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

MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 20, No. 4, Fall 2018, pp. 704–720
ISSN 1523-4614 (print), ISSN 1526-5498 (online)

On-Demand Service Platforms

Terry A. Taylor[a]

[a] Haas School of Business, University of California, Berkeley, California 94720
Contact: taylor@haas.berkeley.edu, http://orcid.org/0000-0002-1488-8923 (TAT)

Received: January 23, 2016
Revised: December 30, 2016; August 22, 2017
Accepted: August 23, 2017
Published Online in Articles in Advance: July 23, 2018

https://doi.org/10.1287/msom.2017.0678

Abstract. An on-demand service platform connects waiting-time-sensitive customers with independent service providers (agents). This paper examines how two defining features of an on-demand service platform—delay sensitivity and agent independence—impact the platform's optimal per-service price and wage. Delay sensitivity reduces expected utility for customers and agents, which suggests that the platform should respond by decreasing the price (to encourage participation of customers) and increasing the wage (to encourage participation of agents). These intuitive price and wage prescriptions are valid in a benchmark setting without uncertainty in the customers' valuation or the agents' opportunity costs. However, uncertainty in either dimension can reverse the prescriptions: Delay sensitivity increases the optimal price when customer valuation uncertainty is moderate. Delay sensitivity *decreases* the optimal wage when agent opportunity cost uncertainty is high and expected opportunity cost is moderate. Under agent opportunity cost uncertainty, agent independence *decreases* the price. Under customer valuation uncertainty, agent independence *increases* the price if and only if valuation uncertainty is sufficiently high.

Supplemental Material: The online appendix is available at https://doi.org/10.1287/msom.2017.0678.

Keywords: service operations • incentives and contracting • game theory • queueing theory

1. Introduction

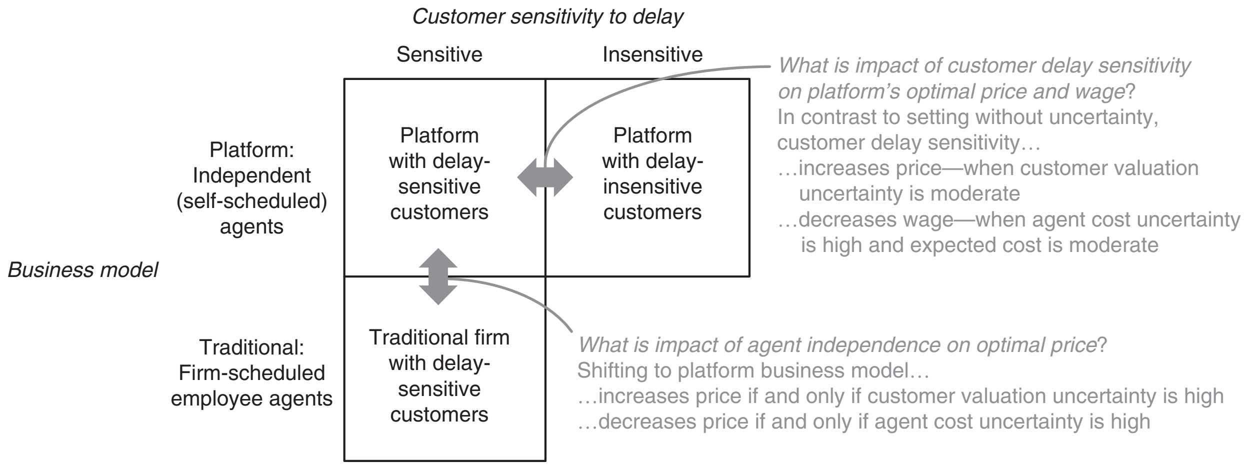
Recent years have witnessed the emergence and rapid growth of platforms for on-demand services. Examples include restaurant food delivery (e.g., Caviar, DoorDash), consumer goods delivery (e.g., UberRush, Go-Mart), and taxi-style transportation (e.g., Fasten, Go-Jek, Lyft, Uber; Roose 2014, Kokalitcheva 2015, MacMillan 2015, Shoot 2015, Watanabe 2016). These services are *on-demand* in the sense that upon experiencing a need for service, a customer desires service immediately and is sensitive to delay. In this way, on-demand service platforms are distinct from scheduled service platforms which book appointments in advance (e.g., Amazon Home Services) (Dowdle 2015).

A platform connects customers seeking service with *independent agents* that provide the service. In each of the preceding examples, an agent is an independent contractor who receives a payment from the platform for each service completion. The agent is independent in the sense that she decides whether and when to work. The platform business model is distinct from the traditional firm–employee business model, wherein the firm determines when its employees work and pays them a salary or hourly rate rather than a piece rate. Examples of on-demand services provided via the firm–employee model are food delivery (e.g., Munchery); pickup, packaging, and shipping (e.g., Shyp); and town-car transportation (Bensinger 2015).

This paper explores two key features of on-demand service platforms: First, upon experiencing a need for service, waiting-time-sensitive customers choose whether to seek service. Second, independent agents choose whether to work. Two elements that connect these customer and agent decisions are agent idleness and scale economies. The fraction of time an agent working for an on-demand service platform is idle can be significant; because an independent agent is not compensated for idle time, the fraction of idle time she anticipates significantly impacts her decision of whether to work (Singer and Isaac 2015, Steinmetz 2015). Although agents prefer this fraction be small, customers prefer that it be large because greater agent availability reduces customers' congestion-driven delay. Whereas idleness hurts agents but benefits customers, both groups benefit from scale economies. Indeed, platforms point to scale economies from pooling efficiencies as one of their primary advantages over small-scale firms (e.g., an individual restaurant or retailer providing its own delivery service) in the provision of on-demand services (Huet 2015, Shoot 2015).

The purpose of this paper is to provide insight into how on-demand service platforms should set per-service prices and wages (see Figure 1). By comparing the setting with (self-scheduled) independent agents to one with (firm-scheduled) employee agents, we address the following question: What is the impact of agent independence on the optimal price? Because of

Figure 1. Impact of Agent Independence and Customer Sensitivity to Delay



the on-demand nature of the service, customers are sensitive to delay. By comparing this base case to a benchmark case where customers are insensitive to delay, we address the following question: What is the impact of delay sensitivity on the platform's optimal price and wage? Customer insensitivity to delay captures the extreme case of customer patience.

To address these two questions, we employ a queueing model in which the customer arrival rate and number of servers (agents) are endogenous. The platform commits to a per-service price and wage prior to the resolution of uncertainty in the customers' valuation for receiving service and in the agents' opportunity costs. Customers have a common valuation for receiving service, and an agent participates if the expected revenue generated from participation exceeds her opportunity cost.

Delay sensitivity reduces expected utility for customers (directly, through waiting) and agents (indirectly, through idleness), which suggests that the platform should respond to delay sensitivity by decreasing the per-service price (to encourage participation of customers) and increasing the per-service wage (to encourage participation of agents). These intuitive price and wage prescriptions are valid in a benchmark setting without uncertainty in the customers' valuation or the agents' opportunity costs. This paper's first contribution is to identify and explain driving forces that cause these prescriptions to break. Moderate customer valuation uncertainty causes the price prescription to break. High opportunity cost uncertainty coupled with a moderate expected opportunity cost causes the wage prescription to break.

The second contribution is to identify and explain driving forces that lead agent independence to either increase or decrease the optimal price. Agent opportunity cost uncertainty causes agent independence to decrease the price. High (low) customer valuation uncertainty causes agent independence to increase (decrease) the price.

A practice of prominent on-demand service platforms that has attracted research (Cachon et al. 2017, Banerjee et al. 2015) is the adjusting of prices (and wages) in real time based on the system state. However, of the previously identified on-demand service platforms, only two, Lyft and Uber, do so. Platforms avoid real-time pricing, in part, because of customer resistance to the practice (MacMillan 2015). To complement the real-time pricing research and to provide insight for platforms that commit to prices and wages in advance, we focus on that setting. More broadly, our work is related to four areas of literature: research on platforms not providing on-demand service, pricing in queueing systems, incentives for agents' capacity-related decisions, and on-demand service platforms.

On-demand service platforms are but one of several platform types in the sharing economy (see Table 1). Product-sharing platforms (e.g., Airbnb, Turo) connect customers seeking to rent assets (e.g., apartments, cars) with owners. Benjaafar et al. (2018), Fraiberger and Sundararajan (2015), and Jiang and Tian (2018) examine individuals' decisions to purchase or rent assets in the presence of a product-sharing platform. Li et al. (2016) studies owners' pricing behavior empirically. Product-sharing platforms differ from on-demand service platforms in several respects: customers book in

Table 1. Sharing Economy Platforms

Platform type	Price-setting party	Nature of offering	
		Differentiation	Timing
On-demand service	Platform	Undifferentiated	On-demand
Freelancing	Skilled agent	Agent specific	On-demand
Product sharing	Asset-owning agent	Agent specific	Scheduled

advance, the platform does not determine prices, and the offering to customers exhibits heterogeneity. The latter two differentiating characteristics are shared by freelancing platforms (e.g., Upwork), which connect customers seeking professional services, such as software development, with skilled agents. When heterogeneity is pronounced, a key role of the platform is matching customers to agents in way that ensures fit between the customer's needs and the agent's capabilities. Allon et al. (2012), Arnosti et al. (2014), and Hu and Zhou (2015) provide insight into the design of matching mechanisms. Several papers (e.g., Snir and Hitt 2003, Moreno and Terwiesch 2014) study freelancing platforms empirically, examining agents' bidding behavior and reputation mechanisms. Heterogeneity, matching, and agent reputation are not of central importance for the on-demand service platforms that motivate our work. Whereas each software development project has distinct characteristics and requires specific skills, an on-demand service (e.g., transportation of a person or product) is generic and does not. Finally, modeling and empirical work (e.g., Einav et al. 2016, Zervas et al. 2017) examines the impact of platform entry on incumbents.

The literature on pricing in queueing systems is extensive, dating back to Naor (1969); see Hassin and Haviv (2003) and Hassin (2016) for reviews. A key feature this literature and our work have in common is that customer demand is sensitive not only to price, but also to the service level (i.e., delay) experienced by customers. Our work employs the customer behavior model of Chen and Frank (2004), which examines pricing and capacity (service rate) decisions in an $M/M/1$ system (i.e., Poisson arrival process; exponential service distribution; one server; first-come, first-served scheduling policy) in a setting where customers are homogenous and the queue is unobservable. Many papers consider less restrictive settings, examining pricing and capacity decisions with a single class of heterogeneous customers (e.g., Mendelson 1985, Maglaras and Zeevi 2003, Kumar and Randhawa 2010) or pricing and scheduling decisions with multiple customer classes (e.g., Mendelson and Whang 1990, Afèche 2013, Afèche and Pavlin 2016, Nazerzadeh and Randhawa 2017, Maglaras et al. 2018). Several papers examine dynamic price and/or lead-time quotation (e.g., Plambeck 2004, Çelik and Maglaras 2008, Ata and Olsen 2013). Some research examines pricing decisions when service quality increases in service time (e.g., Anand et al. 2011), customers are uninformed about service quality (e.g., Debo et al. 2013), or customers purchase subscriptions (e.g., Cachon and Feldman 2011). The aforementioned pricing-in-queues papers assume capacity is exogenous or under the control of the system manager, whereas we focus on the setting where capacity is determined by the strategic behavior of agents.

The literature on incentives for agents' capacity-related decisions is extensive. For a review of the supply chain contracting literature, where a buyer's incentives influence quantity decisions of supplier agents, see Cachon (2003). More relevant to our work, Gilbert and Weng (1998), Cachon and Zhang (2007), Gopalakrishnan et al. (2016), and Zhan and Ward (2015) examine how a system manager's rule for allocating customers to agents and/or the per-service wage influences agents' service-rate decisions in queueing systems. Closer to our work, which examines agents' decisions of whether to work, Ibrahim (2018) examines a setting in which the system manager determines the number of employee agents and wages. Providing agents discretion over when to work can increase or decrease the optimal number of agents. The aforementioned capacity incentives in queueing system papers assume prices and customer arrivals are exogenous, although Ibrahim (2018) and Zhan and Ward (2015) allow for customer abandonment. In contrast, we focus on the setting where price and customer arrivals are endogenous.

Closest to our work are papers examining on-demand service platforms. Banerjee et al. (2015), Cachon et al. (2017), and Gurvich et al. (2018) focus on agents' decisions of whether to work in settings with uncertain demand and heterogeneous agent opportunity costs. Banerjee et al. (2015) and Cachon et al. (2017) focus on comparing static versus dynamic (i.e., demand- or system-state-contingent) prices and wages. In Banerjee et al. (2015), the wage is an exogenous fraction of the price, whereas in Cachon et al. (2017) and our work, both the price and wage are endogenous. Cachon et al. (2017) finds that "surge pricing," wherein the wage is a fixed percentage of a demand-contingent price, performs well. In contrast, Banerjee et al. (2015) finds that static pricing performs well. Cachon et al. (2017) examines an agent's long-run decision of whether to join the platform, as well as the short-run decision of whether to work, whereas Banerjee et al. (2015) and our work only focus on the latter. To obtain insights in a rich queueing model, Banerjee et al. (2015) employs approximations and large system asymptotics; in contrast, our work employs a stylized model that allows for exact analysis. In Banerjee et al. (2015), customers are heterogeneous, seek service if and only if their valuation exceeds the price, and are infinitely impatient (they abandon if not served immediately); in contrast, in our work, customers are homogeneous, finitely patient, and seek service if and only if their expected utility from doing so is nonnegative.

In Gurvich et al. (2018), the platform determines the number of agents, market-condition-contingent price and wage, and a cap on the number of agents allowed to work; agents decide whether to work after observing the market condition. Gurvich et al. (2018) finds that for a fixed agent pool, agent independence reduces the number of working agents and increases the optimal price. Our findings are consistent with the former, but not the latter. The reason for the difference in price findings is, in part, that in the newsvendor-style model of Gurvich et al. (2018) (and that of Cachon et al. 2017), demand does not depend on the service level, whereas in our queueing model it does; see Section 3.1. Finally, Kabra et al. (2016) empirically studies the extent to which a temporary price reduction or wage increase is useful in stimulating growth of an on-demand service platform.

2. Model

A platform connects customers seeking service with independent agents. First, the platform (it) commits to the per-service price p that each customer (he) pays the platform and the per-service wage ω the platform pays each agent (she). Second, uncertainty regarding customers' valuation of the service and agents' opportunity costs is resolved. Each agent observes the customers' valuation and agents' opportunity costs and decides whether to work (participate). Third, each customer observes the customers' valuation and the expected waiting time and decides the probability of seeking service upon experiencing a need for service. All parties are risk neutral. A natural timescale is a period of a few hours, during which the parameters are reasonably stationary (e.g., Tuesday 6:00 P.M. to 9:00 P.M.). To employ a consistent interpretation of uncertainty through this paper, we use weather: Adverse weather conditions tend to increase both customers' valuation of service (e.g., food delivery, transportation) and agents' (e.g., drivers') opportunity costs (Scheiber 2015). The assumption that customers observe the expected waiting time may be reasonable to the extent that platforms provide this information, as is the case for the on-demand service platforms in Section 1. The assumption that agents have common information regarding valuations and costs may be reasonable to the extent that these are determined by factors that are commonly observed (e.g., weather conditions).

