This quantity \( \Delta \) is not directly of interest for optimizing \( p \), as it ignores feedback effects. However, it turns out that \( \Delta \) captures relevant information for optimizing \( p \) and, given our generic model structure, we find that as the system size \( n \) grows

\[
\lim_{n \to \infty} \frac{d\mu_D(p)}{dp} = \left( 1 - \frac{d\mu_D(p)}{dp} \frac{p}{\mu_D(p)} \right) \left\{ \frac{D}{n} < \mu_D(p) \right\} \Delta_D(p) = 0.
\]

(2.5)

where all terms in (2.5) other than \( \Delta \) and the actual average sensitivity \( d\mu_D(p)/dp \) (which accounts for feedback) are readily observable from the data. The upshot is that, once we can estimate \( \Delta \) via local randomization (2.3), we can also get estimates of \( d\mu_D(p)/dp \) by solving (2.5), and finally obtain noisy \( \hat{\Gamma} \) gradients for \( d\chi(p)/dp \). A formal statement is given in Theorem 2.

We can thus use unobstrusive randomization as in (2.3) to estimate the gradient \( d\chi(p)/dp \) around any chosen price \( p \). The final step of our proposed approach, local experimentation, is to use these gradients to guide any first-order optimization method. For example, one could use gradient descent and, at each time period \( k \) estimate \( d\chi(p_k)/dp_k \) as \( \hat{\Gamma}_k \), and then set \( p_{k+1} = p_k + \eta \hat{\Gamma}_k \). In our experiments, we use variant of stochastic gradient descent called adagrad [Duchi et al., 2011]; in principle, one could also use more sophisticated methods that rely on acceleration [Cohen et al., 2018].

Figure 2 shows results on a simple simulation experiment in the setting of Figure 1, where the scaled demand \( D/n \) follows a beta(15, 35) distribution. We initialize the system at \( p_1 = 30 \), and then each day run price perturbations as in (2.3) to guide a price update using adagrad. We see that the system quickly converges to a near-optimal price of around 17.

We also compare our results to what one could obtain using the baseline of global experimentation, where we randomize the price \( p_k \sim \text{Uniform}(10, 30) \) in each time step and measure resulting utility \( U_k \), and then choose the final payment \( \hat{p} \) by maximizing a smooth estimate of the expectation of \( U_k \) given \( p_k \). The left panel of Figure 3 shows the resulting \( (p_k, U_k) \) pairs, as well as the resulting \( \hat{p} \). As seen in the right panel of Figure 2, the final \( \hat{p} \)
Figure 2: Results from learning $p$ via local experimentation. The left panel shows the evolution of the $p_k$ over time. The right panel compares the value of $p_k$ averaged over the last 50 steps of our algorithm to the value of $p$ that optimizes mean scaled utility, and a payment $\hat{p}$ learned via global experimentation.

Figure 3: Results from learning $p$ via global experimentation. The left panel shows pairs $(p_k, U_k)$ resulting from daily experiments, along with an estimate of the optimal $p$. 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 $D_k$).
obtained via this method is a reasonable estimate of the optimal \( p \). However, the price of experimentation incurred for finding this \( \hat{p} \) is huge: As shown in the right panel of Figure 3, after the first few days, the global experimentation approach systematically achieves lower daily utilities \( U_k \) than local experimentation because it often uses very poor choices of \( p_k \).

In principle, it may be possible to design a smarter version of global experimentation than the random search pursued above; for example, it may be possible to leverage recent advances in Bayesian optimization to rule out poor choices of \( p \) early on [Letham et al., 2018]. However, designing a variant of global experimentation that is competitive with our approach when \( n \) is large appears to be a challenging task. For example, any direct alternative to our method that seeks to estimate the gradient \( d\chi(p)/dp \) via global experimentation would need to undertake perturbations that scales linearly in \( n \) in order to get any finite error bounds (or, equivalently, the average per-supplier perturbation would need to be of constant order as \( n \) gets large). In contrast, local experimentation allows for per-supplier perturbations that scale only a little slower than \( 1/\sqrt{n} \). Thus, any approach to global experimentation that seeks to estimate \( d\chi(p)/dp \) will require unboundedly larger perturbations than our approach as the number of suppliers \( n \) gets large.

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 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, \( n \). We largely suppress this notation when the context is clear. When appropriate, we may make such dependence explicitly using the superscript \( (n) \), e.g., \( q^{(n)} \).

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 that the platform cannot react to). 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 \( a \in \mathcal{A} \):

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

and that \( \mathbb{E} [D \mid A = a] = d_a \) for all \( a \in \mathbb{N} \). It is assumed that each supplier can serve up to one unit of demand. 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 two rounds:
Round 1: Suppliers choose whether to they want to be active. A payment \( p_i \) is announced to supplier \( i \), for \( i = 1, 2, \ldots, n \), with the understanding that the supplier will be compensated with \( p_i \) if eventually matched with demand. A supplier will not be matched 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 2: The platform selects a subset of the active suppliers to satisfy the demand. If supplier \( i \) is selected, they will be paid \( p_i \) and tasked with serving one unit of demand; otherwise, the supplier will not be paid, and no demand will be served in that scenario. The selection procedure is randomized: if \( D \geq T \), then all active suppliers are selected. Otherwise, a subset of \( D \) active suppliers will be selected uniformly at random.

**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, denote by

\[
q_a(p) = \mathbb{E}\left[ (D/T) \land 1 \mid A = a \right] \tag{3.2}
\]

a supplier’s belief of their chance of being matched with a unit of demand should they choose to become active, conditionally on the global state being \( a \). Here, the expectation is calculated with respect to the equilibrium distribution of the number of active suppliers, \( T \). (For now, assume that such equilibrium distribution is well defined, and we will justify its meaning rigorously in a moment.) Then, the probability of supplier \( i \) becoming active is given by

\[
\mu_a(p) = \mathbb{P}(Z_i = 1 \mid A = a) = \mathbb{E}\left[ f_{B_i}(p_i q_a(p)) \right], \tag{3.3}
\]

Here, \( p_i q_a(p) \) is their expected revenue in equilibrium. \( B_i \) 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 \( \mathcal{B} \) conditionally on \( A \). \( f_b(\cdot) \) is a choice function, and \( f_b(x) \) represents the probability 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.

