Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Experimenting in Equilibrium

Stefan Wager, Kuang Xu

To cite this article:
Stefan Wager, Kuang Xu (2021) Experimenting in Equilibrium. Management Science 67(11):6694-6715. https://doi.org/10.1287/mnsc.2020.3844

Published in Management Science on February 16, 2021 as DOI: 10.1287/mnsc.2020.3844.
This article has not been copyedited or formatted. The final version may differ from this version.

http://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Vol. 67, No. 11, November 2021, pp. 6694–6715
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Experimenting in Equilibrium

Stefan Wager,[a] **Kuang Xu**[a]

[a] Graduate School of Business, Stanford University, Stanford, California 94305
Contact: swager@stanford.edu, http://orcid.org/0000-0002-7526-9077 (SW); kuangxu@stanford.edu, http://orcid.org/0000-0002-2221-1648 (KX)

Received: August 27, 2019
Revised: February 23, 2020; June 21, 2020
Accepted: June 27, 2020
Published Online in Articles in Advance: February 16, 2021

https://doi.org/10.1287/mnsc.2020.3844

Abstract. Classical approaches to experimental design assume that intervening on one unit does not affect other units. There are many important settings, however, where this noninterference assumption does not hold, as 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 via mean-field modeling. Our approach enables us to accurately estimate the effect of small changes to system parameters by combining unobtrusive randomization with lightweight 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 we show that our approach enables the platform to optimize p in large systems using vanishingly small perturbations.

History: Accepted by Hamid Nazerzadeh, big data analytics.
Funding: This work was supported by Stanford Global Climate and Energy Project.
Supplemental Material: The e-companion is available at https://doi.org/10.1287/mnsc.2020.3844.

Keywords: experimental design • interference • mean-field model • stochastic system

1. Introduction

Randomized controlled trials are widely used to guide decision making across different domains, ranging from classical industrial and agricultural applications (Fisher 1935) to developmental economics (Banerjee and Duflo 2011) and the modern technology sector (Kohavi et al. 2009, Tang et al. 2010, Athey and Luca 2019). In its most basic form, a randomized trial aims to assess the expected effectiveness of a set of interventions on a population by selecting a small but representative subpopulation of units and assigning to each unit a randomly chosen intervention. For example, in a medical trial, the decision maker may want to compare the effectiveness of a new experimental drug with the current standard of care. To do so they select a set of patients and randomly assign some fraction to the new treatment while others are given the control condition (i.e., current standard of care). The drug is then assessed by comparing the outcomes of treated and control patients. Similar randomized experiments are popular with technology companies, where they are often referred to as A/B tests. In this context, a company would select a small population of its users and expose them to different randomly generated designs; the best design that emerges from the experiment is then deployed to the entire user base at large.

When interpreting the results from randomized trials, it is common to make a "no interference" assumption, whereby we assume that the intervention assigned to any given unit does not affect observed outcomes for other units (Imbens and Rubin 2015); for example, in our medical example, we might assume that giving the experimental treatment to some patients does not affect outcomes for the control patients who are still receiving standard care. Such a lack of interference plays a key role in enabling us to use randomized trials to understand the effect of large-scale policy interventions, as it implies that any effects observed by experimenting on a representative subpopulation should also hold when the same interventions are applied to the overall population at large. However, this noninterference assumption is violated in many important applications, and randomized trials can lead to highly misleading conclusions in the presence of cross-unit interference. We illustrate this problem below using an example of Heckman et al. (1998).

Example 1 (Tuition Subsidies)**.** A policy maker is interested in estimating the effect $\theta(p)$ of offering all high school graduates a fixed subsidy of $\$p$ to attend college. To do so, they might consider running a small randomized controlled trial: Given a small set of study participants, randomly assign half of them to receive a

Published in Management Science on February 16, 2021 as DOI: 10.1287/mnsc.2020.3844.
This article has not been copyedited or formatted. The final version may differ from this version.

subsidy p and half of them not to, and then compare college enrollment rates among those two groups. As argued in Heckman et al. (1998), however, such an approach may badly overestimate the effect of the subsidy on enrollments because it fails to consider overall equilibrium effects on the college wage premium.

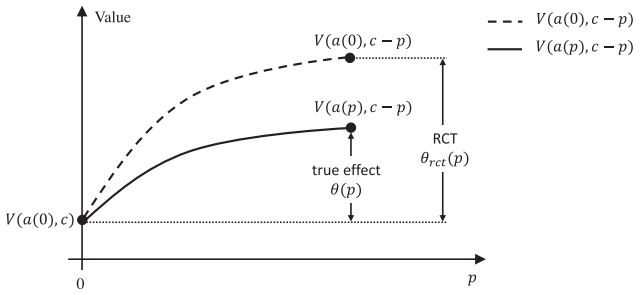
More formally, let $V(a, c)$ denote the average net value of enrolling in college, where a denotes the wage premium resulting from a college degree and c the cost of attendance. In general, we should expect V to be monotonically increasing in a and decreasing in c. The subsidy reduces costs by p and, thus, at first glance makes college more attractive. Where one needs to be careful, however, is in recognizing that the college wage premium a is not set in stone; rather, it is determined by labor market conditions. If more people enroll in college, one may expect the labor market bargaining power of college graduates to diminish and for a to decrease in response. Thus, if we believe the subsidy p increases enrollments, we might expect for $a(p)$ to be a (decreasing) function of p due to equilibrium effects.

We are now ready to illustrate why a simple randomized trial falls short here (Figure 1). The randomized trial only affects a small number of study participants and does not capture changes in a; specifically, it measures $\theta_{rct}(p) = V(a(0), c-p) - V(a(0),c)$. In contrast, the true effect of the subsidy should also reflect its impact on the equilibrium college wage premium, i.e., $\theta(p) = V(a(p), c-p) - V(a(0), c)$. In general, for any subsidy $p > 0$, we should expect $a(p) < a(0)$ and so $V(a(0), c-p) > V(a(p), c-p)$, meaning that the randomized trial will overestimate the effect of the subsidy. On the quantitative front, Heckman et al. (1998) discuss a setup where ignoring equilibrium effects would lead to estimates that are off by an order of magnitude.

1.1. Interference and Clustered Inference

The question of how to run experiments in the presence of cross-unit interference has received considerable attention in the literature. The simplest approach to dealing with interference is to assume that we can divide our experimental samples into disjoint clusters that do not interfere with each other and then to consider inference at the level of these clusters (Hudgens and Halloran 2008, Baird et al. 2018).

Figure 1. Illustration of the Failures of a Randomized Controlled Trial Under the Presence of Cross-Unit Interference in Example 1



One such example involves experimentation in Internet ad auctions, where each auction consists of a keyword along with a set of advertisers who submit competing bids in order for their ads to be displayed when the keyword is queried by a user. There is cross-unit interference because the same advertiser or keyword may appear in multiple auctions. Basse et al. (2016) and Ostrovsky and Schwarz (2011) make the observation that the auction type used for one keyword does not meaningfully affect how advertisers bid for other keywords. They then consider experiments that group auctions into clusters by their keywords and randomize auction formats across these keyword clusters, rather than across advertisers, as a means to avoid problems with interference. More broadly, in the context of tuition subsidies, this idea of cluster-level randomization could correspond to identifying communities that are relatively isolated from each other and randomizing the interventions across communities rather than across individuals; or, in the case of social networking, it could involve deploying different versions of a feature in different countries and hoping that the number of cross-border links is small enough to induce only negligible interference.

The limitation of such cluster-based approaches, however, is that the power of any experiment is limited by the number of noninterfering 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. Recently, several authors have sought to improve on the power of such cluster-based approaches by considering methods that allow interference to be captured by a generic graph, where two units are connected by an edge if the treatment assigned to one unit may affect the other's outcome (Aronow and Samii 2017, Eckles et al. 2017, Athey et al. 2018, Basse et al. 2019, Leung 2020). Even in this general case, however, we typically need to assume that the interference graph is sparse, that is, that most units do not interfere with each other. For example, Leung (2020) assumes that the average degree of the interference graph remains bounded.

1.2. Accounting for Interference via Equilibrium Modeling

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,

Published in Management Science on February 16, 2021 as DOI: 10.1287/mnsc.2020.3844.
This article has not been copyedited or formatted. The final version may differ from this version.

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 be more inclined to participate and reduce the amount of demand available to the remaining suppliers and, thus, reduce their incentives to participate.

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 ith supplier by a small mean-zero shock, that is, $p_i = p + \zeta\varepsilon_i$, where $0 < \zeta \ll 1$ and $\varepsilon_i = \pm 1$ independently and uniformly at random. A reduced form linear regression, one that estimates how the individual random shock $\zeta\varepsilon_i$ affects supplier i's behavior, recovers a certain marginal response function, which captures the supplier's sensitivity to payment changes against a fixed ambient market equilibrium. This marginal response, unfortunately, is not directly relevant for policy design, as it does not take into account the shift in market equilibrium should all suppliers receive the same payment change. However, in the limit where the number of suppliers is large, we show that a mean-field model can be used 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 useful information about the system—at a cost that scales sublinearly 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 2 (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, that is, 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. 2020), 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 3 (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 subpopulation, 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 4 (Renewable Energy Subsidies)**.** In an electricity wholesale 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

Published in Management Science on February 16, 2021 as DOI: 10.1287/mnsc.2020.3844.
This article has not been copyedited or formatted. The final version may differ from this version.

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 nonrenewable 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.

1.3. Related Work

The problem of experimental design under interference has received considerable attention in the statistics literature. For example, Blake and Coey (2014) document failures of the noninterference assumption due to an interaction between treated and control customers in an experiment run by an online marketplace. Blundell et al. (2004) consider the effects of a job search program on employment outcomes and emphasize the importance of considering general equilibrium effects whereby job offers given to program participants may substitute for job offers given to nonparticipants and increased search activity from participants may lower equilibrium wages for less skilled individuals. 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 the auction and thus affect the opportunities available to other advertisers. In all these cases, simple randomized controlled trials would paint a misleading picture about the effect of an overall policy change.

The dominant paradigm for working under interference has focused on robustness to potential interference effects and on defining estimands in settings where some units may be exposed to spillovers from treating other units (Sobel 2006, Hudgens and Halloran 2008, Tchetgen Tchetgen and VanderWeele 2012, Manski 2013, Aronow and Samii 2017, Eckles et al. 2017, Athey et al. 2018, Baird et al. 2018, Basse et al. 2019, Leung 2020). 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.

