

Learning Pay Strategies with Small Samples in Gig Economy Platforms

Arthur Delarue, Zhen Lian, and Tony Qin*

Abstract

Gig economy platforms operate in dynamic labor markets in which workers can choose among multiple job offers in real time. A central challenge 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 general setting in which the global factor 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 improve platform profits by up to 26% in dynamic simulations and provide actionable guidance for gig economy platforms seeking transparent, adaptive worker compensation in competitive spot labor markets.

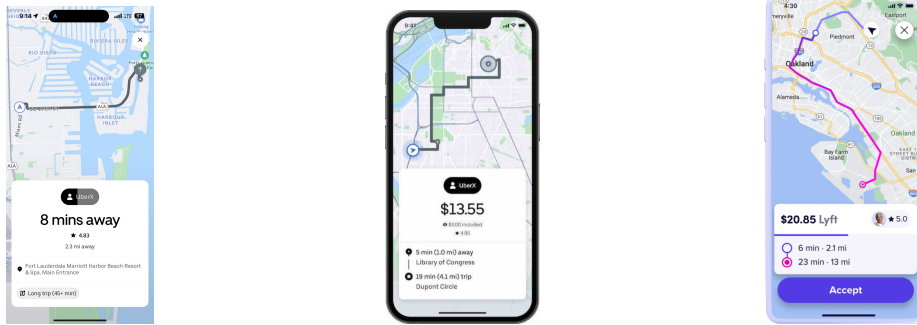
Key words: Gig economy, dynamic pricing, small-sample learning, dynamic programming applications, competition, ride-hailing

*Arthur Delarue: Darden School of Business, University of Virginia. 100 Darden Blvd, Charlottesville, VA 22903. Zhen Lian: Yale School of Management. 165 Whitney Avenue, New Haven, CT 06511. Tony Qin: Santa Clara University. 500 El Camino Real, Santa Clara, CA 95053.

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 Fig. 1a). In 2022, however, both platforms moved to a system where each driver is offered a specific pay for each ride (see Fig. 1b). This change increases transparency for drivers and makes it easier for them to compare offers from competitors (see Fig. 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 exist-

ing 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 workers. 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, 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. The platform knows the distribution of reservation wages but not their realizations; by observing real-time decisions, it can infer the global factor from 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 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 worker heterogeneity by making an initial offer below the optimal myopic pay offer. The optimal pay then weakly increases after a worker rejection and weakly decreases after an acceptance. More surprisingly, we 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 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. For tractability, we assume the global factor follows a two-point distribution in our analysis, but relax this assumption in numerical simulations. We provide an exact dynamic programming algorithm to solve the resulting belief-augmented Markov Decision Process (MDP). Because this algorithm is not very interpretable and suffers from the curse of dimensionality, we also show that strategies that extend the full-information optimal policy remain highly effective. One natural approach is Thompson sampling, in which we sample the global factor from the prior distribution and apply the corresponding full-information optimal policy.

Another approach is to restrict learning to zero or one targeted step, yielding transparent and tractable decision rules that achieve strong performance despite the underlying complexity of the optimal policy. We develop two such heuristics: the first, called “direct-

commit,” irrevocably chooses either the low-global-factor or the high-global-factor optimal pay policy based only on the 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.

We compare the performance of all these pay policies both analytically and numerically, and verify the robustness of our insights in a simulation setting which more closely resembles the dynamic environment of a gig economy platform. Each policy is useful in different settings, depending both on how difficult and on how valuable it is to learn the global factor. We conclude by identifying the value obtained from dynamic pay policies (offering different pay to different workers) over static pay policies. Interestingly, we observe that while dynamic pay is always better for the platform, it does not necessarily make workers worse off.

After reviewing the literature in Section 2, we present our model in Section 3 and solve the full-information setting in Section 4. We describe pay strategies in the general (learning) setting in Section 5 and compare their performance analytically in Section 6 and via simulation in Section 7. Finally, we analyze the value of dynamic over static pay in Section 8.

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. Originally introduced in legacy industries such as airlines (Gallego and Van Ryzin 1994, Bitran and Caldentey 2003), dynamic pricing has been a major focus of researchers and practitioners since the inception of ride-hailing platforms. In a sense, the full-information part of our problem can be viewed as the supply-side counterpart of the classical demand-side pricing problem of Gallego and Van Ryzin (1994): instead of dynamically pricing demand to clear a fixed inventory of perishable goods, we dynamically price supply to fulfill a fixed set of perishable requests. Consistent with this demand-side focus, most of the literature has concentrated 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.

Fewer papers consider the supply side of the pricing problem, which has recently grown more distinct from the demand side as ride-hailing platforms have increasingly separated driver pay from rider fares at the ride level (Chaum 2023). Cachon et al. (2017) study optimal wage contracts for gig economy platforms, finding that commission-based pay with surge pricing are near-optimal. Taylor (2018) establishes the effect of agent independence, delay sensitivity, and uncertainty on both prices and wages in a generic platform. Both approaches assume the platform has full knowledge of supply conditions. In the ride-hailing setting, Guda and Subramanian (2019) examine how communicating surge forecasts affects driver incentives, exploring how drivers can learn from the platform. In contrast, we study how the platform can set wages in order to learn about market conditions (such as external competition) from worker responses.

Our work also seeks to leverage worker heterogeneity, as ride-hailing drivers may have different reservation wages for the same request. Allon et al. (2023) find 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”).

Pricing and learning. In addition to exploiting worker heterogeneity, in our model the platform seeks to leverage sequential offers to learn global information about worker 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). Modern strategies in dynamic pricing (Besbes and Zeevi 2009) and matching (Johari et al. 2021) often rely on a learning component. Qin et al. (2025) provide a review of reinforcement learning approaches to ride-hailing problems, including pricing and incentives. 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 horizon, since ride-hailing and food delivery platforms often operate in nonstationary settings where conditions can change quickly. In a comprehensive review of learning approaches in dynamic pricing, Den Boer (2015) identifies both competition and time-varying conditions as new research directions, which have been explored in recent years by Besbes et al. (2014) and Keskin and Zeevi (2017). However, this work still focuses on asymptotic regret of fairly general policies over long horizons. In contrast, we consider the short-term performance of strategies tailored to a specific platform environment — an approach also employed by recent work in platform experimentation (Bright et al. 2025).

Pricing under competition. Because the global factor captures the effect of competitor behavior, our work also relates to the literature on dynamic pricing under competition. Generic approaches typically involve deriving the equilibrium induced by a particular model of consumer behavior, such as nested logit (Gallego and Wang 2014) or consider-then-choose lexicographic choice (Banerjee et al. 2024). In platforms, Bernstein et al. (2021) study the impact of multi-homing (or multi-app) drivers on price equilibria, while Cohen and Zhang (2022) evaluate the value of semi-cooperative strategies between competing platforms. Tripathy et al. (2023) focus on platform responses to strategic behavior by ride-hailing drivers. 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.

Sequential take-it-or-leave-it offers. The core decisions in ride-hailing platforms are matching and pricing. 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. A platform trying to find a driver for a ride may thus offer it sequentially to multiple drivers at different prices. To simplify our analysis, we adopt a similar approach as Yan et al. (2025) (who tackle the related problem of *shared* or *pooled* rides) and consider identical requests so as to effectively separate the matching problem from the pricing problem.

The 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.

Building on this economics literature, the question of *optimizing* sequential posted price mechanisms is of increasing importance in operations, as such approaches are increasingly used in ride-hailing, food delivery, and freight marketplaces (Chen et al. 2021). Cohen et al. (2023) study the design of airfare price ladders (which restrict the markups airlines impose as the flight date approaches, usually for tractability reasons), which complements our analysis of dominated pay levels in the full-information setting. 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.

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 pay p to serve a request, 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.

An important design choice is that offers are take-it-or-leave-it in the sense that workers only receive at most one offer over the time horizon; hence, workers' dominant strategy is to accept any offer exceeding their true \tilde{W}_i . This choice is less restrictive than it may appear: in practice, the platform can allow workers to re-enter the mechanism after a brief exclusion period; as long as the opportunity cost of waiting during the exclusion period exceeds the possible gain from rejection (which is no greater than the spread of possible pay), workers have no incentive to strategically reject profitable offers. The rolling-horizon simulation in Section 7 illustrates such re-entry dynamics. We discuss this assumption further in Section 8.