We begin with the customer side. Customers are homogenous ex ante and ex post: All customers experiencing a need for service share a common valuation for receiving service $\hat{V}$, a random variable with realization $V \in \{V^h, V^l\}$, and incur delay disutility $c > 0$ per unit time while waiting for service. The customers' valuation per service is high $V^h = v + \delta$ or low $V^l = v - \delta$ with equal probability, where $\delta \in [0, v)$. We refer to δ as customer valuation uncertainty. One can interpret δ as reflecting the degree to which customers' valuations are sensitive to the weather being "good" or "bad."

A customer has the opportunity to seek service from a platform, which has access to $n \in \{0, 1, \ldots, \bar{N}\}$ participating agents, each of which has average service rate μ. For simplicity in exposition, initially suppose n is nonzero. Customers are processed in a first-come, first-served fashion, and the platform over the long term equally allocates requests for service over the n agents. Events triggering a customer's need for service occur at rate Λ. In the spirit of the contention of Banerjee et al. (2015) that ride-sharing platforms are typically supply constrained, we focus on the case where the need for service is abundant,

$$\Lambda > \bar{N}\mu. \tag{1}$$

This focus is consistent with the observation that the growth of on-demand service platforms in many cities is limited by the availability of workers rather than customers (Farrell and Greig 2016, Phillips 2016). If customers, upon experiencing a need for service, seek service with probability q, then the arrival (demand) rate is $\lambda = q\Lambda$. Needs for service occur according to a Poisson process, and service times are exponentially distributed. A customer's expected waiting time in this $M/M/n$ queueing system is

$$W(\lambda, n) = \frac{(\lambda/\mu)^n}{n!(n\mu)(1-\lambda/n\mu)^2} \cdot \left(\sum_{i=0}^{n-1} \frac{(\lambda/\mu)^i}{i!} + \frac{(\lambda/\mu)^n}{n!(1-\lambda/n\mu)}\right)^{-1}, \tag{2}$$

provided that the system is stable, $\lambda/n\mu < 1$.

Each customer decides whether to seek service in step 3, after observing his valuation V, the price p, and the steady-state expected waiting time (2); that is, customers make decisions based on the long-run expected waiting time, rather than on the current state of the queue. A customer seeks service if and only if the expected utility from doing so,

$$V - p - cW(\lambda, n), \tag{3}$$

is nonnegative. After deciding to seek service, the customer does not renege prior to receiving service (i.e., does not abandon the queue). Let $q(V, p, n)$ denote the equilibrium probability of seeking service, and let $\lambda(V, p, n) = q(V, p, n)\Lambda$ denote the equilibrium arrival (demand) rate under valuation V, price p, and n participating agents. It is well known (see, e.g., Hassin and Haviv 2003) that the unique equilibrium $q(V, p, n) = \lambda(V, p, n)/\Lambda$, where $\lambda(V, p, n)$ is the unique value of λ that satisfies

$$V - p - cW(\lambda, n) = 0, \tag{4}$$

provided that $p < V$ and $n \geq 1$. If, instead, either the price exceeds the realized valuation $p \geq V$ or the number of participating agents $n = 0$, so that the platform is not offering service, then the unique equilibrium $q(V, p, 0) = \lambda(V, p, 0) = 0$.

We now turn to the agent side. Agents are homogeneous ex ante and heterogeneous ex post. Agent $i \in \{1, \dots, \bar{N}\}$ has opportunity cost per unit time $\hat{K}_i$, a random variable with realization $K_i \in \{K^H, K^L\}$. Each agent's opportunity cost per unit time is high, $K^H = k + \Delta$, or low, $K^L = k - \Delta$, with equal probability, where $\Delta \in [0, k)$. The correlation of any pair $\{i, j\} \in \{1, \dots, \bar{N}\}^2$, where $i \neq j$, of agents' opportunity costs is $\mathrm{Corr}(\hat{K}_i, \hat{K}_j) = \rho \in [0, 1]$. We refer to Δ as the agent opportunity cost uncertainty. One can interpret Δ as reflecting the degree to which agents are sensitive to the weather being "good" or "bad."

Each agent decides whether to participate in step 2, after observing the agents' opportunity costs $\{K_i\}_{i=1,\dots,\bar{N}}$, the customers' valuation V, and the platform's per-service wage ω and price p. When $n \in \{1, \dots, \bar{N}\}$ agents participate, each agent's demand rate is $\lambda(V, p, n)/n$. If agent i anticipates that $n-1$ other agents participate, then agent i anticipates receiving expected utility at rate

$$\omega \lambda(V, p, n)/n - K_i \tag{5}$$

if she participates. An agent participates if and only if she anticipates receiving nonnegative expected utility from doing so. Let $\hat{N}^L$ denote the number of agents with the low opportunity cost K^L, a random variable with realization $N^L \in \{0, 1, \dots, \bar{N}\}$. Let $\mathcal{N}(V, N^L)$ denote the equilibrium number of participating agents n under valuation V and N^L low-cost agents. We restrict attention to equilibria in which high-cost agents (those with opportunity cost K^H) participate only if all low-cost agents participate. The equilibrium number of participating agents, $\mathcal{N}(V, N^L)$, under $N^L \in \{1, \dots, \bar{N}-1\}$ low-cost agents satisfies

$$\omega \lambda(V, p, \mathcal{N}(V, N^L))/\mathcal{N}(V, N^L) - K^j \geq 0, \tag{6}$$

where $j = H$ if $\mathcal{N}(V, N^L) \in \{N^L + 1, \dots, \bar{N}\}$, and $j = L$ if $\mathcal{N}(V, N^L) \in \{1, \dots, N^L\}$, and

$$\omega \lambda(V, p, \mathcal{N}(V, N^L) + 1)/[\mathcal{N}(V, N^L) + 1] - K^j < 0, \tag{7}$$

where $j = H$ if $\mathcal{N}(V, N^L) \in \{N^L, \dots, \bar{N}-1\}$, and $j = L$ if $\mathcal{N}(V, N^L) \in \{0, \dots, N^L - 1\}$. The equilibrium $\mathcal{N}(V, N^L)$ under $N^L = \bar{N}$ ($N^L = 0$, respectively) satisfies (6), where $j = L$ ($j = H$, respectively) if $\mathcal{N}(V, N^L) \in \{1, \dots, \bar{N}\}$, and (7), where $j = L$ ($j = H$, respectively) if $\mathcal{N}(V, N^L) \in \{0, \dots, \bar{N}-1\}$.

Finally, we turn to the platform side. The platform makes decisions in step 1, in the face of uncertainty regarding customers' valuation of the service and agents' opportunity costs. The platform sets its price and wage to maximize its expected profit rate

$$\max_{(p,\omega)} (p - \omega) E[\lambda(\hat{V}, p, \mathcal{N}(\hat{V}, \hat{N}^L))],$$

where $\lambda(\cdot,\cdot,\cdot)$ satisfies (4), and $\mathcal{N}(\cdot,\cdot)$ satisfies (6) and (7).

To understand the impact of agent independence, we contrast the platform business model with (self-scheduled) independent agents to the firm–employee business model with (firm-scheduled) employee agents. The firm sets the price in step 1. In step 2, reflecting the notion that a firm has more information regarding employees than a platform has regarding independent contractors, the firm observes employee agents' opportunity costs as well as the customers' valuation. The firm decides which employee agents will work, with the firm incurring these agents' costs. This can be implemented by the firm, in step 2, offering to each agent an hourly rate in exchange for commitment by the agent to work. An agent works if and only if she accepts the offer. In the firm–employee business model, the firm sets the price and activates agents to work to maximize its expected profit rate

$$\max_{p} \left\{ pE\left[\max_{\{\mathbf{I}_i\}_{i=1,\dots,\bar{N}}} \left\{ \lambda\left(\hat{V}, p, \sum_{i=1}^{\bar{N}} \mathbf{I}_i\right) - \sum_{i=1}^{\bar{N}} \mathbf{I}_i \hat{K}_i \right\} \right] \right\}, \tag{8}$$

where $\mathbf{I}_i = 1$ if agent i works and $\mathbf{I}_i = 0$ otherwise, and $\lambda(\cdot,\cdot,\cdot)$ satisfies (4). We restrict attention to the parameter range in which the firm activates at least one agent, for any realized customer valuation V and agent opportunity costs $\{K_i\}_{i=1,\dots,\bar{N}}$. Throughout, we restrict attention to the parameter range in which the firm's (platform's) expected profit rate under the optimal price (and wage) is strictly positive.

Our formal results hold when the employment model is instead conceived as capturing a longer-term employment relationship. Reflecting ex ante two-sided commitment, in step 1, the platform offers to each employee agent a salary rate in exchange for a commitment by the agent to work subsequently. An agent works subsequently if and only if she accepted the offer in step 1.

We have assumed that customers observe the long-run (steady-state) expected waiting time, which is roughly consistent with the practice of on-demand service platforms that periodically update the expected waiting time range they provide to customers (e.g., Caviar). To more accurately reflect the practice of platforms that provide real-time expected waiting time information (e.g., Lyft, Uber) would require a more complex model in which, essentially, the queue length was observable to prospective customers.

For expositional simplicity, we have assumed that each agent observes all agents' opportunity costs and the customers' valuation prior to making an (irreversible) participation decision. In reality, an agent may not have perfect information regarding other agents' opportunity costs. Furthermore, an agent's participation decision is fairly easy to reverse: an agent can participate for some initial period to learn her demand rate, to inform her longer-run participation decision. To the extent that the initial period is brief, it is almost costless for the agent to learn her demand rate, which is all that is required for the agent to make her longer-run participation decision.

3. Results

Under price p, wage ω, realized valuation V, and realized number of low-cost agents N^L, the equilibrium in the number of participating agents $\mathcal{N}(V, N^L)$ and demand rate $\lambda(V, p, \mathcal{N}(V, N^L))$, the solution to (4), (6), and (7), is given in the following lemma.

Lemma 1. *An equilibrium in participating agents $\mathcal{N}(V, N^L)$ and demand rate $\lambda(V, p, \mathcal{N}(V, N^L))$ has*

$$\mathcal{N}(V, N^L) = \begin{cases} 0 & \text{if and only if } \omega < K^L/\lambda(V, p, 1), \\ N^L & \text{if and only if } \omega \in [N^L K^L/\lambda(V, p, N^L), \\ & \qquad (N^L+1)K^H/\lambda(V, p, N^L+1)), \\ \bar{N} & \text{if and only if } \omega \geq \bar{N} K^H/\lambda(V, p, \bar{N}) \end{cases}$$

if $N^L \in \{1, \ldots, \bar{N}-1\}$. If $N^L = \bar{N}$, then $\mathcal{N}(V, N^L) = 0$ if and only if $\omega < K^L/\lambda(V, p, 1)$, and $\mathcal{N}(V, N^L) = \bar{N}$ if and only if $\omega \geq \bar{N} K^L/\lambda(V, p, \bar{N})$. If $N^L = 0$, then $\mathcal{N}(V, N^L) = 0$ if and only if $\omega < K^H/\lambda(V, p, 1)$, and $\mathcal{N}(V, N^L) = \bar{N}$ if and only if $\omega \geq \bar{N} K^H/\lambda(V, p, \bar{N})$.

Lemma 1 is straightforward to verify. Proofs of all other formally stated results are in the appendix.

It is natural that, ceteris paribus, participation of an additional agent benefits customers (through reduced waiting times) and/or the platform (through greater customer demand). What is more interesting is that participation of an additional agent benefits currently participating agents: the equilibrium demand allocated to an agent $\lambda(V, p, n)/n$ strictly increases with the number of participating agents n; see Lemma A1(a) in the appendix. We refer to this as the *agent participation externality*. Participation of an additional agent has two effects: First, the improved service level (reduced waiting times) translates into greater total customer demand $\lambda(V, p, n)$. Second, the "competition" among agents over the total demand "pie" $\lambda(V, p, n)$ intensifies in that the pie is being divided into more "slices." If, in violation of inequality (1), the need for service were scarce such that $\lambda(V, p, n) = \Lambda$, then the competition effect would dominate: participation of an additional agent would reduce each agent's slice Λ/n of the fixed-size demand pie. When the need for service is abundant (i.e., inequality (1) holds), the improved service level effect dominates the competition effect because of the scale economies present in queueing systems.

The structure of the equilibrium in participating agents $\mathcal{N}(V, N^L)$—that either 0, N^L or $\bar{N}$ agents participate—is driven by the agent participation externality and the assumption that the ex post heterogeneity in agents' costs is binary. The former implies that either all or no agents of a given cost participate. (The former also implies that the wage ranges in Lemma 1 frequently overlap; indeed there is always overlap when $\bar{N} \geq 2$ and $N^L \neq \bar{N} - 1$. Hence, when $\bar{N} \geq 2$, for any realized valuation V and number of low-cost agents $N^L \neq \bar{N} - 1$ there exist a price and wage (p, ω) such that multiple equilibria exist.)

Lemma 1 points to a risk that platforms face in offering high wages: doing so can sustain undesirable equilibria with few participating agents. This is most easily seen by example. Suppose $\mu = c = K^L = 1$, $K^H = 10$, $V = 20$, $p = 19.9$, $\bar{N} = 15$, and $N^L = 5$. Under wage $\omega = 16$, the equilibria in participating agents are $\mathcal{N}(V, N^L) \in \{5, 15\}$. Under lower wage $\omega = 13$, $\mathcal{N}(V, N^L) = 15$ is the unique equilibrium. To the extent that $\mathcal{N}(V, N^L) = 5$ participating agents emerges as the equilibrium under wage $\omega = 16$, by *reducing* the wage to $\omega = 13$, the platform *increases* the number of participating agents. The agent participation externality implies that sustaining an equilibrium with few agents—with its consequent limited demand per agent—requires a high wage, whereas, under a (moderately) low wage, only equilibria with many agents can be sustained.

In the face of this risk, it is attractive to both the platform and agents to coordinate on the equilibrium with the most agents (formally, each party's expected utility increases with the number of participating agents). In the sequel, we assume that the parties do so, if multiple equilibria exist.

Throughout, to sharpen results, we focus on the special cases wherein uncertainty is present on only the customer or agent side. Lemma 2 characterizes the platform's optimal decisions under *customer valuation uncertainty* (i.e., $\delta > 0$ and $\Delta = 0$), and Lemma 3 under *agent opportunity cost uncertainty* (i.e., $\delta = 0$ and $\Delta > 0$). Let $p^h \in \arg\max_{p \geq 0}\{p\lambda(V^h, p, \bar{N})\}$, $p^l \in \arg\max_{p \geq 0}\{[p - \bar{N}k/\lambda(V^l, p, \bar{N})]\sum_{j \in \{h, l\}} \lambda(V^j, p, \bar{N})/2\}$, and $\omega^j = \bar{N}k/\lambda(V^j, p^j, \bar{N})$ for $j \in \{h, l\}$.