**Assumption 1.** 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 1:

**Example 1 (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: \( f_b(x) = \frac{1}{1 + e^{-\frac{1}{\alpha} x - \beta}} \), where \( \alpha > 0 \) is a parameter. Furthermore, let the private feature \( B_i \) take 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 \( \alpha \). Note that in the limit as \( \alpha \to \infty \), the probability of \( Z_i = 1 \) 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}  Let $S_i$ be the indicator variable that the $i$-th potential supplier was selected to serve a unit of demand. The platform’s utility is defined to be the difference between revenue and total payment:

$$U = R(T, D) - W(T \land D),$$

(3.4)

where $W = \sum_{i=1}^{n} (Z_i \land S_i)p_i / \sum_{i=1}^{n} (Z_i \land S_i)$ is the average payment across all selected suppliers. $R(T, D)$ is the platform’s revenue, with equilibrium active supply size is $T$ and total demand $D$. We assume that the revenue function $R$ is linear, in the sense that

$$R(\alpha T, \alpha D) = \alpha R(T, D), \quad \alpha \in \mathbb{R}_+.$$

(3.5)

As an example, the platform could receive a fixed amount $\gamma$ from each unit of demand served, in which case we have $R(T, D) = \gamma(T \land D)$. The revenue function (2.2) used in our motivating example also satisfies (3.5), but tapers off as $D$ gets very close to $T$ to reflect potential congestion effects. Given this notation, we write the platform’s expected utility in the $n$th system as

$$\chi_{a}^{(n)}(p) = \frac{1}{n}E_n [U \mid A = a], \quad \text{and} \quad \chi^{(n)}(p) = E_n \left[ \chi_{A}^{(n)}(p) \right].$$

(3.6)

The objective of the decision maker (i.e., platform operator) is to set the payment $p$ so as to maximize $\chi(p)$.

**Unit-Level Price Perturbation**  An important instrument that will be used repeatedly throughout the paper is the unit-level payment perturbation. Let $\{\varepsilon_i\}_{i \in \mathbb{N}}$ be a sequence of i.i.d. Bernoulli random variables with $P(\varepsilon_i = -1) = P(\varepsilon_i = +1) = \frac{1}{2}$. Fix $p > \zeta > 0$. We say the payments are $\zeta$-perturbed from $p$,

$$p_i = p + \zeta \varepsilon_i, \quad i \in \mathbb{N}.$$  

(3.7)

In what follows, we will use $q_a(p, \zeta)$ to denote the value of $q_a(p)$ when the payments are $\zeta$-perturbed from $p$. The meanings of $\mu_a(p, \zeta)$, $\chi_{a}^{(n)}(p, \zeta)$, $\chi^{(n)}(p, \zeta)$ are to be understood analogously. We may omit the dependence on $\zeta$, when $\zeta = 0$.

### 3.1 Existence and Uniqueness of Equilibrium Active Supply Size

The model described above assumes the existence and uniqueness of an equilibrium active supply size, $T$. We now provide a formal definition:

**Definition 2** (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 (3.3), the resulting distribution for the number of active suppliers equals that of $T$.

The following results shows that when payments are uniform across all potential suppliers, the equilibrium supply size distribution exists and is unique.

**Lemma 1.** Fix $p > 0$, $\zeta \in [0, p)$, and $a \in \mathcal{A}$. Suppose that the payments are $\zeta$-perturbed from $p$ as in (3.7). Then, conditional on $A = a$, the equilibrium active supply size exists and is unique.
Proof. Throughout, all calculations are based on the conditioning of $A = a$. 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 (3.3). Denote by $X^{(n)}_\mu$ a Binomial random variable with $n$ trials and a success probability $\mu$, independent from the rest of the system, and define
\[
\psi(\mu; p, \zeta) = \mathbb{E} \left[ f_B, (\mu + \varepsilon_i \zeta) \tilde{q}(\mu) \right],
\]
(3.8)
where
\[
\tilde{q}(\mu) = \mathbb{E} \left[ \frac{D}{X^{(n)}_\mu} \wedge 1 \right].
\]
(3.9)
That is, $\psi(\mu; p, \zeta)$ is the probability of a supplier becoming active when offered the payment $p_i = p + \varepsilon_i \zeta$, if they believe that the active supply size is $X^{(n)}_\mu$. Note that if the payments offered are uniform across all suppliers, so are the probabilities of the suppliers becoming active, and as a result the actual active supply size will follow a Binomial distribution. In particular, this implies that $X^{(n)}_\mu$ is an equilibrium active supply size, if and only if it satisfies the following fixed-point equation:
\[
\psi(\mu; p, \zeta) = \mu.
\]
(3.10)
It suffices to show that (3.10) admits a unique solution in the domain $\mu \in [0, 1]$. Because $f_b(\cdot)$ is by construction non-decreasing, it follows that $\psi(\mu; p, \zeta)$ is a continuous function and non-increasing in $\mu$: 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 (3.10), $\psi(\mu; p, \zeta)$, is a non-negative, continuous and non-increasing function over $\mu \in [0, 1]$, which implies that (3.10) admits a unique solution in [0, 1].

\section{Learning via Local Experimentation}

The approach we take in this paper is based on first order optimization: The statistical task is to estimate $\chi^{(n)}(p, \zeta_n)$ at the current value of $p$ as accurately as possible; then, we use this to update the payment using any first-order optimizer that can use noisy gradients. The result below shows that we can in fact accurately estimate gradients of $\chi^{(n)}(p, \zeta_n)$.

