Learning Pay Strategies with Small Samples in Gig Economy Platforms

Arthur Delarue, Zhen Lian, and Tony Qin*

Abstract

Gig economy platforms operate in dynamic and competitive labor markets in which workers can compare and choose among multiple job offers in real time. A central challenge for these platforms is to set worker compensation when individual workers' reservation wages are heterogeneous, and when common market-wide factors such as competitor incentives shift workers' willingness to accept offers. We study a profit-maximizing platform that must serve identical requests by sequentially making take-it-or-leave-it pay offers to a pool of workers, observing only acceptances and rejections. Each worker's reservation wage consists of an individual component and a shared global factor that is initially unknown to the platform. We first characterize the optimal policy in a full-information benchmark where the global factor is known, and show that the problem can be solved efficiently via dynamic programming. The optimal policy leverages worker heterogeneity by initially offering pay below the myopic optimum, increasing pay following rejections and decreasing it following acceptances, and avoiding a set of dominated pay levels that are never optimal. We then analyze the realistic setting in which the global factor is uncertain and must be inferred from a small number of observations. While the exact belief-augmented solution is complex, we develop simple and interpretable heuristics with provable performance guarantees. In particular, we propose a direct-commit policy and a probe-and-commit policy that use little or no learning to adapt pay to market conditions. Our results provide actionable guidance for gig economy platforms seeking to set transparent, adaptive worker compensation in competitive spot labor markets.

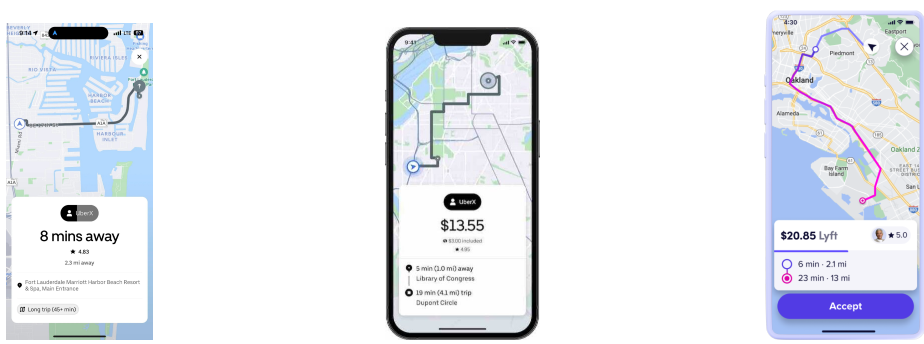
Key words: Gig economy platforms, worker pay, sequential take-it-or-leave-it offers, learning under uncertainty, adaptive policies

*Arthur Delarue: the University of Virginia Darden School of Business. Zhen Lian: Yale School of Management. 165 Whitney Avenue, New Haven, CT 06511. Tony Qin: Santa Clara University

1 Introduction

Gig economy platforms such as Uber, Lyft, and DoorDash operate in highly dynamic and competitive markets where workers often receive job offers from multiple platforms. A key challenge for these platforms is to determine optimal compensation levels, especially as supplier pay has become increasingly independent from consumer pricing (Chaum 2023). This challenge originates from multiple sources: first, workers' valuation of a request may be highly heterogeneous, driven by *individual* factors unknown to the platform (e.g., how soon a food delivery driver is looking to end their shift). Second, there may be unknown *global* factors affecting most or all current workers: for example, Uber introducing a temporary pay incentive will influence drivers' willingness to accept Lyft's offers, effectively raising their reservation wages for reasons that Lyft cannot observe directly.

Our work is particularly motivated by a recent change in the driver experience at Uber and Lyft, the U.S.'s two largest ride-hailing platforms. Until 2022, Uber and Lyft compensated drivers based on a fixed rate per mile and per minute, in addition to surge pricing bonuses, and did not display any pay information at the time of trip match. As a result, drivers did not always have enough information to compute the wages they would receive for completing a particular trip (see Figure 1a). In 2022, however, both platforms moved to a system where each driver is offered a specific pay for each ride (see Figure 1b). This change increases transparency for drivers and makes it easier for them to compare offers from competitors (see Figure 1c). For the platform, more transparency means that offering drivers the right pay becomes paramount.



(a) Before upfront pay (Uber) (b) After upfront pay (Uber) (c) Upfront pay (Lyft)

Figure 1: Upfront Pay at Uber and Lyft.

Notes: Sources (left to right): reddit.com, theverge.com, lyft.com.

In general, identifying the right worker pay is a critical issue whenever a firm receives demand (requests) that it tries to serve using suppliers (workers) on the spot market. Beyond ride-hailing, examples include food delivery platforms, which seek to find a delivery courier for a food order; and freight marketplaces, like Amazon Freight, that complement an existing delivery fleet with a spot market of independent truck operators. All of these contexts share two key properties: first, decisions have to be made quickly, so complex mechanisms such as second-price auctions are impractical; second, real-time market dynamics affect how workers make decisions, and while platforms have abundant information about how the market behaves on average, they can find it hard to predict how individual workers would respond to a specific request at a specific time.

In this paper, we study how a platform can make adaptive decisions on worker pay based on real-time reactions from just a few drivers. We consider a profit-maximizing platform that sequentially makes take-it-or-leave-it job offers with varying pay levels to serve requests using a pool of workers, updating future offers based on observed acceptance and rejection decisions. Serving a request generates fixed revenue for the platform, and a worker accepts an offer only if it exceeds her reservation wage, where the platform knows the distribution of reservation wages but not their realizations. Each driver's reservation wage is modeled as the sum of a shared "global factor", which captures market-wide conditions such as external competition or congestion, and an individual component, which reflects idiosyncratic preferences. By observing real-time decisions, the platform can use rejections to infer whether global effects are influencing worker behavior. We first analyze a full-information setting in which the global factor is known, and then study the more realistic case where it is unknown and must be inferred dynamically.

We first establish that we can optimally solve the full information setting using an efficient dynamic programming algorithm, and we derive key properties of the optimal solution. Our algorithm seeks to leverage excess supply and driver heterogeneity by making an initial offer below the optimal myopic pay offer. The optimal pay then weakly increases after a driver rejection and weakly decreases after an acceptance. More surprisingly, we also find that some pay levels are never offered in any optimal policy because they are "dominated", in the sense that they are too likely to be accepted, yet not profitable enough. We provide a geometric intuition for identifying dominated pay levels using the convex hull of the curve that relates a pay level's acceptance probability to its expected profit.

We then study the general case, in which the global factor is unknown and must be inferred from workers' acceptance and rejection decisions, by providing a dynamic programming algorithm to solve the resulting belief-augmented Markov Decision Process (MDP). However, this exact algorithm is not very interpretable and suffers from the curse of dimensionality. While uncertainty fundamentally increases the complexity of the exact solution, we show that policies built around the full-information setting remain highly effective. In particular, restricting learning to zero or one targeted step yields transparent and tractable decision rules that achieve strong performance despite the underlying complexity of the optimal policy. We therefore develop heuristics with provable guarantees using the insights gleaned from the full information case. We first propose a heuristic we call "direct-commit," in which we commit to either the low-global-factor or the high-global-factor optimal pay policy based only on our initial belief (i.e., with no belief updates regardless of outcomes). We bound this heuristic's performance relative to a clairvoyant optimum, and find it performs well in a range of parameter settings. In the settings where it fails, we propose an alternative heuristic we call "probe-and-commit," in which we use a single initial pay offer to gain as much information as possible about the global factor, update our belief, then commit to the low-global-factor or high-global-factor optimal pay. We find that even one such "probe" can significantly limit the shortcomings of direct commitment. A key challenge of the problem is to balance learning with profit maximization, especially over a short time horizon. Both heuristics seek to restrict learning to just zero or one sample, leading to decisions that can more easily be interpreted and implemented by managers.

2 Related literature

Our work provides a new perspective on pricing and mechanism design in gig economy platforms by combining ideas from several active research streams.

Dynamic pricing in ride-hailing. The question of optimal pricing has been a major focus of researchers and practitioners since the inception of ride-hailing platforms. Most of the literature has focused on pricing rider requests. Castillo et al. (2017) first motivated surge pricing as a way to sufficiently suppress demand relative to supply to prevent distant passenger-driver matches. Hu et al. (2022) compare different surge pricing approaches, while Bimpikis et al. (2019) and Ma et al. (2020) develop more holistic spatial and spatio-temporal pricing models. Özkan (2020) emphasize the importance of jointly optimizing pricing and matching strategies.

Our paper is one of the first to explicitly consider the driver side of the pricing problem, which has recently grown more distinct from the rider side as platforms have increasingly separated driver pay from rider fares at the ride level (Chaum 2023). To simplify our analysis, we adopt a similar approach as Yan et al. (2025) (who tackle the related problem of *shared/pooled* rides) and consider identical requests so as to effectively separate the matching problem from the pricing problem. Even with identical requests, driver reservation wages may vary significantly: Allon et al. (2023) found empirically that drivers' propensity to accept a ride at a particular price depends significantly on drivers' recent driving history (how many hours they have been driving that day and how much they have earned), to say nothing of potential destination preferences (which ride-hailing platforms have recently started considering through schemes like Lyft's "destination mode").

Sequential take-it-or-leave-it offers. In practice, once a driver and passenger have been matched, the ride-hailing platform will offer the driver the opportunity to serve the ride at a particular price. If the driver accepts, the ride is added to their queue. Otherwise, the ride re-enters the matching process and will not be offered to this driver again. In a sense, a platform trying to find a driver for a ride may offer it sequentially to multiple drivers at different prices.

This mechanism of *sequential take-it-or-leave-it offers* has received some interest in the mechanism design literature. Sandholm and Gilpin (2006) study an auction where the seller reveals a sequence of take-it-or-leave-it offers to all buyers and find an equilibrium. Amin et al. (2013) and Vanunts and Drutsa (2018) study a similar mechanism with repeated interactions between just one buyer and one seller and discuss methods to deter strategic behavior. Chawla et al. (2010) and Feldman et al. (2014) study the most similar mechanism to ours, where one or more capacity-constrained goods are offered to buyers sequentially at different prices. They show that under some structural conditions, such a mechanism is near-optimal relative to a Myerson auction.

Optimizing sequential posted price mechanisms is a growing new question in the operations literature, as such mechanisms are increasingly used in ride-hailing, food delivery, and freight marketplaces (Chen et al. 2021). Most recently, Cao et al. (2025) consider a system in which demand requests expire after a certain lead time, prompting the platform to increase supply compensation near the deadline. Though their model does not include a learning component, there are methodological similarities in our use of dynamic programming.

Online learning and bandits. In addition to exploiting driver heterogeneity, in our model the platform seeks to leverage sequential offers to learn global information about driver reservation wages. Our paper therefore relates to the vast literature on online learning and multi-armed bandits (Slivkins et al. 2019), where state-of-the-art methods include Thompson sampling (Russo et al. 2018) and UCB (Garivier and Moulines 2011). Qin et al. (2025) provide a comprehensive review of reinforcement learning approaches to ride-hailing problems, including pricing and incentives. Specialized learning approaches have also been developed for dynamic pricing (Besbes and Zeevi 2009) and matching (Johari et al. 2021). Cohen et al. (2020) study a sequential pricing setting where the platform seeks to learn how customers value different item features over time. Much of this literature focuses on characterizing notions of regret over long time horizons — in contrast, our goal is to learn the shared global factor over a very short-term horizon, since ride-hailing platforms often operate in nonstationary settings where conditions can change quickly. Some related work on *nonstationary* bandits exists (Besbes et al. 2014), though the analysis typically continues to focus on asymptotic regret.

Dynamic pricing under competition. Finally, our work relates to the literature on dynamic pricing under competition. Since the advent of dynamic pricing (Gallego and Van Ryzin 1994), much work has focused on the effect of competition under dynamic pricing agents. Gallego and Wang (2014) study the problem when consumers follow a nested logit model, whereas Banerjee et al. (2024) analyze it under a consider-then-choose lexicographic choice model. Recent works focus on the effect of competition in ride-hailing marketplaces (Cohen and Zhang 2022, Tripathy et al. 2023). Our work differs in that we do not seek to characterize an equilibrium — rather, in our analysis of a generic "global factor," we consider a comparatively short time horizon where competitor actions are exogenous. As such, our work extends to other settings where a platform seeks to learn shared information from its users.

3 Model Description

3.1 Basic Assumptions

We consider a profit-maximizing platform that seeks to serve m identical jobs, or *requests*, using a list of n available workers. Each request generates revenue v for the platform. Worker i has a reservation wage $\tilde{W}_i$: if the platform offers them to serve a request at pay p, they will accept if $p \geq \tilde{W}_i$ and reject otherwise. If they accept, the platform earns profit $(v-p)$ for the request, then offers the next request to the next worker. If they reject, the request is either canceled by the customer (with probability q) or the platform can offer it to the next worker, possibly at a different pay. Any remaining unserved requests after every worker has responded are considered lost. Offers are take-it-or-leave-it in the sense that workers only receive at most one offer over the time horizon.

We assume that worker i's reservation wage can be decomposed as $\tilde{W}_i = W_i + C$, where C designates a *global factor*, while the *individual factors* W_i are i.i.d. and capture workers' idiosyncratic preferences for the considered request. We assume that W_i takes K possible discrete values $w_1, \ldots, w_K$, such that

$$P(W_i = w_j) = f_j, \ \forall \ j \in [K], \tag{1}$$

where $\sum_j f_j = 1$, and we write $\mathcal{W} = \{w_1, \ldots, w_K\}$ as the support of W_i. We denote the cumulative distribution function of W_i by

$$F(p) = \mathbb{P}(W_i \leq p) = \sum_{j, w_j \leq p} f_j. \tag{2}$$

While individual factors are unique to each worker, the global factor influences the reservation wage of *all* workers. Such an effect can arise from competition: for instance, a competing platform offering a bonus amount c to all workers in a given place at a given time can be viewed as setting the global effect to $C = c$ — uniformly raising all workers' reservation wages. Such short-term bonuses are common practice in ride-hailing platforms (Lyft Bonus Times, Uber's Boost+, etc.). The global factor can also capture other effects: for instance, road closures or heavy congestion may make a ride-hailing request less desirable to all drivers, meaning no driver will accept the request unless the compensation increases. For model tractability, we assume that the global factor C follows a two-point distribution, i.e.,

$$C = \begin{cases} c, & \text{w.p. } \mu_0 \\ 0, & \text{w.p. } 1-\mu_0 \end{cases} \tag{3}$$

We say the global factor is "on" when $C = c$ and "off" when $C = 0$. Our model assumes asymmetric information: the platform knows the distribution of the global and individual factors (i.e., F and μ_0), but not its realizations. Meanwhile, workers know their own reservation wage (including both the individual global factor realizations) but have no information about other workers' individual factors.

3.2 The Platform's Problem

Given our setup, the platform must decide what pay to offer each worker, with the goal of maximizing profit from the m requests under consideration. When choosing the pay for worker i, the platform navigates multiple tradeoffs. The first is between profit and service rate: higher pay increases the likelihood of acceptance from worker i, but reduces profit; lower pay increases potential profit, but is more likely to be rejected. The second tradeoff is between profit and learning: since the global effect is known to all workers and influences all of their reservation wages, the platform has the opportunity to gauge the value of C from worker responses, and leverage the obtained information to make better decisions for the remaining workers. The choice of pay offer affects the amount of information that can be gained: for example, the rejection of a low-paying offer may not be as informative as that of a high-paying offer, as workers may reject it regardless of the global factor C.

In this system, a state is characterized by the number of requests m, the number of workers n, and the platform's current belief $\mu \in [0, 1]$ about the unknown global factor C. Given a state (m, n, μ), the probability a pay offer p is accepted is

$$P_{\text{accept}}(p, \mu) = \mu F(p - c) + (1 - \mu)F(p). \tag{4}$$

According to Bayes' rule, if the offer is accepted, the platform's belief updates to

$$\mu'_A(p) = P(C = c|p \text{ is accepted}) = \frac{F(p - c)\mu}{P_{\text{accept}}(p, \mu)}, \tag{5}$$

whereas if the offer is rejected, the platform's belief updates to

$$\mu'_R(p) = P(C = c|p \text{ is rejected}) = \frac{(1 - F(p - c))\mu}{1 - P_{\text{accept}}(p, \mu)}. \tag{6}$$

If a pay offer p is accepted from state (m, n, μ), the system transitions to state $(m - 1, n - 1, \mu'_A(p))$. If a pay offer p is rejected, the system transitions to state $(m, n - 1, \mu'_R(p))$ with probability $(1-q)$ and to state $(m-1, n-1, \mu'_R(p))$ with probability q. We can therefore write the Bellman equation for the optimal pay policy as

$$\begin{aligned} V(m, n, \mu) = \max_{p \geq 0}\ & P_{\text{accept}}(p, \mu)\Big(v - p + V(m-1, n-1, \mu'_A(p))\Big) \\ & + \big(1 - P_{\text{accept}}(p, \mu)\big)\Big[(1-q)V(m, n-1, \mu'_R(p)) + qV(m-1, n-1, \mu'_R(p))\Big]. \end{aligned} \tag{7}$$

We observe that augmenting the state with the platform's belief about the unknown global factor C is a classical way to represent a partially observed MDP, or POMDP, as a standard MDP (Kaelbling et al. 1998). In the following sections we discuss optimal and approximate solutions to the Bellman equation (7).