Lemma 2. *Under customer valuation uncertainty, the platform's optimal price and wage are*

$$(p^*, \omega^*) = \begin{cases} (p^l, \omega^l) & \text{if } \delta \leq \bar{\delta}, \\ (p^h, \omega^h) & \text{if } \delta > \bar{\delta}, \end{cases} \tag{9}$$

where $\bar{\delta} \in (0, v)$.

The agent participation externality implies it is optimal for the platform to induce either all agents to participate or none. Agents are less inclined to participate when the realized customer valuation is low, because this translates to a lower demand rate for each agent, $\lambda(V^l, p, n)/n < \lambda(V^h, p, n)/n$ for $n \in \{1, \ldots, \bar{N}\}$. Consequently, the platform has a choice either to offer a "high wage" $\omega = \bar{N}k/\lambda(V^l, p, \bar{N})$, which induces the agents' participation even when the realized valuation is low, or to offer a "low wage" $\omega = \bar{N}k/\lambda(V^h, p, \bar{N})$, which induces the agents' participation only when the realized valuation is high. When valuation uncertainty is high, low-valuation customers have very low valuations and hence are unattractive to serve; consequently, it is optimal to offer a low wage and serve only high-valuation customers. Conversely, when uncertainty is low, it is optimal to offer a high wage and serve all customers. The platform's optimal price and wage reflect this structure, with the addition that the platform chooses a price that maximizes revenue from high-valuation customers when uncertainty is high and a price that maximizes expected profit from all customers when uncertainty is low.

For use in Lemma 3, let

$$p^A \in \arg\max_{p\geq 0}\{p\lambda(v,p,\bar{N})\}, \qquad R^A = p^A\lambda(v,p^A,\bar{N}),$$

$$R^M(p,n) = p\sum_{j=n}^{\bar{N}}\Pr(N^L = j)\lambda(v,p,j),$$

$$R^M(n) = \max_{p\geq 0} R^M(p,n), \qquad \tilde{k} = \frac{R^A - R^M(1)}{2\bar{N}},$$

$$\Pi^M(p,n) = \left[p - \frac{nK^L}{\lambda(v,p,n)}\right]\sum_{j=n}^{\bar{N}}\Pr(N^L = j)\lambda(v,p,j),$$

$$p^M(n) \in \arg\max_{p\geq 0}\Pi^M(p,n),$$

$$(p^M, n^M) \in \arg\max_{p\geq 0,\, n\in\{1,\ldots,\bar{N}-1\}}\Pi^M(p,n),$$

$$\omega^M = \frac{n^M K^L}{\lambda(v,p^M,n^M)}, \quad \text{and} \quad \omega^j = \frac{\bar{N}K^j}{\lambda(v,p^A,\bar{N})},$$

for $j \in \{H, L\}$.

Lemma 3. *Under agent opportunity cost uncertainty, the platform's optimal price and wage are*

$$(p^*, \omega^*) = \begin{cases} (p^A, \omega^H) & \text{if } \Delta \leq \underline{\Delta}, \\ (p^A, \omega^L) & \text{if } \Delta \in (\underline{\Delta}, \bar{\Delta}], \\ (p^M, \omega^M) & \text{if } \Delta > \bar{\Delta}, \end{cases}$$

where $0 < \underline{\Delta} \leq \bar{\Delta}$. *Furthermore,* $\underline{\Delta} < k$ *if and only if* $k > \tilde{k}$; $\bar{\Delta} < k$ *if and only if* $k > \tilde{k}$ *and* $\rho < 1$.

An agent is more inclined to participate when many other agents are participating (because of the agent participation externality) and when the agent's cost is low. Consequently, the platform has a choice to offer a high wage, $\omega = \bar{N}K^H/\lambda(v,p,\bar{N})$, which induces all agents to participate, or a low wage, $\omega = nK^L/\lambda(v,p,n) < \bar{N}K^H/\lambda(v,p,\bar{N})$, which induces all low-cost agents to participate, provided that there are at least $n \in \{1,\ldots,\bar{N}\}$ of them. If either the expected opportunity cost is small $k \leq \tilde{k}$ or cost uncertainty is small $\Delta < \underline{\Delta}$, then even "high-cost" agents have relatively low cost. Consequently, the platform can induce even high-cost agents to participate at relatively low cost to the platform. (Formally, the "high wage" required to induce such participation is small.) Furthermore, scale economies make it attractive for the platform to induce the participation of all agents.

In contrast, if $k > \tilde{k}$ and $\Delta > \underline{\Delta}$, then high-cost agents have very high cost, and it becomes too costly for the platform to induce their participation. Instead, the platform induces the participation of low-cost agents, provided that there are enough to establish sufficient scale economies. The scale economies are sufficient when only a strict subset of the agents has low cost, provided that their cost is sufficiently low, which occurs when cost uncertainty is very large, $\Delta > \bar{\Delta}$. If, instead, $\Delta < \bar{\Delta}$, then the platform sells to customers only when all $\bar{N}$ agents participate. The platform's optimal price and wage reflect this structure: the platform's price reflects that all agents participate when uncertainty is low, and stochastically fewer agents participate when uncertainty is high.

Lemma A5 in the appendix characterizes the platform's optimal price and wage (p^*, ω^*) in terms of the correlation of agents' opportunity costs ρ: when ρ is small, $(p^*, \omega^*) = (p^A, \omega^H)$; when ρ is moderate, $(p^*, \omega^*) = (p^M, \omega^M)$; and when ρ is large, $(p^*, \omega^*) = (p^A, \omega^L)$. The probability that there are "many" low-cost agents, $\Pr(N^L \geq j)$ where $j \in \{\lceil \bar{N}/2 \rceil + 1, \lceil \bar{N}/2 \rceil + 2, \ldots, \bar{N}\}$, increases with the correlation ρ. Hence, when the correlation is small, setting a low wage $\omega \in \{\omega^L, \omega^M\}$ that induces participation of only low-cost agents is unattractive because the probability that there are enough low-cost agents to achieve substantial scale economies is small; instead, the platform sets a high wage $\omega = \omega^H$ to induce the participation of all agents. As correlation increases, it becomes more attractive to set a low wage $\omega \in \{\omega^L, \omega^M\}$ that induces participation of only low-cost agents. When correlation ρ is very large, the probability that all agents have low cost $\Pr(N^L = \bar{N}) = \rho/2 + (1-\rho)(1/2)^{\bar{N}}$ is substantial, which makes offering the low wage $\omega = \omega^L$ that induces participation only when all agents have low cost attractive.

3.1. Agent Independence

This section characterizes the impact of agent independence on the optimal price. Lemma 4 characterizes the optimal decisions in the firm–employee business model (8).

Lemma 4. *In the firm–employee business model, the following hold*: (a) *under customer valuation uncertainty, the firm's optimal price is* $p_I^* = \arg\max_{p\geq 0}\{p\sum_{j\in\{h,l\}}\lambda(V^j,p,\bar{N})/2\}$, *and the firm activates* $\bar{N}$ *employee agents.* (b) *Under agent opportunity cost uncertainty, the firm's optimal price is* $p_I^* = p^A$, *and the firm activates* $\bar{N}$ *employee agents.*

The agent participation externality implies that customer demand $\lambda(V, p, n)$ is convex in the number of activated agents n. The result that the firm activates $\bar{N}$ employee agents is driven by this convexity and the assumption the firm activates at least one agent for any realized customer valuation and agent opportunity cost. The assumption that the firm activates at least one agent for any realized customer valuation is satisfied if and only if customer valuation uncertainty is sufficiently low $\delta < \delta^a$. The proof of Lemma 4 establishes that $\delta^a \in (\bar{\delta}, v)$.

Proposition 1 characterizes the impact of agent independence on the platform's optimal price under customer valuation uncertainty, and Proposition 2 characterizes it under agent opportunity cost uncertainty.

Proposition 1. *Under customer valuation uncertainty, agent independence decreases the optimal price,*

$$p^* < p_I^*, \tag{10}$$

if valuation uncertainty is low, $\delta \le \bar{\delta}$, and increases the price,

$$p^* > p_I^*, \tag{11}$$

if valuation uncertainty is high, $\delta > \bar{\delta}$.

When valuation uncertainty is low ($\delta \le \bar{\delta}$), it is optimal to serve all customers, whether or not agents are independent. When agents are independent, doing so requires the platform to offer a high wage $\omega = \bar{N}k/\lambda(V^l, p, \bar{N})$, to induce the participation of the agents even when the realized valuation is low. In doing so, the platform cedes rents in expectation to the agents. Decreasing the price allows the platform to reduce the wage and expected rent paid to each agent, $[\lambda(V^h, p, \bar{N})/\lambda(V^l, p, \bar{N}) - 1]k/2$. To see the latter, observe that if the platform charges a high price $p = V^l - \epsilon$, where ϵ is small, then when the realized valuation is low, each agent's demand rate $\lambda(V^l, p, \bar{N})/\bar{N}$ will be very low. Consequently, the platform must offer a very high per-service wage to cover the agent's opportunity cost. This results in very high rent being paid to each agent when the realized valuation is high. Because decreasing the price allows the platform to reduce the expected rents paid to agents, agent independence pushes the platform to decrease its price.

As valuation uncertainty increases, it becomes increasingly costly for the platform to offer the high wage required to induce independent agents to serve low-valuation customers. Consequently, when valuation uncertainty is high ($\delta > \bar{\delta}$), the platform gives up on serving low-valuation customers (and so charges a high price), whereas the firm with employee agents continues to do so (and so charges a low price).

Before addressing the impact of agent independence under agent opportunity cost uncertainty, we discuss a property of the equilibrium demand rate, the price elasticity of demand, $-(\partial\lambda(v, p, n)/\partial p)(p/(\lambda(v, p, n)))$. It is straightforward to show analytically that the price elasticity of demand strictly decreases as the number of agents n increases from 1 to 2. We observed numerically that the price elasticity of demand strictly decreases with the number of agents n in each of the 931 parameter combinations of $\mu = c = 1$, $v \in \{2, 3, \ldots, 20\}$, and $n \in \{2, 3, \ldots, 50\}$.

Proposition 2. *Suppose the price elasticity of demand strictly decreases with the number of agents n. Under agent opportunity cost uncertainty, agent independence decreases the optimal price,*

$$p^* \le p_I^*, \tag{12}$$

where the inequality is strict if and only if cost uncertainty is sufficiently high $\Delta > \bar{\Delta}$.

The intuition for why agent independence strictly decreases the optimal price when cost uncertainty is high is twofold. First, when cost uncertainty is high, the platform induces stochastically fewer agents to participate. The resulting increase in congestion and delay for customers due to the loss of scale economies reduces a customer's expected utility from seeking service, compelling the platform to reduce the price. Second, when the platform induces all low-cost agents to participate provided that there are at least $n \in \{1, \ldots, \bar{N} - 1\}$ of them, the platform (because of the agent participation externality) cedes rents to the low-cost agents when there are strictly more than n of them. Decreasing the price allows the platform to reduce the expected rents paid to agents, for the reasons provided following Proposition 1. (Just as Lemma 3 implies that inequality (12) is strict when agent uncertainty cost is high, Lemma A5 implies the same strict inequality when the correlation of the agents' opportunity costs ρ is moderate.)

Because the platform business model is attracting substantial attention from new ventures launching on-demand services, and because agent independence is a defining feature of this model, it is useful to understand how agent independence shapes the platform's optimal decisions. Propositions 1 and 2, which are summarized at the bottom right of Figure 1, provide guidance to a new venture that is considering offering an on-demand service with a platform business model instead of the traditional firm–employee business model. They also provide guidance to firms with the former business model that are considering shifting to the latter, a path that has been taken by several on-demand service companies (e.g., Instacart, Munchery, Shyp; Bensinger 2015, Scheiber 2015). Some platforms are facing pressure from lawsuits and/or regulatory agencies to convert independent agents into employees, and so may be coerced into making this shift (Bensinger 2015, Scheiber 2015). The price prescriptions depend on the degree of uncertainty in customer valuation and agent opportunity cost. For example, Proposition 1 suggests that a platform shifting to a traditional firm–employee business model may increase its price in one geography (where customer sensitivity to weather states is low) and decrease price in another (where sensitivity is high).

Propositions 1 and 2 contrast with the finding by Gurvich et al. (2018) that agent independence always *increases* the platform's optimal price. Contrasting these results illuminates the forces through which agent independence influences the platform's price. In the Gurvich et al. (2018) price-dependent newsvendor formulation, the effect of agent independence is to increase the marginal cost for agents, with the consequence that the platform induces fewer agents to participate. We observe a similar phenomenon under cost uncertainty: the platform induces stochastically fewer agents to participate than the firm with employees activates. Strikingly, this common phenomenon of agent independence leading to fewer participating agents leads to opposite conclusions about the impact of agent independence on the optimal price—provided that in the queueing formulation the price elasticity of expected demand *decreases* with the number of agents. The driving force behind this divergence is that in the price-dependent newsvendor formulation, the analogous quantity, the price elasticity of expected sales, *increases* with the number of agents. (More formally, in Gurvich et al. 2018, expected sales is $n - n^2/[2D(p)]$, where $D(p)$ is decreasing and $n \in [0, D(p)]$; therefore, the price elasticity of expected sales, $-np[(\partial/\partial p)D(p)]/[2D(p)^2 - D(p)n]$, increases with the number of agents. For the queueing formulation, because every demand is converted into a sale, "demand" and "expected sales" have the same meaning.) When the price elasticity quantity increases (decreases), the revenue-maximizing price decreases (increases).

To build intuition, consider a platform with few agents and a fixed price. Having few agents translates to poor expected service: in the queueing formulation, the lack of scale economies translates to lengthy expected delays for customers seeking service, and in the price-dependent newsvendor formulation, the lack of capacity translates to a low fill rate (i.e., a high fraction of customers seeking service will be unable to obtain it). In the queueing formulation, this poor service discourages customers from seeking service, whereas in the price-dependent newsvendor formulation, the service level does not influence customer decisions to seek service. Consequently, in the queueing formulation, it is optimal for the platform with few agents to set a low price, to encourage customers to seek service despite the poor service level. In the price-dependent newsvendor formulation, it is optimal for the platform with few agents to set a high price, to sell only to customers with very high valuations.

In sum, Gurvich et al. (2018) demonstrate that customer valuation heterogeneity drives agent independence to *increase* the platform's price. In contrast, we demonstrate that service-level-dependent demand drives agent independence to *decrease* the platform's price. To isolate this effect and limit overlap with Gurvich et al. (2018), we have considered a setting with homogenous customers. In a formulation capturing both customer heterogeneity and service-level-dependent demand, we anticipate that agent independence may either increase or decrease the optimal price, depending on the relative strength of these two factors.