As discussed in Section 2 our proposal starts for perturbing individual payments as in (3.7), and then estimating the regression coefficient $\hat{\Delta}$ of market participation $Z_i$ on the perturbation $\zeta_n \varepsilon_i$, i.e.,
\[
\hat{\Delta} = \zeta_n^{-1} \sum_{i=1}^n (Z_i - \bar{Z})(\varepsilon_i - \bar{\varepsilon}) / \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2.
\]
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 prices, namely the price derivative of $\chi$.

\textbf{Theorem 2.} Suppose that the reward function $R$ in (3.4) is twice continuously differentiable in its first argument. Write $\bar{Z} = Z/n$ and $\bar{D} = D/n$. Let
\[
\hat{\Gamma} = \left\{ \frac{d}{dz} R(z, \bar{D}) \right\}_{z = \bar{Z}} - p \mathbb{1} \left( \{\bar{D} \geq \bar{Z}\} \right) \frac{\hat{\Delta}}{1 + 1 \left( \{\bar{Z} \geq \bar{D}\} \right) \bar{\Delta} / \bar{Z}} - (\bar{D} \wedge \bar{Z}).
\]
(4.1)
Then, assuming that the perturbations scale as $\zeta_n = \zeta n^{-\alpha}$ for some $0 < \alpha < 0.5$, we have $\limsup_n \text{Var}_n \left[ \hat{\Gamma} \right] < \infty$ and
\[
\lim_{n \to \infty} \mathbb{E}_n \left[ \hat{\Gamma} \right] = \frac{d}{dp} \chi(p).
\]
(4.2)
Our key use of Theorem 2 then involves optimizing for $p$. Because $\hat{\Gamma}$ is an (asymptotically) unbiased estimate of the gradient of $\chi$ at $p$, we can plug it into any first-order optimization method that allows for noisy gradients. The proposal below is based on adagrad, a form of self-normalizing gradient descent [Duchi et al., 2011]. We need to specify a step size $\eta$. Then, at time step $k$, given a status quo payment $p_k$, we

1. Deploy randomized price perturbations (3.7) to estimate $\hat{\Gamma}_k$ as in (4.1), and
2. Perform a gradient update $p_{k+1} \leftarrow p_k + \eta\hat{\Gamma}_k / G_k$, where $G_k^2 = \sum_{j=1}^{k} \hat{\Gamma}_j^2$.

We can then use standard results to verify convergence of the price sequence $p_k$ to the optimal price $p^*$, such that

$$\lim_{n \to \infty} \chi(p^*) - \frac{1}{K} \sum_{k=1}^{p} \chi(p_k) = \mathcal{O}_p \left( \frac{1}{\sqrt{k}} \right).$$

In some cases, it may be advantageous to use step sizes that decay faster (e.g., as $1/k$ as opposed to $1/\sqrt{k}$) to get price sequences that stabilize more quickly for large values of $k$.

## 5 Proof of Main Results

In this section, we present the proof of our main result, Theorem 2. We begin in Section 5.1 by establishing several regularity results concerning the derivatives of the equilibrium quantities $\mu(p, \zeta)$ and $q(p, \zeta)$ with respect to the perturbation size, $\zeta$ and reference price $p$. Importantly, we show that these bounds hold uniformly over all system sizes $n$. Using these results, we obtain in Section 5.2 upper bounds on the accuracy of various estimators in our local experimentation scheme. Finally, we combine the estimates in Section 5.3 to establish the convergence of $\hat{\Gamma}$ to $d\chi(p)/dp$, thus proving Theorem 2.

### 5.1 Sensitivity in Equilibrium

Denote by $T(p, \zeta)$ the active supply size distribution in equilibrium when the payments are $\zeta$-perturbed from $p$. Note that since $T(p, \zeta)$ remains a Binomial distribution, i.e., $T(p, \zeta) \overset{d}{=} T_{\mu(n)}(p, \zeta)$. Recall that $\mu_a(p, \zeta)$ is the expected fraction of active suppliers in equilibrium when the payments are $\zeta$-perturbed from $p$:

$$\mu_a(p, \zeta) = \mathbb{E}[T(p, \zeta)/n \mid A = a].$$

The following lemma provides some crucial structural properties that we will use subsequently to derive guarantees on the estimators for $\frac{d}{dp} \chi(p)$.

**Lemma 3.** Fix $p > 0$, $a \in A$ and $n \in \mathbb{N}$. $\mu_a(p, \zeta)$ and $q_a(p, \zeta)$ are twice differentiable functions with respect to $\zeta$, and satisfy:

1. $\frac{\partial}{\partial \zeta} [\mu_a(p, \zeta)]_{\zeta=0} = \frac{\partial}{\partial \zeta} [q_a(p, \zeta)]_{\zeta=0} = 0$.

2. Define $g_{e_1}(\zeta) \equiv (p + \epsilon_1 \zeta)q_a(p, \zeta)$. The second derivatives satisfy:

$$\frac{\partial^2}{\partial \zeta^2} [\mu_a(p, \zeta)]_{\zeta=0} = \frac{\mathbb{E}\left[f''_{B_1}(g_{e_1}(0))\right] q_a(p)^2}{1 - \mathbb{E}\left[f'_{B_1}(g_{e_1}(0))\right] q'_a(\mu_a(p)) p},$$

$$\frac{\partial^2}{\partial \zeta^2} [q_a(p, \zeta)]_{\zeta=0} = q'_a(\mu_a(p)) \left( \frac{\mathbb{E}\left[f''_{B_1}(g_{e_1}(0))\right] q_a(p)^2}{1 - \mathbb{E}\left[f'_{B_1}(g_{e_1}(0))\right] q'_a(\mu_a(p)) p} \right) ,$$

where $f_{B_1}$ is a form of self-normalizing gradient descent [Duchi et al., 2011].
derivations are conditional on the event that $A$.