4 Optimal Pay with Known Global Factor

We first consider the setting in which the global factor is known to the platform. For example, the platform knows that its competitor is currently running a promotion, or that a temporary road closure or weather event is under way, affecting all driver reservation wages uniformly. Because the value of C is known, no learning is necessary, and the platform simply seeks to determine how to optimally utilize its pool of n workers. Without loss of generality, we study the problem with $C = 0$ (i.e., $\mu_0 = 0$). The acceptance probability at pay level w becomes $P_{\text{accept}}(w) = F(w)$, and the Bellman equation Eq. (7) simplifies to

$$\begin{aligned} V(m, n) = \max_{p \geq 0} F(p)\Big(v - p + V(m-1, n-1)\Big) & \\ + (1 - F(p))\Big[(1-q)V(m, n-1) + qV(m-1, n-1)\Big]. & \end{aligned} \tag{8}$$

We observe that due to the discrete nature of the reservation wage distribution, any pay offers in an optimal policy must verify $p \in \mathcal{W}$ (if not, observe we can infinitesimally reduce p to increase profit without reducing acceptance probability). Thus we can take the maximum in (8) over $\mathcal{W}$ rather than $\mathbb{R}_+$.

4.1 Convex Hull Filtering and Dynamic Programming Algorithm

Our first result is that, given *any* reservation wage distribution $\{(w_k, f_k)\}_{k=1,\cdots,K}$, it is sufficient to only consider wages w_k that are on the upper convex hull of the immediate expected profit, i.e., the expected profit from a single offer to a single driver. More precisely, we can establish the following theorem.

Theorem 1. *The optimal pay sequence only includes pay levels on the increasing portion of the upper convex hull of the* acceptance-profit *curve, defined as the points* $\{(F(w), F(w)(v-w))\}_{w \in \mathcal{W}}$.

We illustrate Theorem 1 in Figure 2, where we observe a particular distribution of reservation wages (roughly trimodal). Out of 50 possible pay levels, only 17 actually appear on the increasing portion of the upper convex hull of the acceptance-profit curve, and can therefore conceivably appear in an optimal pay policy. We denote these allowed pay levels by $\bar{\mathcal{W}}$.

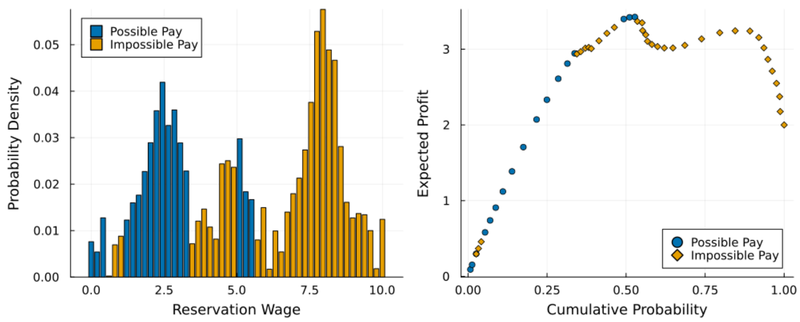


Figure 2: Illustration of possible and impossible pay levels according to Theorem 1.

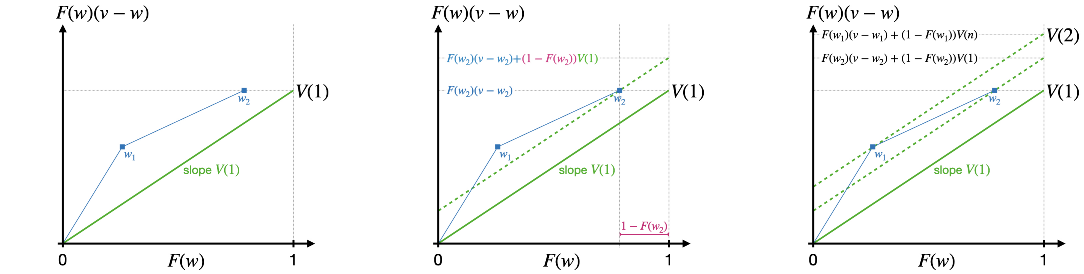
Notes: Assume $v = 12$ and the reservation wages follow the discrete distribution on the left panel. The right panel shows the pay levels that may (resp. may not) appear in an optimal pay sequence because they are (resp. are not) on the increasing portion of the upper convex hull of the acceptance-profit curve.

This result allows the platform to rule out dominated pay levels, i.e., pay levels that are suboptimal even in a one-shot setting. The proof relies on some key observations: first, there exists a point which maximizes immediate expected profit. Any higher pay is thus suboptimal in terms of immediate reward, and, it turns out, remains suboptimal even with multiple drivers and requests. Second, it can be optimal to offer lower pay than the single-shot optimal offer, but only at pay levels on the upper convex hull of the acceptance-profit curve —– this part of the proof is more complex and is graphically illustrated in the appendix. Theorem 1 is useful both computationally (reducing the decision space) as well as managerially — it provides an easy way to rule out pay levels that should never be offered.

4.2 Dynamic Programming Algorithm

Given the Bellman equation in (8), we can clearly compute the optimal pay policy for m drivers and n requests in $O(mn\bar{K})$ time, where $\bar{K} = |\bar{\mathcal{W}}|$ is the number of allowed pay levels after convex hull filtering. Algorithm 1 reduces this runtime to $O(m(n + \bar{K}))$ by leveraging some monotonicity properties of the optimal pay sequence which we analyze more carefully in the next subsection. Algorithm 1 assumes $C = 0$, producing an optimal pay sequence we call π^0. The optimal pay sequence under $C = c$ (denoted by π^c) can be obtained from the same algorithm by replacing v with $v - c$.

We also illustrate one step of Algorithm 1 in Figure 3 on a toy example with two driver



(a) Myopic pay for one worker (b) DP Recursion (c) Best continuation value

Figure 3: Illustration of the dynamic programming algorithm for optimal pay.

Notes: There are two driver reservation wages, w_1 and w_2 and a single request ($m = 1$). All panels depict the acceptance-profit curve. The optimal pay for one worker maximizes immediate expected profit (panel a). For two workers, we can graphically observe the sum of immediate expected profit and continuation value if we offer the first worker w_2 (panel b). We can then compare the total value obtained from offering the first worker w_1 and w_2 (panel c).

types and a single request. While the optimal pay for a single worker is easily obtained from myopic profit maximization, determining the optimal pay for the first of two workers requires considering the continuation value (i.e., the value provided by access to the second worker). The total value $V(2)$ can be visualized using the acceptance-profit curve and a tangent of slope $V(1)$. Interestingly, this behavior also explains the validity of convex hull filtering. A dominated reservation wage on the acceptance-profit curve is never optimal because it is not profitable enough if the continuation value is low, and too likely to be accepted if the continuation value is high.

4.3 Properties of Optimal Pay Sequence

The dynamic programming algorithm described above can easily compute the optimal policy for any number of requests and workers. It is also of interest to characterize properties of the optimal pay policy, as we do in the following theorem.

Theorem 2. *Suppose there are m requests and n workers. Denote the optimal first offer as $p(m, n)$. The following holds true:*

1. *When $m \geq n$, the optimal first offer is constant and always equals the optimal one-shot pay w_{k^*}[1], where*
$$k^* \in \arg\max_{k \in \{1,\ldots,K\}} F(w_k)(v - w_k) \tag{9}$$

2. *When $m < n$, the optimal first offer is weakly increasing in the number of requests m*

[1]Note that $w_{k^*} = w_{\bar{K}}$ in Algorithm 1. In particular, after the convex-hull filtering step, the highest remaining pay level (indexed by $\bar{K}$) coincides with the optimal one-shot pay level, hence $k^* = \bar{K}$.

Algorithm 1 Dynamic Programming Algorithm under $C = 0$

Input: Reservation wage types $\{w_1, \dots, w_K\}$ with acceptance probabilities $F(w_k)$, value v, cancellation prob. $q \in [0, 1)$, requests M, workers N. Define $l_k = F(w_k)(v - w_k)$ for all k.
Step 0: Upper convex hull filtering
Keep only points $(F(w_k), l_k)$ on the upper convex hull (increasing portion), and relabel the remaining types as $\{w_1, \dots, w_{\bar{K}}\}$.
Step 1: Pre-compute Hull Slopes
for $k = 2$ to $\bar{K}$ **do**
    $l'_k \leftarrow \frac{l_k - l_{k-1}}{F(w_k) - F(w_{k-1})}$.
end for
Step 2: Initialization
Initialize $V(0, n) = 0$ for all $n \in \{0, \dots, N\}$ and $V(m, 0) = 0$ for all $m \in \{0, \dots, M\}$.
Step 3: Fill DP Table by row
for $m = 1$ to M **do**
    for $n = 1$ to $\min\{m, N\}$ **do**
        $V(m, n) \leftarrow n \cdot l_{\bar{K}}, \quad p(m, n) \leftarrow w_{\bar{K}}$.
    end for
    Initialize pointer $k = \bar{K}$.
    for $n = m + 1$ to N **do**
        $\Delta \leftarrow V(m, n-1) - V(m-1, n-1)$.
        $\lambda \leftarrow (1 - q)\Delta$.
        while $k > 1$ **and** $\lambda > l'_k$ **do**
            $k \leftarrow k - 1$.
        end while
        $V(m, n) \leftarrow V(m, n-1) + l_k - q\Delta - F(w_k)\lambda$.
        $p(m, n) \leftarrow w_k$.
    end for
end for
Output: Value table $\{V(m, n)\}$ and first-offer policy $\{p(m, n)\}$.

and weakly decreasing in the number of workers n. When m and n both reduce by one, the optimal first offer also weakly decreases. In other words, $p(m, n) \leq p(m, n-1) \leq p(m+1, n)$.

Theorem 2 reveals how the balance between supply and demand affects the optimal pay. When demand and supply are perfectly balanced ($m = n$), there is exactly one worker for each request, and making a low pay offer at the risk of wasting a worker is not worthwhile, as it also risks wasting a request. When demand exceeds supply ($m > n$), workers are even more scarce and the same rationale holds. The consequence is that when $m \geq n$, all workers are offered the optimal one-shot pay w_{k^*}.

However, when supply exceeds demand ($m < n$), the optimal pay depends on the gap between supply and demand, which can vary as workers accept or reject pay offers. After an acceptance, the number of workers and requests each decreases by one: the wedge

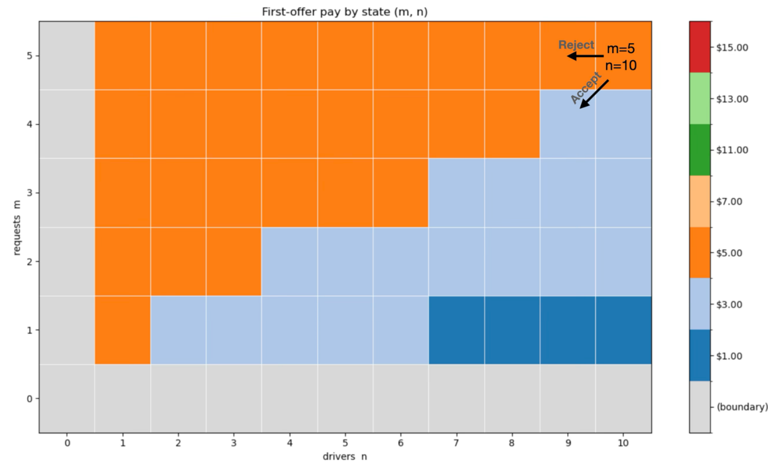


Figure 4: Numerical example: optimal first offer given (m, n)

Notes: Only three pay levels ($5, $3, and $1) can appear in an optimal sequence, as they are the only pay levels on the upper convex hull. All other pay levels are ruled out by Theorem 1. When a worker rejects an offer and the request is not cancelled, the state moves from (m, n) to $(m, n-1)$, illustrated by the horizontal arrow; when a worker accepts an offer or if a request is canceled, the state moves from (m, n) to $(m-1, n-1)$, illustrated by the diagonal arrow.

$(n-m)$ remains the same, but the number of requests decreases. Hence, the supply-demand imbalance, $(n-m)/m$, increases, which decreases the optimal worker pay. In contrast, after a rejection, the number of requests m remains the same, but the wedge $(n-m)$ decreases by one. The supply-demand imbalance $(n-m)/m$ thus also decreases, and the platform should raise its next pay offer. If rejection continues until $n = m$, we return to the balanced setting, and the pay remains at the single-shot optimal pay w_{k^*}, regardless of any future acceptances or rejections. The monotonicity of the optimal first offer is illustrated in Section 4.3.

Another useful observation from Theorem 2 concerns how to compute the optimal pay policy. When the global factor is known, the environment after any rejection is unchanged except for the reduction in number of workers. This means that the optimal policy with m remaining requests and n remaining workers does not depend on how many requests have previously been served or how many workers have previously rejected pay offers. As a result, when the worker pool grows from n to $n+1$, the pay structure for the last n workers remains the same —– the platform need only compute a new pay offer for the first worker — which, by Theorem 2, must be weakly lower than all subsequent offers. This nested structure is helpful in practice because it enhances interpretability of the optimal pay policy.

We next study the *width* of the pay band, denoted by $b(m, n)$, which we define as the difference between the highest and lowest possible offers in the optimal pay sequence for a particular m and n. We can state the following corollary to Theorem 2.

Corollary 1 (Supply-demand effect on the width of pay sequence)**.** *Given m requests and n workers, the highest offer is w_{k^*}, and the lowest offer is $p(1,(n-m)^+ + 1)$. Hence, $b(m,n) = w_{k^*} - p(1,(n-m)^+ + 1)$. As a result,*

1. *if $n \le m$, $b(n,m) = 0$, and*
2. *if $n > m$, $b(m,n)$ only depends on the gap $(n-m)$. As $(n-m)$ decreases, the pay width also decreases, reaching zero when $n = m$.*

As shown by Corollary 1, the width of the pay sequence critically depends on the gap between supply and demand, $(n-m)^+$. Intuitively, the larger the excess supply, the wider the pay sequence, as the platform has more opportunities to identify workers with a lower reservation wage. In addition to supply and demand, the pay sequence width also depends on customer patience, as captured by the request cancellation probability q.

Proposition 1. *Given m requests and n workers, the following statements hold.*

1. *All else being equal, the width of the optimal pay sequence shrinks as q increases.*
2. *When $q \ge q_0$, the width of the optimal pay sequence shrinks to zero, i.e., it is optimal to offer all workers the same pay. Moreover, the threshold is given by*

$$q_0 = 1 - \frac{l'_{k^*-1}}{l_{k^*} + (1 - F(w_{k^*}))l'_{k^*-1}} \tag{10}$$

where k^* is defined by Eq. (9), $l_k \triangleq F(w_k)(v - w_k)$, and $l'_k \triangleq \frac{l_k - l_{k-1}}{F(w_k) - F(w_{k-1})}$.

In other words, the width of the pay sequence is affected by the urgency of the requests. As customers become more impatient, the problem is more likely to terminate due to request cancellation, making the decision more similar to a single-period problem. When the cancellation probability q is so large that it exceeds the probability in Eq. (10), it is optimal to pay all workers the same pay (i.e., the pay that maximizes the platform's immediate reward $F(w_k)(v - w_k)$); the width of the pay sequence thus shrinks to zero. In this setting, there is no room for pay optimization because the risk of losing a request due to a failed match is too great.

5 The General Case: Optimal Pay with Unknown Global Factor

Having characterized the optimal pay policy when the global factor is known, we now turn to the more realistic setting in which it is unknown. Introducing uncertainty fundamentally changes the problem: the optimal policy becomes a belief-based dynamic program that is difficult to compute and interpret. Yet the structural insights from the full-information case remain powerful. We show that policies that reuse the full-information structure—either by committing immediately or after a single informative probe—can achieve strong performance while remaining transparent and computationally tractable.

5.1 Structure of the Optimal Pay Policy

One major challenge in solving the Bellman equation for the belief-augmented MDP (7) is that the value function depends on the platform's current belief about the global factor. Not only do different beliefs about the global factor lead to different actions; the platform's choice of pay for the current period also affects the belief update, and thus the optimal actions in future periods. Therefore, the state of the system must track not only the remaining number of requests m and drivers n, but also the current belief μ, leading to infinite states in the general setting and a potential challenge in solving the belief-augmented MDP. Fortunately, we can adapt the following key result from the Partially Observed MDP (POMDP) literature (Kaelbling et al. 1998).

Proposition 2. *For any m and n, the optimal value function $V(m, n, \mu)$ is a convex, piecewise linear function of μ.*

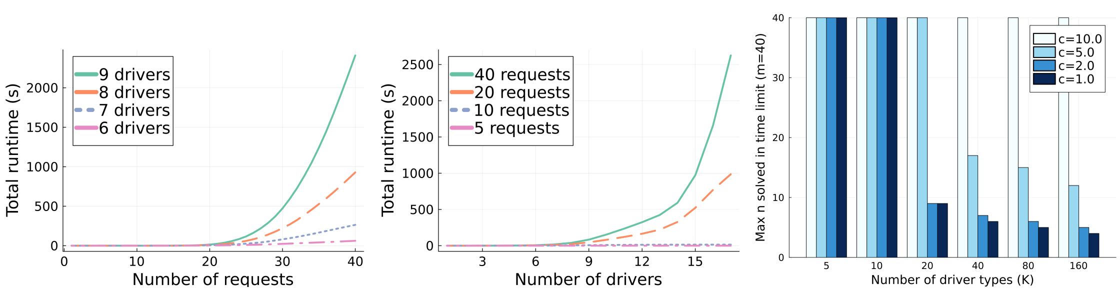
Proposition 2, which we prove in the appendix, enables a compact representation of the original problem: instead of storing a value for every possible belief, we can represent the value function $V(m, n, \mu)$ for each m and n as a finite set of line segments. Solving the Bellman equation (7) to compute each $V(m, n, \mu)$ is then equivalent to computing the maximum of K piecewise linear functions, which we can do efficiently using an adaptation of the line sweep algorithm from Bentley and Ottmann (1979) (see appendix). We illustrate this representation in Fig. 5, where we observe the optimal value function for $m = 1$ request and $n = 3$ workers, along with the optimal pay policy.