Another plausible approach would be to use structural estimation methods, directly estimate the whole underlying system, and subsequently use stochastic optimization to obtain the optimal decision. However, a full-blown structural estimation approach would be infeasible in our problem, because it involves a large number of interacting units, each with unknown features. In particular, as will be clear in Section 3, we consider the interaction among a large number of units, and each unit's behavior depends on a random choice function drawn from a potentially large set of options. The set of problem parameters thus involves the shapes of every possible choice function, as well as sampling distribution with respect to which the function is drawn for each unit. Directly estimating these parameters can be very difficult and, as we show, is not needed if the final goal is to identify the optimal action. Instead, our approach will focus on estimating a small number of key statistics which turn out to be sufficient for performing optimization. Doing so allows us to sidestep the scalability problem of the structural estimation approach and arrive at the optimal action in an efficient manner.

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., Harberger 1964, Chetty 2009). This approach can often be used for practical counterfactual analysis without needing to fit complicated structural models. We are inspired by this approach, and 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 (2014), and in crowd-sourcing systems, including Johari et al. (2017), 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 $x_1, x_2, \ldots$ and obtaining in return noisy versions of the function values $f(x_1), f(x_2), \ldots$ (Spall 2005, Bubeck et al. 2017). A number of bandit methods first generate noisy gradient estimates of the function by

Published in Management Science on February 16, 2021 as DOI: 10.1287/mnsc.2020.3844.
This article has not been copyedited or formatted. The final version may differ from this version.

comparing adjacent function values and subsequently use these estimates in a first-order optimization method (Flaxman et al. 2005, Kleinberg 2005, Jamieson et al. 2012, Ghadimi and Lan 2013, Nesterov and Spokoiny 2017). In our model, this approach would amount to estimating utility gradients via what we call global experimentation, that is, by comparing the empirical utilities observed at two different payment levels. Compared with 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 (Halfin and Whitt 1981, Vvedenskaya et al. 1996, Bramson et al. 2012, Tsitsiklis and Xu 2012, Stolyar 2015) and interacting particle systems (Mézard et al. 1987, Sznitman 1991, Graham and Méléard 1994). A key property of this mean-field limit is that, 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.

Mean-field-inspired approaches have also been used in game theory to analyze equilibria in the presence of a large number of players by assuming that the agents respond to a certain average behavior of the system (Jovanovic and Rosenthal 1988, Hopenhayn 1992, Weintraub et al. 2008, Adlakha et al. 2015); the equilibrium notion we use also falls under this category. In contrast to the existing literature, the main focus of our work lies in using mean-field limits to drive learning and experimentation.

2. Designing Experiments Under Equilibrium Effects

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 $t = 1, \ldots, T$ there are $i = 1, \ldots, n$ 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 $P_{it} \overset{\text{iid}}{\sim} \pi_t$ they commit to pay for each unit of demand served. The suppliers observe both π_t and 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 and furthermore that they do so via stationary reasoning (Hopenhayn 1992). Each supplier first computes $q_{A_t}(\pi_t)$, 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. We refer to this as a stationary model of supplier choice as we implicitly assume that suppliers do not 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 (Weintraub et al. 2008, Chetty 2009, Adlakha et al. 2015).

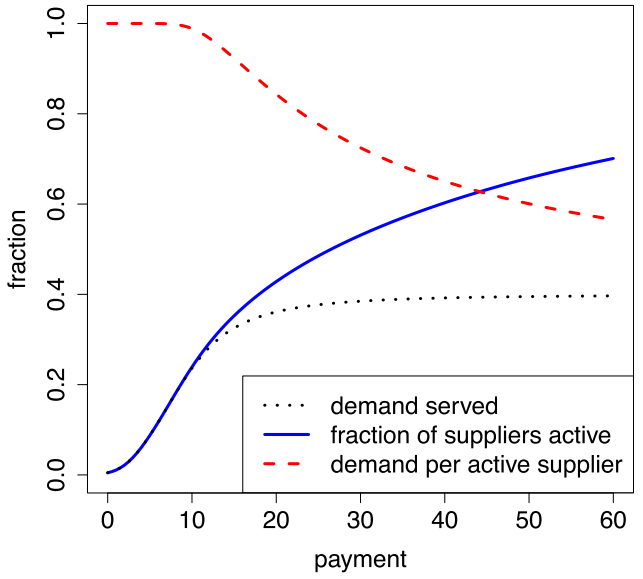
The form of $q(\cdot)$ depends on both the amount of available supply and demand and the efficiency with which supply can be matched with demand (see Section 3 for an example based on a queuing network).

Published in Management Science on February 16, 2021 as DOI: 10.1287/mnsc.2020.3844.
This article has not been copyedited or formatted. The final version may differ from this version.

Finally, the platform's utility U_t is given by the revenue from the demand served minus payments made to suppliers. Figure 2 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 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.

Conversely, our assumption that the platform cannot observe the daily state variable A_t is made to ensure that our learning problem is robust and performs well even if the state variables are unavailable or diff icult to estimate accurately. In practice, of course, it is plausible that a platform may have access to partial—but not full—information about A_t. Here, we focus on the statistically most difficult setting where the platform is oblivious to A_t and thus can only learn how to set p via experimentation, as this setting enables us to establish a crisp separation between different approaches to learning and to highlight our core methodological contributions. However, all methods considered here can be adapted to leverage partial information about A_t, and further work that investigates how best to leverage such information would be of considerable interest.

Figure 2. (Color online) Example of Large-Sample Behavior of Market, Conditionally on a Realization of A



Notes. 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, that is, with number of potential suppliers n growing to infinity such that $\mathbb{E}[D/n \mid A] = 0.4$. Individual supplier preferences are logistic (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 (5) with $L = 8$, visualized in Figure 5.

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 $t = 1, \ldots, T$, the platform chooses a small random fraction of suppliers and offers 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 marketwide 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 $t = 1, \ldots, T$, we randomly chose a payment p_t, made it available to all suppliers, and then observed realized profits U_t. We could then try to 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 $n \to \infty$); 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.[1] 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. Such widespread 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

Published in Management Science on February 16, 2021 as DOI: 10.1287/mnsc.2020.3844.
This article has not been copyedited or formatted. The final version may differ from this version.

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

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

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*, that is, 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).

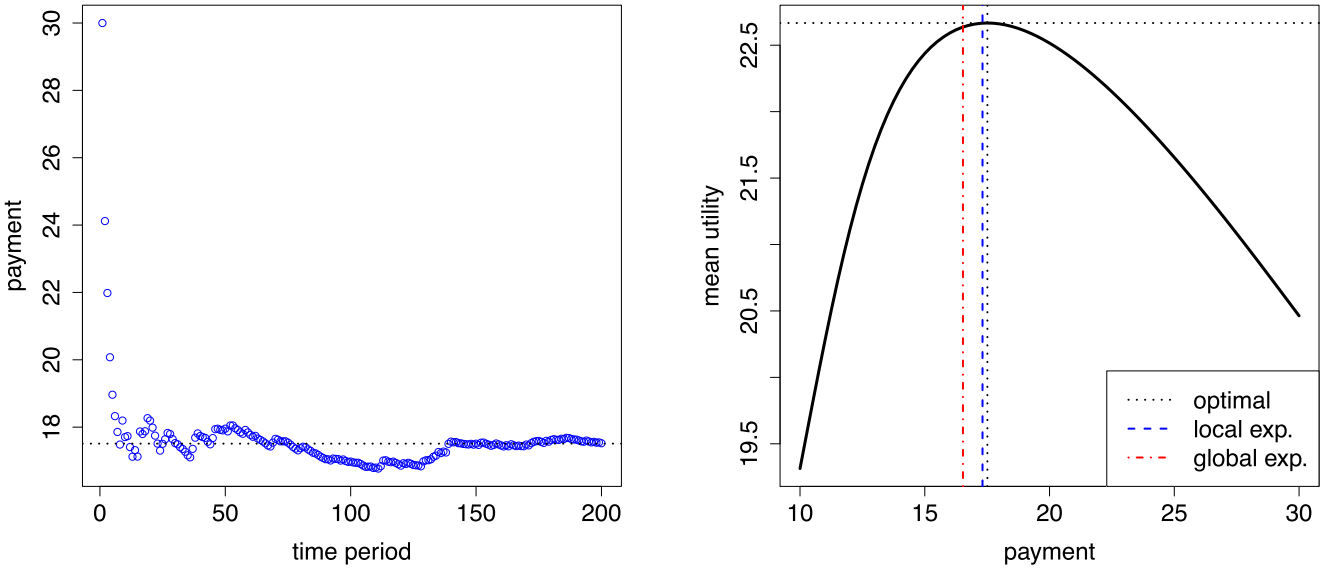
This marginal response function is not directly of interest for optimizing p, as it ignores feedback effects. However, we find that—in our setting—this quantity captures relevant information for optimizing payments. More specifically we show in Section 3.2 that, provided we have good enough understanding 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 (1) to be very small when n is large).

Figure 3 shows results from our local experimentation approach on a simple simulation experiment in the setting of Figure 2, where the scaled demand $\mathbb{E}[D/n \mid A]$ follows a beta(15, 35) distribution. We initialize the system at $p_1 = 30$ and then each day run payment perturbations as in (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.

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, measure the 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 4 shows the resulting (p_t, U_t) pairs, as well as the resulting $\hat{p}$. As seen in the right panel of Figure 3, 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

Figure 3. (Color online) Results from Learning p via Local Experimentation



Notes. The worker preference functions are as in Figure 2; the daily contexts are such that $\mathbb{E}[D/n \mid A] \sim \text{beta}(15, 35)$. The platform utility function is linear as in Lemma 3, with $\gamma = 100$. We learned gradients based on local randomization (1) with $\zeta = 0.5$ and then optimized payments via gradient descent as in (26) 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^*.

Published in Management Science on February 16, 2021 as DOI: 10.1287/mnsc.2020.3844.
This article has not been copyedited or formatted. The final version may differ from this version.

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 4, after the first few days, the global experimentation approach in fact systematically achieves lower daily utilities U_t than local experimentation. In Section 6 we consider further numerical comparisons of local and global experimentation, as well as variants of global exploration that balance exploration and exploitation to improve in-sample regret.