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, \dots, w_K , such that

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

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

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

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.,

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

We say the global factor is “on” when $C = c$ and “off” when $C = 0$. We relax this assumption in numerical experiments in Section 7. 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

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

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

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

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

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

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

$$V(m, n, \mu) = \max_{p \geq 0} P_{\text{accept}}(p, \mu) \left(v - p + V(m - 1, n - 1, \mu'_A(p)) \right) + (1 - P_{\text{accept}}(p, \mu)) \left[(1 - q)V(m, n - 1, \mu'_R(p)) + qV(m - 1, n - 1, \mu'_R(p)) \right]. \quad (7)$$

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

$$V(m, n) = \max_{p \geq 0} F(p) \left(v - p + V(m - 1, n - 1) \right) + (1 - F(p)) \left[(1 - q)V(m, n - 1) + qV(m - 1, n - 1) \right]. \quad (8)$$

Due to the discrete 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 $\{(w_k, f_k)\}_{k=1, \dots, K}$, some wages w_k can never appear in any optimal pay policy.

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 Fig. 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 $\bar{\mathcal{W}}$.

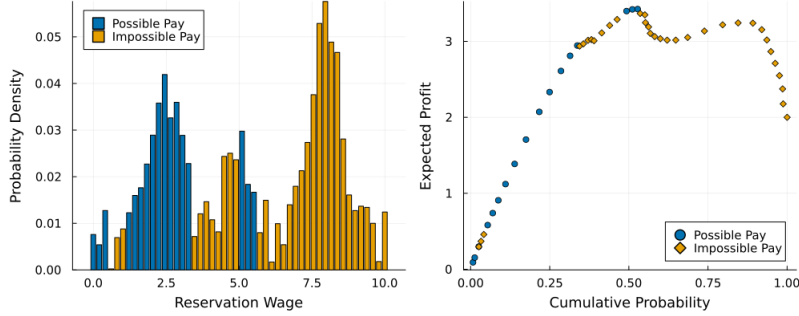


Figure 2: Illustration of possible and impossible pay levels according to Theorem 1.

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 workers 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 workers 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 monotonicity properties of the optimal pay sequence which we analyze more carefully in the next subsection. Because Algorithm 1 assumes $C = 0$, we denote the optimal pay policy it produces by π^0 . The optimal pay sequence under $C = c$ (denoted by π^c) can be obtained analogously by replacing v with $v - c$.

We also illustrate one step of Algorithm 1 in Fig. 3 on a toy example with two worker types and a single request. While the optimal pay for a single worker is easily obtained

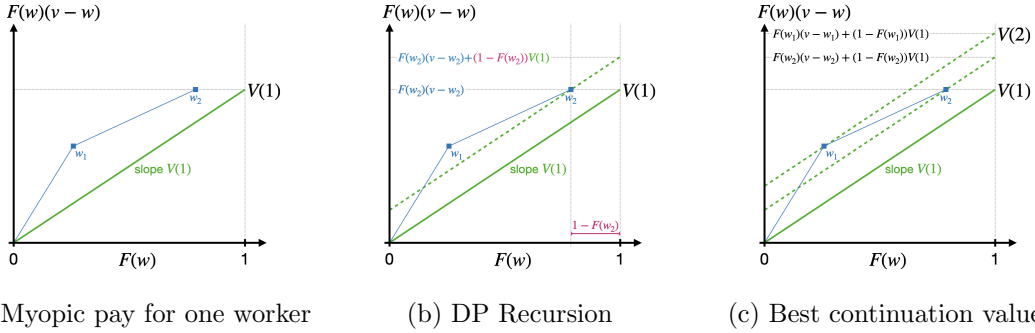


Figure 3: Illustration of the dynamic programming algorithm for optimal pay.

Notes: There are two worker 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).

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 equals the optimal one-shot pay w_{k^*} ¹, where*

$$k^* \in \arg \max_{k \in \{1, \dots, K\}} F(w_k)(v - w_k) \quad (9)$$

2. *When $m < n$, the optimal first offer is weakly increasing in the number of requests m and weakly decreasing in the number of workers n . When m and n both reduce by one,*

¹Note 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 $C = 0$

- 1: **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 .
 - 2: **Step 0: Upper convex hull filtering**
 - 3: Keep only points $(F(w_k), l_k)$ on the upper convex hull (increasing portion), and relabel the remaining types as $\{w_1, \dots, w_{\bar{K}}\}$.
 - 4: **Step 1: Pre-compute Hull Slopes**
 - 5: **for** $k = 2$ to \bar{K} **do**
 - 6: $l'_k \leftarrow \frac{l_k - l_{k-1}}{F(w_k) - F(w_{k-1})}$.
 - 7: **end for**
 - 8: **Step 2: Initialization**
 - 9: Initialize $V(0, n) = 0$ for all $n \in \{0, \dots, N\}$ and $V(m, 0) = 0$ for all $m \in \{0, \dots, M\}$.
 - 10: **Step 3: Fill DP Table by row**
 - 11: **for** $m = 1$ to M **do**
 - 12: **for** $n = 1$ to $\min\{m, N\}$ **do**
 - 13: $V(m, n) \leftarrow n \cdot l_{\bar{K}}, \quad p(m, n) \leftarrow w_{\bar{K}}$.
 - 14: **end for**
 - 15: Initialize pointer $k = \bar{K}$.
 - 16: **for** $n = m + 1$ to N **do**
 - 17: $d \leftarrow V(m, n - 1) - V(m - 1, n - 1)$.
 - 18: $\lambda \leftarrow (1 - q)d$.
 - 19: **while** $k > 1$ **and** $\lambda > l'_k$ **do**
 - 20: $k \leftarrow k - 1$.
 - 21: **end while**
 - 22: $V(m, n) \leftarrow V(m, n - 1) + l_k - qd - F(w_k)\lambda$.
 - 23: $p(m, n) \leftarrow w_k$.
 - 24: **end for**
 - 25: **end for**
 - 26: **Output:** Value table $\{V(m, n)\}$ and first-offer policy $\{p(m, n)\}$.
-

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 request for each worker. Offering low initial pay risks losing both a worker and the request they could have served. 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 ($n - m$) remains the same, but the number of requests decreases. Hence, the supply-

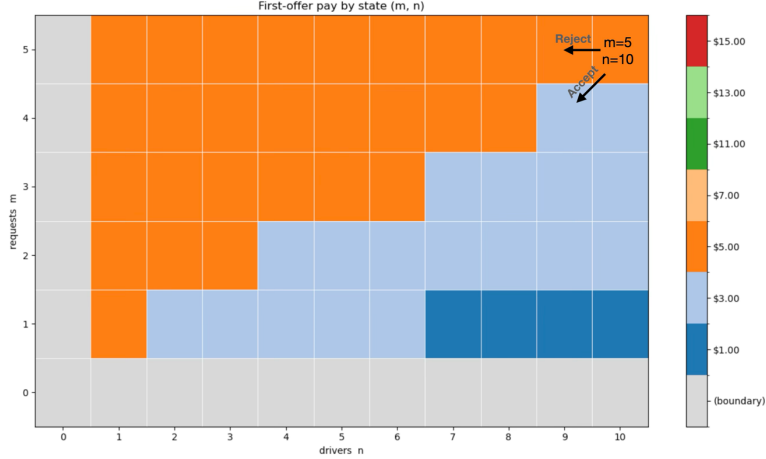


Figure 4: Numerical example: optimal first offer given (m, n)

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.

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 $(n - m)/m$ 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. Section 4.3 illustrates the monotonicity of the optimal first offer.

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 enhances interpretability of the optimal pay policy.

We next study the *width* of the pay sequence, denoted by $b(m, n)$, which we define as the difference between the highest and lowest possible offers that may arise under the optimal

policy starting from state (m, n) . Formally,

$$b(m, n) \triangleq \max p(m', n') - \min p(m', n') \quad (10)$$

where the maximum and minimum are taken over all states (m', n') that may be visited starting from (m, n) under the optimal policy. We have the following results:

Proposition 1. *For any state (m, n) , the width of the optimal pay sequence is*

$$b(m, n) = w_{k^*} - p(1, (n - m)^+ + 1).$$

Moreover, the following hold:

- $b(m, n) = 0$ when demand weakly exceeds supply ($m \geq n$), or when customers are sufficiently impatient ($q \geq q_0$).²
- $b(m, n)$ is weakly increasing in the supply-demand gap $(n - m)^+$ and weakly decreasing in the cancellation probability q .