3.2. Delay Sensitivity

This section characterizes the impact customer delay sensitivity on the platform's optimal price and wage. To understand the impact of customer delay sensitivity, consider the *benchmark case* where the customer is insensitive to delay: Under realized valuation V, price p, demand rate λ, and n participating agents, a customer's expected utility from seeking service is $V - p$, provided that the system is stable, $\lambda/(n\mu) \le 1$. Let (p_0^*, ω_0^*) denote the platform's optimal price and wage in this benchmark case. Lemma 5 characterizes the impact of delay sensitivity on the platform's optimal decisions in the *benchmark setting* with no uncertainty in the customers' valuation or agents' opportunity costs: $\delta = 0$ and $\Delta = 0$.

Lemma 5. *In the benchmark setting with no uncertainty in the customers' valuation or agents' opportunity costs, the following hold*:

(a) *delay sensitivity decreases the platform's optimal price and increases the optimal wage*:

$$p^* < p_0^*, \tag{13}$$

$$\omega^* > \omega_0^*. \tag{14}$$

(b) *If $\bar{N} \in \{1, 2\}$, then the platform's optimal price p^* decreases and the optimal wage ω^* increases with the delay disutility c.*

Because delay sensitivity reduces expected utility for customers (directly, through waiting) and agents (indirectly, through idleness), the platform responds to delay sensitivity by decreasing the per-service price (to encourage participation of customers) and increasing the per-service wage (to encourage participation of agents). While Lemma 5(a) compares the extreme cases of delay insensitive and delay sensitive customers, Lemma 5(b) establishes a parallel result in terms of the delay disutility parameter c. To supplement Lemma 5(b), we conducted a numerical study: we observed that p^* decreases and ω^* increases with c in each of the 960 parameter combinations of $\mu = k = 1$, $v = 10$, $c \in \{1, 2, \ldots, 20\}$, and $\bar{N} \in \{3, 4, \ldots, 50\}$.

The remainder of this section explains how and why the results in Lemma 5 break when there is uncertainty in the customers' valuation or the agents' opportunity costs. Before doing so, we highlight that the assumption that need for service is abundant (inequality (1)) drives several of the benchmark-setting results: when inequality (1) is violated, inequality (13) continues to hold, but there exist parameters under which the other elements of Lemma 5 are reversed.

Lemma 6 characterizes the platform's optimal decisions in the benchmark case where customers are insensitive to delay. Let $\bar{\delta}_0 = (v\mu - k)/(3\mu)$ and $\tilde{\Delta}_0 = (v\mu - k)/3$.

Lemma 6. *In the benchmark case where customers are insensitive to delay, the following hold*:

(a) *under customer valuation uncertainty, the platform's optimal wage is $\omega_0^* = k/\mu$; the optimal price is $p_0^* = V^l$ if $\delta \le \bar{\delta}_0$, and $p_0^* = V^h$ otherwise.*

(b) *Under agent opportunity cost uncertainty, $p_0^* = v$; $\omega_0^* = K^H/\mu$ if $\Delta \le \tilde{\Delta}_0$, and $\omega_0^* = K^L/\mu$ otherwise.*

What is the impact of customer delay sensitivity on the platform's optimal price and wage, under customer valuation uncertainty? As in the benchmark setting without uncertainty, delay sensitivity increases the optimal wage (inequality (14)). Proposition 3(a) establishes that the benchmark-setting result that delay sensitivity decreases the optimal price (inequality (13)) is reversed when customer valuation uncertainty is moderate $\delta \in (\underline{\delta}, \bar{\delta}_0]$. The proof of Proposition 3 shows that $\underline{\delta} < \bar{\delta}_0$ if and only if $\bar{\delta} < \bar{\delta}_0$ and $[p^h - V^l]|_{\delta=\bar{\delta}_0} > 0$. We observed numerically that $\underline{\delta} < \bar{\delta}_0$ in each of the 931 parameter combinations of $\mu = c = k = 1$, $v \in \{2, 3, \ldots, 20\}$, and $\bar{N} \in \{2, 3, \ldots, 50\}$. Similarly, Proposition 3b identifies conditions under which the benchmark-setting result that the platform's optimal wage p^* decreases with the delay disutility c (Lemma 5(b)) is reversed.

Proposition 3. *Under customer valuation uncertainty, the following hold*:

(a) *delay sensitivity increases the platform's optimal price,*

$$p^* > p_0^*, \tag{15}$$

if and only if valuation uncertainty is moderate $\delta \in (\underline{\delta}, \bar{\delta}_0]$.

(b) *If $\delta < \bar{\delta}_0$, then there exist c_l, c_m, and c_h such that $0 < c_l < c_m < c_h$ and $p^*|_{c \in [c_l, c_m]} < p^*|_{c \in (c_m, c_h]}$.*

The implication is that a platform interested in how customer delay sensitivity impacts its price decision should be wary of employing a naive intuition that such sensitivity compels a lower price; the opposite is true when customer valuation uncertainty is moderate. To understand why delay sensitivity *increases* the platform's optimal price in this case, observe that as valuation uncertainty increases, it becomes increasingly costly for the platform to offer the high wage required to induce agents to serve low-valuation customers. Furthermore, delay sensitivity reduces the marginal revenue from serving low-valuation customers. Consequently, when valuation uncertainty is moderate, the platform facing delay-sensitive customers gives up on serving low-valuation customers (and so charges a high price), whereas the platform facing delay-insensitive customers continues to do so (and so charges a low price). (In contrast, if valuation uncertainty is extreme ($\delta \le \underline{\delta}$ or $\delta > \bar{\delta}_0$), then the platform's optimal decision of whether to give up on serving low-valuation customers is the same regardless of whether customers are sensitive to delay. Consequently, the benchmark-setting result, inequality (13), continues to hold.)

What is the impact of customer delay sensitivity on the platform's optimal price and wage, under agent opportunity cost uncertainty? As in the benchmark setting without uncertainty, delay sensitivity decreases the optimal price (inequality (13)). Proposition 4(a) establishes that the benchmark-setting result that delay sensitivity increases the optimal wage (inequality (14)) is reversed when opportunity cost uncertainty Δ is high and the expected cost is moderate $k \in (\underline{k}, \bar{k})$, where $\underline{k} = \min(\Pr(N^L < \bar{N})R^A,\ R^A - R^M(p^M(1), 1))/(2\bar{N})$ and $\bar{k} = v\mu/4$. A sufficient condition for $\underline{k} < \bar{k}$ is that correlation $\rho > \bar{\rho}$ for some $\bar{\rho} < 1$ when $\bar{N} \ge 2$; $\underline{k} < \bar{k}$ if $\bar{N} = 1$. We observed numerically that $\underline{k} < \bar{k}$ in each of the 2,793 parameter combinations of $\mu = c = k = 1$, $\rho \in \{0, 0.25, 0.75\}$, $v \in \{2, 3, \ldots, 20\}$, and $\bar{N} \in \{2, 3, \ldots, 50\}$. Similarly, Proposition 4(b) identifies conditions under which the benchmark-setting result that the platform's optimal wage ω^* increases with the delay disutility c (Lemma 5(b)) is reversed. Let $\check{\Delta}_0 = (v\mu - k)\Pr(N^L < \bar{N})/[1 + \Pr(N^L = \bar{N})]$.

Proposition 4. *Under agent opportunity cost uncertainty, the following hold*:

(a) *if the expected agent opportunity cost is moderate, $k \in (\underline{k}, \bar{k})$, and cost uncertainty is high, $\Delta > \tilde{\Delta}$, where $\tilde{\Delta} < k$, then delay sensitivity decreases the platform's optimal wage,*

$$\omega^* < \omega_0^*.$$

(b) *There exists $\bar{\rho} < 1$ such that if $\rho > \bar{\rho}$ and $\Delta < \check{\Delta}_0$, then there exist c_l, c_m, and c_h such that $0 < c_l < c_m < c_h$ and $\omega^*|_{c \in [c_l, c_m]} > \omega^*|_{c \in (c_m, c_h)}$.*

The implication is that a platform interested in how customer delay sensitivity impacts its wage decision should be wary of employing a naive intuition that such sensitivity, by reducing the utilization of agents, compels a higher wage. To understand when and why this intuition fails, consider the case where agent cost uncertainty is high, $\Delta > \tilde{\Delta}$. Then it is quite costly for the platform to induce high-cost agents to participate. When the expected agent opportunity cost is sufficiently high, $k > \underline{k}$, in the base case with delay-sensitive customers, the marginal revenue generated

Table 2. Examples of Platforms' On-Demand Service Offerings

Customer sensitivity to delay	
High	Low
Restaurant food delivery	Consumer goods delivery
Taxi-style transportation	Home services

by an incremental agent is too low for the platform to incur the high cost to induce high-cost agents to participate, and the platform offers a low wage. Eliminating customer delay sensitivity increases this marginal revenue, which makes it optimal for the platform to offer the high wage required to induce all agents to participate, provided that the expected agent opportunity cost is not too high, $k < \bar{k}$. Thus, if the agent opportunity cost uncertainty is high, $\Delta > \tilde{\Delta}$, and the expected opportunity cost is moderate, $k \in (\underline{k}, \bar{k})$, then delay sensitivity *decreases* the optimal wage.

Propositions 3 and 4, which are summarized at the top right of Figure 1, may provide directional guidance for how a platform should change its price and wage when its customers' delay sensitivity changes. For example, customers' sensitivity to delay may increase as customers grow acclimated to the service and become more demanding. Alternatively, a platform may shift its offering to a service that differs in customer delay sensitivity. (Table 2 classifies on-demand service offerings by customer delay sensitivity.) For example, the platform Sidecar initially provided taxi-style transportation, but later shifted to consumer goods delivery, where customers tend to be less sensitive to delay (MacMillan 2015, Wang 2015). Similarly, the platform Postmates initially provided restaurant food delivery, but later expanded to consumer goods delivery, where customers tend to be less sensitive to delay (Ruggless 2015). The platform TaskRabbit initially provided home services (e.g., repair, cleaning) in a scheduled fashion, but later shifted to offering services on-demand (Solomon 2016).

4. Discussion

This paper examines how two defining features of an on-demand service platform—customer delay sensitivity and agent independence—shape the platform's price and wage decisions. By reducing expected utility for customers and agents, customer delay sensitivity decreases the optimal price and increases the optimal wage—provided that the customers' valuation and agents' opportunity costs are deterministic. However, uncertainty in either dimension can reverse these results: Delay sensitivity *increases* the optimal price when customer valuation uncertainty is moderate. Delay sensitivity *decreases* the optimal wage when agent opportunity cost uncertainty is high and expected opportunity cost is moderate. The intuition is that delay sensitivity decreases the marginal revenue from serving low-valuation customers and from inducing high-cost agents to participate. Consequently, delay sensitivity prompts the platform to give up on serving low-valuation customers (and thus charge a high price) and give up on inducing high-cost agents to participate (and thus offer a low wage).

Two forces push agent independence to *decrease* the price. First, by reducing agent idleness, decreasing the price allows the platform to reduce the wage and expected rent paid to each agent. Second, because agent independence makes it more costly for the platform to induce agent participation, the platform induces fewer agents to work. Compensating customers for the degraded service pushes the platform to decrease the price. A distinct force pushes agent independence to *increase* the price: Agent independence makes it costly for the platform to induce agents to serve low-valuation customers, which pushes the platform to give up on serving these customers (and so charge a high price). The collective result of these three forces is that under agent opportunity cost uncertainty, agent independence decreases the price, strictly so if cost uncertainty is high. Under customer valuation uncertainty, agent independence strictly decreases the price if and only if valuation uncertainty is sufficiently low.

Thus, a central message is that agent independence strictly decreases the price if agent opportunity cost uncertainty is high or valuation uncertainty is low. This message, established when uncertainty is present on either the customer or agent side, is preserved when there is joint uncertainty of the following form: when the weather is "bad," all customers have a high valuation, and all agents have a high opportunity cost; when the weather is "good," all customers have a low valuation, and all agents have a low opportunity cost. Then, under high opportunity cost uncertainty and low valuation uncertainty, agent independence decreases the price. See the online supplement for a formal analysis and proof.

Our assumption that the platform must commit to its price and wage in advance drives some of our results. If, instead, the platform sets its price and wage after observing the system state, the benchmark result that delay sensitivity decreases the optimal price is restored. The results under opportunity cost uncertainty are robust: delay sensitivity can increase or decrease the optimal wage, and agent independence decreases the optimal price. However, the results under valuation uncertainty differ: the optimal price is invariant to whether agents are independent.

Acknowledgments

The author is grateful to Costis Maglaras, Erica Plambeck, and Ramandeep Randhawa as well as seminar participants at

Boston College, Boston University, Duke University, Emory University, London Business School, Santa Clara University, Stanford University, University College London, University of Minnesota, University of Pennsylvania, and University of Utah, for helpful comments.

Appendix

Lemma A1, which provides properties of the equilibrium demand rate $\lambda(V,p,n)$ under n agents and realized valuation V, is useful in the proofs of Lemmas A2, A4, 3, 4 and 6 and Propositions 2 and 3.

Lemma A1. (a) *The equilibrium demand allocated to each agent $\lambda(V,p,n)/n$ strictly increases with the number of agents n.*

(b) $(\partial/\partial p)\lambda(V,p,n)<0$, $(\partial^2/\partial p^2)\lambda(V,p,n)<0$, $(\partial/\partial V)\cdot\lambda(V,p,n)>0$, *and* $(\partial^2/\partial p\partial V)\lambda(V,p,n)>0$.

(c) *The price elasticity of demand,* $-(\partial\lambda(V,p,n)/\partial p)\cdot(p/(\lambda(V,p,n)))$, *strictly decreases with the customers' valuation V.*

Proof of Lemma A1. (a) For the $M/M/n$ system, the traffic intensity is $\lambda/n\mu$. Thus, if $\lambda=\theta n$, where $\theta\in(0,\mu)$, then the traffic intensity is θ/μ. For fixed traffic intensity, expected waiting time $W(\theta n,n)$ strictly decreases with the number of agents n. Let $\gamma(n)$ denote the value of γ that satisfies $V-p-cW(\gamma n,n)=0$. Note $\gamma(n)=\lambda(V,p,n)/n$. For $n_2>n_1$,

$$W(\gamma(n_2)n_2,n_2)=(V-p)/c=W(\gamma(n_1)n_1,n_1)>W(\gamma(n_1)n_2,n_2),\qquad\text{(A.1)}$$

where the equalities follow from the definition of $\gamma(\cdot)$ and the inequality follows because $W(\theta n,n)$ strictly decreases with n. Inequality (A.1), together with the fact that $W(\theta n,n)$ strictly increases with θ, implies that $\gamma(n_2)>\gamma(n_1)$.