Unfortunately, even though the line sweep algorithm we develop scales linearly in the number of line segments in the piecewise linear representation of $V(m, n, \mu)$, that number of line segments can grow exponentially in the number of requests and drivers, leading to tractability challenges in many situations. We summarize these challenges in Figure 6. We observe that the runtime of the exact algorithm quickly rises to minutes and then hours when the number of drivers and requests enters the low double-digits. Tractability is also affected by the parameters of the problem. A lower value of c (meaning that the global effect is small and thus difficult to learn) leads to much longer runtimes than a higher value of c. In practice, computing a pay policy should be completed in seconds so as not to increase waiting times for customers.

In addition to being intractable, the exact approach described here is not very interpretable, as there may be many different optimal pay sequences depending both on the initial belief

Figure 5: Illustration of the optimal value function and pay offers ($m = 1$, $n = 3$)..

Notes: We assume that $v = 12$, $c = 5$ and individual worker reservation wages are uniformly distributed on $\{0, 2.5, 5, 7.5, 10\}$. We see that the value function is indeed piecewise linear and convex in the global factor belief μ. In this example, the optimal pay for the last worker happens to be independent of the belief.



(a) Effect of n ($K = 20, c = 2$). (b) Effect of m ($K = 40, c = 5$). (c) Effect of c and K.

Figure 6: Tractability analysis of exact DP algorithm.

Notes: We assume that driver reservation wages are uniformly distributed among K equally spaced value between 0 and 10, with $v = 20$. All runtimes are computed on a single laptop (2025 Macbook pro with M5 chip). Panels (a) and (b) show the effect of the number of requests and the number of drivers on runtimes for different values of K and c. Panel (c) fixes a one-hour time limit and seeks to compute the optimal value functions for all $m, n \leq 40$. This is possible for small numbers of driver types, but the maximum number of drivers the method can handle within the time limit quickly drops to the single digits as driver heterogeneity increases.

and on the particular acceptance realizations from drivers. In the following sections, we describe simpler approaches with provable guarantees that may provide more actionable practical insights.

5.2 Clairvoyant Benchmark and Regret

We first leverage the work of the previous section to establish a *clairvoyant benchmark*. Let $V^0(m,n)$ and $V^c(m,n)$ denote the optimal value functions when the global factor is known to be $C=0$ and $C=c$, respectively. In other words, $V^r(m,n)$ represents the maximum expected total reward achievable when the platform knows the regime $r \in \{0,c\}$.

We define the clairvoyant benchmark as

$$V^{\mathrm{clair}}(m,n) \ \triangleq\ (1-\mu_0)\, V^0(m,n) + \mu_0\, V^c(m,n),$$

which corresponds to the expected payoff of a platform that observes the true global status at time 0 before making any decisions.

For any policy π that does not observe the global status and operates under the prior belief μ_0, let $V^\pi(m,n \mid \mu_0)$ denote its ex-ante expected value, where the expectation is taken over both the realization of C and the randomness induced by the policy. We evaluate policies by their *expected regret* relative to the clairvoyant benchmark, defined as

$$\mathrm{Regret}(\pi) \ \triangleq\ V^{\mathrm{clair}}(m,n) - V^\pi(m,n \mid \mu_0).$$

This notion of regret captures the performance loss due to uncertainty about the global status and serves as the primary metric for comparing policies throughout the remainder of the paper.

5.3 Direct-Commit: a No-Learning Approach

Given the full-information policies characterized in Section 4, a natural question is how much the platform loses when it chooses to simply ignore the global factor. In particular, we seek to quantify the performance of applying the policy π^0 when the true regime is $C=c$, or applying π^c when the true regime is $C=0$, by bounding the regret induced by such misspecified commitments.

5.3.1 Expected regret under full-information policies

Recall that π^0 and π^c represent the optimal policy under full global information, when $C=0$ and $C=c$, respectively (Section 4). For any policy π, we write $V^0(\pi;m,n)$ and $V^c(\pi;m,n)$ for the expected value obtained when the platform executes policy π while the true global factor is $C=0$ or $C=c$, respectively.

We now establish two robustness results showing that, when the global factor c is small, the performance loss from applying the "wrong" benchmark policy is bounded. We first consider the case in which the platform optimizes as if $C = c$ (i.e., applies π^c), but the true global factor is $C = 0$. Intuitively, this corresponds to the platform being overly conservative in its pay decisions. The following proposition shows that, when the global factor c is small, the resulting performance loss is limited.

Proposition 3. *Let π^c be the optimal policy computed under the assumption $C = c$. Suppose the true global factor is $C = 0$. Then for any number of requests m and workers n,*

$$V^0(m,n) - V^0(\pi^c; m,n) \le c \min\{m,n\}, \tag{11}$$

the right-hand side of which goes to zero as $c \to 0$.

We next consider the opposite case, in which the platform optimizes under the assumption that the global factor is $C = 0$, when in fact $C = c$.

Proposition 4. *Let π^0 be the optimal policy computed under the assumption $C = 0$. Suppose the true global factor is $C = c$. Define $\bar{v} = v - \min_k w_k$. Then for any number of requests m and workers n,*

$$V^c(m,n) - V^c(\pi^0; m,n) \le \beta \min\{m,n\} \tag{12}$$

where $\beta \triangleq \min\{c, (\bar{v} - c)^+\}$ and goes to zero as $c \to 0$ or $c \to \bar{v}$.

Similar to Proposition 3, when c is small, applying π^0 under $C = c$ has limited impact. Moreover, when c approaches $\bar{v}$, the expected loss also vanishes to zero. This is because as c approaches $\bar{v}$, the profit margin from $C = c$ approaches zero; in this case, even under full information, the platform cannot make any profit. Hence, applying the "wrong" policy does not have any loss.

5.3.2 Algorithm: Direct-Commit (DC)

Motivated by the robustness bounds in the previous subsection, we now introduce a simple *direct-commit* (DC) algorithm that performs well when the global factor c is small. Rather than attempting to learn the global status through experimentation, DC commits upfront to one of the two benchmark policies, π^0 or π^c, based on the prior belief about C. At a high level, DC compares the expected loss from committing to each benchmark policy and selects the one with the smaller worst-case regret under the prior belief μ_0.

We formalize DC as follows.

Definition 1 (Direct-Commit)**.** *Let π^{DC} denote the policy that, given prior belief μ_0, applies π^0 if*

$$\mu_0 \le \frac{c}{c + \beta}, \tag{13}$$

and applies π^c otherwise.

Recall that $\beta = \min\{c, (\bar{v} - c)^+\}$. The threshold in (13) balances the expected loss bound associated with the two forms of misspecification. The following proposition shows that Direct-Commit inherits a bounded-regret guarantee from the regime-wise robustness bounds established in the previous subsection.

Proposition 5 (Regret bound for Direct-Commit)**.** *With prior belief μ_0, the Direct-Commit policy π^{DC} satisfies*

$$\mathrm{Regret}(\pi^{\mathrm{DC}}) \triangleq V^{\mathrm{clair}}(m,n) - V^{\pi^{\mathrm{DC}}}(m,n \mid \mu_0) \leq L(\mu_0; c) \cdot \min\{m,n\}. \tag{14}$$

where $L(\mu_0) = \min\{\mu_0\beta, (1-\mu_0)c\}$ and $\beta = \min\{c, (\bar{v} - c)^+\}$.

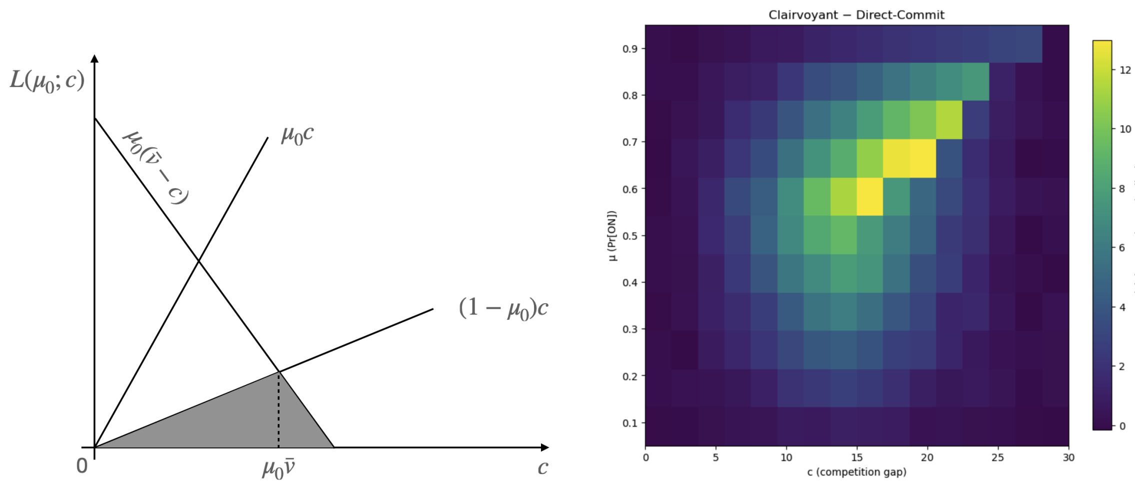
Proposition 5 formalizes the performance guarantee of Direct-Commit. When the global factor c is small, committing to a single benchmark policy incurs only a limited loss, even if the assumed global status is incorrect. In such a regime, learning is difficult but not particularly helpful, so a simple no-learning policy can perform nearly as well as the clairvoyant benchmark.

More interestingly, Proposition 5 shows that when c approaches $\bar{v}$, Direct-Commit is also optimal. Intuitively, as c approaches $\bar{v}$, the maximum profit margin from $C = c$ vanishes to zero; if $C = c$ is no longer a profitable scenario, then there is no value in learning; simply ignoring the possibility of $C = c$ and applying π^0 is the optimal policy.

Fig. 7 illustrates the regret behavior of the Direct-Commit (DC) algorithm. Fig. 7a depicts the constant multiplier $L(\mu_0; c)$ for the regret upper bound in Eq. (14) as a function of the global factor magnitude c, for a fixed prior $\mu_0 \geq 0.5$. The shaded region represents the maximum possible expected regret per transaction incurred by DC. The term is maximized at $c = \mu_0\bar{v}$, which corresponds to the point at which the expected loss from applying π^0 when $C = c$ equals the expected loss from applying π^c when $C = 0$. At this point, the platform is effectively indifferent between committing to either benchmark policy ex ante, making the commitment decision maximally difficult.

The regret bound per transaction at this point equals $(1 - \mu_0)\mu_0\bar{v}$, which is maximized when $\mu_0 = 0.5$. This highlights that the worst-case performance of DC arises precisely when the prior belief is least informative. Fig. 7b corroborates this insight empirically: the average regret from simulation is largest along the diagonal $c = \mu_0\bar{v}$, appearing as a bright band in the heatmap, and the largest regret (bright yellow) is near $\mu_0 = 0.5$.

Taken together, the two panels reveal an inherent limitation of Direct-Commit. While DC performs well when the global factor is either small or large, it underperforms in regions where the prior belief does not clearly favor either benchmark policy. In such settings, the inability to acquire additional information leads to nontrivial regret. Motivated by



(a) $L(\mu_0; c)$ as a function of c ($\mu_0 \geq 0.5$)

(b) Average regret from simualtion

Figure 7: Theoretical regret bound behavior and average regret from simulation for Direct-Commit

Note: Parameters for the simulation: $M = 20$, $N = 20$, $v = 30$, types:$\{0, 5, 10, 15, 20\}$, CDF:$\{0.10, 0.30, 0.55, 0.80, 1.0\}$, 20,000 iterations.

this observation, we next introduce the *Probe-and-Commit* algorithm, which augments DC with an initial learning step to refine the platform's belief before committing.

5.4 Probe-and-Commit: a One-Step Learning Approach

We start by introducing the formal definition of Probe-and-Commit, for a probe value of p.

Definition 2 (Probe-and-Commit with probe price p)**.** *Fix a probe price p and prior belief μ_0. The Probe-and-Commit policy $\pi^{\mathrm{PC}}(p)$ operates as the following steps:*

1. *Probe. In the first period, the platform offers payment p to the first worker and observes the acceptance outcome $Y \in \{A, R\}$.*
2. *Belief update. Update the posterior belief after observing Y is updated according to Bayes' rule, $\mu'_A(p)$ and $\mu'_R(p)$, defined by Eq.* (5) *and Eq.* (6)*, respectively.*
3. *Commit. Conditional on the posterior belief μ_Y, the platform applies the Direct-Commit rule: it commits to policy π^0 if $\mu_Y \leq \frac{c}{c+\beta}$, and to policy π^c otherwise, where $\beta = \min\{c, (\bar{v} - c)^+\}$.*
4. *Continuation. The platform applies the selected benchmark policy for all remaining periods.*

To meaningfully talk about a "good" probe, we first present the performance bound for any Probe-and-Commit type of policy, given an arbitrary probe p:

Proposition 6 (Probe-and-Commit performance bound)**.** *Consider a probe value p. Then the expected regret from Probe-and-Commit with probe p is bounded by:*

$$
\begin{aligned}
\text{Regret}(\pi^{PC}(p)) &= V^{clair}(m,n) - V^{\pi^{PC}(p)}(m,n|\mu_0) && (15)\\
&\leq \underbrace{(\bar{v} - P_{accept}(p,\mu_0)(v-p))}_{\textit{immediate loss}} + \underbrace{\min\{m,n-1\}\cdot \mathbb{E}\left[L(\mu'_Y(p);c)\right]}_{\textit{continuation loss}}. && (16)
\end{aligned}
$$

where $Y \in \{A,R\}$, and $\mu'_A(p)$ and $\mu'_R(p)$ are defined by Eq. (5) *and Eq.* (6)*, respectively.*

The decomposition in Eq. (16) separates the regret from probing into an *immediate loss* term and a *continuation loss* term that depends on the posterior belief induced by the probe. This structure allows us to characterize the probe that minimizes the regret upper bound in a nice form.

Proposition 7 (Optimal probe for minimizing the regret bound)**.** *Fix prior belief μ_0 and c. Then the choice of probe p that minimizes the regret upper bound in Eq.* (16) *is equivalently given by*

$$p^\star \in \arg\max_p \underbrace{P_{\text{accept}}(p,\mu_0)(v-p)}_{\textit{immediate return}} + \underbrace{\min\{m,n-1\}\cdot G(p)}_{\textit{Information gain}} \tag{17}$$

where

$$G(p) \triangleq \big(c(1-\mu_0)\,F(p) - \beta\,\mu_0\,F(p-c)\big)^+. \tag{18}$$

Recall that β captures the scale of penalty associated with applying π^0 when $C=c$, while c plays an analogous role for applying π^c when $C=0$. The function $G(p)$ captures the information value of a probe by measuring the likelihood of observing different worker outcomes under the two regimes, weighted by these asymmetric penalty scales. Intuitively, when β is substantially larger than c, the probing decision places greater emphasis on minimizing the term $\beta\mu_0 F(p-c)$, leading to a smaller probe that increases the chance of rejection and therefore raises the likelihood of committing to π^c in the next step. When the penalty scales are equal, maximizing $G(p)$ reduces to maximizing $(F(p) - F(p-c))$, i.e., the gap between acceptance probabilities under the two regimes.

As a side note, mathematically, $G(p)$ is directly related to the continuation loss through the identity $\mathbb{E}\left[L(\mu'_Y(p);c)\right] = c(1-\mu_0) - G(p)$.

A special case: strong global factor The value of the Probe-and-Commit policy is perhaps most apparent in the special case where the global effect dominates individual heterogeneity, i.e., $c > w_K - w_1$. In this case, the reservation wage distributions under $C = 0$ and $C = c$ do not overlap, implying that the global status C can be identified from a single probe.

Corollary 2. *Assume $c > w_K - w_1$. Then at probe $p = w_K$, we have $\mathbb{E}[L(\mu'_Y(w_K); c)] = 0$, and therefore the continuation loss term in Proposition 6 vanishes. Moreover,*

$$\text{Regret}(\pi^{PC}(w_K)) \leq \bar{v} - (1 - \mu_0)(v - w_K),$$

which is independent of m, n, and c (within the strong-global-factor regime).

Corollary 2 highlights a sharp regime change in the strong global factor case. When $c > w_K - w_1$, probing at the highest type $p = w_K$ perfectly separates the two regimes: under $C = 0$ the offer is always accepted, while under $C = c$ it is always rejected. Formally, $F(w_K) = 1$ and $F(w_K - c) = 0$, which implies $P_{\text{acc}}(w_K, \mu_0) = 1 - \mu_0$ and $\mu'_A(w_K) = 0$, $\mu'_R(w_K) = 1$. As a result, the global status is identified after a single observation, and the platform commits to the correct benchmark policy thereafter. Consequently, the continuation loss term in Proposition 6 vanishes, and the regret of Probe-and-Commit is entirely driven by the mismatch in the first-period payoff. Because this mismatch is bounded by a constant independent of (m, n), the relative cost of probing becomes negligible as the horizon grows. This makes Probe-and-Commit particularly attractive in environments where the global effect is sufficiently strong compared to individual heterogeneity.

5.5 Comparing Policies

Fig. 8 compares the performance of Direct-Commit with Probe-and-Commit, myopic commitment to π^0, and myopic commitment to π^c, using the clairvoyant policy as a benchmark. Direct-Commit performs well when the global factor c is small, and in some regions even outperforms Probe-and-Commit. Its weakest performance occurs near the threshold in Eq. (13), where the prior belief μ_0 and the magnitude of c make the policy choice particularly sensitive. In these cases, Direct-Commit may commit to π^c even though $C = 0$, leading to excessive pay. In contrast, Probe-and-Commit performs better when c is large, as even a single informative probe substantially improves the likelihood of selecting the correct policy. The managerial takeaways are clear: when the magnitude of the global effect is small, learning the global status is hard but not very important, so we can safely commit to a no-learning pay policy. As the magnitude of the global effect increases, learning becomes increasingly desirable, and even a single probing offer can significantly improve platform profit.