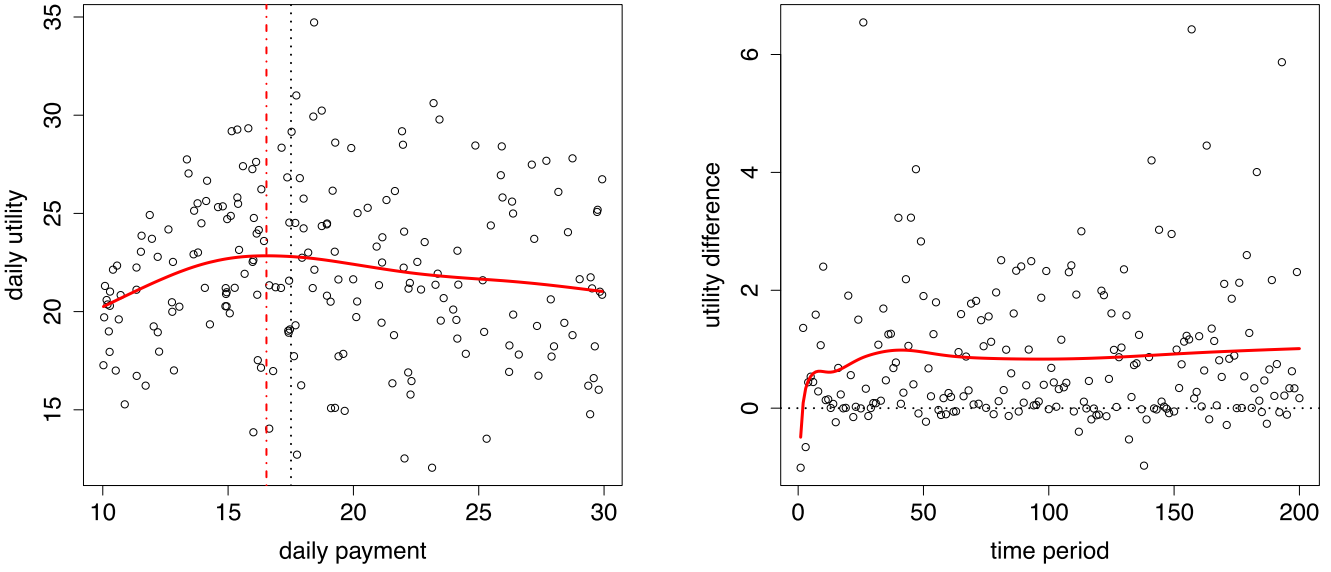
Remark 1 (Relationships to Batched Bandits)**.** Our model bears resemblance to batched multiarm bandits (Perchet et al. 2016, Esfandiari et al. 2019, Gao et al. 2019) and batched online optimization (Duchi et al. 2018, Bubeck et al. 2019), where an analyst sequentially picks multiple arms to pull for one batch at a time. In particular, administering an intervention to a unit in our model could be seen as analogous to pulling one arm in batched bandits or sampling an unknown function at a particular point in batched online optimization. There is, however, a fundamental distinction between our model and the predominant model for batched bandits. Existing work on batched bandits does not allow for interference within batches: The action assigned to one unit in a batch does not directly affect the outcome observed for another unit in the batch. In contrast, the presence of cross-unit interference within batches (or, for us, within days) is at the heart of our model: The outcome of a unit not only depends on their own intervention, but also on the interventions experienced by other units on the same day. Thus, existing results on batched bandits and online optimization cannot be used to reason about how best to deploy heterogeneous incentives to different suppliers in order to converge to a good choice of p in our setting.

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 nth 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), for example, $q^{(n)}$. 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 the distribution $D \sim F_a$. We further assume that the demand scales proportionally with respect to the market

Figure 4. (Color online) Results from Learning p via Global Experimentation



Notes. 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).

Published in Management Science on February 16, 2021 as DOI: 10.1287/mnsc.2020.3844.
This article has not been copyedited or formatted. The final version may differ from this version.

size n and that it concentrates after rescaling by $1/n$. In particular, we assume that there exists $\{d_a\}_{a\in\mathcal{A}} \subset \mathbb{R}_+$ such that, for all $a \in \mathcal{A}$, $\mathbb{E}[D/n \mid A = a] = d_a$ for all $n \in \mathbb{N}$,

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

and

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

and as $n \to \infty$. In general, we will use the subscript 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 $P_i \overset{\text{iid}}{\sim} \pi$ for $i = 1, 2, \ldots, n$. 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 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 ith 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 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

$$\Omega(d, t) \overset{\Delta}{=} \mathbb{E}[S_i \mid D = d, T = t], \tag{3}$$

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

1. $\omega(\cdot)$ is smooth, concave, and nondecreasing.
2. $\lim_{x\to 0} \omega(x) = 0$ and $\lim_{x\to\infty} \omega(x) \le 1$.
3. $\lim_{x\to 0} \omega'(x) \le 1$.

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 $\lim_{x\to\infty} \omega(x) \le 1$ asserts that the maximum capacity of all active suppliers be bounded after normalization.

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

1. *$\Omega(d, t)$ is nondecreasing in d and nonincreasing in t.*
2. *There exists a bounded error function $l : \mathbb{R}_+^2 \to \mathbb{R}_+$ with*

$$|l(d, t)| = o\left(1/\sqrt{t} + 1/\sqrt{d}\right), \tag{4}$$

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 5).

Example 6 (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, $L \in \mathbb{N}$, $L \ge 2$. 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.

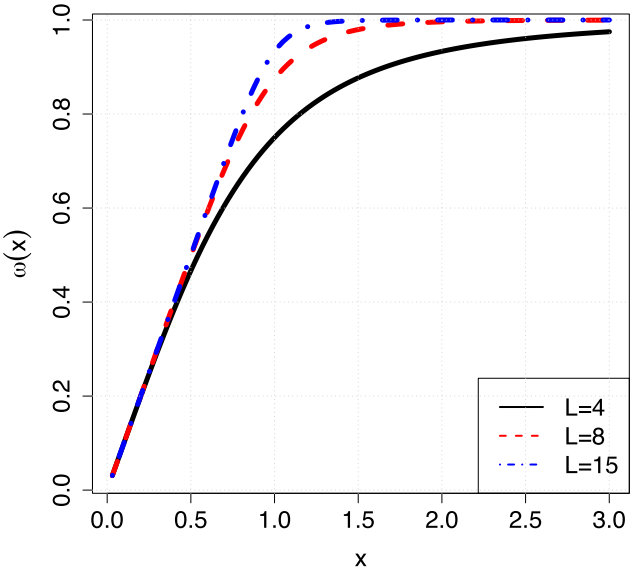
Within this model, each active supplier effectively functions as an $M/M/1$ queue with service rate 1 and arrival rate D/T. Due to the capacity limit at L, some requests may be dropped if they are assigned to a queue currently at capacity. Using the theory of $M/M/1$ queues, it is not difficult to show that (e.g., Spencer et al. 2014, equation 5.6) 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

$$\omega(x) = \begin{cases} \dfrac{x - x^L}{1 - x^L}, & x \neq 1, \\ 1 - \frac{1}{L}, & x = 1. \end{cases} \tag{5}$$

Numerical examples of $\omega(\cdot)$ are given in Figure 5. Note that $\omega(\cdot)$ satisfies all conditions in Definition 5 and is hence a regular allocation function. Finally, we may

Published in Management Science on February 16, 2021 as DOI: 10.1287/mnsc.2020.3844.
This article has not been copyedited or formatted. The final version may differ from this version.

Figure 5. (Color online) Examples of the Regular Allocation Function $\omega(\cdot)$ in Example 6 Under Different Values of Capacity L



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 to become active. In particular, the choice of supplier i becoming active is given as follows, where T is the equilibrium number of active suppliers:

$$\begin{aligned}\mu_a^{(n)}(\pi) &\triangleq \mathbb{P}_\pi[Z_i = 1 \mid A = a] \\ &= \mathbb{E}_\pi\big[\, f_{B_i}(P_i \mathbb{E}_\pi[\Omega(D,T) \mid A = a]) \,|\, A = a\big].\end{aligned} \tag{6}$$

Here, $\mathbb{E}_\pi[\Omega(D,T)\,|\,A=a]$ is the expected amount of demand served by each supplier given the platform's choice of π, and thus, $P_i\,\mathbb{E}_\pi[\Omega(D,T)\,|\,A=a]$ is the expected revenue of the ith supplier in equilibrium.[2] Note that the choice model (6) is stationary in that each supplier only considers the average behavior of other marketplace participants when choosing whether to enter. In particular, suppliers do not consider the effect of their own entry decision on the system or combinatorial interactions between other marketplace participants. Similar types of stationary assumptions, also known as mean-field or oblivious equilibrium, are common in game-theoretic models involving a large number of players where each player's influence on the overall system dynamic is vanishingly small (Hopenhayn 1992, Weintraub et al. 2008) and can be formally justified by showing how stationary equilibrium converges to the true Nash equilibrium in the limit as the system size tends to infinity (Adlakha et al. 2015).

Here, 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 from a set $\mathcal{B}$, and are independent and identically distributed (i.i.d.) with a distribution that 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 choice functions in the family $\{f_b(\cdot)\}_{b\in\mathcal{B}}$ satisfy certain regularity properties detailed below.

Assumption 2. *We assume that supplier choices are determined by the stationary choice model* (6)*. Furthermore, for all $b \in \mathcal{B}$, we assume that the choice function $f_b(\cdot)$ takes values in $[0,1]$, is monotonically nondecreasing and twice differentiable with a uniformly bounded second derivative.*

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

Example 7 (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:

$$\mathbb{P}[Z_i = 1 \,|\, P_i,\, \pi,\, A] = \frac{1}{1 + e^{-\alpha\left(P_i \mathbb{E}_\pi[\Omega(D,T)\,|\,A] - B_i\right)}}, \tag{7}$$

where $\alpha > 0$ is a parameter and the private feature B_i takes values in $\mathbb{R}_+$ and represents the breakeven 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 breakeven cost. The sensitivity of such dependence is modeled by the parameter α. Note that, in the limit as $\alpha \to \infty$, 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 breakeven threshold B_i.