As shown by Proposition 1, the width of the pay sequence critically depends on the gap between supply and demand, $(n - m)^+$. The larger the excess supply, the wider the pay sequence, as the platform has more opportunities to identify workers with a lower reservation wage. The pay sequence width also depends on customer patience, as captured by the request cancellation probability q . 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. (11), 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. We will revisit the question of the value provided by dynamic versus static pay strategies in Section 8.

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

²The threshold q_0 is defined as

$$q_0 = \frac{l_{k^*}}{l_{k^*} + l'_{k^*}} \quad (11)$$

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

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 workers 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 by adapting 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 workers, leading to tractability challenges in many situations. We summarize these challenges in Fig. 6. We observe that the runtime of the exact algorithm quickly rises to minutes and then hours when the number of workers 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 and on the particular acceptance realizations from workers. In the following sections, we describe simpler approaches with provable guarantees that may provide more actionable

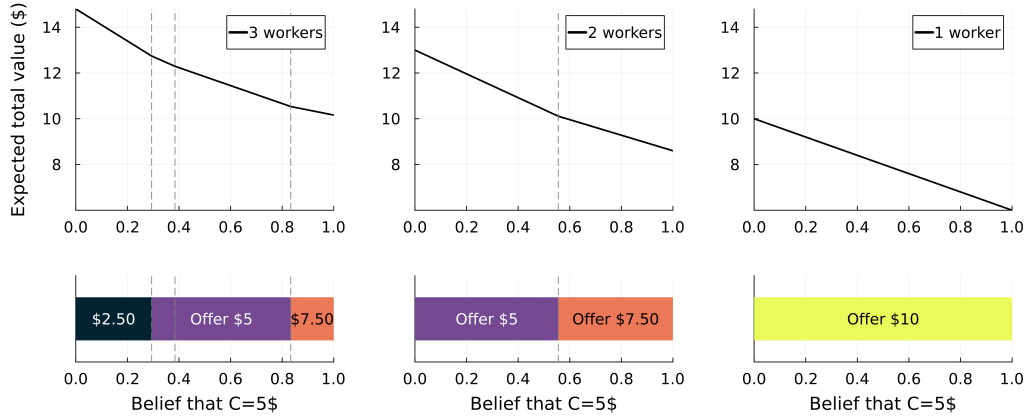
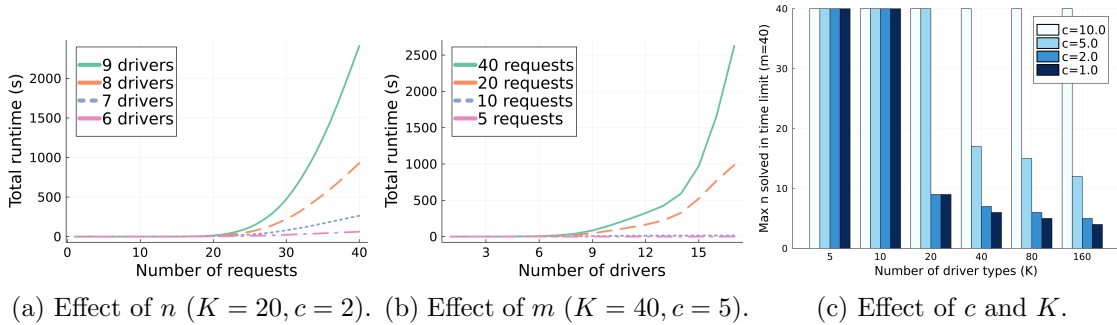


Figure 5: Illustration of the optimal value function and pay offers ($m = 1$, $n = 3$).

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 worker reservation wages are uniformly distributed among K equally spaced values 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 workers 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 worker types, but the maximum number of workers the method can handle within the time limit quickly drops to the single digits as worker heterogeneity increases.

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

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

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 $V^\pi(m, n | \mu_0)$ 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

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

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 Thompson Sampling

A natural approach to learning the global factor is Thompson Sampling (TS), which maintains a posterior belief over the unknown parameter and repeatedly samples from this belief to guide decisions. In our setting, Thompson Sampling naturally integrates with the full-information dynamic program developed in Section 4. Below we introduce this integrated algorithm.

Definition 1 (DP-based Thompson Sampling). *At state (m, n, μ) , the DP-based Thompson Sampling policy proceeds as follows:*

1. Sample $\tilde{C} \sim \text{Bernoulli}(\mu)$.
2. Solve the full-information problem assuming $C = \tilde{C}$ using Algorithm 1 (DP from Section 4).
3. Offer the first pay level prescribed by the resulting policy.
4. Observe the worker's acceptance or rejection and update the belief μ using Eq. (5) and Eq. (6).

Under DP-based Thompson Sampling, the platform samples a competition status from the current belief and behaves as if that sampled state were correct. Over time, belief updates gradually steer the policy toward the correct regime.

While Thompson Sampling has strong theoretical guarantees in classical multi-armed bandit problems, our setting differs substantially from that framework. In particular, actions influence both immediate payoffs and the information obtained from the observed acceptance/rejection decisions through the underlying dynamic program. As a result, existing regret guarantees for bandits do not directly apply. We therefore treat Thompson Sampling as a natural learning benchmark and evaluate its performance through theoretical comparisons and numerical experiments in the next section.

5.4 Direct-Commit: a No-Learning Approach

Building on the full-information policies characterized in Section 4, a natural question is how much the platform loses if it commits to a policy without learning the true competition status.

We introduce a no-learning policy called “Direct-Commit” (DC), which 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 misspecification loss from committing to each benchmark policy and selects the one with the smaller worst-case regret under the prior belief μ_0 .

Definition 2 (Direct-Commit). *Let π^{DC} denote the policy that, given prior belief μ_0 , applies π^0 if*

$$\mu_0 \leq \frac{c}{c + \beta}, \text{ where } \beta = \min\{c, (\bar{v} - c)^+\} \quad (12)$$

and applies π^c otherwise.

The threshold in (12) balances the expected loss bound associated with the two forms of misspecification (applying π^0 when the true regime is $C = c$ and applying π^c when the true regime is $C = 0$), which leads to the following bounded-regret guarantee for the Direct-Commit policy.

Theorem 3 (Regret bound for Direct-Commit). *With prior belief μ_0 , the Direct-Commit policy π^{DC} satisfies*

$$\text{Regret}(\pi^{\text{DC}}) \triangleq V^{\text{clair}}(m, n) - V^{\pi^{\text{DC}}}(m, n \mid \mu_0) \leq L(\mu_0; c) \cdot \min\{m, n\}. \quad (13)$$

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

Theorem 3 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, Theorem 3 shows that when c approaches \bar{v} , Direct-Commit is also near-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 near-optimal.

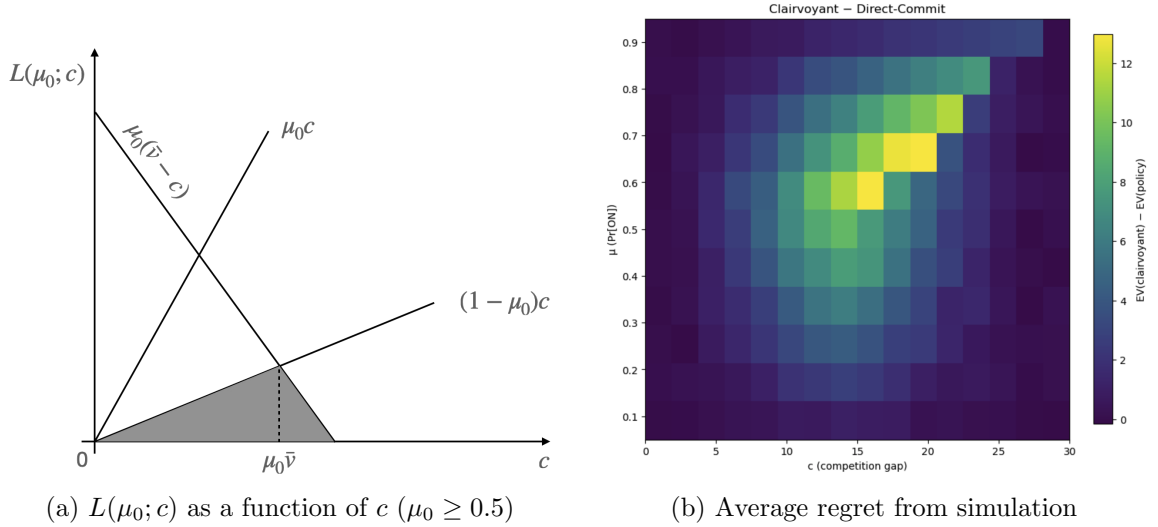


Figure 7: Theoretical regret bound behavior and average regret from simulation for Direct-Commit

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.