6 Value of Pay Flexibility to the Platform

In our model, a key driver of value to the platform is the ability to vary pay offers to different workers in order to increase profits. A natural line of inquiry asks how much such

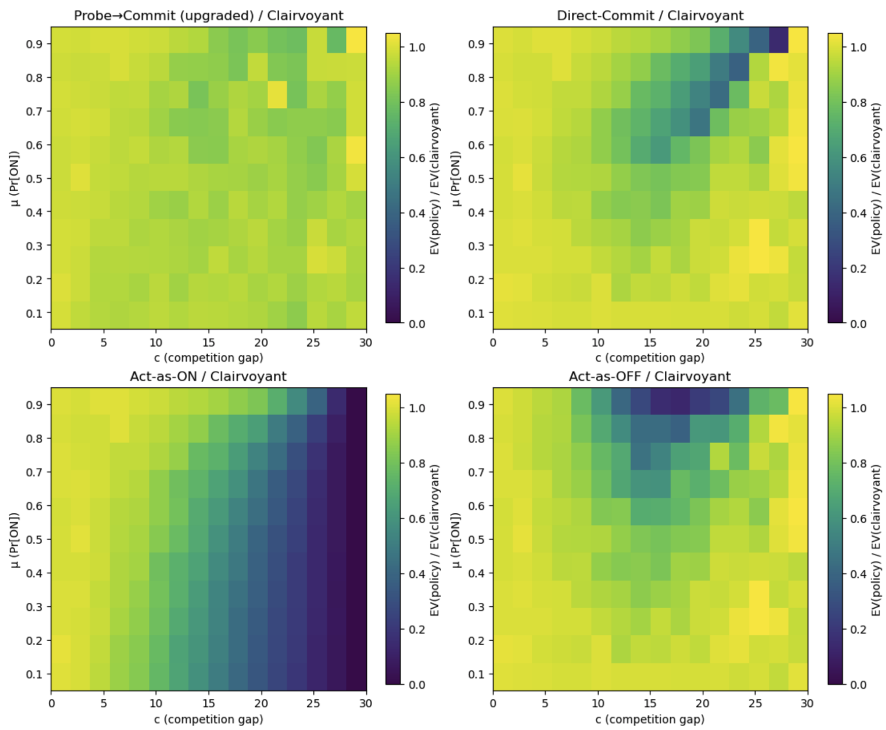


Figure 8: Performance comparion across multiple policies, relative to the clairvoyant benchmark

Note: Parameters: $M = 4$, $N = 7$, $v = 30$. Types $t \in \{0, 1, \ldots, 20\}$ with $\Pr(T \leq t) = t/20$. Grid: $\mu \in \{0.05, 0.14, \ldots, 0.95\}$, $c \in \{0, 2, \ldots, 30\}$. Simulations: 500.

dynamic pay policies benefit the platform relative to a static pay policy. This question is particularly relevant in practice, since a single-pay policy has attractive properties beyond profit, such as fairness and implementability. We first consider the setting where $C = 0$ deterministically, then the general setting where C is unknown and c is large.

6.1 Known Global Factor

We revisit the setting where $C = 0$ from Section 4. We first establish properties of the optimal single-pay policy in the following lemma.

Lemma 1. *Let $p_s(m,n)$ designate the optimal single-pay policy for m requests and n drivers. Then $p_s(m,\infty) := \lim_{n\to\infty} p_s(m,n) = w_{k_s}$, where*

$$k_s = \arg\max_{k\in[K]} \frac{F(w_k)(v - w_k)}{1-(1-q)(1-F(w_k))}.$$

Because the optimal single-pay policy with infinite drivers is the same for any finite number m of requests, we denote this policy by $p_s(\infty)$ to simplify notation. If we denote by $V_s(m,n)$ the value obtained from the optimal static policy, we observe that

$$V_s(m,\infty) = m\frac{F(w_{k_s})(v - w_{k_s})}{1-(1-q)(1-F(w_{k_s}))} = mV_s(1,\infty).$$

Lemma 1 provides a closed form expression of the optimal static pay policy when the number of drivers grows large relative to the number of requests. To interpret this closed form more readily, we first examine it for extreme values of the cancelation rate q.

Corollary 3. *If riders are maximally patient ($q = 0$), the optimal single pay policy with infinite drivers is to offer the minimal acceptable pay, i.e., $p_s(\infty) = w_1$. If riders are maximally impatient ($q = 1$), the optimal single pay policy with infinite drivers is to offer the single-period profit-maximizing pay, i.e., $p_s(\infty) = w_{k^*}$.*

The closed form in Lemma 1 can therefore be seen as interpolating between maximally patient requests, where the optimal static pay is simply the lowest acceptable pay — as the probability this offer is accepted tends to one as the number of drivers tends to infinity; and minimally patient requests, where the optimal static pay is simply myopic. Moderate values of q lead to optimal static pay policies between w_1 and w_{k^*}.

The static pay policy with infinite drivers also turns out to be a useful lower bound on the optimal pay policy, as described in the following lemma.

Lemma 2. *For any m, n, it holds that $p(m,n) \geq p_s(\infty)$. Consequently, the maximum pay width for m requests can be bounded as $\max_n B(m,n) \leq w_{k^*} - p_s(\infty)$.*

One interpretation of Lemma 2 is that, while switching from a static to a dynamic policy obviously increases the pay width, it does not automatically make every driver worse off. When many drivers are available, the optimal static policy will lead to low pay for all, while the optimal dynamic policy will increase pay as more drivers reject the offer.

Using Lemmas 1 and 2, we can establish the following result regarding the value of a flexible pay policy to the platform.

Proposition 8. *The value of flexibility, defined as $VoF(m,n) := V^*(m,n) - V_s(m,n)$, is characterized by the following properties:*

1. *Static pay is optimal for one driver: $VoF(m,1) = 0$.*
2. *Static pay is optimal for infinite drivers: $\lim_{n\to\infty} VoF(m,n) = 0$.*
3. *For any m, if $p_s(\infty) \neq w_{k^*}$, there exists n such that $VoF(m,n) > 0$.*

Unsurprisingly, adopting a dynamic policy is most effective for intermediate n. When n is very small, the optimal pay is nearly myopic, while when n is very large, the optimal pay is nearly static (except for the last few workers). Adopting our proposed optimal pay policy is most valuable when n is in between these two extremes.

6.2 Large Unknown Global Factor

Unlike in the known-status setting, learning and exploiting the global status usually *requires* some pay changes. This means that the value of a flexible pay policy does not tend to zero in general as the number of drivers tends to infinity.

Proposition 9 (Value of Flexibility)**.** *Let the initial belief $\mu_0 \in (0,1)$ and assume that $w_K \leq c$ (no overlap case). There exists m sufficiently large such that $\lim_{n\to\infty} VoF(m,n) > 0$.*

The proof of Proposition 9 bounds the value of flexibility away from zero for large enough m and n. The general lower bound is cumbersome to interpret, but we can obtain a simplified version in the following special case.

Corollary 4. *Assume that $q = 0$ and $w_1 = 0$, then $\lim_{n\to\infty} VoF(m,n) > 0$ if the following condition is satisfied:*

$$\frac{c}{c + (m-1)(v-c)} < \mu_0 < 1.$$

Corollary 4 gives clearer insights as to the value of a dynamic pay policy. Such a policy is worth it when the global factor c is small, the number of requests m is large, or the platform's initial belief that $C = c$ is high. Conversely, a dynamic policy may not be worth the implementation effort if the initial belief μ_0 is low, or if $v - c$ is small — this corresponds to the case where the cost of offering competitive pay to match the global factor erodes the platform's profit margins so much that learning the global status becomes worthless.

7 Conclusion and Discussion

Our analysis focuses on a specific regime of the gig-economy marketplace: an over-supplied setting where the number of available workers (supply) exceeds the immediate number of requests (demand). In this context, the main lever for net-return maximization is leveraging supply heterogeneity. Specifically, workers exhibit differences in their reservation wages, driven by idiosyncratic preferences and unobserved global factors (e.g., outside options or competitor incentives). By sequentially offering workers with a dynamic pay policy, the platform can learn the market status and select workers strategically, thereby minimizing the cost of service. In this perspective, the heterogeneity lies in the cost to serve rather than the value of the service itself.

This approach contrasts some of the recent literature, particularly studies utilizing Reinforcement Learning (RL) for dispatch and pricing. As reviewed by Qin et al. (2025), many RL-based approaches in ridesharing marketplaces work most effectively in a fundamentally under-supplied setting, where request volume exceeds driver availability. In such symmetrical or demand-heavy environments, the optimization focus shifts to leveraging demand heterogeneity. For instance, requests may differ significantly in their future value to the platform, such as a trip destined for a high-demand area that minimizes a driver's subsequent idle time or "deadheading". In these frameworks, the platform's strategic lever is not minimizing the immediate service cost, but rather allocating the scarce supply to the most valuable demand to optimize long-term platform objectives, e.g., total bookings.

While leveraging demand heterogeneity effectively utilizes scarce supply, our work highlights that supply-side optimization is equally critical when supply is abundant but heterogeneous. In markets like airport queues or during off-peak hours, the ability to dynamically compensate workers based on real-time acceptance behavior allows the platform to capture efficiency opportunities that would otherwise be lost to uniform pricing strategies. Thus, our work complements the existing literature by addressing the other side of the coin: optimizing labor acquisition costs when the constraint is not the number of workers, but the efficiency of the service cost.

References

Allon G, Cohen MC, Sinchaisri WP (2023) The impact of behavioral and economic drivers on gig economy workers. *Manufacturing & Service Operations Management* 25(4):1376–1393.

Amin K, Rostamizadeh A, Syed U (2013) Learning prices for repeated auctions with strategic buyers. *Advances in neural information processing systems* 26.

Banerjee S, Hssaine C, Kamble V (2024) Price competition under a consider-then-choose model with lexicographic choice. *arXiv preprint arXiv:2408.10429* .

Bentley, Ottmann (1979) Algorithms for reporting and counting geometric intersections. *IEEE Transactions on computers* 100(9):643–647.

Besbes O, Gur Y, Zeevi A (2014) Stochastic multi-armed-bandit problem with non-stationary rewards. *Advances in neural information processing systems* 27.

Besbes O, Zeevi A (2009) Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. *Operations research* 57(6):1407–1420.

Bimpikis K, Candogan O, Saban D (2019) Spatial pricing in ride-sharing networks. *Operations Research* 67(3):744–769.

Cao Y, Kleywegt A, Wang H (2025) Dynamic pricing for two-sided marketplaces with offer expiration. *Management Science* .

Castillo JC, Knoepfle D, Weyl G (2017) Surge pricing solves the wild goose chase. *Proceedings of the 2017 ACM Conference on Economics and Computation*, 241–242.

Chaum M (2023) Understanding upfront fares. URL `https://medium.com/uber-under-the-hood/understanding-upfr`

Chawla S, Hartline JD, Malec DL, Sivan B (2010) Multi-parameter mechanism design and sequential posted pricing. *Proceedings of the forty-second ACM symposium on Theory of computing*, 311–320.

Chen R, Kim S, Wang H, Wang X (2021) Posted price versus hybrid mechanisms in freight transportation marketplaces. *arXiv e-prints* arXiv–2106.

Cohen MC, Lobel I, Paes Leme R (2020) Feature-based dynamic pricing. *Management Science* 66(11):4921–4943.

Cohen MC, Zhang R (2022) Competition and coopetition for two-sided platforms. *Production and Operations Management* 31(5):1997–2014.

Feldman M, Gravin N, Lucier B (2014) Combinatorial auctions via posted prices. *Proceedings of the twenty-sixth annual ACM-SIAM symposium on Discrete algorithms*, 123–135 (SIAM).

Gallego G, Van Ryzin G (1994) Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management science* 40(8):999–1020.

Gallego G, Wang R (2014) Multiproduct price optimization and competition under the nested logit model with product-differentiated price sensitivities. *Operations Research* 62(2):450–461.

Garivier A, Moulines E (2011) On upper-confidence bound policies for switching bandit problems. *International conference on algorithmic learning theory*, 174–188 (Springer).

Hu B, Hu M, Zhu H (2022) Surge pricing and two-sided temporal responses in ride hailing. *Manufacturing & Service Operations Management* 24(1):91–109.

Johari R, Kamble V, Kanoria Y (2021) Matching while learning. *Operations Research* 69(2):655–681.

Kaelbling LP, Littman ML, Cassandra AR (1998) Planning and acting in partially observable stochastic domains. *Artificial intelligence* 101(1-2):99–134.

Ma H, Fang F, Parkes DC (2020) Spatio-temporal pricing for ridesharing platforms. *ACM SIGecom Exchanges* 18(2):53–57.

Özkan E (2020) Joint pricing and matching in ride-sharing systems. *European Journal of Operational Research* 287(3):1149–1160.

Qin Z, Tang X, Li Q, Zhu H, Ye J (2025) *Reinforcement Learning in the Ridesharing Marketplace* (Springer).

Russo DJ, Van Roy B, Kazerouni A, Osband I, Wen Z, et al. (2018) A tutorial on thompson sampling. *Foundations and Trends® in Machine Learning* 11(1):1–96.

Sandholm T, Gilpin A (2006) Sequences of take-it-or-leave-it offers: Near-optimal auctions without full valuation revelation. *Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems*, 1127–1134.

Slivkins A, et al. (2019) Introduction to multi-armed bandits. *Foundations and Trends® in Machine Learning* 12(1-2):1–286.

Tripathy M, Bai J, Heese HS (2023) Driver collusion in ride-hailing platforms. *Decision Sciences* 54(4):434–446.

Vanunts A, Drutsa A (2018) Optimal pricing in repeated posted-price auctions. *arXiv preprint arXiv:1805.02574* .

Yan C, Yan J, Shen Y (2025) Pricing shared rides. *Operations Research* .

Appendices

A Proofs

Proof of Theorem 1 and Theorem 2.
We present the proofs of Theorem 1 and Theorem 2 together, as they rely on the same set of auxiliary lemmas. For convenience, we first state Theorem 3, which consolidates the main conclusions of Theorem 1 and Theorem 2, and then provide a unified proof.

Theorem 3 (Multiple requests)**.** *Suppose there are m requests and n workers. Denote the optimal first offer given m requests and n workers as $p(m,n)$. The following statements hold true:*

1. *It is only optimal to offer pay on the upper convex hull of $(F(p), F(p)(v-p))$, $p \in \{w_1, w_2, \ldots, w_K\}$. In other words, Theorem 1 continues to hold.*
2. *The optimal first offer is weakly increasing in the number of requests and weakly decreasing in the number of workers. When the number of requests and workers both reduce by one, the optimal first offer decreases. In other words, $w(m,n) \leq w(m,n-1) \leq w(m+1,n)$.*
3. *Any offer in an optimal pay sequence is bounded by the optimal one-shot pay. That is, $p(m,n) \leq w_{k^*}$ for any m, n. (w_{k^*} is defined in Eq. (9)).*
4. *When the number of requests is equal to or exceeds the number of workers, the optimal first pay is the optimal one-shot pay. That is, $p(m,n) = w_{k^*}$ when $m \geq n$. Furthermore, the value function $V(m,n) = n \cdot l_{k^*}$.*
5. *All results above continue to hold when a request may be cancelled after each rejection with a constant probability q.*

Proof of Theorem 3

Proof. Proof. The proof relies on the properties of the value function $V(m,n)$. Hence, we start by introducing Lemma 3:

Lemma 3. *Let $d_m^{(n)} = V(m,n) - V(m-1,n)$. In other words, $d_m^{(n)}$ represents the marginal benefit of adding an additional request when there are $(m-1)$ requests and n workers. The following statements hold:*

1. *$d_m^{(n)}$ is weakly decreasing in m for all $m \geq 1$.*
2. *$d_m^{(n)}$ is weakly increasing in n for all $n \geq 0$.*
3. *$d_m^{(n+1)} \leq d_{m-1}^{(n)}$ for any $m \geq 1$, $n \geq 0$.*

4. All properties above hold when there is a cancellation probability q after each rejection.

Proof. Proof. We first prove the weak concavity of $V(m,n)$ in m. Proof by induction.

Base case: When $n = 0$, $V(m,0) = 0$ for all m. Concavity verified.

Induction hypothesis: For some $N > 0$, $V(m,N) - V(m-1,N)$ is weakly decreasing in m for all $m \geq 1$.

Induction step: Our goal is to show that $V(m,N+1) - V(m-1,N+1)$ is weakly decreasing in m given the induction hypothesis. For simplicity, we define the following notation for the marginal difference in the value functions, holding $n = N$ fixed:

$$d_m = V(m,N) - V(m-1,N) = -\delta_m \tag{19}$$

In other words, the induction hypothesis assumes d_m is weakly decreasing in m. Equivalently, it assumes that δ_m is weakly increasing in m.