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

$$U = R(D,\, T) - \sum_{i=1}^{n} P_i Z_i S_i, \tag{8}$$

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. Analogously 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.

Published in Management Science on February 16, 2021 as DOI: 10.1287/mnsc.2020.3844.
This article has not been copyedited or formatted. The final version may differ from this version.

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

$$R(d,t) = (r(d/t) - l(d,t))t, \quad \text{for all } t,d \in \mathbb{R}_+, \tag{9}$$

where $r : \mathbb{R}_+ \to \mathbb{R}_+$ *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 nth system as

$$u_a^{(n)}(\pi) = \frac{1}{n}\mathbb{E}_n[U \mid A = a], \quad \text{and} \quad u^{(n)}(\pi) = \mathbb{E}_n\left[u_A^{(n)}(\pi)\right]. \tag{10}$$

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

$$P_i = p + \zeta\varepsilon_i, \quad i \in \mathbb{N}. \tag{11}$$

In what follows, we will use $\pi_{p,\zeta}$ to denote the payment distribution when payments are ζ-perturbed from p and use $\mu_a^{(n)}(p,\zeta)$ to denote $\mu_a^{(n)}(\pi_{p,\zeta})$. The meanings of $\mu_a^{(n)}(p,\zeta)$, $u_a^{(n)}(p,\zeta)$, and so on 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 $\mu_a^{(n)}(\pi_{p,0})$.

Remark 2 (What Does the Platform Know?)**.** Our model assumes that the platform has detailed knowledge of the allocation mechanics, but cannot anticipate the behaviors of market participants that drive supply and demand. More specifically, we assume that the platform knows the regular allocation function ω (Definition 5), and its prelimit version, Ω, from (3); the limiting platform utility function r (Assumption 3), and its prelimit version, R, from (8); as well as the payment scheme it chooses to use, that is, p, ζ, and the realizations of the random perturbations, $\{\varepsilon_i\}_{i=1,\dots,n}$. However, the platform cannot anticipate the global state A, the demand D, or the distribution of supplier choice functions $f_{B_i}(\cdot)$; rather, all it can do is collect after-the-fact measurements of D and $\{Z_i\}_{i=1,\dots,n}$, the set of active suppliers. Finally, we implicitly assume that the global state A_t has no effect on the system beyond their assigned time period and that the platform knows this fact.

This modeling choice reflects an understanding that it is realistic for a platform to have a good handle on the mechanics of the marketplace it controls, but it is implausible for it to have an in-depth understanding of the beliefs and preferences of all marketplace participants. For example, in the case of ride sharing, it is plausible that a platform could get good at modeling congestion, but it is less plausible that the platform could fully understand and anticipate how all its drivers may respond to various policy changes.

The fact that we take the platform to be completely oblivious to the global state A_t puts us in an extreme setting, where the platform's learning must be purely driven by randomization in p. We chose this extreme setting primarily for two reasons. First, it crystallizes the difficulty of the learning problem and highlights the value of local experimentation relative to global experimentation baselines. Second, a platform's knowledge of A_t, if any, is likely to be noisy and inaccurate, and it is often difficult for a platform to learn efficiently in practice by matching historical data using noisy estimates of their corresponding contexts. Therefore, it is of considerable practical value to devise a robust learning algorithm that works well without relying on the platform's ability to infer the global state A_t.

In practice, of course, the platform may have some information about the global state A; for example, we may assume that the platform observes a set of covariates X that capture some aspects of A (e.g., we could have $X = \Xi(A)$ for some lossy function Ξ). In such a setting, the information X could be used for variance reduction and/or learning better policies that exploit heterogeneity explained by X. It would be of considerable interest to study a covariate-enriched variant of our approach that allows the platform to use such information to learn better policies; however, we leave this line of investigation to follow-up work.

3.1. Mean-Field Asymptotics

The stochastic model described above in general admits complex dynamics that are not amenable to exact analysis. Fortunately, we show in this subsection that, in the mean-field limit where the number of suppliers is large, various key equilibrium quantities converge to tractable objects described by a mean-field model. To start, we first provide a formal definition of the equilibrium active supply size, T, and verify existence and uniqueness.

Definition 8 (Active Supply Size in Equilibrium)**.** We say that a random variable T is an *equilibrium supply size* if,

Published in Management Science on February 16, 2021 as DOI: 10.1287/mnsc.2020.3844.
This article has not been copyedited or formatted. The final version may differ from this version.

when all suppliers make activation choices according to (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 we 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 (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:

$$q_a^{(n)}(\mu) = \mathbb{E}[\Omega(D, X)|A = a], \quad X \sim \text{Binomial}(n, \mu). \tag{12}$$

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

$$\lim_{n\to\infty} \mu_a^{(n)}(p) = \mu_a(p), \tag{13}$$

$$\lim_{n\to\infty} q_a^{(n)}(\mu) = \omega(d_a/\mu), \tag{14}$$

$$\lim_{n\to\infty} u_a^{(n)}(p) = u_a(p) = (r(d_a/\mu_a(p)) - p\omega(d_a/\mu_a(p)))\mu_a(p), \tag{15}$$

$$\lim_{n\to\infty} \left(q_a^{(n)}\right)'(\mu) = -\omega'(d_a/\mu)\frac{d_a}{\mu^2}, \tag{16}$$

where $\omega(\cdot)$ and $r(\cdot)$ are described in Definition 5 and Assumption 3, respectively. In (13), the limit $\mu_a(p)$ is the only solution to $\mu = \mathbb{E}[f_{B_1}(p\omega(d_a/\mu))\,|\,A = a]$.

Finally, the following result, proven in Appendix B.1, 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: $f_a(x) = \mathbb{E}[f_{B_1}(x)\,|\,A = a]$. 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 $\overline{x} = \sup_{p\in(c_0,\gamma)} pq_a(\mu_a(p))$. The average choice function $f_a(\cdot)$ satisfies:*
   a. *$f_a(\cdot)$ is strongly concave in the domain $(\underline{x}, \overline{x})$;*
   b. *$f_a(\underline{x}) - f_a'(\underline{x})\underline{x} \geq 0$, or equivalently, that there exists a differentiable, nonnegative 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 9 (Marginal Response Function)**.** Fix $n \in \mathbb{N}$, $a \in \mathcal{A}$, and $p > 0$. The marginal response function is defined by

$$\Delta_a^{(n)}(p) = q_a^{(n)}\left(\mu_a^{(n)}\right)(p)\,\mathbb{E}\left[f_{B_1}'\left(pq_a^{(n)}\left(\mu_a^{(n)}(p)\right)\right)\,|\,A = a\right]. \tag{17}$$

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 $n \to \infty$, 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 (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*

$$\frac{d}{dp}\mu_a^{(n)}(p) = \frac{\Delta_a^{(n)}(p)}{1 - p\Delta_a^{(n)}(p)\,q_a^{(n)'}\left(\mu_a^{(n)}(p)\right)/q_a^{(n)}\left(\mu_a^{(n)}(p)\right)} \quad \text{for any } n \geq 1. \tag{18}$$

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

$$\lim_{n\to\infty} \Delta_a^{(n)}(p) = \Delta_a(p) \overset{\Delta}{=} \omega(d_a/\mu_a(p))\mathbb{E}\left[f_{B_1}'(p\omega(d_a/\mu_a(p)))\,|\,A = a\right], \tag{19}$$

$$\lim_{n\to\infty} \frac{d}{dp}\mu_a^{(n)}(p) = \mu_a'(p) = \Delta_a(p)\left/\left(1 + \frac{pd_a\Delta_a(p)\omega'(d_a/\mu_a(p))}{\mu_a(p)^2\omega(d_a/\mu_a(p))}\right)\right. . \tag{20}$$

Published in Management Science on February 16, 2021 as DOI: 10.1287/mnsc.2020.3844.
This article has not been copyedited or formatted. The final version may differ from this version.

In addition to powering our approach to experimentation, the result of Lemma 4 also provides qualitative insights about the drivers of interference in our model. If there were no interference among the suppliers, then the gradient $(d/dp)\mu_a(p)$ would have coincided with the marginal response $\Delta_a(p)$; but due to interference, the gradient is attenuated by an interference factor $1 + R_a(p)$, where

$$R_a(p) = \Sigma_a^{\Delta}(p)\Sigma_a^{\Omega}(p), \quad \underbrace{\Sigma_a^{\Delta}(p) = \frac{p\Delta_a(p)}{\mu_a(p)}}_{\text{scaled marginal sensitivity}},$$
$$\underbrace{\Sigma_a^{\Omega}(p) = \frac{d_a}{\mu_a(p)}\frac{\omega'\left(d_a/\mu_a(p)\right)}{\omega\left(d_a/\mu_a(p)\right)}}_{\text{scaled matching elasticity}}. \tag{21}$$

We thus observe the following:

- The interference factor is negligible when the "scaled marginal sensitivity" $\Sigma_a^{\Delta}(p)$ is small, that is, the marginal response function is small relative to the current supply $\mu_a(p)$. Note that $p\Delta_a(p)$ is a scale-free version of our marginal response function that is invariant to rescaling p.
- The interference factor is negligible when the "scaled matching elasticity" $\Sigma_a^{\Omega}(p)$ is small, that is, the elasticity of the matching function $\omega(\cdot)$ is small relative to the current ratio of supply to demand $\mu_a(p)/d_a$. In particular, because $\omega(\cdot)$ is concave and bounded by assumption, we can verify that $\Sigma_a^{\Omega}(p)$ is small whenever demand far exceeds supply, that is, $d_a/\mu_a(p) \gg 1$ (see Proposition 5 stated below and proven in Appendix B.2).
- The interference factor is nonnegligible when neither of the above conditions hold.

These observations are aligned with what one might have anticipated based on qualitative arguments. For example, interference effects clearly cannot matter if marketplace participants are overall unresponsive to changes in p, and this is exactly what we found in the first bullet point. Meanwhile, one might have expected for the effect of interference to be more pronounced when there is more intense competition among the suppliers than when there is enough demand to keep all suppliers busy, and this conjecture is well in line with our finding in the second bullet point.