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. (13) 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, making the 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 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.

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

A limitation of Direct-Commit arises from its inability to learn the true competition status. This suggests a simple improvement: before committing to a policy, the platform may first gather information about the global factor. We introduce a “Probe-and-Commit” policy that uses the first offer to elicit information about the global factor, then commits to the corresponding policy π^0 or π^c for the remaining periods. The key design question is how to choose the probe price so that it balances immediate payoff and the informativeness of the resulting acceptance or rejection.

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

Definition 3 (Probe-and-Commit with probe price p). *Fix a probe price p and prior belief μ_0 . The Probe-and-Commit policy $\pi^{\text{PC}}(p)$ operates as follows:*

1. *Probe.* In the first period, offer pay level p to the first worker and observes the acceptance outcome $Y \in \{A, R\}$.
2. *Belief update.* After observing Y , update the posterior belief $\mu'_A(p)$ and $\mu'_R(p)$ (defined by Eq. (5) and Eq. (6)).
3. *Commit.* Conditional on the posterior belief μ'_Y , apply the Direct-Commit rule: commit to policy π^0 if $\mu'_Y \leq \frac{c}{c+\beta}$, and π^c otherwise, where $\beta = \min\{c, (\bar{v} - c)^+\}$.
4. *Continuation.* Apply the selected policy for all remaining periods.

The design of the Probe-and-Commit policy allows us to establish the following regret bound.

Theorem 4 (Regret bound for Probe-and-Commit). *Consider the Probe-and-Commit policy $\pi^{\text{PC}}(p)$ with probe value p . The expected regret relative to the clairvoyant benchmark satisfies*

$$\text{Regret}(\pi^{\text{PC}}(p)) = V^{\text{clair}}(m, n) - V^{\pi^{\text{PC}}(p)}(m, n \mid \mu_0) \tag{14}$$

$$\leq \underbrace{(\bar{v} - P_{\text{accept}}(p, \mu_0)(v - p))}_{\text{immediate loss}} + \underbrace{\min\{m, n - 1\} \mathbb{E}[L(\mu'_Y(p); c)]}_{\text{continuation loss}}, \tag{15}$$

where $Y \in \{A, R\}$ and $\mu'_A(p), \mu'_R(p)$ are defined in Eqs. (5) and (6). Moreover, define

$$G(p) \triangleq c(1 - \mu_0)F(p) - \beta\mu_0F(p - c).$$

Then the continuation loss depends on p only through $G(p)$ and is weakly decreasing in $G(p)$; hence minimizing the continuation loss is equivalent to maximizing $G(p)$.

Theorem 4 shows that choosing the probe involves a tradeoff between immediate loss and the continuation loss. The first term captures the lost payoff from using the probe instead of the lowest possible pay; the second term reflects the value of information generated by the probe for future decisions.

Theorem 4 also reveals the key forces driving the continuation loss through the representation of $G(p)$. 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 $\beta\mu_0F(p - c)$, 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.

6 Comparing Pay Strategies

Section 5 introduces a menu of pay strategies which vary in both complexity and potential to learn the global factor. We now introduce some guidelines for practitioners to select a pay policy depending on problem parameters.

6.1 Summary of managerial insights

We view the choice of policy as depending on two key dimensions: the effective horizon (i.e., the number of workers and requests that share the same global status C), and the strength of the global factor relative to worker heterogeneity. Intuitively, more complex learning policies such as Thompson sampling tend to be worthwhile only if the global factor remains in effect for long enough. Conversely, no-learning or low-learning approaches may be preferable when the effective horizon is shorter or when the global factor is strong and thus easier to learn. We summarize our insights in Table 1. The remainder of this section provides the analytical and numerical evidence underlying these recommendations.

Table 1: Recommended learning policies across environments

Environment	Recommended Policy	Key Reason
Short horizon	Direct-Commit (DC)	Learning opportunities limited, exploration costs dominate.
Moderate horizon	Probe-and-Commit (PC)	One informative probe quickly identifies the regime.
Very long horizon	DP-Based Thompson Sampling (TS)	Gradual learning becomes worthwhile over many periods.
Strong global factor	Probe-and-Commit (PC)	One probe is highly informative, enabling near-clairvoyant performance.

6.2 Probe-and-Commit vs. Thompson Sampling

To obtain a tractable comparison between Probe-and-Commit and DP-based Thompson Sampling, we consider the setting with no heterogeneity in workers’ reservation wages. This isolates the learning aspect of the two approaches. The setting can also be interpreted as one in which the global competition status dominates worker heterogeneity. We establish the following result:

Proposition 3. *Assume all workers share the same reservation wage. Moreover, all m requests share the same global factor $C \in \{0, c\}$, and $n > m$. Let μ_0 denote the platform’s prior belief that $C = c$. Then the following hold:*

- $\text{Regret}(\pi^{PC}(0)) = 0$. That is, Probe-and-Commit with probe $p = 0$ is optimal.
- $\text{Regret}(\pi^{TS}) = cp_\epsilon \cdot \frac{1-\mu_0^m}{1-\mu_0}$, where $p_\epsilon = \mu_0(1 - \mu_0)$ denotes the probability of an uninformative acceptance. This occurs when the platform samples $C = c$ while the true state is $C = 0$, leading it to offer the higher payment c , which is accepted regardless of the competition status and therefore does not reveal information about the regime.

Proposition 3 indicates that when worker heterogeneity is small and the global competition status is strong, Probe-and-Commit performs well by leveraging the ability to perfectly separate the two regimes and complete learning with a single offer. In contrast, Thompson Sampling tends to overpay in this setting. Such overpaying has two costs: first, it reduces the profit margin of the current offer; second, it makes the observation less informative, since a high pay leads to acceptance regardless of the regime. This highlights the advantage of Probe-and-Commit: when the global factor is strong, it is preferable to exploit the problem structure rather than rely on the more gradual learning process of a generic method such as Thompson Sampling.

One can also observe that, all else equal, as m increases, the regret of Thompson Sam-

pling decreases and converges to the lower bound $cp_\epsilon/(1-p_\epsilon)$. In other words, Thompson Sampling performs worst when the effective horizon is short (i.e., when m is small). This again highlights a limitation of Thompson Sampling in short-horizon environments: learning through sampling is too slow relative to the effective horizon, and by the time the global status is learned, most of the decisions have already been made.

6.3 Probe-and-Commit vs. Direct-Commit

We now return to the general setting with heterogeneous reservation wages. The value of the Probe-and-Commit policy becomes particularly clear in the special case where the global effect dominates individual heterogeneity, i.e., $c > w_K - w_1$. In this regime, 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. As a result, the regret of the probe-and-commit policy simplifies as follows.