Now by Eq. (8), we can rewrite the value function as

$$V(m,N+1) = \max_k F(w_k)(v - w_k + V(m-1,N) - V(m,N)) + V(m,N) \tag{20}$$

$$= \max_k F(w_k)(v - w_k + \delta_m) + V(m,N) \tag{21}$$

Note that the terms concerning the maximization problem are only in the first part of the equation, i.e. $\max_k F(w_k)(v-w_k+\delta_m)$. Moreover, recall the definition $l_k = F(w_k)(v-w_k)$. Furthermore, we define $q_k = F(w_k)$ which implies $0 \leq q_k \leq 1$, and define:

$$H(\Delta) = \max_k \; l_k + q_k\Delta \tag{22}$$

Note that $H(\Delta)$ is a piecewise linear function and therefore has first order derivative everywhere. Moreover, since $0 \leq q_k \leq 1$ for all k, $H(\Delta)$ is weakly increasing and 1-Lipschitz. Hence, it satisfies the following inequality for any $x < y$:

$$|H(x) - H(y) \leq |x - y| \tag{23}$$

$$H(x) - H(y) \leq x - y \tag{24}$$

The second inequality is by the monotonicity. Plugging in $H(\cdot)$ to the value function, we have

$$V(m,N+1) = H(\delta_m) + V(m,N) \tag{25}$$

Hence,

$$\begin{aligned}
&V(m,N+1) - V(m-1,N+1) \\
&= H(\delta_m) + V(m,N) - H(\delta_{m-1}) - V(m-1,N) \\
&= H(\delta_m) - H(\delta_{m-1}) + (V(m,N) - V(m-1,N)) \\
&= H(\delta_m) - H(\delta_{m-1}) + d_m
\end{aligned}$$

By the induction hypothesis, δ_m is weakly increasing in m. Hence, $\delta_m \geq \delta_{m-1}$. We can then apply the inequality Eq. (23) and obtain:

$$\begin{aligned} H(\delta_m) - H(\delta_{m-1}) + d_m &\leq \delta_m - \delta_{m-1} + d_m && (26) \\ &= -\delta_{m-1} && (27) \\ &= d_{m-1} && (28) \end{aligned}$$

Therefore, we have shown that $V(m, N+1) - V(m-1, N+1) \leq d_{m-1}$. Next, we show that $V(m-1, N+1) - V(m-2, N+1) \geq d_{m-1}$. Apply Eq. (25) again, we have

$$V(m-1, N+1) - V(m-2, N+1) = H(\delta_{m-1}) - H(\delta_{m-2}) + d_{m-1}$$

Again, by the induction hypothesis, $\delta_{m-1} \geq \delta_{m-2}$. Since H is weakly increasing, we have

$$V(m-1, N+1) - V(m-2, N+1) \geq d_{m-1}$$

which implies that

$$V(m-1, N+1) - V(m-2, N+1) \geq V(m, N+1) - V(m-1, N+1)$$

This concluds the proof of the weak concavity of $V(m, n)$ in m.

Next, we prove that $V(m, n) - V(m-1, n)$ is weakly increasing in n. Recall that in the previous proof, we have defined d_m to represent the gap $V(m, n) - V(m-1, n)$ when $n = N$. Here, we extend the definition to incorporate n as a parameter. More precisely, let

$$d_m^{(n)} = V(m, n) - V(m-1, n)$$

and

$$\delta_m^{(n)} = V(m-1, n) - V(m, n) = -d_m^{(n)} \tag{29}$$

Then we have

$$V(m, n+1) = V(m, n) + H(\delta_m^{(n)})$$
$$V(m-1, n+1) = V(m-1, n) + H(\delta_{m-1}^{(n)})$$

Hence, we have

$$V(m, n+1) - V(m-1, n+1) = V(m, n) - V(m-1, n) + H(\delta_m^{(n)}) - H(\delta_{m-1}^{(n)})$$

$$d_m^{(n+1)} = d_m^{(n)} + H(\delta_m^{(n)}) - H(\delta_{m-1}^{(n)})$$

Since we have proved the weak concavity of $V(m, n)$ in m, we have

$$\delta_m^{(n)} \geq \delta_{m-1}^{(n)}$$

And we have shown that H is weakly increasing. Hence,

$$d_m^{(n+1)} = d_m^{(n)} + H(\delta_m^{(n)}) - H(\delta_{m-1}^{(n)}) \geq d_m^{(n)}$$

which verifies that $V(m, n+1) - V(m-1, n+1) \geq V(m, n) - V(m-1, n)$.

Last but not least, we prove that $d_m^{(n+1)} \leq d_{m-1}^{(n)}$. Applying the lipschitz property to function H, we have

$$H(\delta_m^{(n)}) - H(\delta_{m-1}^{(n)}) \leq \delta_m^{(n)} - \delta_{m-1}^{(n)}$$

Hence,

$$\begin{aligned}
d_m^{(n+1)} &= d_m^{(n)} + H(\delta_m^{(n)}) - H(\delta_{m-1}^{(n)}) && (30) \\
&\leq d_m^{(n)} + \delta_m^{(n)} - \delta_{m-1}^{(n)} && (31) \\
&= -\delta_{m-1}^{(n)} && (32) \\
&= d_{m-1}^{(n)} && (33)
\end{aligned}$$

which concludes the proof.

Extension to constant cancellation rate q. Next, we prove that all results above hold up when there is a probability of cancellation q after a rejection. The value function Eq. (8) in this case is given by

$$\begin{aligned}
V(m, n) = \max_{k \in [K]} F(w_k)\big((v - w_k) + V(m-1, n-1)\big) \\
+ (1 - F(w_k))\big(qV(m-1, n-1) + (1-q)V(m, n-1)\big).
\end{aligned} \quad (34)$$

Recall that a core proof technique in the previous setting with cancellation probability $q = 0$ is the definition of the function H (Eq. (22)), which we have shown to be non-decreasing and 1-Lipschitz. With the cancellation probability, we extend the definition of H to a new function H_c and show that all the rest proof arguments hold up.

By rearranging the terms in Eq. (34), we have

$$\begin{aligned}
V(m, n) = \max_{k \in [K]} F(w_k)\big((v - w_k) + (1-q)(V(m-1, n-1) - V(m, n-1))\big) \\
+ \big(qV(m-1, n-1) + (1-q)V(m, n-1)\big).
\end{aligned}$$

Now, define

$$H_c(\Delta) \triangleq q\Delta + H((1-q)\Delta) = q\Delta + \max_k \{l_k + q_k(1-q)\Delta\} \quad (35)$$

The second equality is the definition of function H. Recall that $l_k = F(w_k)(v - w_k)$ and $q_k = F(w_k)$. Then we can rewrite the value function $V(m, n)$ as:

$$\begin{aligned} V(m,n) &= \max_{k\in[K]}\{l_k + q_k(1-q)\delta_m^{(n-1)}\} + V(m,n-1) + q\delta_m^{(n-1)} \\ &= H_c(\delta_m^{(n-1)}) + V(m,n-1) \end{aligned}$$

where $\delta_m^{(n-1)}$ is defined Eq. (29) and represents the marginal benefit of adding one more request given $(n-1)$ workers. One can find that the above equation has the same format as Eq. (25); furthermore, we can show that the function H_c also has the same properties as function H in that it is non-decreasing and 1-Lipschitz. To see why: first, it is easy to confirm the monotonicity; given that q is the cancellation probability and therefore between 0 and 1, and that H is non-decreasing, H_c is the sum of two non-decreasing functions and therefore also non-decreasing. To verify that it is 1-Lipschitz, let $\delta_1 > \delta_2$, then we have

$$\begin{aligned} H_c(\delta_1) - H_c(\delta_2) &= q(\delta_1 - \delta_2) + H((1-q)\delta_1) - H((1-q)\delta_2) \\ &\le q(\delta_1 - \delta_2) + (1-q)(\delta_1 - \delta_2) \\ &= \delta_1 - \delta_2 \end{aligned}$$

which proves the property. Hence, all proofs can go through by replacing function H with H_c. $\square$ $\square$

Next, we prove each property in Theorem 3 in order. For brevity, we directly assume a cancellation probability of $q \in [0, 1)$. Hence, readers should refer to the extended value function Eq. (34).

Upper convex hull. Next, we start with the proof for Item 1 and first show that only the convex hull of $(F(w_k), F(w_k)(v - p_k)$ may be offered in the optimal policy for any (m, n). For ease of exposition we continue to use the notation $q_k = F(w_k)$ and $l_k = F(w_k)(v - p_k)$.

Consider a point (q_j, l_j) that is below the convex hull. Then there must exist two points (q_a, l_a) and (q_b, l_b) and weight $\lambda \in (0, 1)$ such that

$$q_j = \lambda q_a + (1-\lambda)q_b,\ l_j < \lambda l_a + (1-\lambda)l_b$$

Then it must be true that for any δ,

$$\begin{aligned} l_j + q_j\delta &< \lambda l_a + (1-\lambda)l_b + q_j\delta && (36) \\ &= \lambda l_a + (1-\lambda)l_b + (\lambda q_a + (1-\lambda)q_b)\delta && (37) \\ &= \lambda(l_a + q_a\delta) + (1-\lambda)(l_b + q_b\delta) && (38) \\ &\le \max\{l_a + q_a\delta,\ l_b + q_b\delta\} && (39) \end{aligned}$$

Hence, $(l_j + q_j\delta)$ is strictly dominated by either $(l_a + q_a\delta)$ or $(l_b + q_b\delta)$. Recall that the value function is given by

$$V(m,n) = \max_k\{l_k + q_k(1-q)\delta_m^{(n-1)}\} + V(m,n-1)$$

Hence, an offer with (q_j, l_j) can never appear in the optimal policy. This concludes our proof for the convex hull.

To show that the candidate solution must be on the *upper* convex hull (i.e. the increasing portion of the convex hull), consider a point (q_i, l_i) on the convex hull but not on the increasing portion. Thus, there must exist (q_a, l_a) such that $q_a < q_i$ and $l_a > l_i$. Since $\delta_m^{(n-1)} \le 0$, we must have

$$l_a + q_a(1-q)\delta_m^{(n-1)} > l_i + q_i(1-q)\delta_m^{(n-1)}$$

which means (q_i, l_i) is strictly dominated and cannot appear in the optimal sequence.

Offer monotonicity. Here we prove the monotonicity of the optimal first offer. The core of the proof is to show that the sign of $(p(m_1,n_1) - p(m_2,n_2))$ for any given m_1, m_2, n_1, n_2 is the opposite of the sign of $(d_{m_1}^{(n_1-1)} - d_{m_2}^{(n_2-1)})$. Once we have shown this, combining with Lemma 3, we obtain the monotonicity results in the theorem.

We present a graphical proof. Note that by the definition of the value function, the optimal pay under any (m,n) verifies

$$w(m,n) = \arg\max_k\{l_k + q_k(1-q)\delta_m^{(n-1)}\} = \max_k\{l_k - q_k(1-q)d_m^{(n-1)}\}$$

Define function $f(x; q_k, l_k) = l_k + d_m^{(n-1)}(1-q)(x - q_k)$. Then at $x = 0$, $f(0; q_k, l_k) = l_k - d_m^{(n-1)}(1-q)q_k$, which is exactly the function to be maximized within the brackets. Hence, $w(m,n)$ is the point on the upper convex hull (q_k, l_k) that leads to the highest $f(0)$. Fig. 9 illustrates how the optimal pay is determined by the slope d. As the slope of the line increases, the highest intercept happens at a (weakly) smaller q_k. Combined with Lemma 3, we conclude the following:

- $d_m^{(n-1)} \ge d_{m+1}^{(n-1)}$ implies that $w(m,n) \le w(m+1,n)$.
- $d_m^{(n-1)} \le d_m^{(n)}$ implies that $w(m,n) \ge p(m,n+1)$.
- $d_m^{(n)} \le d_{m-1}^{(n-1)}$ implies that $p(m,n+1) \ge w(m-1,n)$.

Offer upper bound. This is directly implied by the offer monotonicity. Given any number of requests and workers (m,n), by the monotonicity, it must be true that

$$w(m,n) \le w(m,n-1) \le \cdots \le p(m,1)$$

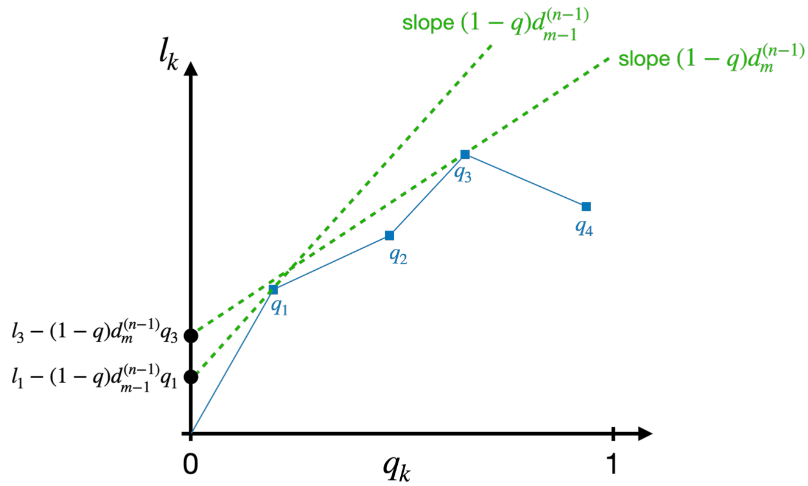


Figure 9: Graphical proof of offer monotonicity

When there is only one worker,

$$V(m,1) = \max_k F(w_k)(v - w_k) = l_{k^*}$$

and hence the optimal offer $p(m,1) = w_{k^*}$. Combined with monotonicity, we derive the upper bound in the statement.

Optimal pay when $m \geq n$. Consistent with previous proofs, we directly consider a cancellation probability q. We prove the statement by induction.

Base case. When $n = 1$ and $m \geq 1$, $V(m,1) = \max_k F(w_k)(v - w_k) = l_{k^*}$ and $p(m,1) = w_{k^*}$, which has been verified in the proof for "Offer upper bound".

Induction hypothesis. For some $N > 0$, $V(m,N) = N \cdot l_{k^*}$ and $w(m,n) = w_{k^*}$ for any $m \geq N$.

Induction step. Our goal is to show that for $n = N+1$, $V(m, N+1) = (N+1) \cdot l_{k^*}$ and $w(m,n) = w_{k^*}$ for any $m \geq N+1$. By definition,

$$\begin{aligned} V(m, N+1) = \max_{k \in [K]} \; & F(w_k)\big((v - w_k) + (1-q)\big(V(m-1,N) - V(m,N)\big)\big) \\ & + \big(qV(m-1,N) + (1-q)V(m,N)\big). \end{aligned}$$

Since $m \geq N+1$, $m-1 \geq N$, we can applying the induction hypothesis and obtain

$$V(m, N+1) = \max_{k \in [K]} F(w_k)(v - w_k) + N l_{k^*} \tag{40}$$
$$= l_{k^*} + N l_{k^*} \tag{41}$$
$$= (N+1) l_{k^*} \tag{42}$$

It is also easy to verify that $p(m, N+1) = w_{k^*}$, as the terms related to k is just $F(w_k)(v - w_k)$. $\square$ $\square$

Proof of Proposition 1

Proof. Proof. Define the scalar

$$V_\infty \triangleq \frac{l_{k^*}}{q + (1-q)F(\bar{w})}$$

and the condition

$$(1-q)V_\infty \leq l'_{k^*} \tag{43}$$

Fix any $m \geq 1$ and $n \geq 1$. With cancellation probability $q \in [0, 1)$, the DP recursion Eq. (34) can be written in the following form

$$V(m, n) = V(m, n-1) + q\, d_m^{(n-1)} + \max_{k \in [\bar{K}]} \Big\{ l_k - (1-q) F(w_k)\, d_m^{(n-1)} \Big\}, \tag{44}$$

where we use the marginal notation from Lemma 3:

$$d_m^{(n-1)} \triangleq V(m, n-1) - V(m-1, n-1).$$

Hence the optimal first offer at (m, n) is

$$k(m, n) \in \arg\max_{k \in [\bar{K}]} \Big\{ l_k - \lambda(m, n)\, F(w_k) \Big\}, \qquad \lambda(m, n) \triangleq (1-q) d_m^{(n-1)}.$$

By the convex-hull characterization used in the proof of Theorem 3, the maximizer equals $k^\star$ (i.e., the top hull point) whenever

$$\lambda(m, n) \leq l'_{k^*}. \tag{45}$$

Therefore, it suffices to show that

$$(1-q) d_m^{(n-1)} \leq l'_{k^*} \quad \text{for all } m, n. \tag{46}$$

Step 1: reduce to bounding the $m = 1$ marginal. By Lemma 3(1), for each fixed $n-1$, the marginal $d_m^{(n-1)}$ is weakly decreasing in m. Hence, for all $m \geq 1$,

$$d_m^{(n-1)} \leq d_1^{(n-1)} = V(1, n-1) - V(0, n-1) = V(1, n-1). \tag{47}$$

Thus it is enough to upper bound $V(1, n)$ uniformly over n.

Step 2: a uniform upper bound on $V(1, n)$. Consider the constant-offer policy that offers $\bar{w}$ to every worker. Let U_n be its achieved value with one request and n workers. Then $U_0 = 0$ and for $n \geq 1$,

$$U_n = l_{k^*} + (1 - F(\bar{w}))(1 - q)\, U_{n-1}. \tag{48}$$

Since $V(1, n)$ is the *optimal* value, $V(1, n) \geq U_n$; moreover, we will only use that U_n provides a simple scalar recursion that yields a uniform bound on $V(1, n)$ through the same one-step decomposition. Define $\rho \triangleq (1 - F(\bar{w}))(1 - q) \in [0, 1)$. Iterating Eq. (48) gives

$$U_n = l_{k^*} \sum_{t=0}^{n-1} \rho^t \leq \frac{l_{k^*}}{1 - \rho} = \frac{l_{k^*}}{q + (1 - q)F(\bar{w})} = V_\infty. \tag{49}$$

Consequently, $V(1, n) \leq V_\infty$ for all n (the right-hand side is the value of the corresponding infinite-horizon geometric recursion).

Combining Eq. (47) and the uniform bound $V(1, n-1) \leq V_\infty$ yields

$$(1 - q)d_m^{(n-1)} \leq (1 - q)V_\infty.$$

Therefore, if Eq. (43) holds, then Eq. (45) holds for every (m, n), so the maximizer in Eq. (44) is always $k^\star$. Hence the optimal offer is constant and equals $\bar{w}$ at every state.

Finally, Eq. (43) is equivalent to

$$(1 - q)\frac{l_{k^*}}{q + (1 - q)F(\bar{w})} \leq l'_{k^*},$$

which rearranges to Eq. (10). This completes the proof. □ □

Proof of Lemma 1.