Proof. Some remarks on notation. Throughout the proof, we will assume that the global state $a$ remains fixed, and suppress it from our notation with the understanding that all derivations are conditional on the event that $A = a$. For a function $f(p, \zeta)$, we will use the short-hand $\frac{\partial}{\partial \zeta} f(p, 0)$ to denote $\frac{\partial}{\partial \zeta} [f(p, \zeta)]_{\zeta=0}$, and the same convention applies to higher order derivatives.

We begin with a useful proposition showing that $\bar{q}'(\mu)$ is uniformly bounded over $n$. The proof is given in Appendix A.1.

**Proposition 4.** Suppose $A = a$ and $\mu \in (0, 1)$, Then, there exists $c_0 > 0$, such that

$$- c_0 \leq \bar{q}'(\mu) \leq 0, \quad \forall n \in \mathbb{N}. \quad (5.4)$$

We first prove that $\frac{\partial}{\partial \zeta} \mu(p, 0) = 0$. By (3.10), $\mu(p, \zeta)$ satisfies

$$\frac{\partial}{\partial \zeta} \mu(p, \zeta) = \frac{\partial}{\partial \zeta} \psi(\mu(p, \zeta), \zeta), \quad k \in \mathbb{N}. \quad (5.5)$$

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

$$\psi(\mu(p, \zeta); \zeta) - \psi(\mu(p, 0); 0) \\
\overset{(a)}{=} \frac{1}{2} \left( \mathbb{E} \left[ \sigma( (p + \zeta) q(p, \zeta) - B_1) \right] + \mathbb{E} \left[ \sigma( (p - \zeta) q(p, \zeta) - B_1) \right] \right) \\
- \mathbb{E} \left[ \sigma( p q(p, 0) - B_1) \right] \\
= \frac{1}{2} \left( \mathbb{E} \left[ \sigma( (p + \zeta) q(p, \zeta) - B_1) \right] - \mathbb{E} \left[ \sigma( p q(p, 0) - B_1) \right] \right) \\
+ \frac{1}{2} \left( \mathbb{E} \left[ \sigma( (p - \zeta) q(p, \zeta) - B_1) \right] - \mathbb{E} \left[ \sigma( p q(p, 0) - B_1) \right] \right), \quad (5.6)$$

where $(a)$ follows from the definition of $\zeta$-perturbation ((3.7)) and the independence of $\{\varepsilon_i\}_{i \in \mathbb{N}}$ from the rest of the system. Since both $\sigma_\beta$ and $\bar{q}$ are bounded, for the first term on the right-hand side of (5.6), it is not difficult to show using the dominated convergence theorem that there exists $c > 0$ such that\(^1\)

$$\mathbb{E} \left[ \sigma( (p + \zeta) q(p, \zeta) - B_1) \right] - \mathbb{E} \left[ \sigma( p q(p, 0) - B_1) \right] \in d_\psi(p) \zeta \pm c \zeta^2, \quad (5.7)$$

for all sufficiently small $\zeta$, where $d_\psi(p) := \frac{\partial}{\partial \zeta} \mathbb{E} \left[ \sigma( p q(p, 0) - B_1) \right]$. Applying the same argument to the second term in (5.6), we have that there exists $c$, such that for all sufficiently small $\zeta$

$$\psi(\mu(p, \zeta); \zeta) - \psi(\mu(p, 0); 0) \in \frac{1}{2} \left( d_\psi(p) \zeta - d_\psi(p) \zeta \right) \pm c \zeta^2, \quad (5.8)$$

which further implies that

$$\frac{\partial}{\partial \zeta} \mu(p, 0) = 0. \quad (5.9)$$

\(^1\)Notation: $x \equiv y \pm z \leftrightarrow x \in [y - z, y + z]$. 

13
For the derivative of $q$, note that by chain rule, we have

$$
\frac{\partial}{\partial \zeta} q(p, \zeta) = \frac{\partial}{\partial \zeta} \mathbb{E} \left[ \frac{D}{T(\mu(p, \zeta))} \wedge 1 \right] = \tilde{q}'(\mu(p, \zeta)) \frac{\partial}{\partial \zeta} \mu(p, \zeta). \tag{5.10}
$$

Since $\frac{\partial}{\partial \zeta} \mu(p, 0) = 0$, and $\tilde{q}'(\mu)$ is finite by Proposition 4, we have that $\frac{\partial}{\partial \zeta} q(p, 0) = 0$. This proves the first claim of Lemma 3.

For the second claim, recall the chain rule of differentiation: for differentiable functions $f$ and $g$, we have that $\frac{d}{dx} f(g(x)) = f'(g(x))g'(x)$. Applying the chain rule to (5.5), we have that

$$
\frac{\partial^2}{\partial \zeta^2} \mu(p, \zeta) = \mathbb{E} \left[ \frac{\partial^2}{\partial \zeta^2} f_B(\xi_1(\zeta)) \right] = \mathbb{E} \left[ f''_B(\xi_1(\zeta)) \xi'_{\xi_1}(\zeta)^2 + f'_B(\xi_1(\zeta)) g''_1(\zeta) \right]. \tag{5.11}
$$

Note that

$$
g'_{\xi_1}(\zeta) = \frac{\partial}{\partial \zeta} (p + \varepsilon \zeta) q(p, \zeta) = p \frac{\partial}{\partial \zeta} q(p, \zeta) + \varepsilon_1 q(p, \zeta) + \varepsilon_1 \hat{\zeta} \frac{\partial}{\partial \zeta} q(p, \zeta), \tag{5.12}
$$

and

$$
g''_{\xi_1}(\zeta) = p \frac{\partial^2}{\partial \zeta^2} q(p, \zeta) + 2 \varepsilon \frac{\partial}{\partial \zeta} q(p, \zeta) + \varepsilon_1 \hat{\zeta} \frac{\partial^2}{\partial \zeta^2} q(p, \zeta). \tag{5.13}
$$