Corollary 1. *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 Theorem 4 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 1 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 Theorem 4 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. In particular, we can show the following result:

Proposition 4. *It holds that*

$$\text{Regret}(\pi^{DC}) > \text{Regret}(\pi^{PC})$$

as long as $\min\{m, n\} > \frac{\bar{v} - (1 - \mu_0)(v - w_K)}{\delta(\mu_0)}$, where

$$\delta(\mu_0) = \begin{cases} \mu_0 \max_{k \in [K]} F(w_k)(v - w_k - c) & , \text{ if } \mu_0 \leq \frac{c}{c+\beta} \\ (1 - \mu_0)(c - (w_K - w_1)) & , \text{ if } \mu_0 > \frac{c}{c+\beta}. \end{cases}$$

As long as the system is large enough, Probe-and-Commit clearly dominates Direct-Commit when the global effect is sufficiently strong compared to individual heterogeneity. In the

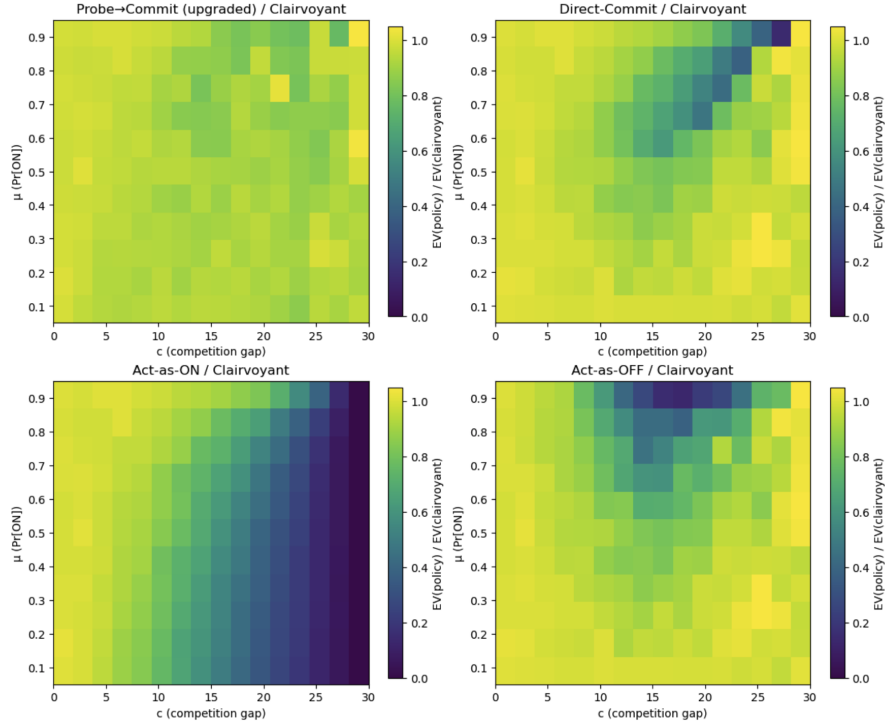


Figure 8: Performance comparison across multiple policies, relative to the clairvoyant benchmark

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

general setting, comparing the two approaches is more difficult, as the result also depends on the value of the prior belief μ_0 .

Numerical comparison. Fig. 8 compares the performance of Direct-Commit with Probe-and-Commit, along with simple myopic commitment to π^0 (act as if the global factor is “off”), and myopic commitment to π^c (act as if the global factor is “on”), 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. (12), 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.

7 Simulation Studies in a Dynamic Environment

Our analysis so far has relied on a static setting with a fixed number of requests m and workers n . However, gig economy platforms are typically much more dynamic environments, with requests and workers stochastically entering and exiting the platform. The goal of this section is to explore the performance of our different policies in such a dynamic environment. We observe that our insights in Table 1 are robust to this more realistic setting, with no-learning policies performing better when learning is harder or less valuable.

7.1 Dynamic Platform Model and Rolling Horizon Policies

We consider a fixed horizon of T periods (or time steps), which we refer to as an episode. We let m_t and n_t designate the number of active requests and workers at the beginning of each period; the episode begins with m_1 pending requests and n_1 available workers. In time step t , new requests arrive following a Poisson distribution with parameter λ_r , and new workers arrive following a Poisson distribution with parameter λ_w . Each of the n_t current workers independently exits with probability δ .

As long as $n_t > 0$ and $m_t > 0$, the platform makes a pay offer to one worker in period t . Pay dynamics follow the model described in Section 3. Each worker’s reservation wage is the sum of an i.i.d. individual effect and a shared global effect. If the worker accepts an offer with pay p , the platform obtains revenue $v - p$ and the worker and the fulfilled request both exit the platform. If the worker rejects the offer, they exit the platform, and the request is canceled with probability q .

Though designed for a fixed value of m and n , all of the pay policies described so far can be adapted to this new dynamic setting using a rolling-horizon framework. At every time step, we first observe the current state of the platform (m_t, n_t) ; we then apply the first pay level specified by the policy given m_t requests and n_t workers, denoted by $\pi(m_t, n_t)$. The policies then only differ in how they select the dynamic programming table from which to read the initial pay offer. We first consider two **dynamic programming (DP)** policies in which we either always use π^0 or always use π^c — these policies correspond to guessing the value of the global factor and then using the optimal full-information pay policy. We also consider **direct-commit (DC)** and **probe-and-commit (PC)** policies, in which we commit to π^0 or π^c in each episode based either on our initial belief or on a single high-information probing offer. The **Thompson Sampling (TS)** policy samples the current value of C from its prior belief, applies the corresponding pay $\pi^0(m_t, n_t)$ or $\pi^c(m_t, n_t)$, then updates its belief based on worker response. The **belief-augmented DP** policy uses the optimal pay given its current belief (computed using the methods described in

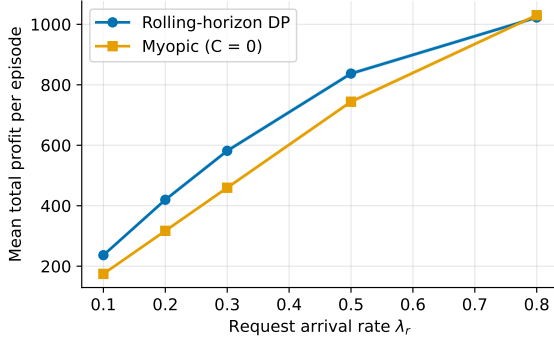


Figure 9: Comparing the performance of adaptive and myopic pay strategies under known global factor.

Notes:

Section 5.1), then updates its belief. The **clairvoyant** policy knows the current value of C and correspondingly applies either π^0 or π^c . We compare all these policies to two **myopic** policies (assuming $C = 0$ and $C = c$) in which the platform simply tries to maximize immediate expected profit (e.g., $F(p)(v - p)$ when $C = 0$).

Throughout the simulations, we let $T = 100$ and $K = 5$: each individual reservation wage is sampled from $\{0, 5, 10, 15, 20\}$ with the probabilities $\{0.1, 0.2, 0.25, 0.25, 0.2\}$. We also fix $m_1 = 2$, $n_1 = 20$, $\lambda_w = 1$, $\delta = 0.05$, and $q = 0.1$. The value of serving a request is fixed at $v = 30$. We vary other parameters as needed. We show an example of a single episode in Fig. 14 in the appendix. Results throughout this section are averaged over 500 episodes. Standard errors are provided when large enough to be visible.

7.2 Policy Performance

We first compare the performance of adaptive pay strategies relative to myopic strategies when the platform knows the global factor $C = 0$. In this case, all our dynamic policies collapse to π^0 , which we can compare to the optimal myopic pay under $C = 0$. We show results for varying λ_r in Fig. 9. When λ_r grows large, the platform becomes undersupplied and the optimal adaptive pay is simply myopic. However, small to moderate values of λ_r lead to large profit increases from adaptive pay (up to 26% when $\lambda_r = 0.3$).

Regret relative to clairvoyant policy. We fix $\lambda_r = 0.3$ and compare all policies when $C = 8$ and $\mu_0 = 0.3$. Fig. 10 shows each’s policy average regret compared to the clairvoyant policy. We observe that even when we blindly guess the value of C , the DP approach reduces regret by a factor of more than two compared to the corresponding myopic pay. The best no-learning approach (direct-commit) achieves a regret of 7.5%. The simplest learning approach (probe-and-commit) achieves 5.1%, with Thompson Sampling at 1.6%

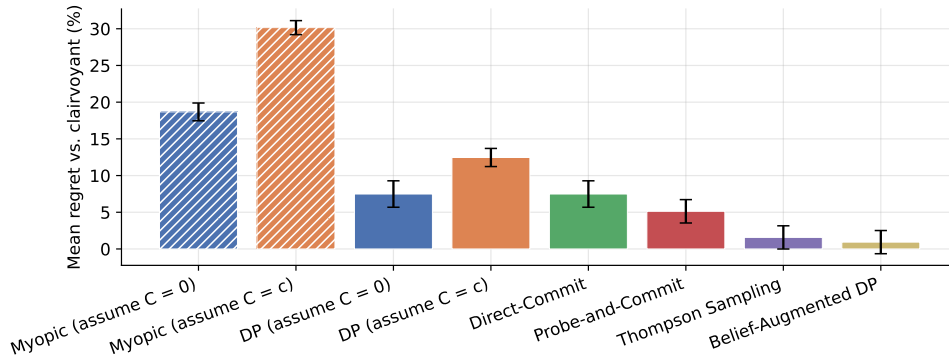


Figure 10: Regret comparison across adaptive policies in general setting.