Proof. Proof. Assume M and N are fixed and consider a fixed pay policy, i.e., $p(m, n) = w_k$ for all $0 \leq m \leq M$, $1 \leq n \leq N$. Let $V_k(m, n)$ designate the value of this optimal policy for a particular m and n. To simplify notation, let $l_k = F(w_k)(v - w_k)$, and let $a_k = (1 - q)(1 - F(w_k))$. We claim that we can write:

$$V_k(m, n) = \begin{cases} l_k \left[m \frac{1 - a_k^n}{1 - a_k} + \sum_{i=1}^{m-1} b_i(m, n) a_k^{n-i} \right] & \text{if } m < n \\ n l_k & \text{if } m \geq n \end{cases} \tag{50}$$

where

$$|b_i(m,n)| \le n^{m-1}.$$

We begin by proving the second case by induction over n. First, observe that if $n=1$, for any $m \ge 1$ we have $V_k(m,1) = l_k$. By the Bellman equation, we can write:

$$V_k(m, n+1) = l_k + a_k V_k(m,n) + (1-a_k)V_k(m-1,n).$$

Assume we indeed have $V_k(m,n) = nl_k$ for some n and all $m \ge n$. Then let $m \ge n+1$ (meaning $m-1 \ge n$) and observe:

$$V_k(m,n+1) = l_k + a_k V_k(m,n) + (1-a_k)V_k(m-1,n) = l_k + a_k n l_k + (1-a_k) n l_k = (n+1) l_k.$$

This completes the proof of the second case of (50). We now prove the first case by induction over m. We will need to work out two base cases explicitly to simplify the induction step later on. First, let $m=1$. In this case, we can write

$$V_k(1,n) = l_k\left(1 + a_k + \ldots + a_k^{n-1}\right) = l_k \sum_{i=0}^{n-1} a_k^i = l_k \frac{1-a_k^n}{1-a_k},$$

which matches (50) for $m=1$. Now let $m=2$. We claim that $b_1(m,n) = -n$, i.e., for $n > 2$,

$$V_k(2,n) = l_k\left[2\frac{1-a_k^n}{1-a_k} - n a_k^{n-1}\right].$$

We can do so by induction over n. Observe that

$$\begin{aligned}
V_k(2,3) &= l_k + a_k V_k(2,2) + (1-a_k)V_k(1,2) \\
&= l_k + 2a_k l_k + l_k(1-a_k^2) \\
&= l_k\left[2 + 2a_k + 2a_k^2 - 3a_k^2\right] \\
&= l_k\left[2\frac{1-a_k^3}{1-a_k} - 3a_k^2\right],
\end{aligned}$$

which proves the base case $n=3$. Now assume the claim is true for $n \ge 2$. Then we can write

$$\begin{aligned}
V_k(2,n+1) &= l_k + a_k l_k\left[2\frac{1-a_k^n}{1-a_k} - n a_k^{n-1}\right] + (1-a_k) l_k \frac{1-a_k^n}{1-a_k} \\
&= l_k\left[1 + 2\frac{a_k - a_k^{n+1}}{1-a_k} - n a_k^n + 1 - a_k^n\right] \\
&= l_k\left[2\frac{1-a_k^{n+1}}{1-a_k} - (n+1)a_k^n\right],
\end{aligned}$$

which completes the proof by induction over n. Now, let $m \geq 3$ and assume the first case of (50) holds for every (m_0, n) with $m_0 \leq m-1$. We first check the base case $m = n$:

$$\begin{aligned}
V_k(m,m) &= ml_k \\
&= l_k \left[m(1 + a_k + a_k^2 + \ldots + a_k^{n-1}) - m(a_k + a_k^2 + \ldots + a_k^{n-1})\right] \\
&= l_k \left[m\frac{1-a_k^n}{1-a_k} - \sum_{i=1}^{n-1} m a_k^{n-i}\right] \\
&= l_k \left[m\frac{1-a_k^n}{1-a_k} - \sum_{i=1}^{m-1} n a_k^{n-i}\right],
\end{aligned}$$

with $b_i(m,m) = -n$ clearly satisfying $|b_i(m,n)| \leq n^{m-1}$. We can now assume the claim holds for some $n \geq m$ and observe that it also holds for $n+1$:

$$\begin{aligned}
&V_k(m,n+1) = l_k + a_k V_k(m,n) + (1-a_k)V_k(m-1,n) \\
&\frac{V_k(m,n+1)}{l_k} = 1 + a_k\left[m\frac{1-a_k^n}{1-a_k} + \sum_{i=1}^{m-1} b_i(m,n)a_k^{n-i}\right] + (1-a_k)\left[(m-1)\frac{1-a_k^n}{1-a_k} + \sum_{i=1}^{m-2} b_i(m-1,n)a_k^{n-i}\right] \\
&\qquad = 1 + (m-1) + m\frac{a_k - a_k^{n+1}}{1-a_k} - (m-1)a_k^n + \sum_{i=1}^{m-1} b_i(m,n)a_k^{n+1-i} + (1-a_k)\sum_{i=1}^{m-2} b_i(m-1,n)a_k^{n-i} \\
&\qquad = m\frac{1-a_k^{n+1}}{1-a_k} - (m-1)a_k^n + \sum_{i=1}^{m-1} b_i(m,n)a_k^{n+1-i} + \sum_{i=2}^{m-1} b_{i-1}(m-1,n)a_k^{n+1-i} - \sum_{i=1}^{m-2} b_i(m-1,n)a_k^{n+1-i} \\
&\qquad = m\frac{1-a_k^{n+1}}{1-a_k} + \sum_{i=1}^{m-1} b_i(m,n+1)a_k^{n+1-i},
\end{aligned}$$

where

$$b_i(m,n+1) = \begin{cases} b_i(m,n) - (m-1) - b_i(m-1,n), & \text{if } i = 1, \\ b_i(m,n) + b_{i-1}(m-1,n) - b_i(m-1,n), & \text{if } 2 \leq i \leq m-2, \\ b_i(m,n) + b_{i-1}(m-1,n), & \text{if } i = m-1. \end{cases}$$

If $i = 1$, then

$$|b_i(m,n+1)| \leq |b_i(m,n)| + |b_i(m-1,n)| + (m-1) \leq n^{m-1} + n^{m-2} + (m-1) \leq (n+1)^{m-1},$$

where the penultimate inequality follows from the induction hypothesis and the last one holds because $m \geq 3$. Similarly, if $2 \leq i \leq m-2$, then

$$|b_i(m,n+1)| \leq |b_i(m,n)| + |b_{i-1}(m-1,n)| + |b_i(m-1,n)| \leq n^{m-1} + 2n^{m-2} \leq (n+1)^{m-1},$$

where the last inequality clearly holds whenever $n \geq m \geq 3$. Finally, when $i = m+1$,

$$|b_i(m,n+1)| \leq |b_i(m,n)| + |b_{i-1}(m-1,n)| \leq n^{m-1} + n^{m-2} \leq (n+1)^{m-1},$$

which completes the induction step. Thus the fixed pay policy $p(m,n) = w_k$ achieves value $V_k(m,n)$ as defined in (50). By the convergence properties of geometric series, and observing that $a_k \leq 1$ for all k, we can fix m and let n tend to infinity to obtain:

$$V_k(m,\infty) := \lim_{n\to\infty} V_k(m,n) = ml_k \frac{1}{1-a_k} = \frac{mF(w_k)(v-w_k)}{1-(1-q)(1-F(w_k))} = mV_k(1,\infty).$$

Assume without loss of generality that k_s is the unique maximizer of the equation in Lemma 1. Then, there must exist ε such that for all $k \neq k_s$,

$$\frac{F(w_{k_s})(v-w_{k_s})}{1-(1-q)(1-F(w_{k_s}))} > \frac{F(w_k)(v-w_k)}{1-(1-q)(1-F(w_k))} + \varepsilon.$$

By the convergence property of geometric series, there must also exist N_0 large enough such that for every k and for every $n \geq N_0$,

$$\left| m\frac{F(w_k)(v-w_k)}{1-(1-q)(1-F(w_k))} - V_k(m,n) \right| < \varepsilon.$$

Therefore, for any $n \geq N_0$, $V_{k_s}(m,n) > V_k(m,n)$ for all $k \neq k_s$, which completes the proof. □ □

Proof of Lemma 2

Proof. Proof. Assume for a contradiction that there exists m,n such that $w(m,n) < w_s(\infty)$. This means that there exists $k < k_s$ such that

$$\frac{F(w_{k_s})(v-w_{k_s})}{1-(1-q)(1-F(w_{k_s}))} > \frac{F(w_k)(v-w_k)}{1-(1-q)(1-F(w_k))}, \tag{51}$$

but

$$\begin{aligned} F(w_k)(v-w_k) + a_k V(m,n-1) + (1-a_k)V(m-1,n-1) > \\ F(w_{k_s})(v-w_{k_s}) + +a_{k_s}V(m,n-1) + (1-a_{k_s})V(m-1,n-1), \end{aligned}$$

where once again $a_k = (1-q)(1-F(w_k))$. We can re-arrange this inequality as

$$(a_k - a_{k_s})(V(m,n-1) - V(m-1,n-1)) > F(w_{k_s})(v-w_{k_s}) - F(w_k)(v-w_k),$$

and using (51) to lower bound the right-hand side we obtain

$$\begin{aligned} (a_k - a_{k_s})(V(m,n-1) - V(m-1,n-1)) &> F(w_k)(v-w_k)\left(\frac{1-(1-q)(1-F(w_{k_s}))}{1-(1-q)(1-F(w_k))} - 1\right) \\ &= \frac{F(w_k)(v-w_k)}{1-a_k}(a_k - a_{k_s}). \end{aligned}$$

Because $k < k_s$, we know that $F(w_k) < F(w_{k_s})$, implying $a_k > a_{k_s}$, meaning we can write

$$
\begin{aligned}
V(m, n-1) - V(m-1, n-1) &> \frac{F(w_k)(v - w_k)}{1 - a_k} \\
V(m, n-1) &> F(w_k)(v - w_k) + a_k V(m, n-1) + (1 - a_k) V(m-1, n-1) \\
V(m, n-1) &> V(m, n),
\end{aligned}
$$

which violates the monotonicity of the optimal value function.□ □

Proof of Proposition 8

Proof. Proof. The first statement in the proposition follows immediately from the definition of the single-pay and optimal policies.

For the second statement, we already know by definition that for any m, n, $V^*(m, n) \geq V_s(m, n)$. All we need to show is that $V^*(m, n) \leq V_s(m, \infty)$ for all m, n, which we do by induction.

Base cases:

$$
\begin{aligned}
V^*(m, 1) = F(w_{k^*})(v - w_{k^*}) &\leq \frac{F(w_{k^*})(v - w_{k^*})}{1 - (1 - q)(1 - F(w_{k^*}))} \\
&\leq \max_k \frac{F(w_k)(v - w_k)}{1 - (1 - q)(1 - F(w_k))} = V_s(1, \infty) \leq V_s(m, \infty),
\end{aligned}
$$

and $V^*(0, n) = 0 = V_s(0, \infty)$.

Induction step: Fix $m \geq 1$ and $n \geq 1$. Assume that $V^*(m_0, n_0) \leq V_s(m, \infty)$ if $m_0 \leq m - 1$ or $m_0 = m$ and $n_0 \leq n - 1$. Then we can write

$$
\begin{aligned}
V^*(m, n) &= \max_k F(w_k)(v - w_k) + a_k V^*(m, n-1) + (1 - a_k) V^*(m-1, n-1) \\
&\leq F(w(m, n))(v - w(m, n)) + a_{m,n} m V_s(1, \infty) + (1 - a_{m,n})(m - 1) V_s(1, \infty) \\
&\leq F(w_{k_s})(v - w_{k_s}) \frac{1 - a_{m,n}}{1 - a_{k_s}} + a_{m,n} V_s(1, \infty) + (m - 1) V_s(1, \infty) \\
&= m V_s(1, \infty) = V_s(m, \infty),
\end{aligned}
$$

which completes the proof of the second statement.

Finally, we turn to the proof of the third statement. Without loss of generality, assume the only possible pay offers are w_{k_s} and w_{k^*}. From the proof of Lemma 1, we know that because $w_s(\infty) \neq w_{k^*}$, there exists n_0 such that $V_{k_s}(m, n_0) > \bar{V}(m, n_0)$, where $\bar{V}(\cdot)$ is shorthand for $V_k(\cdot)$ when $w_k = w_{k^*}$. Consider the smallest such n_0. We know $n_0 > 1$ since w_{k^*} is always optimal if $n = 1$. Now, for $n = n_0$, consider the adaptive policy that offers

w_{k_s} to the first driver, then only offers w_{k^*} if the first request does not cancel, otherwise offers the optimal pay sequence with $m-1$ requests and n_0-1 drivers. Its value is given by

$$\hat{V}(m, n_0) = F(w_{k_s})(v - w_{k_s}) + a_{k_s}\bar{V}(m, n_0 - 1) + (1 - a_{k,s})V^*(m - 1, n_0 - 1),$$

By the definition of n_0, we know that $\bar{V}(m, n_0 - 1) > V_{k_s}(m, n_0 - 1)$. From optimality of $V^*(\cdot)$, we additionally know that $V^*(m - 1, n_0 - 1) \geq V_{k_s}(m - 1, n_0 - 1)$. We therefore conclude

$$V^*(m, n_0) \geq \hat{V}(m, n_0) > F(w_{k_s})(v - w_{k_s}) + a_{k_s}V_{k_s}(m, n_0 - 1) + (1 - a_{k,s})V_{k_s}(m - 1, n_0 - 1) = V_{k_s}(m, n_0),$$

which establishes that $\mathrm{VoF}(m, n_0) > 0$. □ □

A.1 Proofs for the general setting.

Proof. Proof of Proposition 2. We can prove the result by backwards induction. Note that in this proof we use w as a decision variable instead of p. First, observe that with one driver remaining and any number of requests $1 \leq m_0 \leq m$, we can write:

$$\begin{aligned} V(m, 1, \mu) &= \max_{w \in \mathcal{W}} \{P_{\text{accept}}(w, \mu) \cdot (v - w)\} \\ &= \max_{w \in \mathcal{W}} \{[(1 - \mu) \cdot F(w) + \mu \cdot F(w - c)] \cdot (v - w)\}. \end{aligned}$$

The right-hand side is the maximum of $|\mathcal{W}|$ linear functions of μ, therefore, $V(m_0, 1, \mu)$ is a piecewise linear convex function for any $1 \leq m_0 \leq m$. Now assume that $V(1, n', \mu)$ is indeed piecewise linear and convex for all $n' < n_0$. We can write

$$V(1, n_0, \mu) = \max_{w \in \mathcal{W}} \left\{P_{\text{accept}}(w, \mu) \cdot (v - w) + (1 - P_{\text{accept}}(w, \mu)) \cdot (1 - q)V(1, n_0 - 1, \mu'_R(w))\right\}$$

We know that for all $w \in \mathcal{W}$, the first term $P_{\text{accept}}(w, \mu) \cdot (v - w)$ is a linear function of μ. We then turn our attention to the second term $T(\mu, w) = (1 - P_{\text{accept}}(w, \mu))(1 - q)V(1, n_0 - 1, \mu'_R(w))$. By the induction hypothesis, there exist M pairs of coefficients (c_j, d_j) such that

$$T(\mu, w) = (1 - P_{\text{accept}}(w, \mu))(1 - q) \max_{j \in [M]}(c_j\mu'_R(w) + d_j(1 - \mu'_R(w))),$$

which we can expand by recalling that from Bayes' rule,

$$\mu' = \frac{\mu \cdot \Pr(W > w - c)}{\mu \cdot \Pr(W > w - c) + (1 - \mu) \cdot \Pr(W > w)} = \frac{\mu(1 - F(w - c))}{\mu(1 - F(w - c)) + (1 - \mu)(1 - F(w))}.$$

Furthermore, notice that we can also write

$$1 - P_{\text{accept}}(w, \mu) = 1 - (1 - \mu) \cdot F(w) - \mu \cdot F(w - c) = \mu(1 - F(w - c)) + (1 - \mu)(1 - F(w)),$$

which is exactly the denominator of μ'. Therefore, we can simplify

$$T(\mu, w) = (1-q) \max_{j \in [M]} (1 - F(w-c))c_j \mu + (1 - F(w))d_j(1-\mu).$$

This is a piecewise linear convex function of μ, which completes the proof of the second base case. We can now move to the general induction step. Assume that $V(m', n', \mu)$ is a convex piecewise linear function of μ for all (m', n') such that either $n' < n_0$ or $n' = n_0$ and $m' \leq m_0$. Then we can write

$$\begin{aligned} V(m, n, \mu) = \max_{p \in \mathcal{W}} & P_{\text{accept}}(p, \mu) \left(v - p + V(m-1, n-1, \mu'_A(p))\right) \\ & + (1 - P_{\text{accept}}(p, \mu)) \left[(1-q)V(m, n-1, \mu'_R(p)) + qV(m-1, n-1, \mu'_R(p))\right]. \end{aligned}$$

We can apply the same reasoning as when $m = 1$ to observe that the first and second terms are each piecewise linear convex function, and therefore the maximum remains a piecewise linear convex function. □

Proof of Proposition 3

Proof. Proof. We start the proof by first introducing the following lemma:

Lemma 4. *Fix $c > 0$. For each (m, n), let $V^0(m, n; v)$ and $V^0(m, n; v - c)$ be the optimal values under the environment of $C = 0$ and per request values v and $v - c$, respectively. Define*

$$\Delta(m, n) = V^0(m, n; v) - V^0(m, n; v - c) \tag{52}$$

Then for all m, n,

$$0 \leq \Delta(m, n) \leq c \min\{m, n\} \tag{53}$$

Proof. Proof of Lemma 4 We proceed by induction on $m + n$.