By chain rule, we have

$$
\frac{\partial^2}{\partial \zeta^2} q(p, 0) = \tilde{q}''(\mu(p, 0)) \left( \frac{\partial}{\partial \zeta} \mu(p, 0) \right)^2 + \tilde{q}'(\mu(p, 0)) \frac{\partial^2}{\partial \zeta^2} \mu(p, 0)
= \tilde{q}'(\mu(p, 0)) \frac{\partial^2}{\partial \zeta^2} \mu(p, 0), \tag{5.14}
$$

where the last step follows from the fact that $\frac{\partial}{\partial \zeta} \mu(p, 0) = 0$. Applying (5.10) and (5.14) to (5.12) and (5.13), we have

$$
g'_{\xi_1}(0) = p \frac{\partial}{\partial \zeta} q(p, 0) + \varepsilon_1 q(p, 0) + 0 = \varepsilon_1 q(p), \tag{5.15}
$$

and

$$
g''_{\xi_1}(0) = p \frac{\partial^2}{\partial \zeta^2} q(p, 0) + 2 \varepsilon \frac{\partial}{\partial \zeta} q(p, 0) + 0 = p \tilde{q}'(\mu(p)) \frac{\partial^2}{\partial \zeta^2} \mu(p, 0). \tag{5.16}
$$

Substituting the expressions for $g'_{\xi_1}(0)$ and $g''_{\xi_1}(0)$ into (5.11), we obtain:

$$
\frac{\partial^2}{\partial \zeta^2} \mu(p, 0) = \mathbb{E} \left[ f''_B(\xi_1(0)) \varepsilon_1 q(p) + f'_B(\xi_1(0)) \hat{\zeta} \frac{\partial}{\partial \zeta} \mu(p, 0) \right]
= \mathbb{E} \left[ f''_B(\xi_1(0)) \right] q(p)^2 + \mathbb{E} \left[ f'_B(\xi_1(0)) \right] p \tilde{q}'(\mu(p)) \frac{\partial^2}{\partial \zeta^2} \mu(p, 0), \tag{5.17}
$$

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

$$
\frac{\partial^2}{\partial \zeta^2} \mu(p, 0) = \frac{\mathbb{E} \left[ f''_B(\xi_1(0)) \right] q(p)^2}{1 - \mathbb{E} \left[ f'_B(\xi_1(0)) \right] \tilde{q}'(\mu(p)) p}, \tag{5.18}
$$

14
and by (5.14), we have
\[
\frac{\partial^2}{\partial^2 \zeta} q(p, 0) = \frac{\partial^2}{\partial^2 \zeta} \mu(p, 0) = \frac{\partial^2}{\partial^2 \zeta} \mu(p, 0) = \frac{\partial^2}{\partial^2 \zeta} \left( \frac{E \left[ f''_{B_1} (g_{z_1}(0)) \right] q(p)^2}{1 - E \left[ f'_{B_1} (g_{z_1}(0)) \right] q(p)} \right). 
\] (5.19)

This proves the second claim of Lemma 3.

Finally, we check the uniform boundedness of the second derivatives with respect to \( n \). The boundedness of \( \frac{\partial^2}{\partial^2 \zeta} \mu(p, 0) \) follows from the fact that \( f'_q \) is non-negative and \( \tilde{q}' \) non-positive (Proposition 8). Therefore, the term \( E \left[ f'_{B_1} (g_{z_1}(0)) \right] \tilde{q}'(\mu(p))p \) is non-positive. By (5.18), this implies the uniform boundedness of \( \frac{\partial^2}{\partial^2 \zeta} \mu(p, 0) \). For \( \frac{\partial^2}{\partial^2 \zeta} q(p, 0) \), note that by Proposition 4, \( \tilde{q}'(\mu(p)) \) is non-positive and bounded, and this establishes the boundedness of \( \frac{\partial^2}{\partial^2 \zeta} q(p, 0) \).

5.2 Estimation

Lemma 5. Define
\[
\Delta_a^{(n)}(p) = q_a^{(n)}(p)E_n \left[ f'_{B_1} (pq_a^{(n)}(p)) \right]. 
\] (5.20)

Fix \( a \in A \) and \( p > 0 \). There exists a constant \( C > 0 \) such that, for every \( \zeta \) and \( n \)
\[
\left| \frac{1}{\zeta} \text{Cov}_n \left[ Z_i, \epsilon_i | A = a \right] - \Delta_a^{(n)}(p) \right| \leq C. 
\] (5.21)

Proof. Given \( \zeta \)-perturbed payments \( p_t = p + \zeta \epsilon_i \), we have
\[
\text{Cov}_n \left[ Z_i, \epsilon_i | A = a \right] = E_n \left[ \epsilon_i f_{B_1} \left( (p + \zeta \epsilon_i) q_a^{(n)}(p, \zeta) \right) \right]. 
\]

We can then take the limit \( \zeta \to 0 \), and verify that there exists \( C > 0 \) such that
\[
\left| E_n \left[ \frac{1}{\zeta} \epsilon_i f_{B_1} \left( (p + \zeta \epsilon_i) q_a^{(n)}(p, \zeta) \right) \right] - q_a^{(n)}(p, 0)E_n \left[ f'_{B_1} (pq_a^{(n)}(p, 0)) \right] \right| \leq C \] (5.22)

Here, we used the fact that
\[
\frac{\partial}{\partial \zeta} \left[ q_a^{(n)}(p, \zeta) \right] \bigg|_{\zeta = 0} = 0 \] (5.23)

by Lemma 3, both \( f_{B_1}(\cdot) \) and \( q_a^{(n)}(p, \cdot) \) are twice differentiable with bounded second derivatives (Lemma 3) uniformly over \( n \), and \( \epsilon_i \) has variance 1. \( \square \)

The main consequence of Lemma 5 is that, for small perturbations \( \zeta \), we can consistently recover \( \Delta_a^{(n)} \) by regressing market participation \( Z_i \) against the exogenous shocks \( \epsilon_i \). The following result provides a formal statement in a large-sample regime where the size of the perturbation \( \zeta_n \) decays as \( n \) gets large.