Notes: We assume $c = 8$ and $\mu_0 = 0.3$.

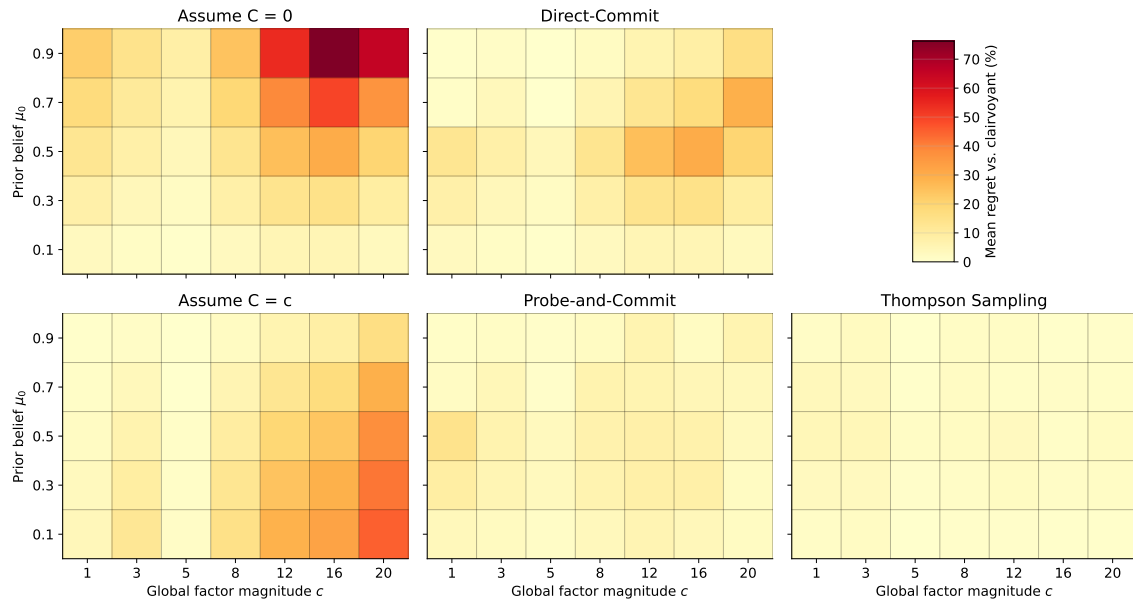


Figure 11: Regret of adaptive policies under varying global factor magnitude (c) and likelihood (μ_0).

Notes:

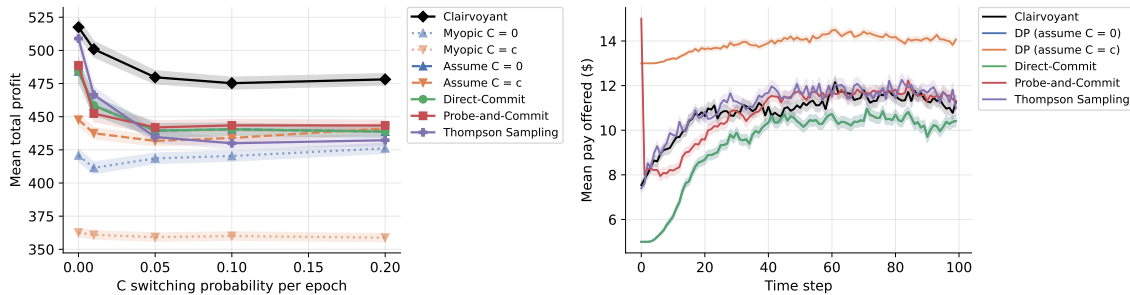
and the belief-augmented DP approach at 0.9%. As expected, the “optimal” approach performs the best, though with a significant computational overhead (over 2 hours). This intractability may make the approach less appealing in practice, though we observe that almost all the computation occurs when creating the dynamic programming table (which could in principle be pre-computed). Meanwhile, a simple “one-shot learning” approach is within striking distance of more advanced sampling or optimization approaches.

In the appendix (Table 2) we present a more detailed assessment of the platform behavior under these different policies. In particular, we observe that direct-commit performs the best of any policy when correctly guessing $C = 0$, but the worst of any policy otherwise, mostly because its pay is simply too low given the high global factor. A single probe mildly erodes profit when $C = 0$ but massively improves it when $C = c$. Compared to Thompson Sampling and the belief-augmented MDP method, probe-and-commit tends to lead to slightly higher pay when $C = 0$, and slightly lower pay when $C = c$. Correspondingly, it fulfills slightly more requests when $C = 0$ and slightly fewer when $C = c$.

Fig. 11 expands the regret analysis to different values of c and μ_0 . We observe that while direct-commit achieves “best of both worlds” performance relative to simply guessing whether $C = 0$ or $C = c$, it can have significant regret when the global factor magnitude and likelihood are large. A single informative probe can significantly alleviate this risk, achieving regret performance close to Thompson Sampling with no learning beyond the first pay offer.

Robustness to global factor dynamics. Our analysis so far has relied on a two-point distribution for the global factor. We now study two settings which depart from this assumption. First, instead of sampling the global factor once per episode, we let it evolve according to a two-state Markov chain: we let $C = c$ in the first period with probability μ_0 , and in each time step it either remains the same with probability $1 - \rho$ or switches (from 0 to c or c to 0) with probability ρ . None of our pay strategies are aware of this switching behavior (i.e., they remain designed with a fixed C in mind), but the clairvoyant baseline is altered to have access to the current value of C in each time step.

Fig. 12a shows the profit obtained from each pay strategy as a function of the switching probability ρ . We observe that the no-learning and low-learning approaches, including probe-and-commit, direct-commit, and, surprisingly, guessing that $C = c$, are more robust to these new dynamics than Thompson Sampling. Fig. 12b shows that this robustness can materialize in different ways in the particular case when $\rho = 0.1$. Direct-commit sets lower average pay, serving about 2 fewer requests than Thompson Sampling (23.6 vs. 25.7), but at higher profit per request. Conversely, guessing that $C = c$ leads to higher average pay and serves about 2 more requests (27.5 vs. 25.7) at lower profit per request. Finally, probe-and-commit uses one high pay offer to gain information about the global factor: on average, this allows it to set prices lower than Thompson Sampling for the first 10-20 time



(a) Profit as a function of switching probability ρ (b) Average pay per time step under $\rho = 0.1$

Figure 12

Notes:

steps, serving almost the same number of requests (25.5 vs. 25.7) at a slightly higher profit. Intuitively, low-learning approaches are reasonable in setting where learning is inherently less valuable due to system drift.

We also consider a “noisy” global factor model in the appendix, where C is no longer sampled from a two-point distribution but from a mixture of (clipped) Gaussians. We show the performance of our various pay policies (which still assume a two-point distribution) as a function of global factor variance in Fig. 15 in the appendix, where we do not observe significant deterioration from model misspecification. Overall, our simulation results in a dynamic environment largely confirm the analytical results from Section 6.

8 The Value of Dynamic Pay

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 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. It also allows us to relax the assumption that workers do not return to the platform after a rejection, as it precludes strategic behavior.

8.1 Optimal Static Pay

So far, the only static policy we have considered is myopic, in which we offer every worker the pay that maximizes single-period expected profit. Over a given horizon with m requests and n workers, myopic pay may not even be the optimal static policy. We compare numerically the expected profit of optimal static pay and myopic pay in a particular setting in Figure 13. We observe that though the optimal static pay is suboptimal (relative to a clairvoyant dynamic baseline), it can often outperform myopic pay. Adaptive pay strategies

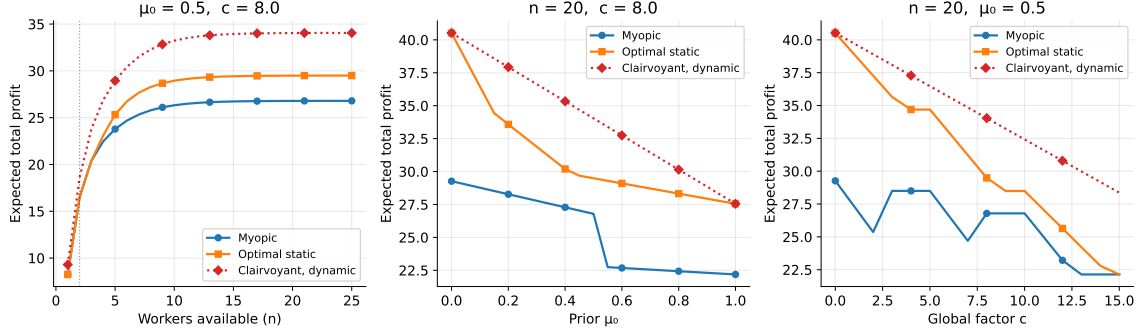


Figure 13

Notes: The reservation wage distribution is the same as in Section 7, and we fix $m = 2$.

can therefore be valuable even if they only adapt once (at the beginning of the episode) rather than for every worker.