Base cases. If $m = 0$ or $n = 0$, no requests can be completed. Hence,

$$V^0(m, n; v) - V^0(m, n; v - c) = 0$$

which satisfies Eq. (53).

Induction hypothesis. Assume that for all (m', n') with $m' + n' < m + n$,

$$0 \leq \Delta(m', n') \leq c \min\{m', n'\}$$

Algorithm 2 Line sweep algorithm for maximum of piecewise linear (PWL) functions.

```
function MaxN(f_1, ..., f_k)                                  ▷ Upper envelope of k PWL functions on [0, 1]
    if k = 1 then return f_1
    else if k = 2 then return MaxTwo(f_1, f_2)
    else return MaxN(MaxTwo(f_1, f_2), ..., MaxTwo(f_{k-1}, f_k))
    end if
end function
function MaxTwo(f, g)                                         ▷ Upper envelope of two PWL functions on [0, 1]
    Input: Piecewise-linear functions f = (x^f, y^f) and g = (x^g, y^g)     ▷ (breakpoints, values)
    B ← [ ],  V ← [ ],                                        ▷ Envelope breakpoints and values
    i ← 1,  j ← 1
    f_larger ← (y^f_1 ≥ y^g_1)
    while i < |x^f| and j < |y^f| do
        if f_larger then AddPoint(B, V, x^f_i, y^f_i, ε)
        else AddPoint(B, V, x^g_j, y^g_j, ε)
        end if
        (x_int, y_int) ← FindIntersection(f, g, i, j, ε)      ▷ Find intersection of line segments using basic geometry.
        if inter ≠ Nothing then
            AddPoint(B, V, x_int, y_int, ε)
            f_larger ← ¬f_larger
            i ← i + 1,  j ← j + 1
        else
            if R_f < R_g − ε then i ← i + 1
            else if R_g < R_f − ε then j ← j + 1
            else i ← i + 1,  j ← j + 1
            end if
        end if
    end while
    Add the last point of f or g depending on which is larger. return (B, V)
end function
function AddPoint(B, V, x, y, ε)
    Only add x to B and y to V if x is sufficiently different (more than ε) from the last x added to B.
end function
```

Induction step. Under $C = 0$ and value v, the Bellman equation is given by

$$
\begin{aligned}
V^0(m,n;v) = \max_w F(w)(v - w + V^0(m-1,n-1;v)) \\
+ (1 - F(w))((1-q)V^0(m,n-1;v) + qV^0(m-1,n-1;v))
\end{aligned}
$$

Moreover, define

$$
\begin{aligned}
f_v(w) = F(w)(v - w + V^0(m-1,n-1;v)) + \\
(1 - F(w))((1-q)V^0(m,n-1;v) + qV^0(m-1,n-1;v))
\end{aligned}
$$

and similarly

$$f_{v-c}(w) = F(w)(v - c - w + V^0(m-1, n-1; v-c)) + \\ (1 - F(w))((1-q)V^0(m, n-1; v-c) + qV^0(m-1, n-1; v-c))$$

Then

$$V^0(m, n; v) = \max_w f_v(w),\ V^0(m, n; v-c) = \max_w f_{v-c}(w)$$

Then for any fixed w,

$$\begin{aligned}
&f_v(w) - f_{v-c}(w) \\
=&F(w)\big(c + V^0(m-1, n-1; v) - V^0(m-1, n-1; v-c)\big) \\
+&(1 - F(w))\big((1-q)(V^0(m, n-1; v) - V^0(m, n-1; v-c)) \\
+&q(V^0(m-1, n-1, v) - V^0(m-1, n-1; v-c)\big) \\
=&F(w)(c + \Delta(m-1, n-1)) + (1 - F(w))((1-q)\Delta(m, n-1) + q\Delta(m-1, n-1))
\end{aligned}$$

By the induction hypothesis,

$$0 \le \Delta(m-1, n-1) \le c \min\{m-1, n-1\},\ 0 \le \Delta(m, n-1) \le c \min\{m, n-1\}$$

Since $F(w) \in [0, 1]$, the expression above is nonnegative for every w, hence

$$f_v(w) \ge f_{v-c}(w), \forall w$$

Hence,

$$\Delta(m, n) = \max_w f_v(w) - \max_w f_{v-c}(w) \ge 0$$

For the upper bound, it must hold that

$$\max_w f_v(w) - \max_w f_{v-c}(w) \le \max_w(f_v(w) - f_{v-c}(w))$$

Therefore,

$$\Delta(m, n) \le \max_w\{F(w)(c + \Delta(m-1, n-1)) \\ + (1 - F(w))((1-q)\Delta(m, n-1) + q\Delta(m-1, n-1))\}$$

Applying the induction hypothesis, we have

$$\begin{aligned}
\Delta(m, n) \le& \max_w\{F(w)(c + c\min\{m-1, n-1\}) \\
&+ (1 - F(w))((1-q)c\min\{m, n-1\} + qc\min\{m-1, n-1\})\} \\
&\le c \cdot \max_w\{F(w)\min\{m, n\} + (1 - F(w))((1-q)\min\{m, n-1\} \\
&+ q\min\{m-1, n-1\})\}
\end{aligned}$$

Consider two subcases: (1) $m \geq n$. Then the right-hand side is equal to

$$c \cdot \max_w \{F(w)n + (1 - F(w))(n-1)\} \leq c \cdot n = c \min\{m, n\}$$

(2) $m \leq n - 1$. Then the right-hand side is equal tos

$$c \cdot \max_w \{F(w)m + (1 - F(w))(m-q)\} \leq c \cdot m = c \min\{m, n\}$$

This completes the induction. □ □

Now we prove the remaining statements. The idea is to show that $V^0(\pi^c; m, n) = V^0(m, n; v-c)$. Then we can directly apply Lemma 4 to prove Eq. (11).

Under $C = c$, the one-step Q function can be written as

$$\begin{aligned}
Q(w; m, n|C = c, v) &= F(w-c)(v - w + V^c(m-1, n-1)) \\
&+ (1 - F(w-c))((1-q)V^c(m, n-1) + qV^c(m-1, n-1)) \\
&= F(w-c)(v-w) + \big(F(w-c) + (1 - F(w-c))q\big)V^c(m-1, n-1) \\
&+ (1 - F(w-c))(1-q)V^c(m, n-1)
\end{aligned}$$

Define $x = w - c$, we can rewrite the above function as

$$Q(w; m, n|C = c, v) \tag{54}$$
$$= F(x)(v - x - c) + \big(F(x) + (1 - F(x))q\big)V^c(m-1, n-1) \tag{55}$$
$$+ (1 - F(x))(1-q)V^c(m, n-1)) \tag{56}$$

but this expression is exactly the Bellman equation for $C = 0$ with request value $(v - c)$. Hence, we have

$$V^c(m, n; v) = V^0(m, n; v - c), \ \forall m, n$$

and therefore, the policy $(\pi^c - c)$ is the optimal policy under $C = 0$ and request value $(v - c)$. Now consider $V^0(\pi^c; m, n)$, with request value v. For any policy π,

$$V^0(\pi; m, n; v) \geq V^c(\Pi; m, n, v)$$

This inequality holds because for any fixed policy π, the acceptance probabilities are higher $(F(w) \geq F(w-c))$ under $C = 0$, while immediate payoffs and continuation values are otherwise the same. Applying the inequality to π^c yields

$$V^0(\pi^c; m, n; v) \geq V^c(\pi^c; m, n; v) = V^0(\pi^c - c; m, n; v - c) = V^0(m, n; v - c)$$

Then by Lemma 4, we have

$$V^0(m, n; v) - V^0(\pi^c; m, n; v) \leq V^0(m, n; v) - V^0(m, n; v - c) \leq c \min\{m, n\}$$

This concludes the proof. □ □

Proof of Proposition 4

Proof. Proof. First, proving that $V^c(m,n) - V^c(\pi^0; m,n) \leq (v - w_{min} - c)\min\{m,n\}$ is trivial: the highest possible margin from a request is $v - w_{min} - c$, because v is the price and $(w_{min} + c)$ is the lowest possible acceptable pay when $C = c$; $\min\{m,n\}$ is the highest possible number of accepted requests. Hence, $V^c(m,n) \leq (v - w_{min} - c)\min\{m,n\}$, which implies the inequality.

To prove $V^c(m,n) - V^c(\pi^0; m,n) \leq c\min\{m,n\}$, we proceed in three steps.

Step 1. Write

$$V^c(m,n;v) - V^c(\pi^0;m,n;v) = \underbrace{\left(V^c(m,n;v) - V^0(m,n;v)\right)}_{=:T_1} + \underbrace{\left(V^0(m,n;v) - V^c(\pi^0;m,n;v)\right)}_{=:T_2}.$$

We claim that $T_1 \leq 0$. Indeed, by monotonicity of the value function in the per-match reward parameter v,

$$V^0(m,n;v') \leq V^0(m,n;v) \quad \text{whenever } v' \leq v. \tag{57}$$

Moreover, as established in the model (and used throughout), the shift identity holds:

$$V^c(m,n;v) = V^0(m,n;v-c). \tag{58}$$

Therefore,

$$T_1 = V^0(m,n;v-c) - V^0(m,n;v) \leq 0,$$

by (57). Consequently,

$$V^c(m,n;v) - V^c(\pi^0;m,n;v) \leq T_2. \tag{59}$$

Step 2. Define a transformed policy $\tilde{\pi}^0$ for the baseline environment $C = 0$ as follows: for every history/state at which π^0 would offer wage w, policy $\tilde{\pi}^0$ offers wage $w - c$. Consider any fixed sample path of acceptance/rejection outcomes generated by π^0 under $C = c$. Under policy $\tilde{\pi}^0$ in the baseline environment $C = 0$, the acceptance probability at each offer equals

$$F\big((w-c) - 0\big) = F(w-c),$$

which matches the acceptance probability under $C = c$ when offering w. Moreover, whenever an acceptance occurs, the instantaneous profit under $(C = c, v)$ with wage w equals

$$(v - w),$$

while the instantaneous profit under $(C = 0, v - c)$ with wage $(w - c)$ equals

$$(v - c) - (w - c) = v - w.$$

Hence,

$$V^c(\pi^0; m, n; v) = V^0(\tilde{\pi}^0; m, n; v - c). \tag{60}$$

Therefore,

$$T_2 = V^0(m, n; v) - V^0(\tilde{\pi}^0; m, n; v - c).$$

Insert and subtract $V^0(\tilde{\pi}^0; m, n; v)$:

$$T_2 = \Big(V^0(m, n; v) - V^0(\tilde{\pi}^0; m, n; v)\Big) + \Big(V^0(\tilde{\pi}^0; m, n; v) - V^0(\tilde{\pi}^0; m, n; v - c)\Big). \tag{61}$$

The first term in (61) is nonpositive because $V^0(m, n; v)$ is the optimal value at $(C = 0, v)$:

$$V^0(m, n; v) = \sup_{Pi} V^0(\pi; m, n; v) \geq V^0(\tilde{\pi}^0; m, n; v).$$

Hence,

$$T_2 \leq V^0(\tilde{\pi}^0; m, n; v) - V^0(\tilde{\pi}^0; m, n; v - c). \tag{62}$$

It remains to bound the right-hand side for a fixed policy.

Step 3. We prove the following lemma.

Lemma 5. *Fix any admissible policy π under $C = 0$. Then for all $m, n \in \mathbb{Z}_+$ and all $c \geq 0$,*

$$V^0(\pi; m, n; v) - V^0(\pi; m, n; v - c) \leq c \min\{m, n\}. \tag{63}$$

Proof of Lemma 5 For fixed π, define

$$D_\pi(m, n) := V^0(\pi; m, n; v) - V^0(\pi; m, n; v - c).$$

We prove by induction on $m + n$ that $D_\pi(m, n) \leq c \min\{m, n\}$.

Base cases. If $m = 0$ or $n = 0$, no matches are possible, so $V^0(\pi; m, n; v) = V^0(\pi; m, n; v - c) = 0$ and hence $D_\pi(m, n) = 0 \leq c \min\{m, n\}$.

Induction step. Fix (m, n) with $m, n \geq 1$, and suppose $D_\pi(m', n') \leq c \min\{m', n'\}$ holds for all (m', n') with $m' + n' < m + n$. Consider the first offer made by policy π at state (m, n), which prescribes some wage w for the current driver. Let $p := F(w)$ denote the acceptance probability (since $C = 0$). The fixed-policy DP recursion gives:

$$V^0(\pi; m, n; v) = p\Big((v - w) + V^0(\pi; m - 1, n - 1; v)\Big)$$
$$+ (1 - p)V^0(\pi; m, n - 1; v),$$
$$V^0(\pi; m, n; v - c) = p\Big((v - c - w) + V^0(\pi; m - 1, n - 1; v - c)\Big)$$
$$+ (1 - p)V^0(\pi; m, n - 1; v - c).$$

Subtracting yields

$$D_\pi(m,n) = p\,c + p\,D_\pi(m-1,n-1) + (1-p)\,D_\pi(m,n-1). \tag{64}$$

By the induction hypothesis,

$$D_\pi(m-1,n-1) \le c\min\{m-1,n-1\} = c(\min\{m,n\}-1),$$

and

$$D_\pi(m,n-1) \le c\min\{m,n-1\} \le c\min\{m,n\}.$$

Let $h := \min\{m,n\}$. Plugging these bounds into (64) gives

$$\begin{aligned} D_\pi(m,n) &\le p\,c + p\,c(h-1) + (1-p)\,ch && (65)\\ &= c\Big(p + p(h-1) + (1-p)h\Big) && (66)\\ &= c\,h && (67)\\ &= c\ \min\{m,n\}. && (68) \end{aligned}$$

This completes the induction and proves (63). □

Applying the lemma (63) to $Pi = \tilde{\pi}^0$ and combining with (62) yields

$$T_2 \le c\ \min\{m,n\}.$$

Finally, by (59),

$$V^c(m,n;v) - V^c(\pi^0;m,n;v) \le c\ \min\{m,n\},$$

as claimed.□ □

Proof of Proposition 5

Proof. Proof. Fix (m,n) and prior belief μ_0. Consider the two benchmark policies π^0 and π^c.

If the platform commits to π^0, then its regret relative to the clairvoyant benchmark is

$$\begin{aligned} V^{\text{clair}}(m,n) - V^{\pi^0}(m,n \mid \mu_0) = (1-\mu_0)\big(V^0(m,n) - V^0(\pi^0;m,n)\big)\\ + \mu_0\big(V^c(m,n) - V^c(\pi^0;m,n)\big) \end{aligned}$$

The first term is 0 since π^0 is optimal under $C=0$, and by Proposition 4 the second term is at most $\mu_0\,\beta\ \min\{m,n\}$. Hence

$$\text{Regret}(\pi^0) \le \mu_0\,\beta\ \min\{m,n\}.$$

If instead the platform commits to π^c, then similarly

$$V^{\text{clair}}(m,n) - V^{\pi^c}(m,n \mid \mu_0) = (1-\mu_0)\big(V^0(m,n) - V^0(\pi^c; m,n)\big) \\ + \mu_0\big(V^c(m,n) - V^c(\pi^c; m,n)\big)$$

The second term is 0 since π^c is optimal under $C = c$, and by Proposition 3 the first term is at most $(1-\mu_0)\, c\, \min\{m,n\}$. Hence

$$\text{Regret}(\pi^c) \le (1-\mu_0)\, c\, \min\{m,n\}.$$

By Definition 1, π^{DC} selects π^0 when $\mu_0 \le \frac{c}{c+\beta}$ and π^c otherwise, i.e., it picks the option with the smaller of the two upper bounds above. Therefore

$$\text{Regret}(\pi^{\text{DC}}) \le \min\{\mu_0\beta, (1-\mu_0)c\} \cdot \min\{m,n\} = L(\mu_0; c) \cdot \min\{m,n\},$$

which proves the result. □ □

Proof of Proposition 6

We first introduce with an equivalent and alternative statement of Proposition 6 that has more mathematical details. We then directly prove this statement. Proposition 6 is then implied.

Full statement of Proposition 6:

Proposition 10. *Assume the feasible offer set satisfies $\mathcal{A} \subseteq [w_{\min}, v]$, and define $\bar{v} \triangleq v - w_{\min}$. Fix a probe offer $p \in \mathcal{A}$. Let $Y \in \{A, R\}$ denote the probe outcome, and define the posterior belief after the probe*

$$\mu_1(Y) = \begin{cases} \mu'_A(p), & \text{if } Y = A, \\ \mu'_R(p), & \text{if } Y = R. \end{cases}$$

Let (m_Y, n_Y) denote the remaining numbers of requests and workers after the probe:

$$(m_Y, n_Y) = \begin{cases} (m-1, n-1), & \text{if } Y = A, \\ (m, n-1), & \text{with prob. } 1-q \text{ if } Y = R, \\ (m-1, n-1), & \text{with prob. } q \text{ if } Y = R. \end{cases}$$

Let

$$V^{\text{clair}}(m,n) = \mu_0 V^c(m,n) + (1-\mu_0) V^0(m,n)$$

be the clairvoyant benchmark. Let $V_p^{PC}(m,n)$ be the expected value of Probe-and-Commit with probe p, i.e., it offers p to the first worker, updates its belief to $\mu_1(Y)$, and then *commits to the Direct-Commit policy (choosing between π^0 and π^c) based on $\mu_1(Y)$ for the remaining subproblem (m_Y, n_Y).*

Then the loss of Probe-and-Commit relative to clairvoyant satisfies

$$
\begin{aligned}
& V^{\text{clair}}(m,n) - V_p^{PC}(m,n) && (69)\\
& \le \bar{v} - P_{accept}(p,\mu_0)(v-p) + \mathbb{E}\left[\min\{\mu_1(Y)\beta,\, (1-\mu_1(Y))c\} \cdot \min\{m_Y, n_Y\}\right] && (70)\\
& \le \bar{v} - P_{accept}(p,\mu_0)(v-p) + \min\{m,n\} \cdot \mathbb{E}\left[\min\{\mu_1(Y)\beta,\, (1-\mu_1(Y))c\}\right]. && (71)
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
\mathbb{E}\left[\min\{\mu_1(Y)\beta,\, (1-\mu_1(Y))c\}\right] = P_{\text{acc}}(p,\mu_0)\, & \min\{\mu_A'(p)\beta,\, (1-\mu_A'(p))c\}\\
& + \big(1 - P_{\text{acc}}(p,\mu_0)\big)\, \min\{\mu_R'(p)\beta,\, (1-\mu_R'(p))c\},
\end{aligned}
$$

where $P_{\text{acc}}(p,\mu_0) = F(p-c)\mu_0 + F(p)(1-\mu_0)$.