Lemma 6. Fix \( a \in A \) and \( p \in \mathbb{R}_+ \). Suppose we have a sequence of problems indexed by \( n \), and we set a perturbation size \( \zeta_n = \zeta n^{-\alpha} \) for some \( 0 < \alpha < 0.5 \). Let
\[
\tilde{\Delta} = \zeta_n^{-1} \text{Cov}[Z_i, \epsilon_i] / \text{Var}[\epsilon_i] \] (5.24)
be the scaled regression coefficient of \( Z_i \) on \( \varepsilon_i \). Then, as \( n \) grows to infinity, the following limit exists
\[
\lim_{n \to \infty} \mu_a^{(n)}(p) = \mu_a^{\infty}(p), \quad (5.25)
\]
and
\[
\left\{ \zeta_n \sqrt{n} \left( \hat{\Delta} - \Delta_a^{(n)}(p) \right) \mid A = a \right\} \Rightarrow \mathcal{N} \left( 0, \mu_a^{\infty}(p)(1 - \mu_a^{\infty}(p)) \right), \quad (5.26)
\]

**Proof.** To show that the limit of \( \mu_a^{(n)}(p) \) exists as \( n \to \infty \), recall from (3.10) that \( \mu_a^{(n)}(p) \) is the solution to
\[
\mu = \mathbb{E} \left[ f_{B_1} \left( (p + \varepsilon_i \zeta_n) \hat{q}(\mu) \right) \right],
\]
where \( \hat{q}_a^{(n)}(\mu) = \mathbb{E} \left[ (D/X^{(n)}_{\mu}) \wedge 1 \mid A = a \right] \). By the assumption on \( D \) (cf. (3.1)) and the fact that \( X_{\mu}^{(n)} \) is a Binomial random variable with \( n \) trials and mean \( n \mu \), it is not difficult to show that \( \hat{q}_a^{(n)}(\mu) \) is a continuous function that converges uniformly over compact intervals to the function
\[
\hat{q}_a^{\infty}(\mu) := \frac{d_a}{\mu} \wedge 1
\]
as \( n \to \infty \). Since \( \zeta_n \to 0 \) as \( n \to \infty \), the right-hand side of (5.27) converges to \( \mathbb{E} \left[ f_{B_1} \left( p \hat{q}_a^{\infty}(\mu) \right) \right] \) uniformly over compact intervals. This proves that \( \mu_a^{(n)}(p) \) converges to \( \mu_a^{\infty}(p) \) as \( n \to \infty \), the latter being defined as the unique solution to the balance equation:
\[
\mu = \mathbb{E} \left[ f_{B_1} \left( p \hat{q}_a^{\infty}(\mu) \right) \right].
\]

Next, we note that when \( n \) gets large the effect of the payment perturbations on any single potential supplier gets infinitesimally small, and so the covariance of \( Z_i \) and \( \varepsilon_i \) decays to zero. Standard linear regression theory thus implies that
\[
\zeta_n \sqrt{n} \left( \hat{\Delta} - \mathbb{E} \left[ \hat{\Delta} \mid A = a \right] \right) / \text{Var}_n \left[ Z_i \mid A = a \right]^{1/2} \Rightarrow \mathcal{N} \left( 0, 1 \right),
\]
and furthermore our independent sampling design implies that reading off the relevant variance from the limiting Bernoulli variable \( Z_i \), we have \( \lim_{n \to \infty} \text{Var}_n \left[ Z_i \mid A = a \right] = \mu_a^{\infty}(p)(1 - \mu_a^{\infty}(p)) \). Finally
\[
\mathbb{E} \left[ \hat{\Delta} \mid A = a \right] = \zeta_n^{-1} \text{Cov}_n \left[ Z_i, \varepsilon_i \mid A = a \right],
\]
and so the desired claim then follows by noting (5.21).

Finally, we will use \( \hat{\Delta} \) to construct an estimator for the derivative of the fraction of active suppliers, \( \frac{d}{dp} \mu_a^{(n)}(p) \).

**Lemma 7.** Fix \( a \in \mathcal{A} \) and \( p \in \mathbb{R}_+ \), and consider the sequence of problems and \( \zeta_n \) as defined in Lemma 6. Then,
\[
\lim_{n \to \infty} \mathbb{E} \left[ \left( \hat{\gamma} - \frac{d}{dp} \mu_a^{(n)}(p) \right)^2 \mid A = a \right] = 0, \quad \hat{\gamma} = \frac{\hat{\Delta}}{1 + 1 \left( \{ Z > D \} \right) p \hat{\Delta}/Z}. \quad (5.30)
\]

**Proof.** We first note that, for any \( a \in \mathcal{A} \) and \( p \in \mathbb{R}_+ \), we have
\[
\frac{d}{dp} \mu_a(p) = \frac{\Delta_a^{(n)}(p)}{1 - p \Delta_a^{(n)}(p) \hat{q}_a^{(n)}(\mu_a^{(n)}(p)) / \hat{q}_a^{(n)}(p)}, \quad (5.31)
\]
where $\Delta^{(n)}_a(p)$ is defined in Lemma 5. We will verify this fact at the end of the proof. Furthermore, we note that—outside a set of measure zero—the ratio $\hat{q}^{(n)'}(\mu^{(n)}_a(p)) / q^{(n)}_a(p)$ concentrates to $1/\mu^{(n)}_a(p)$ conditionally on $A = a$, and we can accurately estimate this latter quantity as $1/\hat{Z}$. The desired result then follows immediately from Lemma 6 (the fact that $\hat{\Delta}$ is also consistent in squared error also follows from the argument of Lemma 6).