(b) By the implicit function theorem, $(\partial/\partial p)\lambda(V,p,n)=-1/[c(\partial/\partial\lambda)W(\lambda,n)]|_{\lambda=\lambda(V,p,n)}<0$ and $(\partial^2/\partial p^2)\lambda(V,p,n)=-(\partial^2/\partial\lambda^2)W(\lambda,n)/\{c^2\times[(\partial/\partial\lambda)W(\lambda,n)]^3\}|_{\lambda=\lambda(v,p,n)}<0$, where the inequalities follow because $W(\lambda,n)$ is a strictly convex function, strictly increasing in λ. Furthermore, we have $(\partial/\partial V)\lambda(V,p,n)=-(\partial/\partial p)\lambda(V,p,n)$ and $(\partial^2/\partial p\partial V)\lambda(V,p,n)=-(\partial^2/\partial p^2)\lambda(V,p,n)$.

(c) Note

$$\frac{\partial}{\partial V}\left[-\frac{\partial\lambda(V,p,n)}{\partial p}\frac{p}{\lambda(V,p,n)}\right]=\frac{p}{\lambda(V,p,n)^2}\left[\frac{\partial\lambda(V,p,n)}{\partial V}\frac{\partial\lambda(V,p,n)}{\partial p}-\lambda(V,p,n)\frac{\partial^2\lambda(V,p,n)}{\partial p\partial V}\right]<0,$$

where the inequality follows from part (b). □

For use in the statements and/or proofs of Lemmas A2, 2, and 4 and Proposition 3, let $\Pi^a(p)=p\sum_{j\in\{h,l\}}\lambda(V^j,p,\bar N)/2-\bar Nk$, $\Pi^h(p)=[p\lambda(V^h,p,\bar N)-\bar Nk]/2$ and $\Pi^l(p)=[p-\bar Nk/\lambda(V^l,p,\bar N)]\times\sum_{j\in\{h,l\}}\lambda(V^j,p,\bar N)/2$. Let $p^j\in\arg\max_{p\ge0}\Pi^j(p)$ and $\Pi^j=\Pi^j(p^j)$ for $j\in\{a,h,l\}$. (Note these definitions of p^h and p^l are consistent with those provided in body of the paper.) Lemma A2, which establishes properties of these functions and quantities, is useful in the proofs of Lemmas 2, A3, and 4 and Proposition 1.

Lemma A2. (a) *$\Pi^j(p)$ is strictly concave for $j\in\{a,h\}$.*

(b) $p^h>p^a>p^l$.

(c) $(\partial/\partial\delta)\Pi^h>\max((\partial/\partial\delta)\Pi^a,(\partial/\partial\delta)\Pi^l)$.

Proof of Lemma A2. (a) Note that

$$(\partial^2/\partial p^2)\Pi^a(p)=\sum_{j\in\{h,l\}}\big[p(\partial^2/\partial p^2)\lambda(V^j,p,\bar N)/2+(\partial/\partial p)\lambda(V^j,p,\bar N)\big]<0,$$

where the inequality follows from Lemma A1(b). That $\Pi^h(p)$ is strictly concave in p follows by parallel argument.

(b) Note that

$$(\partial/\partial p)\Pi^h(p)=\lambda(V^h,p,\bar N)\{1+[(\partial/\partial p)\lambda(V^h,p,\bar N)]p/\lambda(V^h,p,\bar N)\}/2$$
$$(\partial/\partial p)\Pi^a(p)=(\partial/\partial p)\Pi^h(p)+\lambda(V^l,p,\bar N)\cdot\{1+[(\partial/\partial p)\lambda(V^l,p,\bar N)]p/\lambda(V^l,p,\bar N)\}/2.$$

Therefore,

$$\begin{aligned}(\partial/\partial p)\Pi^a(p)|_{p=p^h}&=\lambda(V^l,p^h,\bar N)\{1+[(\partial/\partial p)\lambda(V^l,p^h,\bar N)]p^h/\lambda(V^l,p^h,\bar N)\}/2\\&<\lambda(V^h,p^h,\bar N)\{1+[(\partial/\partial p)\lambda(V^h,p^h,\bar N)]p^h/\lambda(V^h,p^h,\bar N)\}/2\\&=0,\end{aligned}\qquad\text{(A.2)}$$

where the inequality follows from Lemma A1(b) and A1(c), and the equalities follow because p^h satisfies the first order condition $(\partial/\partial p)\Pi^h(p)=0$. Because $\Pi^a(p)$ is strictly concave in p (by part (a)), (A.2) implies $p^h>p^a$. It remains to show that $p^a>p^l$. Note that $\Pi^l(p)=\Pi^a(p)-f(p)$, where $f(p)=[\lambda(V^h,p,\bar N)/\lambda(V^l,p,\bar N)-1]\bar Nk/2$. Furthermore,

$$\frac{\partial f(p)}{\partial p}=\frac{\lambda(V^h,p,\bar N)}{p\lambda(V^l,p,\bar N)}\left[\frac{\partial\lambda(V^h,p,\bar N)}{\partial p}\frac{p}{\lambda(V^h,p,\bar N)}-\frac{\partial\lambda(V^l,p,\bar N)}{\partial p}\frac{p}{\lambda(V^l,p,\bar N)}\right]\frac{\bar Nk}{2}>0,\qquad\text{(A.3)}$$

where the inequality follows from Lemma A1(c). It immediately follows that $(\partial/\partial p)\Pi^l(p)|_{p=p^a}<0$, so $p^a\ne p^l$. Furthermore, $f(p^a)-f(p^l)>\Pi^a(p^a)-\Pi^a(p^l)>0$, where the first inequality holds because p^l maximizes $\Pi^l(p)$, and the second inequality holds because p^a maximizes $\Pi^a(p)$. Together, $f(p^a)>f(p^l)$ and inequality (A.3) imply $p^a>p^l$.

(c) Next we show that $(\partial/\partial\delta)\Pi^h>(\partial/\partial\delta)\Pi^a$. To see this, observe that

$$(\partial/\partial p)[p(\partial/\partial V)\lambda(V^h,p,\bar N)]=(\partial/\partial V)\lambda(V^h,p,\bar N)+p(\partial^2/\partial p\partial V)\lambda(V^h,p,\bar N)>0,\qquad\text{(A.4)}$$

where the inequality follows by Lemma A1(b). Furthermore,

$$\begin{aligned}(\partial/\partial\delta)\Pi^h&=p^h(\partial/\partial V)\lambda(V^h,p^h,\bar N)/2\\&>p^a(\partial/\partial V)\lambda(V^h,p^a,\bar N)/2\\&>p^a[(\partial/\partial V)\lambda(V^h,p^a,\bar N)-(\partial/\partial V)\lambda(V^l,p^a,\bar N)]/2\\&=(\partial/\partial\delta)\Pi^a,\end{aligned}$$

where the first inequality follows from (A.4) and $p^h>p^a$ (from part (b)), and the second inequality follows from Lemma A1(b). It remains to show that $(\partial/\partial\delta)\Pi^h>(\partial/\partial\delta)\Pi^l$. Note that

$$\begin{aligned}(\partial/\partial\delta)\Pi^h&=p^h(\partial/\partial V)(V^h,p^h,\bar N)/2>p^l[(\partial/\partial V)(V^h,p^l,\bar N)/2\\&>\{p^l(\partial/\partial V)(V^h,p^l,\bar N)-p^l(\partial/\partial V)(V^l,p^l,\bar N)\\&\quad-\bar Nk\sum_{(i,j)\in\{h,l\}^2,\,i\ne j}[(\partial/\partial V)(V^i,p^l,\bar N)]\\&\quad\times\lambda(V^j,p^l,\bar N)/\lambda(V^l,p^l,\bar N)^2\}/2=(\partial/\partial\delta)\Pi^l,\end{aligned}$$

where the first inequality follows from (A.4) and $p^h > p^l$ (from part (b)), and the second inequality follows from Lemma A1(b). □

Lemma A3, which characterizes the platform's optimal price and wage as a function of the delay disutility c, is useful in the proof of Proposition 3(b). Let $\bar{\delta}_0 = (v\mu - k)/(3\mu)$.

Lemma A3. *Under customer valuation uncertainty, if $\delta < \bar{\delta}_0$, then there exist $\underline{c}$ and $\bar{c}$ such that $0 < \underline{c} \le \bar{c}$ and the platform's optimal price and wage are $(p^*, \omega^*) = (p^l, \omega^l)$ if $c \le \underline{c}$ and $(p^*, \omega^*) = (p^h, \omega^h)$ if $c > \bar{c}$.*

Let $\bar{\delta}$ denote the value of δ such that $\Pi^h = \Pi^l$; the proof of Lemmas 2 and A3 establishes that $\bar{\delta}$ is unique.

Proof of Lemmas 2 and A3. The proof proceeds in three steps. The first step establishes that, in terms of identifying an optimal price and wage, the set of prices and wages can be reduced to two candidates. The second step completes the proof of Lemma 2, and the third step completes the proof Lemma A3.

Step 1. We first characterize the equilibrium number of participating agents $N(V, N^L)$ under various wages, prices, and realized valuations V. Because all agents have opportunity cost k, we drop the second argument of $\mathcal{N}(V^l, N^L)$. If price $p < V^l$ and wage $\omega \ge \bar{N}k/\lambda(V^l, p, \bar{N})$, then $\mathcal{N}(V) = \bar{N}$ for $V \in \{V^h, V^l\}$. If $p < V^h$ and $\omega \in [\bar{N}k/\lambda(V^h, p, \bar{N}), \bar{N}k/\lambda(V^l, p, \bar{N}))$, then $\mathcal{N}(V^h) = \bar{N}$ and $\mathcal{N}(V^l) = 0$. Otherwise, $\mathcal{N}(V) = 0$, for $V \in \{V^h, V^l\}$. Consequently, we can restrict attention to two wages $\omega \in \{\bar{N}k/\lambda(V^j, p, \bar{N})\}_{j\in\{h,l\}}$. Under $\omega = \bar{N}k/\lambda(V^j, p, \bar{N})$, the platform's objective function, $\Pi^j(p)$, is maximized at price $p = p^j$ for $j \in \{h, l\}$. Thus, the set of prices and wages can be reduced to two candidates: $(p, \omega) = (p^j, \omega^j)$, with corresponding expected profit rate Π^j for $j \in \{h, l\}$.

Step 2. As $\delta \to 0$, $\Pi^h \to [p^A\lambda(v, p^A, \bar{N}) - \bar{N}k]/2$ and $\Pi^l \to p^A\lambda(v, p^A, \bar{N}) - \bar{N}k$, where $p^A \in \arg\max_{p\ge 0}\{p\lambda(v, p, \bar{N})\}$. Because the platform's expected profit rate under the optimal price and wage, $\max(\Pi^h, \Pi^l)$, is strictly positive, $p^A\lambda(v, p^A, \bar{N}) - \bar{N}k > 0$; therefore, as $\delta \to 0$, $\Pi^h - \Pi^l \to -[p^A\lambda(v, p^A, \bar{N}) - \bar{N}k]/2 < 0$. As $\delta \to v$, $\Pi^h \to [p^h\lambda(2v, p^h, \bar{N}) - \bar{N}k]/2 > 0$ and $\Pi^l \to -\infty$ because $\lambda(V^l, p^l, \bar{N}) \to 0$. These observations, together with the inequality $(\partial/\partial\delta)\Pi^h > (\partial/\partial\delta)\Pi^l$ (by Lemma A2(c)) imply that there exists a unique $\delta \in (0, v)$, namely, $\bar{\delta}$, such that $\Pi^h = \Pi^l$. Furthermore, $\Pi^h \le \Pi^l$ if and only if $\delta \le \bar{\delta}$.

Step 3. As $c \to 0$, $\Pi^l \to (V^l\mu - k)\bar{N}$ and $\Pi^h \to (V^h\mu - k)\bar{N}/2$. Therefore, $\delta < \bar{\delta}_0$ implies $\lim_{c\to 0}[\Pi^l - \Pi^h] > 0$. Because Π^h and Π^l are continuous in c, there exists $\underline{c} > 0$ such that $\Pi^l > \Pi^h$ if $c \in (0, \underline{c})$. Let $c^a = \sup\{c\colon \Pi^a \ge \Pi^h\}$. With a little effort, one can show that our assumption that the firm activates at least one employee agent, for any realized valuation V, implies that $c \le c^a$. Because Π^h and Π^a are continuous in c, at $c = c^a$, $\Pi^h = \Pi^a$. Therefore, because $\Pi^a > \Pi^l$, at $c = c^a$, $\Pi^h > \Pi^l$. Because Π^h and Π^l are continuous in c, there exists $\bar{c} < c^a$ such that if $c \in (\bar{c}, c^a]$, then $\Pi^h > \Pi^l$. □

For use in the statements and/or proofs of Lemmas A4 and 3 and Propositions 2 and 4, let $\Pi^H = R^A - \bar{N}K^H$, $\Pi^L = \Pr(N^L = \bar{N})[R^A - \bar{N}K^L]$, and $\Pi^M = \Pi^M(p^M, n^M)$. Lemma A4 is useful in the proof of Lemma 3.

Lemma A4. (a) $\lim_{\Delta\to 0}[\Pi^H - \max(\Pi^L, \Pi^M)] > 0$.

(b) $\lim_{\Delta\to k}[\Pi^M - \Pi^H] > 0$ *if and only if $k > \tilde{k}$. Furthermore, $\lim_{\Delta\to k}[\Pi^M - \Pi^L] \ge 0$, where the inequality is strict if and only if $\rho < 1$.*

Proof of Lemma A4. (a) Note $\lim_{\Delta\to 0}\Pi^H = R^A - \bar{N}k$ and $\lim_{\Delta\to 0}\Pi^L = \Pr(N^L = \bar{N})[R^A - \bar{N}k]$. Because $\Pr(N^L = \bar{N}) < 1$, $\lim_{\Delta\to 0}[\Pi^H - \Pi^L] > 0$. It remains to show that $\lim_{\Delta\to 0}[\Pi^H - \Pi^M] > 0$. Recall that $n^M < \bar{N}$. Note

$$\begin{aligned}\lim_{\Delta\to 0}\Pi^M &= \max_{p\ge 0}\left\{[p - n^M k/\lambda(v, p, n^M)]\sum_{j=n^M}^{\bar{N}}\Pr(N^L = j)\lambda(v, p, j)\right\}\\ &\le \sum_{j=n^M}^{\bar{N}}\Pr(N^L = j)\max_{p\ge 0}\{[p - n^M k/\lambda(v, p, n^M)]\lambda(v, p, j)\}\\ &< \sum_{j=n^M}^{\bar{N}}\Pr(N^L = j)\max_{p\ge 0}\{p\lambda(v, p, j) - jk\}\\ &< \sum_{j=n^M}^{\bar{N}}\Pr(N^L = j)\max_{p\ge 0}\{p\lambda(v, p, \bar{N}) - \bar{N}k\} \le \lim_{\Delta\to 0}\Pi^H,\end{aligned}$$

where the second inequality follows because $n^M/\lambda(v, p, n^M) > j/\lambda(v, p, j)$ for $j > n^M$ (from Lemma A1(a)), and the third inequality follows because $j[p\lambda(v, p, j)/j - k] < \bar{N}[p\lambda(v, p, \bar{N})/\bar{N} - k]$ for $j < \bar{N}$, which holds because $\lambda(v, p, j)/j < \lambda(v, p, \bar{N})/\bar{N}$ (from Lemma A1(a)).