Though it is intractable to characterize the optimal static pay in closed form for any m and n , we can do so easily in the limit when the number of workers grows very large.

Lemma 1. *Let $p_s(\mu_0, m, n)$ designate the optimal single-pay policy for m requests and n workers. Then $p_s(\mu_0, m, \infty) := \lim_{n \rightarrow \infty} p_s(\mu_0, m, n)$ satisfies*

$$p_s(\mu_0, m, \infty) = \arg \max_p (1 - \mu_0) \frac{F(p)(v - p)}{1 - (1 - q)(1 - F(p))} + \mu_0 \frac{F(p - c)(v - p)}{1 - (1 - q)(1 - F(p - c))}. \quad (16)$$

The first term in (16) represents the expected profit when $C = 0$ and the second represents the expected profit when $C = c$. In a slight abuse of notation, we let the second term be zero when $F(p - c) = 0$ to avoid undefined fractions. Furthermore, we observe that the optimal single-pay policy with infinite workers is the same for any finite number m of requests and denote it by $p_s(\mu_0, \infty)$ to simplify notation. To interpret Lemma 1 more readily, we first examine its implications for extreme values of the cancellation rate q .

Corollary 2. *In the limit with infinite workers, the optimal static pay policy offers*

- the optimal expected myopic pay when requests are maximally impatient, i.e., when $q = 1$,

$$p_s(\mu_0, \infty) = \arg \max_p [(1 - \mu_0)F(p) + \mu_0 F(p - c)](v - p),$$

- the “minimal acceptable pay” when requests are maximally patient, i.e., when $q = 0$,

$$p_s(\mu_0, \infty) = \begin{cases} w_1, & \text{if } \mu_0 < \frac{c}{v - w_1}, \\ w_1 + c, & \text{otherwise.} \end{cases}$$

The closed form in Lemma 1 interpolates between these two regimes in the limit with many workers. When requests are impatient (large q), the optimal policy (static or not) becomes myopic (see Proposition 1), where the single-period expected profit is computed using the true distribution of reservation wages, which is simply a mixture of $F(p)$ and $F(p - c)$ with weights $(\mu_0, 1 - \mu_0)$. Conversely, when requests are patient (small q), the optimal static pay is the lowest acceptable pay (which could be w_1 or $w_1 + c$ depending on the platform's initial belief). Indeed, the probability this offer is accepted tends to one as the number of workers tends to infinity.

Moderate values of q lead to optimal static pay policies between these two extremes. For example, when $\mu_0 = 0$ (known global factor), the optimal static pay is always between w_1 and w_{k^*} for any q . In this case, it turns out that the static pay policy with infinite workers is also a lower bound on the optimal dynamic pay policy, as described in the following result.

Proposition 5. *Let $\mu_0 = 0$. For any m, n , it holds that $p(m, n) \geq p_s(0, \infty)$. Consequently, the maximum pay width for m requests can be bounded as $\max_n B(m, n) \leq w_{k^*} - p_s(0, \infty)$.*

One interpretation of Proposition 5 is that, while switching from a static to a dynamic policy obviously increases the pay width, it does not automatically make every worker worse off. When many workers are available, the optimal static policy will lead to low pay for all, while the optimal dynamic policy will increase pay as more workers reject the offer.

8.2 Value of Dynamic Pay

We now study the gap between the value of the optimal dynamic policy $V^*(\mu_0, m, n)$ and the value of the optimal static policy $V_s(\mu_0, m, n)$.

Theorem 5 (Value of Flexibility). *The value of flexibility, defined as $\text{VoF}(\mu_0, m, n) := V^*(\mu_0, m, n) - V_s(\mu_0, m, n)$, is characterized by the following properties:*

1. *If $n \leq m$, static pay is optimal: $\text{VoF}(\mu_0, m, n) = 0$.*
2. *If $n > m$:*
 - *If $\mu_0 \in \{0, 1\}$, static pay is optimal for infinite workers: $\lim_{n \rightarrow \infty} \text{VoF}(m, n) = 0$. Furthermore, for any m , if $p_s(\mu_0, \infty) \neq w_{k^*}$, there exists n such that $\text{VoF}(m, n) > 0$.*
 - *If $\mu_0 \in (0, 1)$ and $w_k \leq c$, there exists m sufficiently large such that $\lim_{n \rightarrow \infty} \text{VoF}(m, n) > 0$.*

Theorem 5 highlights the differences between the settings where the global factor is known and where it must be learned. When it is known, adopting a dynamic policy is most effective for intermediate n . Indeed, the optimal pay is myopic when n is small, and nearly

static when n is large (except for the last few workers). Adopting a dynamic pay policy is most valuable when n is in between these two extremes.

In contrast, learning and exploiting the global status usually *requires* some pay changes. This means that the value of dynamic pay does not tend to zero in general as the number of workers tends to infinity. Note that the proof of the last statement requires $w_K \leq c$ for the statement to hold for any $\mu_0 \in (0, 1)$. If c is small and μ_0 is small, the global factor has limited impact on pay and the platform operates more like the known-global-factor setting.

In proving the last statement of Theorem 5, we bound the value of flexibility away from zero for large enough m and n . The general lower bound in the appendix is cumbersome to interpret, but we can obtain a simplified version in the following special case.

Corollary 3. *Assume that $\mu_0 \in (0, 1)$, $q = 0$, $w_1 = 0$, and $w_K \leq c$, then $\lim_{n \rightarrow \infty} VoF(\mu_0, m, n) > 0$ if the following condition is satisfied:*

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

Corollary 3 gives clearer insights as to the value of a dynamic pay policy. Such a policy is worth it when 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.

9 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 with much of the recent literature on learning approaches for platform operations. Many reinforcement learning approaches in ridesharing marketplaces work most effectively in an under-supplied setting, where request volume exceeds driver

availability (Qin et al. 2025). In symmetric 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 will minimize a driver’s subsequent idle time. In this case, the platform’s strategic lever is no longer minimizing service cost, but instead allocating scarce supply to the most valuable demand to optimize long-term platform objectives.

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.
- Bernstein F, DeCroix GA, Keskin NB (2021) Competition between two-sided platforms under demand and supply congestion effects. *Manufacturing & Service Operations Management* 23(5):1043–1061.
- 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.
- Bitran G, Caldentey R (2003) An overview of pricing models for revenue management. *Manufacturing & Service Operations Management* 5(3):203–229.
- Bright I, Delarue A, Lobel I (2025) Reducing marketplace interference bias via shadow prices. *Management Science* 71(8):7094–7112.
- Cachon GP, Daniels KM, Lobel R (2017) The role of surge pricing on a service platform with self-scheduling capacity. *Manufacturing & Service Operations Management* 19(3):368–384.

- 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-upfront-fares>
- 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, Jacquillat A, Serpa JC, Benborhoum M (2023) Managing airfares under competition: Insights from a field experiment. *Management Science* 69(10):6076–6108.
- 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 cooperation for two-sided platforms. *Production and Operations Management* 31(5):1997–2014.
- Den Boer AV (2015) Dynamic pricing and learning: Historical origins, current research, and new directions. *Surveys in operations research and management science* 20(1):1–18.
- 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).
- Guda H, Subramanian U (2019) Your uber is arriving: Managing on-demand workers through surge pricing, forecast communication, and worker incentives. *Management Science* 65(5):1995–2014.
- 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.
- Keskin NB, Zeevi A (2017) Chasing demand: Learning and earning in a changing environment. *Mathematics of Operations Research* 42(2):277–307.
- 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.
- Taylor TA (2018) On-demand service platforms. *Manufacturing & Service Operations Management* 20(4):704–720.
- 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 Additional Simulation Results

Fig. 14 shows the number of workers and requests in one particular episode under the direct-commit policy. We observe the number of active workers varies from 0 to 20 while the number of requests varies from 2 to 8.