It now remains to verify (5.31). To simplify notation, we suppress the dependence on $(n)$ for the rest of this proof. Define

$$b = \mathbb{E} \left[ f'_{B_1} (pq_a(\mu_a(p))) \right].$$

By (3.10), we have

$$\mu'_a(p) = \frac{d}{dp} \mathbb{E} [f_{B_1} (pq_a(\mu_a(p)))]$$

$$= \mathbb{E} \left[ f'_{B_1} (pq_a(\mu_a(p))) (\hat{q}_a(\mu_a(p)) + \hat{p}_a(\mu_a(p)) \mu'_a(p)) \right]$$

$$= b_n q_a(p) + pb_n \hat{q}_a(\mu_a(p)) \mu'_a(p).$$

(5.33)

Note that, by definition,

$$b_n = \Delta_a(p)/q_a(p),$$

and the desired result follows.

5.3 Proof of Theorem 2

Conditionally on $A = a$, we have $Z \sim \text{Binomial}(\pi, n)$ with $\pi = \mu_A(p)$. Now recall that, if $X$ is any random variable drawn from an exponential family with natural parameter $\theta$ and sufficient statistic $H(X)$, then for any function $f$ with $\mathbb{E}_\theta [|f'(X)|] < \infty$ we have [Hudson, 1978]

$$\frac{d}{d\theta} \mathbb{E}_\theta [f(X)] = \text{Cov}_\theta [f(X), H(X)].$$

(5.35)

The binomial distribution is an exponential family with sufficient statistic $Z$ and natural parameter $\log(\pi/(1 - \pi))$, and so we can use the chain rule to check that, in our setting

$$\frac{d}{dp} \mathbb{E} [f(Z)] = \mathbb{E} \left[ \frac{(d/ dp) \mu_A(p)}{\mu_A(p)(1 - \mu_A(p))} \left( n \text{Cov} [f(Z), Z | A] \right) \right],$$

(5.36)

again for any $f$ with an absolutely integrable derivative. We now apply this result to different components of $(d/ dp) \chi^{(n)}_A(p, \zeta_n)$. First, by twice differentiability of the reward function $R$, we can use Taylor approximation to check that

$$\lim_{n \to \infty} \mathbb{E} \left[ \frac{d}{dp} \mathbb{E} [R(\tilde{Z}, \tilde{D}) | A] \right]$$

$$= \mathbb{E} \left[ \frac{(d/ dp) \mu_A(p)}{\mu_A(p)(1 - \mu_A(p))} \lim_{n \to \infty} n \text{Cov} [R(\tilde{Z}, \tilde{D}), \tilde{Z} | A] \right]$$

$$= \mathbb{E} \left[ \frac{d}{dp} \mu_A(p) \left\{ \frac{d}{dz} R(z, \tilde{D}) \right\}_{z=\mu_A(p)} \right].$$

17
and moreover
\[ \lim_{n \to \infty} \mathbb{E} \left[ \left\{ \frac{d}{dz} R(z, D) \right\}_{z=\bar{z}} \right] = \mathbb{E} \left[ \left\{ \frac{d}{dz} R(z, \bar{D}) \right\}_{z=\mu_A(p)} \right]. \]

Similarly, for almost all values of \( \mu_A(p) \),
\[ \lim_{n \to \infty} \frac{d}{dp} \mathbb{E} \left[ \bar{Z} \wedge \bar{D} \right] = \mathbb{E} \left[ \frac{d\mu_A(p)}{dp} 1 \left( \{ \bar{Z} \leq \bar{D} \} \right) \right]. \quad (5.37) \]

We also note that \( \mathbb{E} [W \mid A, T, D] \to p \), and \( (d/dp)\mathbb{E} [W \mid A, T, D] \to 1 \). The final challenge is that all the above quantities involve unknown terms \( (d/dp)\mu_A(p) \). However, from Lemma 7, we know that \( \hat{Y} \) is a consistent estimate of \( (d/dp)\mu_A(p) \), and the desired result follows.

References

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.

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.

Raj Chetty. Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. *Annual Review of Economics*, 1(1):451–488, 2009.

Michael B Cohen, Jelena Diakonikolas, and Lorenzo Orecchia. On acceleration with noise-corrupted gradients. *arXiv preprint arXiv:1805.12591*, 2018.
John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(Jul):2121–2159, 2011.

John C Duchi, Peter L Bartlett, and Martin J Wainwright. Randomized smoothing for stochastic optimization. *SIAM Journal on Optimization*, 22(2):674–701, 2012.

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.

Ronald A Fisher. *The Design of Experiments*. Oliver and Boyd, Edinburgh, 1935.

Saeed Ghadimi and Guanghui Lan. Stochastic first- and zeroth-order methods for nonconvex stochastic programming. *SIAM Journal on Optimization*, 23(4):2341–2368, 2013.

Carl Graham and Sylvie Mélédard. 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.

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

H Malcolm Hudson. A natural identity for exponential families with applications in multi-parameter estimation. *The Annals of Statistics*, 6(3):473–484, 1978.

Guido W. Imbens and Joshua D. Angrist. Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475, 1994.

Guido W Imbens and Donald B Rubin. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press, 2015.

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, 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, 2017.

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.
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.

Michael Ostrovsky and Michael Schwarz. Reserve prices in internet advertising auctions: a field experiment. *EC*, 11:59–60, 2011.

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.

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 4

Proof. The fact that $q'(p)$ is non-positive follows directly from the definition. We will use the following result; the proof is given in Appendix A.2.