(b) Note $\lim_{\Delta\to k}\Pi^H = R^A - 2\bar{N}k$ and $\lim_{\Delta\to k}\Pi^M = R^M(1)$. Therefore, $\lim_{\Delta\to k}[\Pi^M - \Pi^H] > 0$ if and only if $k > \tilde{k}$. Furthermore, $\lim_{\Delta\to k}\Pi^L = \Pr(N^L = \bar{N})R^A = \max_{p\ge 0}R^M(p, \bar{N})$, and $\lim_{\Delta\to k}\Pi^M = \max_{p\ge 0}R^M(p, 1)$. Because $R^M(p, 1) \ge R^M(p, \bar{N})$, where the inequality is strict if and only if $\rho < 1$, $\lim_{\Delta\to k}[\Pi^M - \Pi^L] \ge 0$, where the inequality is strict if and only if $\rho < 1$. □

Lemma A5. *Under agent opportunity cost uncertainty, the platform's optimal price and wage are $(p^*, \omega^*) = (p^A, \omega^H)$ if $\rho \le \rho_l$, $(p^*, \omega^*) = (p^M, \omega^M)$ if $\rho \in (\rho_l, \rho_h]$, and $(p^*, \omega^*) = (p^A, \omega^L)$ if $\rho > \rho_h$, where $\rho_l \le \rho_h$.*

Lemma A6, which characterizes the platform's optimal price and wage as a function of the delay disutility c, is useful in the proof of Proposition 4(b). Let $\bar{\Delta}_0 = (v\mu - k)\Pr(N^L < \bar{N})/[1 + \Pr(N^L = \bar{N})]$.

Lemma A6. *Under agent opportunity cost uncertainty, there exists $\bar{\rho} < 1$ and $\bar{c}$ such that if $\rho > \bar{\rho}$, then the platform's optimal price is $p^* = p^A$, and the platform's optimal wage is $\omega^* = \omega^H$ if $c \le \bar{c}$ and $\omega^* = \omega^L$ if $c > \bar{c}$. Furthermore, $\bar{c} > 0$ if and only if $\Delta < \bar{\Delta}_0$.*

Proof of Lemmas 3, A5, and A6. The proof proceeds in four steps. The first step establishes that, in terms of identifying an optimal price and wage, the set of prices and wages can be reduced to three candidates. The second step completes the proof of Lemma 3, the third step completes the proof Lemma A5, and the fourth step completes the proof Lemma A6.

Step 1. We first characterize the equilibrium number of participating agents $\mathcal{N}(v, N^L)$ under various wages and realizations of N^L, for any price $p < v$. If the wage $\omega \ge \bar{N}K^H/\lambda(v, p, \bar{N})$, then $\mathcal{N}(v, N^L) = \bar{N}$. If $\omega \in [nK^L/\lambda(v, p, n), \bar{N}K^H/\lambda(v, p, \bar{N}))$ for $n \in \{1, \ldots, \bar{N}\}$ and $N^L \ge n$, then $\mathcal{N}(v, N^L) = N^L$ (because all low-cost and no high-cost agents participate). Otherwise, $\mathcal{N}(v, N^L) = 0$. Consequently, we can restrict attention to wages $\omega \in \{\bar{N}K^H/\lambda(v, p, \bar{N}), \bar{N}K^L/\lambda(v, p, \bar{N}), nK^L/\lambda(v, p, n)\}$, where $n \in \{1, \ldots, \bar{N} - 1\}$. Under $\omega = \bar{N}K^H/\lambda(v, p, \bar{N})$, the platform's objective function, $p\lambda(v, p, \bar{N}) - \bar{N}K^H$, is maximized at price $p = p^A$, yielding expected profit rate Π^H. Similarly, under $\omega = \bar{N}K^L/\lambda(v,p,\bar{N})$, the platform's objective function, $\Pr(N^L=\bar{N})[p\lambda(v,p,\bar{N}) - \bar{N}K^L]$, is maximized at price $p=p^A$, yielding expected profit rate Π^L. Under $\omega = nK^L/\lambda(v,p,n) < \bar{N}K^H/\lambda(v,p,\bar{N})$, where $n\in\{1,\dots,\bar{N}-1\}$, the platform's objective function, $\Pi^M(p,n)$, is maximized at $(p,n)=(p^M,n^M)$, yielding expected profit rate Π^M. Note that if $n^MK^L/\lambda(v,p^M,n^M) \ge \bar{N}K^H/\lambda(v,p,\bar{N})$, then $\Pi^H > \Pi^M$. To summarize, the set of prices and wages can be reduced to three candidates: $(p,\omega)=(p^A,\omega^j)$, with corresponding expected profit rate Π^j, for $j\in\{H,L\}$, and $(p,\omega)=(p^M,\omega^M)$, with corresponding rate Π^M.

Step 2. Suppose $k\le\tilde{k}$. Then, $\lim_{\Delta\to k}[\Pi^H - \max(\Pi^M,\Pi^L)] \ge 0$ (by Lemma A4(b)). Because Π^H strictly decreases with Δ and Π^j strictly increases with Δ for $j\in\{L,M\}$, this implies that $\Pi^H > \max(\Pi^M,\Pi^L)$ for $\Delta\in(0,k)$; that is, $(p^*,\omega^*)=(p^A,\omega^H)$ for $\Delta\in(0,k)$. For the remainder of this proof, suppose instead that $k>\tilde{k}$. Note $\Pr(N^L=\bar{N}) = \rho/2+(1-\rho)(1/2)^{\bar{N}}$. Suppose $\rho=1$. Then $\Pr(N^L=\bar{N})=\Pr(N^L=0)=1/2$. Furthermore,

$$\begin{aligned}\Pi^M &= [p^M - n^MK^L/\lambda(v,p^M,n^M)]\lambda(v,p^M,\bar{N})/2\\ &< [p^M\lambda(v,p^M,\bar{N}) - \bar{N}K^L]/2\\ &\le [p^A\lambda(v,p^A,\bar{N}) - \bar{N}K^L]/2 = \Pi^L,\end{aligned} \tag{A.5}$$

where the first inequality follows because $n^M<\bar{N}$ implies $\lambda(v,p^M,n^M)/n^M < \lambda(v,p^M,\bar{N})/\bar{N}$ (by Lemma A1(a)). Inequality (A.5) implies that $p^*=p^A$. Because Π^H strictly decreases with Δ, Π^L strictly increases with Δ, $\lim_{\Delta\to0}[\Pi^H-\Pi^L]>0$ (by Lemma A4(a)), and $\lim_{\Delta\to k}[\Pi^H-\Pi^L]<0$ (by Lemma A4(b)), there exists $\underline{\Delta}\in(0,k)$ such that $\Pi^H\ge\Pi^L$ if and only if $\Delta\le\underline{\Delta}$; that is, $\omega^*=\omega^H$ if $\Delta\le\underline{\Delta}$, and $\omega^*=\omega^L$ otherwise. For the remainder of this proof suppose instead that $\rho<1$. First, we establish that

$$(\partial/\partial\Delta)\Pi^M > (\partial/\partial\Delta)\Pi^L. \tag{A.6}$$

To do so, we establish that $(\partial/\partial\Delta)\Pi^M(p^M(n),n)) > (\partial/\partial\Delta)\Pi^L$ for $n\in\{1,\dots,\bar{N}-1\}$. Note

$$\begin{aligned}&(\partial/\partial\Delta)\Pi^M(p^M(n),n)\\ &\quad=\sum_{j=n}^{\bar{N}}\Pr(N^L=j)\lambda(v,p^M(n),j)n/\lambda(v,p^M(n),n)\\ &\quad>\Pr(N^L=\bar{N})\lambda(v,p^M(n),\bar{N})n/\lambda(v,p^M(n),n)\\ &\quad>\Pr(N^L=\bar{N})\bar{N} = (\partial/\partial\Delta)\Pi^L,\end{aligned} \tag{A.7}$$

where the second inequality holds because $n<\bar{N}$ implies $\lambda(v,p^M(n),n)/n < \lambda(v,p^M(n),\bar{N})/\bar{N}$ (by Lemma A1(a)). Inequality (A.7) implies (A.6). Because Π^H strictly decreases with Δ, $\max(\Pi^L,\Pi^M)$ strictly increases with Δ, $\lim_{\Delta\to0}[\Pi^H-\max(\Pi^L,\Pi^M)]>0$ (by Lemma A4(a)), and $\lim_{\Delta\to k}[\Pi^H-\max(\Pi^L,\Pi^M)]<0$ (by Lemma A4(b)), there exists $\underline{\Delta}\in(0,k)$ such that $\Pi^H\ge\max(\Pi^L,\Pi^M)$ if and only if $\Delta\le\underline{\Delta}$. Therefore, if $\Delta\le\underline{\Delta}$, then $(p^*,\omega^*)=(p^A,\omega^H)$. Because inequality (A.6) holds and $\lim_{\Delta\to k}[\Pi^M-\Pi^L]>0$ (by Lemma A4(b)), there exists $\check{\Delta}<k$ such that $\Pi^M>\Pi^L$ if and only if $\Delta>\check{\Delta}$. Let $\bar{\Delta}=\max(\underline{\Delta},\check{\Delta})$. If $\Delta\in(\underline{\Delta},\bar{\Delta}]$, then $(p^*,\omega^*)=(p^A,\omega^L)$. If $\Delta>\bar{\Delta}$, then $(p^*,\omega^*)=(p^M,\omega^M)$.

Step 3. It is sufficient to show that

$$0=(\partial/\partial\rho)\Pi^H < (\partial/\partial\rho)\Pi^M < (\partial/\partial\rho)\Pi^L. \tag{A.8}$$

The equality is immediate. Because $\Pi^M = \Gamma\{\sum_{j=n^M}^{\bar{N}-1}\binom{\bar{N}}{j}(1-\rho)\cdot(1/2)^{\bar{N}}\lambda(v,p^M,j) + [\rho/2+(1-\rho)(1/2)^{\bar{N}}]\lambda(v,p^M,\bar{N})\}$, where $\Gamma=[p^M-n^MK^L/\lambda(v,p^M,n^M)]$,

$$\begin{aligned}&(\partial/\partial\rho)\Pi^M\\ &\quad=\Gamma\left\{[1/2-(1/2)^{\bar{N}}]\lambda(v,p^M,\bar{N}) - (1/2)^{\bar{N}}\sum_{j=n^M}^{\bar{N}-1}\binom{\bar{N}}{j}\lambda(v,p^M,j)\right\}\\ &\quad>\Gamma\left\{1/2-(1/2)^{\bar{N}}-(1/2)^{\bar{N}}\sum_{j=1}^{\bar{N}-1}\binom{\bar{N}}{j}j/\bar{N}\right\}\lambda(v,p^M,\bar{N})=0,\end{aligned} \tag{A.9}$$

where the inequality follows because $\lambda(V,p,n)/n$ strictly increases with n (by Lemma A1(a)) and because $n^M\ge1$. This establishes the first inequality in (A.8). Because $\Pi^L=[\rho/2+(1-\rho)(1/2)^{\bar{N}}][p^A\lambda(v,p^A,\bar{N})-\bar{N}K^L]$,

$$\begin{aligned}(\partial/\partial\rho)\Pi^L &= [1/2-(1/2)^{\bar{N}}][p^A\lambda(v,p^A,\bar{N})-\bar{N}K^L]\\ &> [1/2-(1/2)^{\bar{N}}][p^M\lambda(v,p^M,\bar{N})\\ &\quad - n^MK^L\lambda(v,p^M,\bar{N})/\lambda(v,p^M,n^M)] > (\partial/\partial\rho)\Pi^M,\end{aligned}$$

where the first inequality follows because $\lambda(V,p,n)/n$ strictly increases with n (by Lemma A1(a)), and the second inequality follows from (A.9). This completes the proof of (A.8).

Step 4. First, we show that there exists $\bar{\rho}<1$ such that if $\rho>\bar{\rho}$, then $\Pi^L>\Pi^M$. We establish in the proof of Proposition 2 that $p^A\ne p^M$. Therefore, $\Pi^L > [\rho/2+(1-\rho)\cdot(1/2)^{\bar{N}}][p^M\lambda(v,p^M,\bar{N})-\bar{N}K^L]$, which implies

$$\begin{aligned}\Pi^L-\Pi^M &> [\rho/2+(1-\rho)(1/2)^{\bar{N}}]\\ &\quad\cdot[n^M\lambda(v,p^M,\bar{N})/\lambda(v,p^M,n^M)-\bar{N}]K^L\\ &\quad-(1-\rho)\Gamma\sum_{j=n^M}^{\bar{N}-1}\binom{\bar{N}}{j}(1/2)^{\bar{N}}\lambda(v,p^M,j).\end{aligned} \tag{A.10}$$

Because $n^M<\bar{N}$, $n^M/\lambda(v,p^M,n^M) > \bar{N}/\lambda(v,p^M,\bar{N})$ (by Lemma A1(a)), which implies that the second term in square brackets on the right-hand side of (A.10) is strictly positive. Therefore, $\lim_{\rho\to1}[\Pi^L-\Pi^M]>0$. Because Π^j is continuous in ρ for $j\in\{H,L\}$, this implies that there exists $\bar{\rho}<1$ such that if $\rho>\bar{\rho}$, then $\Pi^L>\Pi^M$. For the remainder of this proof, suppose that $\rho>\bar{\rho}$. This implies $p^*=p^A$. Note that

$$\begin{aligned}(\partial/\partial c)\Pi^H &= (\partial/\partial c)R^A < \Pr(N^L=\bar{N})(\partial/\partial c)R^A\\ &= (\partial/\partial c)\Pi^L.\end{aligned} \tag{A.11}$$

Our assumption that the platform's expected profit rate is strictly positive implies that $R^A>\bar{N}K^L$. Because R^A is strictly decreasing in c and $\lim_{c\to\infty}R^A=0$, there exists c^H such that $R^A>\bar{N}K^H$ if and only if $c<c^H$. Furthermore, (A.11) implies that there exists $\bar{c}<c^H$ such that $\Pi^H<\Pi^L$ if and only if $c>\bar{c}$. Note that $\lim_{c\to0}[\Pi^H-\Pi^L] = \bar{N}\{(v\mu-k)\Pr(N^L<\bar{N}) - \Delta[1+\Pr(N^L=\bar{N})]\}$, which is strictly positive if and only if $\Delta<\check{\Delta}_0$. Therefore, $\bar{c}>0$ if and only if $\Delta<\check{\Delta}_0$. ◻

Let δ^a denote the value of δ such that $\Pi^h=\Pi^a$; the proof of Lemma 4 establishes that δ^a is unique.