Proposition 5. *Let $g : \mathbb{R}_+ \to \mathbb{R}_+$ be concave with piecewise continuous derivative g'. Then $xg'(x) \le g(x)$ for all $x > 0$. If moreover $0 < \lim_{x\to\infty} g(x) < \infty$, then $\lim_{x\to\infty} xg'(x)/g(x) = 0$.*

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. First, we show in Section 4.1 that our local experimentation approach enables us to construct an asymptotically accurate estimate of the utility gradient at a given payment p in the mean-field limit as $n \to \infty$. Then, we use these gradient estimates to update the payment using a form of gradient ascent and show that their performance is superior to what can be achieved via classical continuous-armed bandit and zeroth-order optimization algorithms. Specifically, we establish in Section 4.2 an $O(1/T)$ upper bound for the rate of convergence to the optimal platform utility under our algorithm. In Section 4.3, we study the cost of the local experimentation needed to estimate utility gradients, and we verify that it scales sublinearly in n. Finally in Section 4.4, we compare our results to those available to classical continuous-armed bandits and show that it is not possible to achieve the $O(1/T)$ convergence rate within the classical bandit framework. Throughout this section, we focus on optimizing utility in the mean-field limit, while verifying that finite-n errors have an asymptotically vanishing effect on learning.

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 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 $\overline{Z} = \frac{1}{n}\sum_{i=1}^n Z_i$ and $\overline{D} = D/n$. As discussed in Section 2 our proposal starts by perturbing individual payments as in (11) and then estimating the regression coefficient $\widehat{\Delta}$ of market participation Z_i on the perturbation $\zeta_n\varepsilon_i$, that is,

$$\widehat{\Delta} = \zeta_n^{-1}\sum_{i=1}^n (Z_i - \overline{Z})(\varepsilon_i - \bar{\varepsilon}) \Big/ \sum_{i=1}^n .(\varepsilon_i - \bar{\varepsilon})^2. \tag{22}$$

Our first result below relates this quantity $\widehat{\Delta}$ that 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 6. *Suppose the conditions of Lemma 1 hold. Let*

$$\widehat{\Upsilon} = \widehat{\Delta}\Big/\left(1 + \frac{p\overline{D}\widehat{\Delta}\omega'\left(\overline{D}/\overline{Z}\right)}{\overline{Z}^2\omega\left(\overline{D}/\overline{Z}\right)}\right), \tag{23}$$

Published in Management Science on February 16, 2021 as DOI: 10.1287/mnsc.2020.3844.
This article has not been copyedited or formatted. The final version may differ from this version.

and let

$$\widehat{\Gamma} = \widehat{\Upsilon}\Big[r\left(\overline{D}/\overline{Z}\right) - p\omega\left(\overline{D}/\overline{Z}\right) - \left(r'\left(\overline{D}/\overline{Z}\right) - p\omega'\left(\overline{D}/\overline{Z}\right)\right)\overline{D}/\overline{Z}\Big] - \omega\left(\overline{D}/\overline{Z}\right)\overline{Z}. \tag{24}$$

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

$$\lim_{n\to\infty} \mathbb{P}\left[\left|\widehat{\Gamma} - \frac{d}{dp} u_A(p)\right| > \varepsilon\right] = 0, \tag{25}$$

for any $\varepsilon > 0$.

Remark 3 (Population-Wide Experimentation and Symmetric Perturbation)**.** It is instructive to note that our experimentation scheme in (11) has two distinguishing features that depart from a conventional approach to A/B testing that would give a small subset of suppliers an ε increase in payment while keeping payments in the rest of the population unchanged. First, our perturbation is symmetric across the units (zero-mean perturbation), whereas in classical settings those in a treatment group may receive asymmetric, and possibly identical, treatments. Second, we experiment across the entire population as opposed to a small subpopulation.

These features are in fact deliberate and interdependent, and the rationales are as follows. The perturbations being symmetric ensures that our experimentation scheme does not meaningfully shift the overall supply-demand equilibrium ($\mu_a^{(n)}(p)$), which in turn allows us to circumvent the impact of cross-unit interference. Moreover, as we show in Section 4.3, the symmetric perturbations lead to a small cost of experimentation: Roughly speaking, the effect of paying half of the population ε more is roughly neutralized by simultaneously paying the other half ε less. Meanwhile, the fact that we experiment on the whole population enables us to attain reasonable power using small enough perturbations ε such as not to be biased by the curvature of the supplier-specific choice functions $f_b(\cdot)$.

If perturbing the price carries a large fixed cost and the analyst wishes to apply local experimentation only to a small set of the population, one could also consider letting the random shocks ϵ_i be equal to 0 with nonzero probability, and restrict the rest of the analysis on the subpopulation who has received a nonzero shock. This would also amount to a valid local experimentation scheme. However, we note that the power of our approach to estimate the marginal response function depends on $\mathrm{Var}[\epsilon_i]$, and so, in order the maintain a given level of power (i.e., to ensure that $\mathrm{Var}[\epsilon_i] = 1$), the platform would need to use larger shocks ε_i for those suppliers that receive nonzero shocks. This in turn may expose the analyst to larger approximation errors from linearly approximating a curved function.

4.2. A First-Order Algorithm

Our key use of Theorem 6 involves optimizing for a utility-maximizing p. At every time period t, $\widehat{\Gamma}_t$ 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 $t = 1, 2, \ldots$, we do the following:

1. Deploy randomized payment perturbations (11) around p_t to estimate $\widehat{\Gamma}_t$ as in (24).
2. Perform a gradient update[3]

$$p_{t+1} = \arg\min_p \left\{\frac{1}{2\eta}\sum_{s=1}^{t} s(p - p_s)^2 - \theta_t p : p \in I\right\}, \qquad \theta_t = \sum_{s=1}^{t} s\widehat{\Gamma}_s. \tag{26}$$

The following result shows that if we run our method for T time periods in a large marketplace 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 $1/t$.[4]

Theorem 7. *Under the conditions of Theorem 6, 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* (26) *with step size* $\eta > \sigma^{-1}$ *and that the gradients of u are bounded, that is,* $|u_a'(p)| < M$ *for all* $p \in I$ *and* $a \in \mathcal{A}$*. Then*

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{1}{T}\sum_{t=1}^{T} t\left(u_{A_t}(p) - u_{A_t}(p_t)\right) \le \frac{\eta M^2}{2}\right] = 1, \tag{27}$$

for any $p \in I$ *and* $T \ge 1$.

The above result does not make any distributional assumptions on the contexts A_t; rather, (27) bounds the regret of our payment sequence $p_1, p_2, \ldots$ 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's are independent and identically distributed, Theorem 7 also implies that an appropriate average of our learned

Published in Management Science on February 16, 2021 as DOI: 10.1287/mnsc.2020.3844.
This article has not been copyedited or formatted. The final version may differ from this version.

payments is consistent for the optimal payment via online-to-batch conversion (Cesa-Bianchi et al. 2004).

Corollary 8. *Under the conditions of Theorem 7, suppose moreover that the A_t's are independent and identically distributed and let $u(p) = \mathbb{E}[u_{A_t}(p)]$. Then, for any $\delta > 0$,*

$$\limsup_{n\to\infty} \mathbb{P}\left[(p^* - \bar{p}_T)^2 \le \frac{\eta M^2}{\sigma T}\left(16\log\left(\delta^{-1}\right) + 4\right)\right] \ge 1 - \delta, \tag{28}$$

where $p^* = \arg\max\{u(p) : p \in I\}$ 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;[5] the result below, however, shows that this excess cost decays quadratically in the magnitude of payment perturbations ζ.

Theorem 9. *Under the conditions of Theorem 6 there are constants C, $\alpha > 0$ such that*

$$\frac{1}{T}\sum_{t=1}^{T}\left(u_{A_t}(p_t) - u_{A_t}(p_t, \zeta)\right) \le C\zeta^2 \quad \text{for all} \quad 0 \le \zeta < \alpha. \tag{29}$$

Recall that, as the market size gets large, Theorem 6 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 9, 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 Rates for Global Experimentation

As discussed above, our local experimentation approach makes two departures from the classical literature on experimental design under interference, including Sobel (2006), Hudgens and Halloran (2008), Manski (2013), Tchetgen Tchetgen and VanderWeele (2012), Aronow and Samii (2017), Eckles et al. (2017), Athey et al. (2018), Baird et al. (2018), Basse et al. (2019), and Leung (2020). First we use mean-field equilibrium modeling to capture and correct for interference effects; second, we operationalize our approach in a dynamic setting where a decision maker wants to tune a decision variable while controlling realized regret while learning.

To highlight the value of mean-field equilibrium modeling, we compare our result from Theorem 7 to what can be achieved via the global experimentation baseline that is tailored to sequential decision making, but does not use equilibrium modeling: Each day $t = 1, \ldots, T$, global experimentation chooses a payment p_t given to all workers on that day and then observes the corresponding reward U_t. Analogously to the random saturation design discussed in Baird et al. (2018) and Hudgens and Halloran (2008), global experimentation does not suffer any bias due to interference because there is no cross-day interference in our model. The downside of global experimentation is that, unlike our equilibrium-modeling-based approach, it does not provide the analyst any direct information about gradients $u_{A_t}'(p_t)$, and this severely limits the ability of global experimentation to effectively discover a good choice of p.

To understand the limits of global experimentation we turn to the literature on continuous-armed bandits (or zeroth-order optimization), which has established strong lower bounds for closely related problems. Shamir (2013) considers the following setting: We have a sequential decision-making problem where, in each time period, the analyst gets to choose p_t from a bounded interval I and observes a reward U_t with $\mathbb{E}[U_t \mid p_t] = u(p_t)$ and $\mathrm{Var}[U_t \mid p_t] = 1$; the goal is to choose a sequence p_t that makes the regret $\sum_{t=1}^{T}(u(p^*) - u(p_t))$ small, where p^* is the maximizer of $u(\cdot)$ over the interval I. Shamir (2013) then shows that, even if $u(\cdot)$ is strongly concave, no algorithm can achieve expected regret that grows slower than $\sqrt{T}$; and, in fact, this result holds even if $u(\cdot)$ is known a priori to be a quadratic with unit curvature. Further results in this line of work are given in Bubeck et al. (2017). We also note closely related results by Keskin and Zeevi (2014), who establish a $\sqrt{T}$ lower bound on regret for pricing under a linear demand model (note that, with linear demand, the seller's profit is quadratic), and by Nambiar et al. (2019), who propose a global experimentation scheme driven by random perturbations to p_t that could be used to achieve $\sqrt{T}$ regret in our model.