**Proposition 8.** Let $X$ and $Y$ be two random variables taking values in $\mathbb{Z}_+$, such that $Y \leq X$ (X stochastically dominates Y). Let $f$ and $g$ be two functions defined on $\mathbb{Z}_+$ such that $f - g$ is a monotonically non-increasing function over $\mathbb{Z}_+$. Then, we have that

$$\mathbb{E} [f(X) - f(Y)] \leq \mathbb{E} [g(X) - g(Y)].$$  \hspace{1cm} (A.1)
For now, we will use the convention that $\frac{d}{\varepsilon} := 0$ for the remainder of the proof. Fix $\mu \in (0, 1)$, $\varepsilon \in (0, 1 - \mu)$, and $d \in \mathbb{N}$. Applying Proposition 8 with $X = X_{\mu + \varepsilon}$, $Y = X_{\mu}$, $f(x) = \frac{d}{\varepsilon}$ and $g(x) = \frac{d}{\varepsilon} \wedge 1$, we have that

$$
\frac{d}{d\mu} \left( \frac{d}{X_{\mu} \wedge 1} \right) = \lim_{\varepsilon \downarrow 0} \frac{\mathbb{E} \left[ g(X_{\mu + \varepsilon}) \right] - \mathbb{E} \left[ g(X_{\mu}) \right]}{\varepsilon} \\
\geq \lim_{\varepsilon \downarrow 0} \frac{\mathbb{E} \left[ f(X_{\mu + \varepsilon}) \right] - \mathbb{E} \left[ f(X_{\mu}) \right]}{\varepsilon} \\
= \frac{d}{d\mu} \left( \frac{d}{X_{\mu}} \right). \quad (A.2)
$$

Next, we show that

$$
\frac{d}{d\mu} \left( \frac{d}{X_{\mu}} \right) \geq \frac{d_a}{\mu^2}. \quad (A.3)
$$

To this end, define

$$p(\mu, x) = \mathbb{P}(X_{\mu} = x) = \binom{n}{x} \mu^x (1 - \mu)^{n-x}. \quad (A.4)
$$

We have that

$$
\frac{d}{d\mu} p(\mu, x) = \binom{n}{x} \mu^{x-1} (1 - \mu)^{n-x-1} (x - n\mu) = \frac{1}{\mu(1 - \mu)} (x - n\mu) p(\mu, x). \quad (A.5)
$$

This implies

$$
\frac{d}{d\mu} \mathbb{E} \left[ \frac{d}{X_{\mu}/n} \right] = \sum_{x=1}^{\infty} \frac{d}{x/n} \frac{d}{d\mu} p(\mu, x) \\
= \frac{d}{n\mu(1 - \mu)} \sum_{x=1}^{\infty} \frac{1}{x/n} (x - n\mu) p(\mu, x) \\
= \frac{d}{n\mu(1 - \mu)} \mathbb{E} \left[ \frac{1}{X_{\mu}/n} (X_{\mu} - n\mu) \right] \\
= \frac{d}{n\mu^2(1 - \mu)} \mathbb{E} \left[ \frac{1}{X_{\mu}/n} (X_{\mu} - n\mu) \right] \\
\geq (a) \frac{d}{n\mu^2(1 - \mu)} \mathbb{E} \left[ \left( 1 - \frac{X_{\mu} - n\mu}{n\mu} \right) (X_{\mu} - n\mu) \right] \\
= \frac{d}{n\mu^2(1 - \mu)} \mathbb{E} \left[ (X_{\mu} - n\mu) - \mathbb{E} \left[ (X_{\mu} - n\mu)^2 \right] \right] \\
\geq (b) \frac{d}{n\mu^2(1 - \mu)} \left( 1 - \mu \right) \\
= - \frac{d}{n\mu^2}, \quad (A.6)
$$

where step (a) follows from the Taylor approximation of $\frac{1}{x}$ for $x \in (0, 1)$, and (b) from the fact that $\mathbb{E} \left[ X_{\mu} - n\mu \right] = 0$ and the variance of $X_{\mu}$ is $n\mu(1 - \mu)$. Combining (A.2) and (A.6), we have that

$$
\mathbb{E} \left[ \frac{D/n}{\mu^2} \mid A = a \right] = - \frac{d_a}{\mu^2}. \quad (A.7)
$$
Finally, note that we have thus far used the convention that \( \frac{d}{d} := 0 \). This assumption simplifies the analysis but leads to a small inconsistency when \( X_\mu = 0 \), since \( d \wedge 1 \) should have been 1, than 0. However, note that as \( n \) grows, the probability of \( X_\mu = 0 \) vanishes, it is not difficult to show that the boundedness of \( \tilde{q}'(\cdot) \) should continue to hold.

A.2 Proof of Proposition 8

Proof. Denote by \( p_X \) and \( p_Y \) the probability mass functions of \( X \) and \( Y \), respectively.

\[
\begin{align*}
\mathbb{E} [f(X) - f(Y)] &- \mathbb{E} [g(X) - g(Y)] \\
= & \left( \sum_{l \in \mathbb{Z}_+} f(l)p_X(l) - f(l)p_Y(l) \right) - \left( \sum_{l \in \mathbb{Z}_+} g(l)p_X(l) - g(l)p_Y(l) \right) \\
= & \left( \sum_{l \in \mathbb{Z}_+} (f - g)(l)p_X(l) \right) - \left( \sum_{l \in \mathbb{Z}_+} (f - g)(l)p_Y(l) \right) \\
= & \mathbb{E} [(f - g)(X)] - \mathbb{E} [(f - g)(Y)] \\
\leq & 0,
\end{align*}
\]

where the last inequality follows from the fact that, since \( Y \leq X \) and \( (f - g)(\cdot) \) is monotonically non-increasing, we have that \( (f - g)(X) \leq (f - g)(Y) \).

\( \Box \)