Proof of Lemma 4. (a) After setting price p and observing realized valuation V, the firm's employee-agent activation problem is $\max_{n\in\{0,1,\dots,\bar{N}\}}\{p\lambda(V,p,n)-nk\}$. Because $\lambda(V,p,n)/n$ strictly increases with n (by Lemma A1(a)),

$$\underset{n\in\{0,1,\dots,\bar{N}\}}{\arg\max}\{p\lambda(V,p,n)-nk\} = \begin{cases}\bar{N} & \text{if } k<p\lambda(V,p,\bar{N})/\bar{N},\\ 0 & \text{otherwise.}\end{cases}$$

If $k \geq p\lambda(V^h, p, \bar{N})/\bar{N}$, then it is never optimal for the firm to activate an agent. Because, by assumption, the firm activates at least one agent for any realized customer valuation, this implies that there exists $p \geq 0$ such that $k < p\lambda(V^h, p, \bar{N})/\bar{N}$ and that the optimal p satisfies this inequality. Under p satisfying $k < p\lambda(V^l, p, \bar{N})/\bar{N}$, the firm's objective function, $\Pi^a(p)$, is maximized at $p = p^a$, yielding expected profit rate Π^a. Under p satisfying $k \in [p\lambda(V^l, p, \bar{N})/\bar{N}, p\lambda(V^h, p, \bar{N})/\bar{N})$, the firm's objective function, $\Pi^h(p)$, is maximized at $p = p^h$, yielding expected profit rate Π^h. As $\delta \to 0$, $\Pi^h \to [p^A\lambda(v, p^A, \bar{N}) - \bar{N}k]/2$ and $\Pi^a \to p^A\lambda(v, p^A, \bar{N}) - \bar{N}k$. Because the firm's expected profit rate under the optimal price, $\max(\Pi^a, \Pi^h)$, is strictly positive, $p^A\lambda(v, p^A, \bar{N}) - \bar{N}k > 0$; therefore, as $\delta \to 0$, $\Pi^h - \Pi^a \to -[p^A\lambda(v, p^A, \bar{N}) - \bar{N}k]/2 < 0$. As $\delta \to v$, $\Pi^h \to [p^B\lambda(2v, p^B, \bar{N}) - \bar{N}k]/2$, $\Pi^a \to p^B\lambda(2v, p^B, \bar{N})/2 - \bar{N}k$, and $\Pi^h - \Pi^a \to \bar{N}k/2 > 0$, where, $p^B \in \arg\max_{p\geq 0}\{p\lambda(2v, p, \bar{N})\}$. These observations, together with the inequality $(\partial/\partial\delta)\Pi^h > (\partial/\partial\delta)\Pi^a$ (by Lemma A2(c)) imply that there exists a unique $\delta \in (0, v)$, namely, δ^a, such that $\Pi^h = \Pi^a$. Furthermore, $\Pi^a > \Pi^h$ if and only if $\delta < \delta^a$. Our assumption that the firm activates at least one agent for any realized customer valuation V is satisfied if and only if $\Pi^a > \Pi^h$. We conclude that for $\delta < \delta^a$, the firm's optimal price $p_I^* = p^a$, and the firm activates $\bar{N}$ agents. To see that $\delta^a > \bar{\delta}$, observe that at $\delta = \bar{\delta}$, $\Pi^h = \Pi^l < \Pi^a$, where the equality follows from the definition of $\bar{\delta}$, and the inequality follows from $\Pi^l(p) < \Pi^a(p)$ for $p \in (0, V^H)$.

(b) Because $\lambda(v, p, n)/n$ strictly increases with n (by Lemma A1(a)),

$$\begin{aligned}&\underset{n\in\{0,1,\ldots,\tilde{n}\}}{\arg\max}\{p\lambda(v,p,n) - nK\}\\ &\quad = \begin{cases}\tilde{n} & \text{if } K < p\lambda(v,p,\tilde{n})/\tilde{n},\\ 0 & \text{otherwise},\end{cases}\end{aligned} \tag{A.12}$$

where $\tilde{n} \in \{1, 2, \ldots, \bar{N}\}$ and $K \in \{K^H, K^L\}$. If $K^H \geq p_I^*\lambda(v, p_I^*, \bar{N})/\bar{N}$, then under the optimal price p_I^* it is optimal for the firm to activate zero agents when $N^L = 0$. Because, by assumption, the firm activates at least one agent, it must be that

$$K^H < p_I^*\lambda(v, p_I^*, \bar{N})/\bar{N}. \tag{A.13}$$

After setting price $p = p_I^*$ and observing realized costs $\{K_i\}_{i=1,\ldots,\bar{N}}$, the firm's employee agent activation problem is $\max_{n\in\{0,1,\ldots,\bar{N}\}}\{p_I^*\lambda(v, p_I^*, n) - \max(n - N^L, 0)K^H - \min(N^L, n)K^L\}$. Next we show that it is optimal for the firm to activate $\bar{N}$ agents, that is,

$$\begin{aligned}&p_I^*\lambda(v,p_I^*,\bar{N}) - (\bar{N} - N^L)K^H - N^LK^L\\ &\quad\geq p_I^*\lambda(v,p_I^*,n) - \max(n - N^L,0)K^H - \min(N^L,n)K^L\end{aligned} \tag{A.14}$$

for $n \in \{0, 1, \ldots, \bar{N} - 1\}$. To see that (A.14) holds for $n \in \{0, 1, \ldots, N^L\}$, observe that in that case the right-hand side of (A.14) simplifies to $p_I^*\lambda(v, p_I^*, n) - nK^L$, which is maximized at $n \in \{0, N^L\}$ (by (A.12)). Inequality (A.13) implies that the left-hand side of (A.14) is strictly positive, so (A.14) holds for $n = 0$. For $n = N^L$, (A.14) simplifies to

$$p_I^*\lambda(v,p_I^*,\bar{N}) - \bar{N}K^H \geq p_I^*\lambda(v,p_I^*,N^L) - N^LK^H. \tag{A.15}$$

That (A.15) holds follows from (A.12), where $\tilde{n} = \bar{N}$ and $K = K^H$, and (A.13). To see that (A.14) holds for $n \in \{N^L + 1, \ldots, \bar{N} - 1\}$, observe that in that case (A.14) simplifies to

$$p_I^*\lambda(v,p_I^*,\bar{N}) - \bar{N}K^H \geq p_I^*\lambda(v,p_I^*,n) - nK^H. \tag{A.16}$$

That (A.16) holds follows from (A.12), where $\tilde{n} = \bar{N}$ and $K = K^H$, and (A.13). □

Proof of Proposition 1. Note that $p_I^* = p^a$ (by Lemma 4(a)). The result then follows from Lemmas A2(b) and 2. □

Proof of Proposition 2. In view of Lemmas 3 and 4(b), to establish the result it is sufficient to show that

$$p^M < p^A. \tag{A.17}$$

Note that for $n \in \{1, \ldots, \bar{N} - 1\}$,

$$\begin{aligned}&(\partial^2/\partial p\partial\Delta)\Pi^M(p,n)\\ &\quad = n\sum_{j=n}^{\bar{N}}\Pr(N^L = j)(\partial/\partial p)[\lambda(v,p,j)/\lambda(v,p,n)]\\ &\quad = [n/\lambda(v,p,n)]\sum_{j=n}^{\bar{N}}\Pr(N^L = j)[\lambda(v,p,j)/p]\\ &\qquad\times\{[(\partial/\partial p)\lambda(v,p,j)]p/\lambda(v,p,j)\\ &\qquad\quad - [(\partial/\partial p)\lambda(v,p,n)]p/\lambda(v,p,n)\} > 0,\end{aligned} \tag{A.18}$$

where the inequality holds because the price elasticity of demand decreases with the number of agents. Let $p^M(n, \Delta) \in \arg\max_{p\geq 0}\Pi^M(p, n, \Delta)$, where $\Pi^M(p, n, \Delta)$ denotes $\Pi^M(p, n)$ expanded to denote dependence on Δ. Next, we show that inequality (A.18) implies, for $\Delta < k$ and $n \in \{1, \ldots, \bar{N} - 1\}$,

$$p^M(n,\Delta) \leq p^M(n,k). \tag{A.19}$$

Suppose, to the contrary, that $p^M(n, \Delta) > p^M(n, k)$. For $\Delta < k$,

$$\begin{aligned}&\Pi^M(p^M(n,\Delta),n,k) - \Pi^M(p^M(n,k),n,k)\\ &\quad > \Pi^M(p^M(n,\Delta),n,\Delta) - \Pi^M(p^M(n,k),n,\Delta) \geq 0,\end{aligned} \tag{A.20}$$

where the first inequality follows from (A.18). Inequality (A.20) implies $\Pi^M(p^M(n, \Delta), n, k) > \Pi^M(p^M(n, k), n, k)$, a contradiction. This establishes (A.19). Next, we show that for $n \in \{1, \ldots, \bar{N} - 1\}$,

$$p^M(n,k) < p^A. \tag{A.21}$$

Note that $p^M(n, k) \in \arg\max_{p\geq 0} R^M(p, n)$. To see that $R^M(p, n)$ is strictly concave in p, observe that $(\partial^2/\partial p^2)R^M(p, n) = \sum_{j=n}^{\bar{N}}\Pr(N^L = j)[p(\partial^2/\partial p^2)\lambda(v, p, j) + 2(\partial/\partial p)\lambda(v, p, j)] < 0$, where the inequality follows from Lemma A1(b). That $p\lambda(v, p, n)$ is strictly concave in p follows by parallel argument. Note

$$\begin{aligned}&(\partial/\partial p)R^M(p,n)\\ &\quad = \sum_{j=n}^{\bar{N}}\Pr(N^L = j)\lambda(v,p,j)\{1 + [(\partial/\partial p)\lambda(v,p,j)]p/\lambda(v,p,j)\}\end{aligned}$$

and

$$\begin{aligned}&(\partial/\partial p)[p\lambda(v,p,\bar{N})]\\ &\quad = \lambda(v,p,\bar{N})\{1 + [(\partial/\partial p)\lambda(v,p,\bar{N})]p/\lambda(v,p,\bar{N})\}\end{aligned}$$

imply

$$\begin{aligned}&(\partial/\partial p)R^M(p,n)|_{p=p^A}\\ &\quad = \sum_{j=n}^{\bar{N}-1}\Pr(N^L = j)\lambda(v,p^A,j)\\ &\qquad\cdot\{1 + [(\partial/\partial p)\lambda(v,p^A,j)]p^A/\lambda(v,p^A,j)\} < 0,\end{aligned} \tag{A.22}$$

where the equality holds because $(\partial/\partial p)[p^A\lambda(v,p^A,\bar{N})]=0$, and the inequality holds because the price elasticity of demand strictly decreases with the number of agents. Because $R^M(p,n)$ is strictly concave in p, (A.22) implies (A.21). Together, (A.19) and (A.21) imply that for $n\in\{1,\ldots,\bar{N}-1\}$,

$$p^M(n,\Delta)<p^A. \tag{A.23}$$

Because, for $\Delta<k$, the optimal $p^M=p^M(n^M,\Delta)$ for some $n^M\in\{1,\ldots,\bar{N}-1\}$, (A.23) implies (A.17). □

Proof of Lemma 5. (a) Consider the benchmark case where customers are insensitive to delay. The unique equilibrium in participating agents and demand rate $(\mathcal{N},\lambda(v,p,\mathcal{N}))=(\bar{N},\bar{N}\mu)$ if $p\le v$ and $\omega\ge k/\mu$, and $(\mathcal{N},\lambda(v,p,\mathcal{N}))=(0,0)$ otherwise. Consequently, the platform's expected profit rate is $(p-\omega)\bar{N}\mu$ if $p\le v$ and $\omega\ge k/\mu$, and is zero otherwise. This implies the platform's optimal price and wage $(p_0^*,\omega_0^*)=(v,k/\mu)$. Consider the base case where customers are sensitive to delay. If $p\ge v$ or $\omega\le k/\mu$, then the unique equilibrium $(\mathcal{N},\lambda(v,p,\mathcal{N}))=(0,0)$ (by Lemma 1), and the platform's expected profit rate is zero. Because, by assumption, the platform's expected profit rate under the optimal price and wage is strictly positive, $p^*<v$ and $\omega^*>k/\mu$. (b) If $p<v$ and $\omega\ge\bar{N}k/\lambda(v,p,\bar{N})$, then $\bar{N}$ agents participate; otherwise, no agents participate. Therefore, under price $p<v$, wage $\omega=\bar{N}k/\lambda(v,p,\bar{N})$ is optimal, and the platform's objective function simplifies to $p\lambda(v,p,\bar{N})-\bar{N}k$, which is maximized at price $p^*\in\arg\max_{p\ge 0}\{p\lambda(v,p,\bar{N})\}$. Using $\lambda(v,p,1)=(v-p)\mu^2/[(v-p)\mu+c]$ and $\lambda(v,p,2)=2\sqrt{(v-p)\mu^3/[(v-p)\mu+c]}$, it is straightforward to show that $p^*|_{\bar{N}=1}=[v\mu+c-\sqrt{c(v\mu+c)}]/\mu$ and $p^*|_{\bar{N}=2}=[4v\mu+3c-\sqrt{c(8v\mu+9c)}]/(4\mu)$, and thus that for $\bar{N}\in\{1,2\}$, $(\partial/\partial c)p^*<0$ and $(\partial/\partial c)\lambda(v,p^*,\bar{N})<0$. the last inequality implies that $\omega^*=\bar{N}k/\lambda(v,p^*,\bar{N})$ increases with c. □

Proof of Lemma 6. (a) The expected equilibrium demand rate $E[\lambda(\hat{V},p,\mathcal{N})]=\bar{N}\mu$ if $\omega\ge k/\mu$ and $p\le V^l$, $E[\lambda(\hat{V},p,\mathcal{N})]=\bar{N}\mu/2$ if $\omega\ge k/\mu$ and $p\in(V^l,V^h]$, and $E[\lambda(\hat{V},p,\mathcal{N})]=0$ otherwise. The result follows. (b) The expected equilibrium demand rate $E[\lambda(v,p,\mathcal{N})]=\bar{N}\mu$ if $p\le v$ and $\omega\ge K^H/\mu$, $E[\lambda(v,p,\mathcal{N})]=\bar{N}\mu/2$ if $p\le v$ and $\omega\in[K^L/\mu,K^H/\mu)$, and $E[\lambda(v,p,\mathcal{N})]=0$ otherwise. The result follows. □