The upshot is that, when the daily reward functions $u_{A_t}(p_t)$ are strongly concave and there is meaningful cross-day noise due to A_t, our approach can achieve cumulative regret on the order of $\log(T)$ (corresponding to a $1/t$ rate of decay in errors), whereas global experimentation cannot improve over $\sqrt{T}$

Published in Management Science on February 16, 2021 as DOI: 10.1287/mnsc.2020.3844.
This article has not been copyedited or formatted. The final version may differ from this version.

regret (corresponding to a $1/\sqrt{t}$ rate of decay in errors). 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.

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 cancels 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, for example, 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 and Limitations

So far, we have focused our discussion on a specific model of a centralized market for freelance labor; but, as outlined in the introduction, we expect the general principles outlined here to be more broadly applicable. A full theory of experimental design powered by mean-field equilibria is beyond the scope of this paper. In this section, however, we take a first step toward a more general theory by presenting two problem settings of considerable practical interest that are amenable to our approach, risk-averse suppliers and surge pricing, and discuss another problem, immunization via vaccines, that does not appear to be amenable to it.

5.1. Equilibrium Modeling via Generalized Earning Functions

In our motivating model for freelance labor, we considered a setting where the platform first chooses a distribution π, 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 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.

Here, we briefly discuss how to extend our approach to allow for risk-averse suppliers and surge pricing. In order to do so, we first define choice models for both problems below and write down balance conditions generalizing (6). Afterward, 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. As discussed further below, what enables us to extend our discussion to these problems is that, in both cases, we can explain the choices of suppliers in terms of a unifying formalism we refer to as generalized earning functions.

Example 10 (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 β(revenue) (Pratt 1964, Holt and Laury 2002). Suppose that $\beta(0) = 0$ and that each worker can serve 0 or 1 units of demand.[6] Then our balance condition (6) becomes

$$\begin{aligned}\mu_a^{(n)}(\pi) &= \mathbb{P}_\pi[Z_i = 1 \mid A = a] \\ &= \mathbb{E}_\pi\left[f_{B_i}\left(\beta(P_i) q_a^{(n)}\left(\mu_a^{(n)}(\pi)\right)\right) \middle| A = a\right].\end{aligned} \tag{30}$$

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 11 (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 (Hall et al. 2015, Cachon et al. 2017). As a simple model, suppose that surge is triggered automatically based on the supply-demand ratio, that is, there is a function $s : \mathbb{R}_+ \to \mathbb{R}_+$ such that, in each period, the ith 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

$$\begin{aligned}\mu_a^{(n)}(\pi) &= \mathbb{P}_\pi[Z_i = 1 \mid A = a] \\ &= \mathbb{E}_\pi\left[f_{B_i}\left(s\left(\frac{d_a}{\mu_a^{(n)}(\pi)}\right) P_i\, q_a^{(n)}\left(\mu_a^{(n)}(\pi)\right)\right) \middle| A = a\right],\end{aligned} \tag{31}$$

Published in Management Science on February 16, 2021 as DOI: 10.1287/mnsc.2020.3844.
This article has not been copyedited or formatted. The final version may differ from this version.

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 (30) and (31) but without the n-superscripts. In this case, we can write both mean-field limits in a unified form via *generalized earning functions*, $\theta : \mathbb{R}_+^2 \to \mathbb{R}_+$, so that the asymptotic balance condition is

$$\begin{aligned}\mu_a(\pi) &= \mathbb{P}_\pi[Z_i = 1 \mid A = a] \\ &= \mathbb{E}_\pi\left[f_{B_i}\left(\theta\left(P_i, q_a\left(\mu_a(\pi)\right)\right)\right) \mid A = a\right].\end{aligned} \tag{32}$$

In the case of (30), we have $\theta_{risk}(p, q) = \beta(p)q$. Meanwhile, for (31), 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 (31) converges to (32) 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, that is, by showing how (32) lets us map from a marginal response function to utility gradients with respect to surge; we leave verification of the conjectured convergence to (32) 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 $\theta(\cdot)$ affect the computation of utility derivative $u_a'(p)$ only through the expression for $\mu_a'(p)$ (cf. the proof of Proposition 12). Hence, we here only focus on expressions for $\mu_a'(p)$.

Now, we can directly check that a reduced form expression as in (1) allows us to estimate the following marginal response function via local randomization:

$$\Delta_a(p) = (\nabla\theta)_1\left(p, q_a\left(\mu_a(p)\right)\right)\mathbb{E}\left[f_{B_1}'\left(\theta\left(p, q_a\left(\mu_a(p)\right)\right)\right)\middle| A = a\right]. \tag{33}$$

where $(\nabla\theta)_i(\cdot,\cdot)$ denotes the ith 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

$$\begin{aligned}\mu_a'(p) = \Delta_a(p) \Bigg/ \Bigg(1 + \frac{(\nabla\theta)_2\left(p, q_a\left(\mu_a(p)\right)\right)}{(\nabla\theta)_1\left(p, q_a\left(\mu_a(p)\right)\right)} \\ \times \omega'\left(\frac{d_a}{\mu_a(p)}\right)\frac{d_a}{\mu_a^2(p)}\,\Delta_a(p)\Bigg).\end{aligned} \tag{34}$$

Note, furthermore, that all quantities in (34) 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 (32) enables us to map an easy-to-estimate marginal response function to $\mu_a'(p)$ via (34). 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.

5.2. Interference and Choice Modeling

Although our approach to interference via equilibrium modeling provides useful insights in many problems of interest, it does not unlock all problems where we want to understand the effects of deploying an intervention at scale in a large system. One prominent example to which our approach does not (at least obviously) apply pertains to the study of vaccine effectiveness in the presence of herd immunity (Hudgens and Halloran 2008, Ogburn and VanderWeele 2017).

Example 12 (Vaccine Effectiveness)**.** We are considering whether to enact a policy that would increase vaccination rates against a contagious disease in a population where only a moderate fraction of people are currently vaccinated. Due to the interaction among people within the same geographical vicinity, the risk of infection for any given individual not only depends on whether they are vaccinated themselves, but also on the overall fraction of infected individuals in the ambient population (which in turn is modulated by the overall fraction of vaccinated individuals). Thus, simple randomized controlled trials cannot be used to consistently estimate the effect of policies that increase the overall rate of vaccination, and instead, methods that explicitly account for interference are required.

The classical way to think about experiments for community-level vaccine immunity is to randomize the fraction of people vaccinated across different disjoint (and thus noninterfering) communities (Hudgens and Halloran 2008, Baird et al. 2018). This approach is directly analogous to the global experimentation baseline considered throughout this paper and naturally leads to the question of whether our approach could be used to design more powerful alternatives.

In analogy to notation used in the rest of the paper, index communities by t and people within communities by i, and let $Z_{it} \in \{0, 1\}$ denote whether the ith person in the tth community gets infected. We write $\mu(p)$ for the expected fraction of people who get infected in a community in which a fraction p of people are vaccinated, and we focus on estimating $d\mu(p)/dp$, that is, the decrease in the overall infection rate that can be achieved by increasing the vaccination probability. In this context, global experimentation seeks to estimate $\mu(p)$ by randomly assigning a single vaccination probability $p_t \in [0, 1]$ to each community, so that each person in community t gets (randomly) vaccinated with probability p_t. In contrast, a local experimentation might consider using individualized randomization probabilities $p_{it} \in [0, 1]$ to get a better handle on $d\mu(p)/dp$.

At first glance, the problem may not appear so different from our leading example. It seems plausible that the above model sketch could be formalized in a way that makes it amenable to mean-field asymptotics.

Published in Management Science on February 16, 2021 as DOI: 10.1287/mnsc.2020.3844.
This article has not been copyedited or formatted. The final version may differ from this version.

Furthermore, in this setting, using symmetric perturbations $p_{it} = p_t \pm \zeta\varepsilon_{it}$ and regressing Z_{it} on ε_{it} should recover a well-defined marginal response function $\Delta(p)$ that, under regularity conditions, corresponds exactly to what is called the (average) direct effect in the statistics literature (Hudgens and Halloran 2008, Sävje et al. 2021).

At this point, however, we appear to get stuck. Unlike in the main examples considered in this paper, there does not seem to be a natural way to map from $\Delta(p)$ to our main quantity of interest, namely, $d\mu(p)/dp$. In the case of modeling freelance labor, we assumed that suppliers only care about expected revenue; thus, once $\Delta(p)$ gave us a handle on how they react to changes in expected revenue due to the platform directly changing p_{it}, we were also able to reason about how they might react to changes in expected revenue due to changes in marketplace conditions that arise from general equilibrium effects. In the case of vaccine effectiveness, however, there is no a priori obvious way to connect the direct effect of vaccinating a specific person to how the same person will react to a change in the overall fraction of the population that is infected. For example, there is presumably a positive association between how much different people benefit from the vaccine directly and how much they benefit from it via herd immunity; however, some people may not be responsive to the vaccine and so have zero direct effect, but will still benefit indirectly from the vaccine via herd immunity. Thus, there appears to be no way to credibly learn about vaccine effectiveness without considering exogenous variation in the fraction of the population that is infected.

An interesting conceptual distinction between all the positive examples presented in this paper and the above vaccination example is that, in the former, interference effects are fundamentally due to choices made by participants in the system. For example, in the case of our model for freelance labor, interference effects arise because suppliers choose not to participate in marketplaces that are too congested. In contrast, in the vaccination example, getting sick is not a choice; it is simply a random event whose probability can be modulated up or down via different vaccination policies and community-level infection levels. The fact that joining a marketplace is a choice, whereas getting sick is not may not matter from the point of view of mean-field asymptotics; however, making assumptions about how suppliers make choices is what lets us credibly connect $\Delta(p)$ with $d\mu(p)/dp$ and proceed with our approach. The role of choice versus pure chance in understanding best practices for statistical estimation has been the topic of a longstanding discussion at the intersection of economics and statistics (e.g., Roy 1951, Heckman 2001, Imbens 2014); and, in this context, our result can be seen as one example where simple choice modeling helps motivate a powerful approach to statistical inference and learning.