Proof. Proof. Fix $p \in \mathcal{A}$. For each $C \in \{0, c\}$, define $V_p^{C,\text{probe}}(m,n)$ as the clairvoyant value when the first offer is forced to be p and thereafter the platform acts optimally knowing C. Define the corresponding ex ante mixture

$$
V_p^{\text{probe}}(m,n) \triangleq \mu_0 V_p^{c,\text{probe}}(m,n) + (1-\mu_0) V_p^{0,\text{probe}}(m,n).
$$

We decompose the Probe-and-Commit loss as

$$
V^{\text{clair}}(m,n) - V_p^{PC}(m,n) = \Big(V^{\text{clair}}(m,n) - V_p^{\text{probe}}(m,n)\Big) + \Big(V_p^{\text{probe}}(m,n) - V_p^{PC}(m,n)\Big). \tag{72}
$$

Step 1 (probe cost). Fix $C \in \{0, c\}$. Since at most one additional request can be fulfilled when moving from $(m-1, n-1)$ to (m,n) and each fulfilled request yields profit at most $v - w_{\min} = \bar{v}$, we have

$$
V^C(m,n) \le V^C(m-1, n-1) + \bar{v}. \tag{73}
$$

On the other hand, under the forced-probe clairvoyant policy, regardless of the probe outcome, the continuation value is at least $V^C(m-1,n-1)$: if the probe is accepted then the next state is $(m-1, n-1)$; if the probe is rejected and not canceled then the next state is $(m, n-1)$ which weakly dominates $(m-1, n-1)$; if it is rejected and canceled then the next state is $(m-1, n-1)$. Moreover, $p \le v$ implies $v - p \ge 0$. Hence,

$$
V_p^{C,\text{probe}}(m,n) \ge V^C(m-1, n-1). \tag{74}
$$

Combining (73) and (74) yields

$$
0 \le V^C(m,n) - V_p^{C,\text{probe}}(m,n) \le \bar{v}.
$$

Averaging over C implies

$$0 \leq V^{\text{clair}}(m,n) - V_p^{\text{probe}}(m,n) \leq \bar{v}. \tag{75}$$

Step 2 (commit cost). By construction, V_p^{probe} and Probe-and-Commit take the same first action p and thus, for any realized probe outcome Y, they transition to the same remaining state (m_Y, n_Y). Conditional on Y, the forced-probe clairvoyant policy continues with the full-information optimal policy π^c, while Probe-and-Commit commits to π^0 or π^c based on the posterior $\mu_1(Y)$. Applying the Direct-Commit bound to the remaining subproblem gives, conditional on Y,

$$V_p^{\text{probe}}(m,n) - V_p^{PC}(m,n) \leq \min\{\mu_1(Y)\beta, (1-\mu_1(Y))c\} \cdot \min\{m_Y, n_Y\}.$$

Taking expectations over the probe outcome (and the cancellation realization when $Y = R$) yields

$$V_p^{\text{probe}}(m,n) - V_p^{PC}(m,n) \leq \mathbb{E}\left[\min\{\mu_1(Y)\beta, (1-\mu_1(Y))c\} \cdot \min\{m_Y, n_Y\}\right]. \tag{76}$$

Step 3 (combine). Combining (72), (75), and (76) proves (70). Finally, since $m_Y \leq m$, $n_Y = n-1$, we have $\min\{m_Y, n_Y\} \leq \min\{m, n-1\}$, we obtain (71). The explicit expansion of the expectation in terms of $P_{\text{acc}}(p, \mu_0)$ follows directly from the definition of $\mu_1(Y)$ and the law of total probability. $\square$

Proof of Proposition 7

We first introduce with an equivalent and alternative statement of Proposition 7 that has more mathematical details. We then directly prove this statement. Proposition 7 is then implied.

Proposition 11. *Let $\beta = \min\{c, (v - w_{\min} - c)^+\}$, and define*

$$a \triangleq \beta\mu_0, \qquad b \triangleq c(1-\mu_0).$$

Fix a probe offer $p \in \mathcal{A}$. Let $Y \in \{A, R\}$ denote the probe outcome, and let $\mu_1 = \mu'_A(p)$ if $Y = A$ and $\mu_1 = \mu'_R(p)$ if $Y = R$. Define the continuation objective

$$\mathcal{R}(p) \triangleq \mathbb{E}[\min\{\mu_1\beta, (1-\mu_1)c\}].$$

Then $\mathcal{R}(p)$ admits the explicit form

$$\mathcal{R}(p) = \min\{aF(p-c),\, bF(p)\} + \min\{a(1-F(p-c)),\, b(1-F(p))\}. \tag{77}$$

Define

$$D(p) \triangleq aF(p-c) - bF(p).$$

Assume $a > b$. Then $\mathcal{R}(p)$ admits the representation

$$\mathcal{R}(p) = b + (D(p))^+ \tag{78}$$

Consequently, an optimal probe for minimizing $\mathcal{R}(p)$ can be selected as follows:

$$p^\star \in \arg\min_{p \in \mathcal{A}} D(p)$$

Moreover, in the discrete K-type model (with support $\{w_1, \dots, w_K\}$), it suffices to search over the finite candidate set $\mathcal{A} \cap \{w_j,\ w_j + c :\ j = 1, \dots, K\}$.

Proof. Proof. *Step 1: Simplifying $\mathcal{R}(p)$.* By the law of total probability,

$$\mathcal{R}(p) = P(A)\min\{\mu'_A(p)\beta, (1 - \mu'_A(p))c\} + P(R)\min\{\mu'_R(p)\beta, (1 - \mu'_R(p))c\}.$$

Using Bayes rule, $\mu'_A(p) = P(C = c \mid A) = P(C = c, A)/P(A)$, hence $P(A)\mu'_A(p) = P(C = c, A)$. Similarly, $P(A)(1 - \mu'_A(p)) = P(C = 0, A)$. Under the model,

$$P(A \mid C = c) = F(p - c), \qquad P(A \mid C = 0) = F(p),$$

so

$$P(C = c, A) = \mu_0 F(p - c), \qquad P(C = 0, A) = (1 - \mu_0)F(p).$$

Therefore,

$$\begin{aligned} P(A)\min\{\mu'_A(p)\beta, (1 - \mu'_A(p))c\} &= \min\{\beta\,\mu_0 F(p - c),\ c(1 - \mu_0)F(p)\} \\ &= \min\{aF(p - c),\, bF(p)\}. \end{aligned}$$

Analogously, since $P(R \mid C = c) = 1 - F(p - c)$ and $P(R \mid C = 0) = 1 - F(p)$, we obtain

$$\begin{aligned} P(R)\min\{\mu'_R(p)\beta, (1 - \mu'_R(p))c\} &= \min\{\beta\,\mu_0(1 - F(p - c)),\ c(1 - \mu_0)(1 - F(p))\} \\ &= \min\{a(1 - F(p - c)),\, b(1 - F(p))\}. \end{aligned}$$

Combining the last two displays yields (77).

Step 2: Piecewise representation when $a > b$. Let $x = F(p - c)$ and $y = F(p)$, so that $0 \le x \le y \le 1$, and define $A = ax$ and $B = by$. Then $D(p) = A - B$ and (77) becomes

$$\mathcal{R}(p) = \min\{A, B\} + \min\{a - A,\ b - B\}. \tag{79}$$

Assume $a > b$ (equivalently $a - b > 0$). We consider three cases.

Case 1: $D(p) \le 0$ (i.e., $A \le B$). Then $\min\{A, B\} = A$. For the second minimum, note that

$$(a - A) - (b - B) = (a - b) - (A - B) = (a - b) - D(p) > 0,$$

so $\min\{a - A, b - B\} = b - B$. Substituting into (79) gives

$$\mathcal{R}(p) = A + (b - B) = b + (A - B) = b + D(p).$$

Case 2: $0 \le D(p) \le a - b$ **(i.e.,** $A \ge B$ **but** $A - B \le a - b$**).** Then $\min\{A, B\} = B$. Moreover,

$$(a - A) - (b - B) = (a - b) - D(p) \ge 0,$$

so $\min\{a - A, b - B\} = b - B$, and hence

$$\mathcal{R}(p) = B + (b - B) = b.$$

Case 3: $D(p) \ge a - b$**.** Then again $\min\{A, B\} = B$, but now

$$(a - A) - (b - B) = (a - b) - D(p) \le 0,$$

so $\min\{a - A, b - B\} = a - A$. Therefore,

$$\mathcal{R}(p) = B + (a - A) = a - (A - B) = a - D(p).$$

Combining the three cases yields (78). The stated selection rule follows immediately from the monotonicity of the right-hand side of (78) in $D(p)$ over each region: it is increasing in D for $D \le 0$, constant for $D \in [0, a - b]$, and decreasing in D for $D \ge a - b$.

Step 3: Finite candidate set under discrete types. Under the discrete support $\{w_1, \ldots, w_K\}$, the pair $(F(p - c), F(p))$ (hence $D(p)$ and $\mathcal{R}(p)$) changes only when p crosses a point in $\{w_j\}$ or $\{w_j + c\}$, implying that $\mathcal{R}(p)$ is piecewise constant in p and it suffices to search the candidate set $\mathcal{A} \cap \{w_j, w_j + c\}_{j=1}^{K}$. □

Proof of Proposition 9

Proof. Proof. Fix m, n and let $S(m, n)$ designate the value obtained from the optimal static policy (i.e., offering the same pay to all drivers for all requests). Additionally, let $PI(m, n)$ designate the value obtained in the perfect-information case, where the platform knows the value of C, i.e.,

$$\begin{aligned} PI(m, n) &= \mu_0 V^*_{C=c}(m, n) + (1 - \mu_0) V^*_{C=0}(m, n) \\ &\ge \mu_0 S_{C=c}(m, n) + (1 - \mu_0) S_{C=0}(m, n). \end{aligned}$$

By Lemma 1, we know there exists n_0 large enough such that for $m \ge 1$ and $n \ge n_0$, we can write

$$\frac{PI(m, n)}{m} = \mu_0 \frac{F(w_s^{C=c}(\infty) - c)(v - w_s^{C=c}(\infty))}{1 - (1 - q)(1 - F(w_s^{C=c}(\infty) - c))} + (1 - \mu_0) \frac{F(w_s^{C=0}(\infty))(v - w_s^{C=0}(\infty))}{1 - (1 - q)(1 - F(w_s^{C=0}(\infty)))}.$$

To simplify notation, we write $w_0 = w_s^{C=0}(\infty)$ and $w_c = w_s^{C=c}(\infty)$. Now consider the two special static policies where we offer only w_0, yielding value $\hat{S}_0(m,n) \leq S(m,n)$, and where we offer only w_c, yielding value $\hat{S}_c(m,n) \leq S(m,n)$. Then we can write

$$\frac{PI(m,n)}{m} - \frac{\hat{S}_0(m,n)}{m} = \mu_0 \left(\frac{F(w_c - c)(v - w_c)}{1 - (1-q)(1 - F(w_c - c))} - \frac{F(w_0 - c)(v - w_0)}{1 - (1-q)(1 - F(w_0 - c))} \right) \geq 0,$$

and similarly

$$\frac{PI(m,n)}{m} - \frac{\hat{S}_c(m,n)}{m} = (1-\mu_0) \left(\frac{F(w_0)(v - w_0)}{1 - (1-q)(1 - F(w_0))} - \frac{F(w_c)(v - w_c)}{1 - (1-q)(1 - F(w_c))} \right) \geq 0.$$

Putting the two equations together, we obtain

$$\begin{aligned} \frac{PI(m,n)}{m} - \frac{S(m,n)}{m} \geq \min \Bigg(&\mu_0 \left(\frac{F(w_c - c)(v - w_c)}{1 - (1-q)(1 - F(w_c - c))} - \frac{F(w_0 - c)(v - w_0)}{1 - (1-q)(1 - F(w_0 - c))} \right), \\ &(1-\mu_0) \left(\frac{F(w_0)(v - w_0)}{1 - (1-q)(1 - F(w_0))} - \frac{F(w_c)(v - w_c)}{1 - (1-q)(1 - F(w_c))} \right) \Bigg). \end{aligned}$$

We can simplify the above by observing that, since $w_K \leq c$, we must have $F(w_0 - c) = 0$ and $F(w_c) = 1$. Thus

$$\begin{aligned} \frac{PI(m,n)}{m} - \frac{S(m,n)}{m} \geq \min \Bigg(&\mu_0 \left(\frac{F(w_c - c)(v - w_c)}{1 - (1-q)(1 - F(w_c - c))} \right), \\ &(1-\mu_0) \left(\frac{F(w_0)(v - w_0)}{1 - (1-q)(1 - F(w_0))} - (v - w_c) \right) \Bigg). \end{aligned} \tag{80}$$

Separately, we seek to upper-bound the gap between the optimal policy value and the perfect-information value, via the probe-and-commit heuristic — recall that $V_{PC}(m,n) \leq V^*(m,n)$, which means $PI(m,n) - V^*(m,n) \leq PI(m,n) - V_{PC}(m,n)$.

We can easily compute the value of the probe-and-commit strategy as

$$V_{PC}(m,n) = (1-\mu_0) \left[(v - w_K) + V^*_{C=0}(m-1, n-1) \right] + \mu_0 \left[q V^*_{C=c}(m-1, n-1) + (1-q) V^*_{C=c}(m, n-1) \right],$$

which for n_0 large enough and $n \geq n_0$, we know from Proposition 8 we can re-write as

$$\begin{aligned} V_{PC}(m,n) = (1-\mu_0) &\left[(v - w_K) + (m-1) \left(\frac{F(w_0)(v - w_0)}{1 - (1-q)(1 - F(w_0))} + \varepsilon \right) \right] + \\ &\mu_0 (m - q) \left(\frac{F(w_c - c)(v - w_c)}{1 - (1-q)(1 - F(w_c - c))} + \varepsilon \right). \end{aligned}$$

Therefore, we obtain

$$\begin{aligned}
PI(m,n) - V^*(m,n) \leq \quad & (1-\mu_0)\left[\frac{F(w_0)(v-w_0)}{1-(1-q)(1-F(w_0))} - (m-1)\varepsilon - (v-w_K)\right] \\
& +\mu_0\left[q\frac{F(w_c-c)(v-w_c)}{1-(1-q)(1-F(w_c-c))} - (m-q)\varepsilon\right] \\
\leq \quad & (1-\mu_0)\left[\frac{F(w_0)(v-w_0)}{1-(1-q)(1-F(w_0))} - (v-w_K)\right] \\
& +\mu_0\left[q\frac{F(w_c-c)(v-w_c)}{1-(1-q)(1-F(w_c-c))}\right]. \quad (81)
\end{aligned}$$

Putting (80) and (81) together, we obtain

$$\begin{aligned}
V^*(m,n) - S(m,n) \geq m \cdot \min\Bigg(& \mu_0\left[\frac{F(w_c-c)\,(v-w_c)}{1-(1-q)\big(1-F(w_c-c)\big)}\right], \\
& (1-\mu_0)\left[\frac{F(w_0)\,(v-w_0)}{1-(1-q)\big(1-F(w_0)\big)} - v + w_c\right]\Bigg) \\
- (1-\mu_0)\left[\frac{F(w_0)\,(v-w_0)}{1-(1-q)\big(1-F(w_0)\big)} - (v-w_K)\right] & - \mu_0\, q\,\frac{F(w_c-c)\,(v-w_c)}{1-(1-q)\big(1-F(w_c-c)\big)}.
\end{aligned} \tag{82}$$

Because $w_0 \neq w_c$ (recall $c > w_K$) , for any μ_0, there exists m large enough to make the right-hand side positive. □

Proof of Corollary 4

Proof. Proof. Let $q = 0$. Then assuming $F(w_0 = 0) > 0$, then $w_0 = 0$, $w_c = c$, and Equation (82) simplifies to

$$\begin{aligned}
V^*(m,n) - S(m,n) &\geq m \cdot \min\left(\mu_0(v-c), (1-\mu_0)(v-v+c)\right) - (1-\mu_0)(v-v+w_K) \\
&\geq m \cdot \min\left(\mu_0(v-c), (1-\mu_0)c\right) - (1-\mu_0)c \\
&\geq m\frac{\mu_0(v-c)(1-\mu_0)c}{\mu_0(v-c)+(1-\mu_0)c} - (1-\mu_0)c \\
&= (1-\mu_0)c\left(\frac{m\mu_0(v-c)}{\mu_0(v-c)+(1-\mu_0)c} - 1\right),
\end{aligned}$$

where the last inequality holds as long as $0 \leq c \leq v$ and $0 < \mu_0 < 1$. In order to ensure the right-hand side is positive, we simply need $\mu_0 < 1$ and

$$\begin{aligned}
\frac{m\mu_0(v-c)}{\mu_0(v-c)+(1-\mu_0)c} &> 1 \\
m\mu_0(v-c)-\mu_0(v-c)+\mu_0 c &> c \\
\mu_0 &> \frac{c}{c+(m-1)(v-c)}.
\end{aligned}$$

□