Proof of Proposition 3. (a) From Lemmas 2 and 6(a), if $\delta\le\min(\bar{\delta}_0,\bar{\delta})$, then $p^*=p^l<V^l=p_0^*$, where the inequality follows because in the base case where customers are sensitive to delay, $\lambda(V^l,p,n)=0$ if $p\ge V^l$. If $\delta>\max(\bar{\delta}_0,\bar{\delta})$, then $p^*=p^h<V^h=p_0^*$, where the inequality follows by parallel argument. Thus, for the remainder of this proof, it only remains to consider the parameter region $\delta\in(\min(\bar{\delta}_0,\bar{\delta}),\max(\bar{\delta}_0,\bar{\delta})]$. If $\bar{\delta}_0\le\bar{\delta}$ and $\delta\in(\bar{\delta}_0,\bar{\delta}]$, then $p^*=p^l<V^h=p_0^*$. Suppose instead for the remainder of the proof that $\bar{\delta}<\bar{\delta}_0$, in which case it only remains to consider $\delta\in(\bar{\delta},\bar{\delta}_0]$, so that $p^*=p^h$ and $p_0^*=V^l$. Next we show that if $[p^h-V^l]|_{\delta=\bar{\delta}_0}\le 0$, then $p^*\le p_0^*$. To do so, we first establish that p^h increases with δ. Let $p^h(\delta)$ denote p^h expanded to denote dependence on δ. Note $p^h(\delta)$ is the unique solution to the first order condition $(\partial/\partial p)\Pi^h(p)=0$ (by Lemma A2a), which can be rewritten as

$$\begin{aligned}&[(\partial/\partial p)\lambda(v+\delta,p^h(\delta),\bar{N})]p^h(\delta)/\lambda(v+\delta,p^h(\delta),\bar{N})+1\\&\quad=0.\end{aligned}\tag{A.24}$$

Note that for $\delta_1<\delta_2$,

$$\begin{aligned}-1&=[(\partial/\partial p)\lambda(v+\delta_1,p^h(\delta_1),\bar{N})]p^h(\delta_1)/\lambda(v+\delta_1,p^h(\delta_1),\bar{N})\\&<[(\partial/\partial p)\lambda(v+\delta_2,p^h(\delta_1),\bar{N})]p^h(\delta_1)/\lambda(v+\delta_2,p^h(\delta_1),\bar{N}),\end{aligned}\tag{A.25}$$

where the inequality follows from Lemma A1(c). Because $\Pi^h(p)$ is strictly concave in p, (A.24) and (A.25) imply that $p^h(\delta)$ strictly increases with δ. If $\delta\in(\bar{\delta},\bar{\delta}_0]$, then $p^*\le p^h|_{\delta=\bar{\delta}_0}\le V^l|_{\delta=\bar{\delta}_0}\le p_0^*$, where the first inequality holds because $p^*=p^h$, which strictly increases with δ, and the last inequality holds because $p_0^*=V^l$, which strictly decreases with δ. Finally, we show that if $[p^h-V^l]|_{\delta=\bar{\delta}_0}>0$, then there exists $\underline{\delta}\in[\bar{\delta},\bar{\delta}_0)$ such that (15) holds if and only if $\delta\in(\underline{\delta},\bar{\delta}_0]$. Because $[p^h-V^l]|_{\delta=\bar{\delta}_0}>0$ and because p^h strictly increases and V^l strictly decreases with δ, there exists $\underline{\delta}\in[\bar{\delta},\bar{\delta}_0)$ such that $p^h\le V^l$ if $\delta\in(\bar{\delta},\underline{\delta}]$ and $p^h>V^l$ if $\delta\in(\underline{\delta},\bar{\delta}_0]$. Because $p^*=p^h$ and $p_0^*=V^l$, we conclude that (15) holds if and only if $\delta\in(\underline{\delta},\bar{\delta}_0]$. (b) Suppose $\delta<\bar{\delta}_0$. Because Π^h and Π^l are continuous in c, Lemma A3 implies that there exist $\underline{c}_l$, c_m, and $\bar{c}_h$ such that $0\le\underline{c}_l<c_m<\bar{c}_h$; $p^*=p^l$ if $c\in[\underline{c}_l,c_m]$; and $p^*=p^h$ if $c\in(c_m,\bar{c}_h]$. Because $p^l<p^h$ (by Lemma A2(b)) and p^j is continuous in c for $j\in\{h,l\}$, there exist $c_l\in[\underline{c}_l,c_m)$ and $c_h\in(c_m,\bar{c}_h]$ such that $p^*|_{c\in[c_l,c_m]}<p^*|_{c\in(c_m,c_h]}$. □

Proof of Proposition 4. (a) We first show that if $k>\underline{k}$, then there exists $\check{\Delta}<k$ such that for $\Delta\in[\check{\Delta},k]$, $\omega^*\in\{\omega^L,\omega^M(1)\}$, where $\omega^M(n)=nK^L/\lambda(v,p^M(n),n)$. If $k>\Pr(N^L<\bar{N})R^A/(2\bar{N})$, then $\lim_{\Delta\to k}[\Pi^L-\Pi^H]=\Pr(N^L=\bar{N})R^A-(R^A-2\bar{N}k)>0$. If $k>[R^A-R^M(p^M(1),1)]/(2\bar{N})$, then $\lim_{\Delta\to k}[\Pi^M-\Pi^H]=R^M(p^M(1),1)-(R^A-2\bar{N}k)>0$. Therefore, if $k>\underline{k}$, then $\lim_{\Delta\to k}[\max(\Pi^L,\Pi^M)-\Pi^H]>0$. The proof of Lemma 3 establishes that if $\max(\Pi^L,\Pi^M)>\Pi^H$, then $\omega^*\in\{\omega^L,\omega^M\}$. Furthermore, for sufficiently large Δ, $n^M=1$. Therefore, because Π^j is continuous in Δ for $j\in\{H,L,M\}$, there exists $\check{\Delta}<k$ such that for $\Delta\in[\check{\Delta},k]$, $\omega^*\in\{\omega^L,\omega^M(1)\}$. Because $k<\bar{k}$, $\tilde{\Delta}_0>k$, and therefore $\omega_0^*=K^H/\mu>k/\mu$ (by Lemma 6b). Because $\omega^*\in\{\omega^L,\omega^M(1)\}$ on $\Delta\in[\check{\Delta},k]$, ω^L and $\omega^M(1)$ are continuous in Δ, and $\lim_{\Delta\to k}\omega^L=\lim_{\Delta\to k}\omega^M(1)=0$, there exists $\tilde{\Delta}<k$ such that if $\Delta>\tilde{\Delta}$, $\omega^*<k/\mu$. (b) By Lemma A6, $\rho>\bar{\rho}$ and $\Delta<\check{\Delta}_0$ imply that $\omega^*=\omega^H$ if $c\le\bar{c}$, and $\omega^*=\omega^L$ if $c>\bar{c}$, where $\bar{c}>0$. Furthermore, $\lim_{c\uparrow\bar{c}}\omega^H>\lim_{c\downarrow\bar{c}}\omega^L$. Therefore, because ω^j is continuous in c for $j\in\{H,L\}$, there exist c_l, c_m, and c_h such that $0<c_l<c_m<c_h$ and $\omega^*|_{c\in[c_l,c_m]}>\omega^*|_{c\in(c_m,c_h)}$. □

References

Afèche P (2013) Incentive-compatible revenue management in queueing systems: Optimal strategic delay. *Manufacturing Service Oper. Management* 15(3):423–443.

Afèche P, Pavlin JM (2016) Optimal price/lead-time menus for queues with customer choice: Segmentation, pooling and strategic delay. *Management Sci.* 62(8):2412–2436.

Allon G, Bassamboo A, Çil EB (2012) Large-scale service marketplaces: The role of the moderating firm. *Management Sci.* 58(10):1854–1872.

Anand KS, Paç MF, Veeraraghavan S (2011) Quality-speed conundrum: Trade-offs in customer-intensive services. *Management Sci.* 57(1):40–56.

Arnosti N, Johari R, Kanoria Y (2014) Managing congestion in decentralized matching markets. Working paper, Stanford University, Stanford, CA.

Ata B, Olsen TL (2013) Congestion-based leadtime quotation and pricing for revenue maximization with heterogeneous customers. *Queueing Systems* 73(1):35–78.

Banerjee S, Riquelme C, Johari R (2015) Pricing in ride-share platforms: A queueing-theoretic approach. Working paper, Stanford University, Stanford, CA.

Benjaafar S, Kong G, Li X, Courcoubetis C (2018) Peer-to-peer product sharing: Implications for ownership, usage, and social welfare in the sharing economy. *Management Sci.*, ePub ahead of print May16, https://doi.org/10.1287/mnsc.2017.2970.

Bensinger G (2015) Startups scramble to define "employee." *Wall Street Journal* (July 30), B1.

Cachon G (2003) Supply chain coordination with contracts. Graves S, De Kok T, eds. *Supply Chain Management*. Handbooks in Operations Research and Management Science, Vol. 11 (Elsevier, Amsterdam), 227–339.

Cachon GP, Feldman P (2011) Pricing services subject to congestion: Charge per-use fees or sell subscriptions. *Manufacturing Service Oper. Management* 13(2):244–260.

Cachon GP, Zhang F (2007) Obtaining fast service in a queueing system via performance-based allocation of demand. *Management Sci.* 53(3):408–420.

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.

Çelik S, Maglaras C (2008) Dynamic pricing and lead-time quotation for a multiclass make-to-order queue. *Management Sci.* 54(6):1132–1146.

Chen H, Frank M (2004) Monopoly pricing when customers queue. *IIE Trans.* 36(6):1–13.

Debo L, Rajan U, Veeraraghavan SK (2013) Signaling quality via long lines and uninformative prices. Working paper, Dartmouth College, Hanover, NH.

Dowdle L (2015) Amazon opportunity. *Landscape Management* 54(8):64–67.

Einav L, Farronato C, Levin J (2016) Peer-to-peer markets. *Annual Rev. Econom.* 8(October):615–635.

Farrell D, Greig F (2016) *The Online Platform Economy: Has Growth Peaked?* (JPMorgan Chase Institute, Washington, DC).

Fraiberger SP, Sundararajan A (2015) Peer-to-peer rental markets in the sharing economy. Working paper, New York University, New York.

Gilbert SM, Weng ZK (1998) Incentive effects favor nonconsolidating queues in a service system: The principal agent perspective. *Management Sci.* 44(12):1662–1669.

Gopalakrishnan R, Doroudi S, Ward A, Wierman A (2016) Routing and staffing when servers are strategic. *Oper. Res.* 64(4): 1033–1050.

Gurvich I, Lariviere M, Moreno-Garcia A (2018) Operations in the on-demand economy: Staffing services with self-scheduling capacity. Hu M, ed. *Sharing Economy: Making Supply Meet Demand.* Springer Series in Supply Chain Management (Springer, Cham, Switzerland). Forthcoming.

Hassin R (2016) *Rational Queueing* (CRC Press, Boca Raton, FL).

Hassin R, Haviv M (2003) *To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems* (Kluwer, Boston).

Hu M, Zhou Y (2015) Dynamic type matching. Working paper, University of Toronto, Toronto.

Huet E (2015) The delivery war is on: Uber expands its on-demand courier service. *Forbes* (October 14), https://www.forbes.com/sites/ellenhuet/2015/10/14/delivery-wars-uber-targets-merchants-expands-its-on-demand-courier-service-uberrush/.

Ibrahim R (2018) Managing queueing systems where capacity is random and customers are impatient. *Production Oper. Management* 27(2):234–250.

Jiang B, Tian L (2018) Collaborative consumption: Strategic and economic implications of product sharing. *Management Sci.* 64(3): 1171–1188.

Kabra A, Belavina E, Girotra K (2016) Designing promotions to scale marketplaces. Working paper, INSEAD, Fontainebleau, France.

Kokalitcheva K (2015) Uber and Lyft face a new challenger in Boston. *Fortune* (October 5), http://fortune.com/2015/10/05/uber-challenger-fasten/.

Kumar S, Randhawa RS (2010) Exploiting market size is service systems. *Manufacturing Service Oper. Management* 12(3):511–526.

Li J, Moreno A, Zhang DJ (2016) Pros vs. Joes: Agent pricing behavior in the sharing economy. Working paper, University of Michigan, Ann Arbor.

MacMillan D (2015) The $50 billion question: Can Uber deliver? *Wall Street Journal* (June 16), A1, A12.

Maglaras C, Zeevi A (2003) Pricing and capacity sizing for systems with shared resources: Approximate solutions and scaling relations. *Management Sci.* 49(8):1018–1038.

Maglaras C, Yao J, Zeevi A (2018) Optimal price and delay differentiation in large-scale queueing systems. *Management Sci.* 64(5):2427–2444.

Mendelson H (1985) Pricing computer services: Queueing effects. *Comm. ACM* 28(3):312–321.

Mendelson H, Whang S (1990) Optimal incentive-compatible priority pricing for the $M/M/1$ queue. *Oper. Res.* 38(5):870–883.

Moreno A, Terwiesch C (2014) Doing business with strangers: Reputation in online service marketplaces. *Inform. Systems Res.* 25(4):865–886.

Naor P (1969) The regulation of queue size by levying tolls. *Econometrica* 37(1):15–24.

Nazerzadeh H, Randhawa RS (2017) Near-optimality of coarse service grades for customer differentiation in queueing systems. *Production Oper. Management* 27(3):578–595.

Phillips R (2016) Marketplace optimization at Uber. Presentation, INFORMS Conference Nashville, November 14, INFORMS, Catonsville, MD.

Plambeck EL (2004) Optimal leadtime differentiation via diffusion approximations. *Oper. Res.* 52(2):213–228.

Roose K (2014) Does Silicon Valley have a contract-worker problem? *NYMag* (September 18), http://nymag.com/daily/intelligencer/2014/09/silicon-valleys-contract-worker-problem.html.

Ruggless R (2015) Postmates takes early lead in delivery game. *Nation's Restaurant News* (May 25), 15.

Scheiber N (2015) Employees, contractors and those in between. *New York Times* (December 11), B1.

Shoot B (2015) Hot food, fast. *Entrepreneur* (August), 68.

Singer N, Isaac M (2015) A dashboard management consultant. *New York Times* (May 10), B1.

Snir EM, Hitt LM (2003) Costly bidding in online markets for IT services. *Management Sci.* 49(11):1504–1520.

Solomon B (2016) TaskRabbit founder steps down as CEO after eight years. *Forbes* (April 14), https://www.forbes.com/sites/briansolomon/2016/04/14/taskrabbit-founder-steps-down-as-ceo-after-eight-years/#23fdfa3025c3.

Steinmetz K (2015) Is the on-demand economy taking workers for a ride? *Time* (August 3), 11–15.

Wang S (2015) No. 3 Sidecar, squeezed out by Uber and Lyft, shutting down. *Daily Herald* (December 31), 1.

Watanabe S (2016) Smartphone apps unleash new wave of bike taxi services. *Nikkei Asian Rev.* (June 24), https://asia.nikkei.com/Business/Smartphone-apps-unleash-new-wave-of-bike-taxi-services.

Zervas G, Proserpio D, Byers J (2017) The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry. *J. Marketing Res.* 54(5):687–705.

Zhan D, Ward A (2015) Incentive based service system design: Staffing and compensation to trade off speed and quality. Working paper, University College London, London.