6. Simulation Results

We now consider a more comprehensive empirical evaluation of the performance of local versus global experimentation, building on 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 $1 \le t \le T$ time periods, fit a spline to the data (as in the left panel of Figure 4), and then deploy the learned policy for the remaining $200 - T$ time periods. We consider the choices $T \in \{40, 60, 80, \ldots, 200\}$. For both methods, we report both in-sample regret, that is, the mean utility shortfall relative to deploying the population-optimal p^* for the T learning periods, as well as future expected regret, that is, the expected utility shortfall from deploying the learned policy $\hat{p}$ after the T learning periods. For local experimentation, we set $\hat{p} = 2\sum_{t=1}^{t} t\, p_t / (T(T+1))$ following Corollary 8, 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 6, 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, $T = 80$).

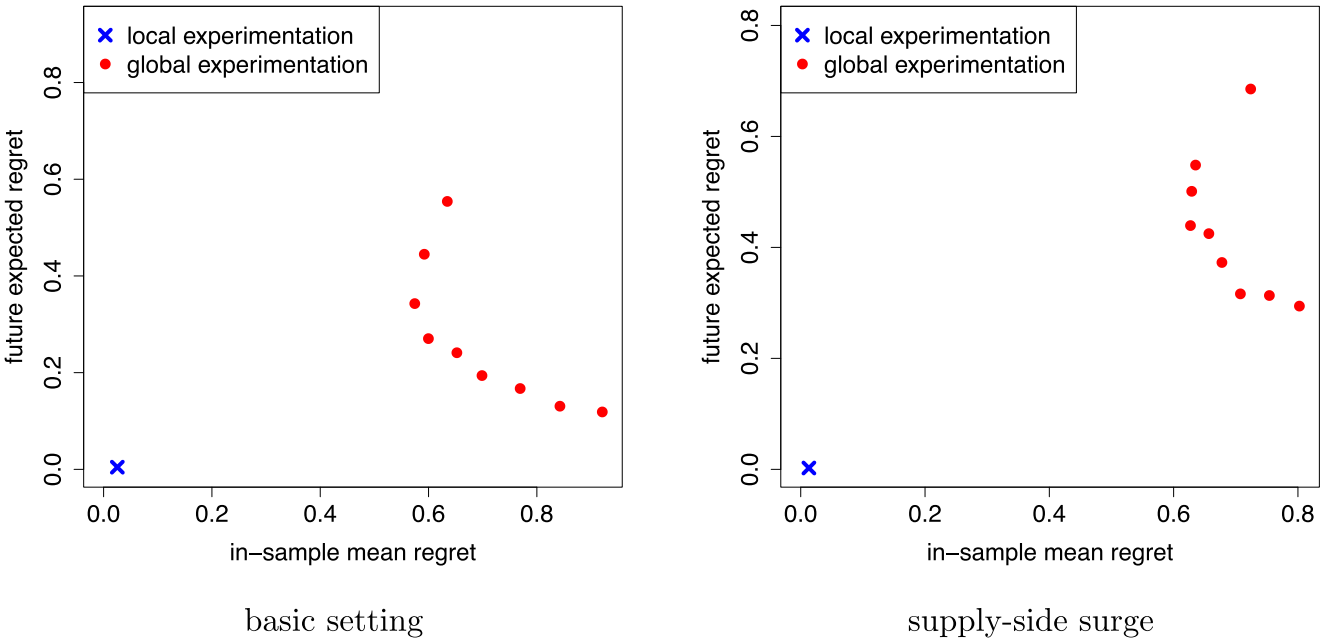
Next, we consider an analogous simulation design, but with supply-side surge pricing. As discussed in Section 5, we assume that the platform makes a public commitment to mechanistically increase supply-side payments by a multiplicative factor $s(D/T)$ once the demand D and supply T are realized, and suppliers take this commitment into account when choosing whether to join the marketplace. Here, we use

$$s(D/T) = \frac{D}{T} \Big/ \omega\left(\frac{D}{T}\right), \tag{35}$$

meaning that, by the properties of $\omega(\cdot)$ as outlined in Definition 5, 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.

Published in Management Science on February 16, 2021 as DOI: 10.1287/mnsc.2020.3844.
This article has not been copyedited or formatted. The final version may differ from this version.

Figure 6. (Color online) Comparison of the Regret of Local and Global Experimentation in the Setting of Section 2, Averaged Across 1,000 Simulation Replications



Notes. 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 (35).

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,

$$u_a(p) = \left(\gamma - ps\left(\frac{d_a}{\mu_a(p)}\right)\right)\omega\left(\frac{d_a}{\mu_a(p)}\right)\mu_a(p). \tag{36}$$

As discussed above, we can estimate the p-derivative of the expected scaled active supply size, $\mu_a'(p)$, by local experimentation via (33) and (34). Moreover, following the argument of Theorem 6, we obtain p-derivatives of $u_a(p)$ via

$$\begin{aligned} u_a(p) = &\left(\gamma - ps\left(\frac{d_a}{\mu_a(p)}\right)\right)\left(-\omega'\left(\frac{d_a}{\mu_a(p)}\right)\frac{d_a}{\mu_a(p)}\right. \\ &\left.+\omega\left(\frac{d_a}{\mu_a(p)}\right)\right)\mu_a'(p) - s\left(\frac{d_a}{\mu_a(p)}\right)\omega\left(\frac{d_a}{\mu_a(p)}\right)\mu_a(p) \\ &+ ps'\left(\frac{d_a}{\mu_a(p)}\right)\frac{d_a}{\mu_a(p)}\omega\left(\frac{d_a}{\mu_a(p)}\right)\mu_a'(p). \end{aligned} \tag{37}$$

We turn this into a feasible estimator by 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 6. 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 (35) decreased the optimal base payment from 17.6 to 15.7, while increasing the 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 (35) relative to not using surge—whereas the regret of local experimentation is less than the effect of adopting surge.

Finally, we note that the global experimentation baseline considered here—namely, our two-phase algorithm that starts with pure exploration and then moves to pure exploitation—is fairly simple, and it is possible that a more sophisticated global experimentation baseline could somewhat improve performance. However, one can check that, under reasonable conditions and provided we explore for the first $\sqrt{T}$ periods, our implemented baseline attains the optimal $\sqrt{T}$ regret rate of Shamir (2013) discussed in Section 4.4. Thus, more sophisticated methods like Bayesian zeroth-order optimization[7] as considered in, for example, Letham et al. (2019) may improve on

Published in Management Science on February 16, 2021 as DOI: 10.1287/mnsc.2020.3844.
This article has not been copyedited or formatted. The final version may differ from this version.

finite sample performance but cannot improve on the overall regret rate of our baselines.[8]

7. Discussion

We introduced a new framework for experimental design in stochastic systems with significant cross-unit interference. The key insight is that, in certain families of models, 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 and equilibrium modeling. 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.

There are some simplifying assumptions we make in this work that can be relaxed or verified in future research. For instance, we have assumed that the demand is exogenous. We expect that an extension of our method can be used to capture scenarios where the demand may, for instance, depend on the supply level. For example, a passenger may be less likely to hail a ride if they know there would be a long wait. Another assumption we made is that the market equilibrium can be reached relatively quickly. While there is recent empirical evidence suggesting that drivers in a ride-sharing platform do respond to payment changes in a manner that takes into account the resulting market equilibrium (Hall et al. 2020), it would be interesting to consider a more realistic model where prices may be updated continuously before a new market equilibrium is fully reached. It is less clear how the current model would apply in this setting, which is likely to require a substantially more sophisticated analysis.

We believe that the general approach proposed in this paper, one that leverages mean-field modeling 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 multidimensional. This could occur in a marketplace which, instead of being fully centralized, consists of a small number of interconnected submarkets. For instance, in a ride-sharing platform, the submarkets 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 submarkets. 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 submarkets, while the total number of submarkets 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 submarkets 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 analyze the effects of local experimentation in these systems. Finally, it would be interesting to investigate whether the local experimentation scheme proposed in this paper can be generalized to estimate higher-order derivatives of the utility function.

Endnotes

[1] Of course, the platform may try to correct for contexts, for example, 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. As discussed above, our goal in this paper is to develop methods for learning that are driven purely by experimentation and that do not rely on the platform being able to accurately observe A_t.

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

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

[4] In (27), 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 $T^{-1}\sum_{t=1}^{T}\left(u_{A_t}(p) - u_{A_t}(p_t)\right) = \mathcal{O}_P(\log(T))$.

[5] This 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 give to suppliers without increasing the amount of demand the platform is able to serve.

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

[7] One potentially promising approach would be to use local experimentation to get gradient estimates $u'_{A_t}(p_t)$ and then incorporate these estimates into a Bayesian learning framework. It is plausible that this could yield practically meaningfully improvements over the first-order approach considered in this paper.

[8] Another class of popular continuous-armed bandit algorithms were introduced by Flaxman et al. (2005) and Kleinberg (2005). These methods estimate derivatives by noisy function evaluations and then use these for gradient descent. However, while desirable due to their transparency and ease, these methods suffer cumulative regret on the order of $T^{3/4}$ in our setting. In our simulations, this class of methods performed worse than the global experimentation baseline we report results for.

References

Abadie A, Diamond A, Hainmueller J (2010) Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *J. Amer. Statist. Assoc.* 105(490):493–505.

Adlakha S, Johari R, Weintraub GY (2015) Equilibria of dynamic games with many players: Existence, approximation, and market structure. *J. Econom. Theory* 156:269–316.

Aronow PM, Samii C (2017) Estimating average causal effects under general interference, with application to a social network experiment. *Ann. Appl. Stat.* 11(4):1912–1947.

Athey S, Luca M (2019) Economists (and economics) in tech companies. *J. Econom. Perspect.* 33(1):209–230.

Athey S, Eckles D, Imbes GW (2018) Exact p-values for network interference. *J. Amer. Statist. Assoc.* 113(521):230–240.

Baird S, Bohren JA, McIntosh C, Özler B (2018) Optimal design of experiments in the presence of interference. *Rev. Econom. Statist.* 100(5):844–860.

Banerjee A, Duflo E (2011) *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty* (Public Affairs, New York).

Basse GW, Feller A, Toulis P (2019) Randomization tests of causal effects under interference. *Biometrika* 106(2):487–494.

Basse GW, Soufiani HA, Lambert D (2016) Randomization and the pernicious effects of limited budgets on auction experiments. *Artificial Intelligence and Statistics*, 1412–1420.

Beck A, Teboulle M (2003) Mirror descent and nonlinear projected subgradient methods for convex optimization. *Oper. Res. Lett.* 31(3):167–175.

Blake T, Coey D (2014) Why marketplace experimentation is harder than it seems: The role of test-control interference. *Proc.15th ACM Conf. Econom. Comput.* (ACM), 567–582.

Blundell R, Dias MC, Meghir C, Van Reenen J (2004) Evaluating the employment impact of a mandatory job search program. *J. Eur. Econom. Assoc.* 2(4):569–606.

Bottou L, Peters J, Candela JQ, Charles DX, Chickering M, Portugaly E, Ray D, Simard PY, Snelson E (2013) Counterfactual reasoning and learning systems: The example of computational advertising. *J. Machine Learn. Res.* 14(1):3207–3260.

Bramson M, Lu Y, Prabhakar B (2012) Asymptotic independence of queues under randomized load balancing. *Queueing Systems* 71(3):247–292.

Bubeck S, Lee YT, Eldan R (2017) Kernel-based methods for bandit convex optimization. *Proc. 49th Annual ACM SIGACT Symp. Theory Comput.* (ACM), 72–85.

Bubeck S, Jiang Q, Lee Y-T, Li Y, Sidford A (2019) Complexity of highly parallel non-smooth convex optimization. *Advances in Neural Information Processing Systems*, 13900–13909.

Cachon GP, Daniels KM, Lobel R (2017) The role of surge pricing on a service platform with self-scheduling capacity. *Manufacturing Service Oper. Management* 19(3):368–384.

CAISO (2009) Renewable resources and the California electric power industry: System operations, wholesale markets and grid planning. California ISO Report, https://www.caiso.com/Documents/RenewableResourcesandCaliforniaElectricPowerIndustry-SystemOperations_WholesaleMarketsandGridPlanning.pdf.

Cesa-Bianchi N, Conconi A, Gentile C (2004) On the generalization ability of on-line learning algorithms. *IEEE Trans. Inform. Theory* 50(9):2050–2057.

Chetty R (2009) Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. *Annual Rev. Econom.* 1(1):451–488.

Duchi J, Ruan F, Yun C (2018) Minimax bounds on stochastic batched convex optimization. *Conference On Learning Theory*, 3065–3162.

Duchi JC, Jordan MI, Wainwright MJ, Wibisono A (2015) Optimal rates for zero-order convex optimization: The power of two function evaluations. *IEEE Trans. Inform. Theory* 61(5):2788–2806.

Eckles D, Karrer B, Ugander J (2017) Design and analysis of experiments in networks: Reducing bias from interference. *J. Causal Inference* 5(1):1–23.

Esfandiari H, Karbasi A, Mehrabian A, Mirrokni V (2019) Regret bounds for batched bandits. Preprint, submitted October 11, https://arxiv.org/abs/1910.04959.

Feng Z, Podimata C, Syrgkanis V (2018) Learning to bid without knowing your value. *Proc. 2018 ACM Conf. Econom. Comput.* (ACM), 505–522.

Fisher RA (1935) *The Design of Experiments* (Oliver and Boyd, Edinburgh).

Flaxman AD, Kalai AT, McMahan HB (2005) Online convex optimization in the bandit setting: gradient descent without a gradient. *Proc. 16th Annual ACM-SIAM Symp. Discrete Algorithms* (Society for Industrial and Applied Mathematics), 385–394.

Gao Z, Han Y, Ren Z, Zhou Z (2019) Batched multi-armed bandits problem. *Advances in Neural Information Processing Systems*, 501–511.

Ghadimi S, Lan G (2013) Stochastic first-and zeroth-order methods for nonconvex stochastic programming. *SIAM J. Optim.* 23(4):2341–2368.

Goh M (2002) Congestion management and electronic road pricing in Singapore. *J. Transport Geography* 10(1):29–38.

Graham C, Méléard S (1994) Chaos hypothesis for a system interacting through shared resources. *Probab. Theory Related Fields* 100(2):157–174.

Halfin S, Whitt W (1981) Heavy-traffic limits for queues with many exponential servers. *Oper. Res.* 29(3):567–588.

Hall J, Kendrick C, Nosko C (2015) The effects of Uber's surge pricing: A case study. Report, The University of Chicago Booth School of Business, Chicago.

Hall JV, Horton JJ, Knoepfle DT (2020) Ride-sharing markets re-equilibrate. Accessed November 11, 2020, https://john-joseph-horton.com/papers/uber_price.pdf.

Harberger AC (1964) The measurement of waste. *Amer. Econom. Rev.* 54(3):58–76.

Heckman JJ (2001) Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture. *J. Political Econ.* 109(4):673–748.

Heckman JJ, Lochner L, Taber C (1998) General-equilibrium treatment effects: A study of tuition policy. *Amer. Econom. Rev.* 88(2):381–386.

Holt CA, Laury SK (2002) Risk aversion and incentive effects. *Amer. Econom. Rev.* 92(5):1644–1655.

Hopenhayn HA (1992) Entry, exit, and firm dynamics in long run equilibrium. *Econometrica* 60(5):1127–1150.

Hudgens MG, Halloran ME (2008) Toward causal inference with interference. *J. Amer. Statist. Assoc.* 103(482):832–842.

Imbens GW (2014) Instrumental variables: An econometrician's perspective. *Statist. Sci.* 29(3):323–358.

Imbens GW, Rubin DB (2015) *Causal Inference in Statistics, Social, and Biomedical Sciences* (Cambridge University Press, New York).

Iyer K, Johari R, Sundararajan M (2014) Mean field equilibria of dynamic auctions with learning. *Management Sci.* 60(12):2949–2970.

Jamieson KG, Nowak R, Recht B (2012) Query complexity of derivative-free optimization. *Advances in Neural Information Processing Systems*, 2672–2680.

Johari R, Kamble V, Kanoria Y (2017) Matching while learning. *Proc. 2017 ACM Conf. Econom. Comput.* (ACM), 119.

Jovanovic B, Rosenthal RW (1988) Anonymous sequential games. *J. Math. Econom.* 17(1):77–87.

Kanoria Y, Nazerzadeh H (2014) Dynamic reserve prices for repeated auctions: Learning from bids. *International Conf. Web Internet Econom.* (Springer), 232.

Keskin NB, Zeevi A (2014) Dynamic pricing with an unknown demand model: Asymptotically optimal semi-myopic policies. *Oper. Res.* 62(5):1142–1167.

Khetan A, Oh S (2016) Achieving budget-optimality with adaptive schemes in crowdsourcing. *Advances in Neural Information Processing Systems*, 4844–4852.

Kleinberg RD (2005) Nearly tight bounds for the continuum-armed bandit problem. *Advances in Neural Information Processing Systems*, 697–704.

Kohavi R, Longbotham R, Sommerfield D, Henne RM (2009) Controlled experiments on the web: Survey and practical guide. *Data Mining Knowledge Discovery* 18(1):140–181.

Letham B, Karrer B, Ottoni G, Bakshy E (2019) Constrained Bayesian optimization with noisy experiments. *Bayesian Anal.* 14(2):495–519.

Leung MP (2020) Treatment and spillover effects under network interference. *Rev. Econom. Stat.* 102(2):368–380.

Manski CF (2013) Identification of treatment response with social interactions. *Econom. J.* 16(1):S1–S23.

Massoulié L, Xu K (2018) On the capacity of information processing systems. *Oper. Res.* 66(2):568–586.

Mézard M, Parisi G, Virasoro M (1987) *Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications* (World Scientific Publishing Company, Singapore).

Nambiar M, Simchi-Levi D, Wang H (2019) Dynamic learning and pricing with model misspecification. *Management Sci.* 65(11):4980–5000.

Nesterov Y, Spokoiny V (2017) Random gradient-free minimization of convex functions. *Foundations Comput. Math.* 17(2):527–566.

Ogburn EL, VanderWeele TJ (2017) Vaccines, contagion, and social networks. *Ann. Appl. Statist.* 11(2):919–948.

Ostrovsky M, Schwarz M (2011) Reserve prices in Internet advertising auctions: A field experiment. *EC '11: Proc. 12th ACM Conf. Electronic Commerce*, 59–60.

Perchet V, Rigollet P, Chassang S, Snowberg E (2016) Batched bandit problems. *Ann. Statist.* 44(2):660–681.

Pratt JW (1964) Risk aversion in the small and in the large. *Econometrica* 32(1–2):122.

Roy AD (1951) Some thoughts on the distribution of earnings. *Oxford Econom. Papers* 3(2):135–146.

Sävje F, Aronow PM, Hudgens MG (2021) Average treatment effects in the presence of unknown interference. *Ann. Statist*. Forthcoming.

Shamir O (2013) On the complexity of bandit and derivative-free stochastic convex optimization. *Conf. Learn. Theory*, 3–24.

Sobel ME (2006) What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference. *J. Amer. Statist. Assoc.* 101(476):1398–1407.

Spall JC (2005) *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control* (John Wiley & Sons, Hoboken, New Jersey).

Spencer J, Sudan M, Xu K (2014) Queuing with future information. *Ann. Appl. Probab.* 24(5):2091–2142.

Stolyar AL (2015) Pull-based load distribution in large-scale heterogeneous service systems. *Queueing Systems* 80(4):341–361.

Sznitman AS (1991) Topics in propagation of chaos. *Ecole d'été de probabilités de Saint-Flour XIX—1989* (Springer, Berlin, Heidelberg), 165–251.

Tang D, Agarwal A, O'Brien D, Meyer M (2010) Overlapping experiment infrastructure: More, better, faster experimentation. *Proc. 16th ACM SIGKDD Internat. Conf. Knowledge Discovery Data Mining* (ACM), 17–26.

Tchetgen Tchetgen EJ, VanderWeele TJ (2012) On causal inference in the presence of interference. *Stat. Methods Medical Res.* 21(1): 55–75.

Train KE (2009) *Discrete Choice Methods with Simulation* (Cambridge University Press, New York).

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

Vvedenskaya ND, Dobrushin RL, Karpelevich FI (1996) Queueing system with selection of the shortest of two queues: An asymptotic approach. *Problemy Peredachi Informatsii* 32(1):20–34.

Weintraub GY, Benkard CL, Roy BV. (2008) Markov perfect industry dynamics with many firms. *Econometrica* 76(6):1375–1411.