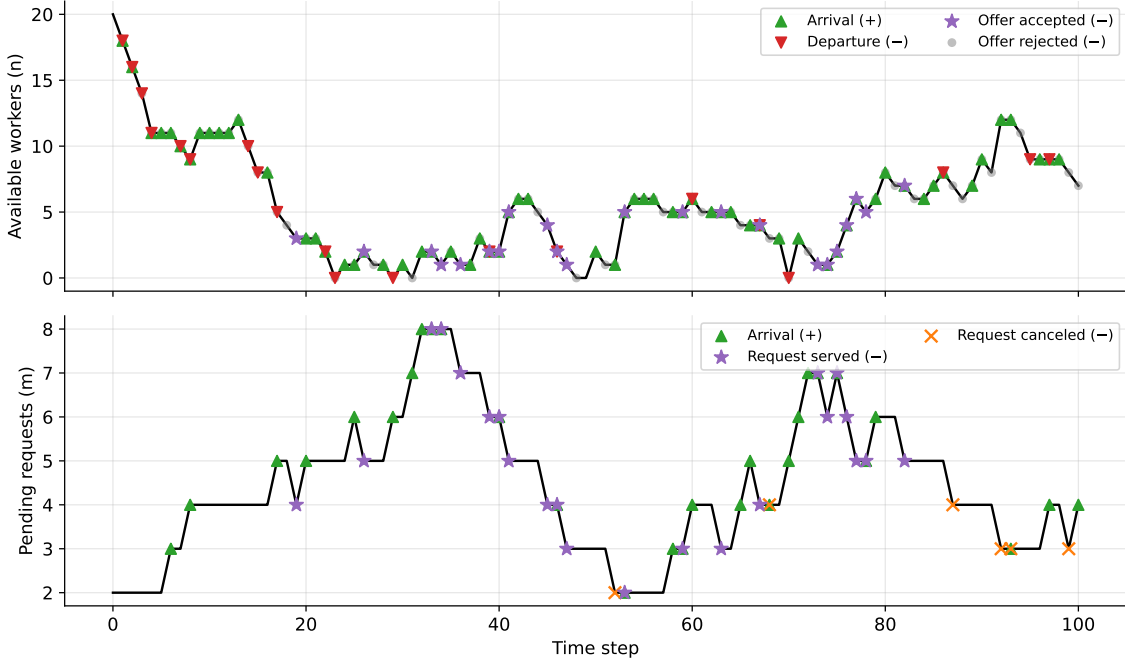


Figure 14: Dynamics of a single episode.

Notes: Top panel depicts the number of workers and bottom panel the number of requests. In this example we fix $c = 8$, $\mu_0 = 0.3$, and $\lambda_r = 0.3$.

Fig. 15 depicts the performance of our various pay strategies when the global factor is no longer sampled from a two-point distribution. Instead, we let $C = \max(0, \hat{C})$, where \hat{C} is sampled with probability μ_0 from a normal distribution with mean 0 and standard deviation σ and with probability $1 - \mu_0$ from a normal distribution with mean c and standard deviation σ . Even though the pay policies still operate under the assumption of a two-point distribution, we don't observe significant performance deterioration. Note that the clairvoyant policy here knows the true value of C in each episode, but is still restricted to the optimal pay offers under $C = 0$ and $C = c$ (choosing based on whether the sampled C is closer to 0 or c).

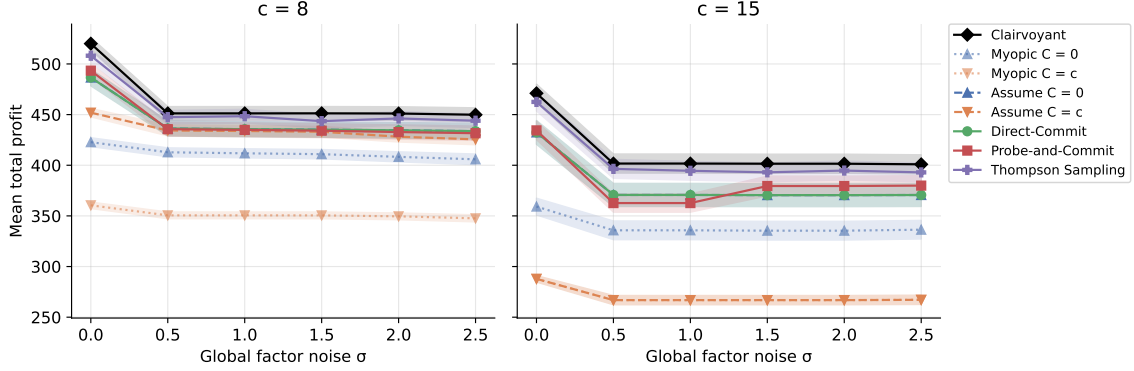


Figure 15: Performance of pay strategies under noisy global factor model.

Notes: We fix $\mu_0 = 0.3$ and $\lambda_r = 0.3$.

Finally, Table 2 shows summary statistics of the platform behavior under the parameters for Fig. 10.

B Proofs: Optimal Pay with Known Global Factor

This section provides proofs for the full-information setting, when the global factor is known to the platform (Section 4). We first establish auxiliary results on the value function, which are then used to prove Theorem 1, Theorem 2, and Proposition 1.

B.1 Auxiliary Results

Lemma 2 (Marginal value). *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. Suppose a request may be cancelled with fixed probability $q \in [0, 1]$ after each rejection. 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 $m \geq 1, n \geq 0$.

Proof. Proof of Lemma 2 We prove each points in Lemma 2 one by one.

Proof of Item 1. Proof by induction.

Base case: When $n = 0$, there are no workers available to serve any request. Hence no request can be matched, and the platform earns zero payoff. Therefore, $V(m, 0) = 0$ for all m , which implies $d_m^{(0)} = 0$ for all m . Hence, $d_m^{(0)}$ is weakly decreasing in m .

Table 2: Per-episode metrics for different pay policies.

(a) $C = 0$

Policy	Profit	Fulfilled	Requests		Pay	
			Unfulfilled	Periods Waiting	Offered	Accepted
Clairvoyant	579.29	26.39	5.42	4.35	6.72	7.83
Myopic ($C = 0$)	459.70	30.65	1.31	1.54	15.00	15.00
Myopic ($C = c$)	367.76	30.65	1.31	1.54	18.00	18.00
DP (assume $C = 0$)	579.29	26.39	5.42	4.35	6.72	7.83
DP (assume $C = c$)	482.62	29.32	2.97	2.47	13.35	13.49
Direct-Commit	579.29	26.39	5.42	4.35	6.72	7.83
Probe-and-Commit	556.40	27.42	4.73	3.76	8.42	9.39
Thompson Sampling	574.73	26.52	5.35	4.14	6.97	8.14
Belief-Augmented DP	578.48	26.49	5.38	4.26	6.82	7.95

(b) $C = 8$

Policy	Profit	Fulfilled	Requests		Pay	
			Unfulfilled	Periods Waiting	Offered	Accepted
Clairvoyant	384.82	26.05	6.02	5.04	14.46	15.11
Myopic ($C = 0$)	335.33	22.36	9.78	9.11	15.00	15.00
Myopic ($C = c$)	340.97	28.41	3.31	2.69	18.00	18.00
DP (assume $C = 0$)	271.09	17.50	14.36	17.54	12.14	14.47
DP (assume $C = c$)	384.82	26.05	6.02	5.04	14.46	15.11
Direct-Commit	271.09	17.50	14.36	17.54	12.14	14.47
Probe-and-Commit	351.70	23.65	8.29	8.40	13.77	14.98
Thompson Sampling	369.82	25.65	6.36	5.78	13.86	15.45
Belief-Augmented DP	372.25	25.80	6.24	5.90	14.10	15.40

Note: Simulation parameters are the same as in Fig. 10. Out of 500 episodes, we sampled $C = 0$ 331 times and $C = 8$ 169 times.

Induction hypothesis: For some $N > 0$, $d_m^{(N)}$ is weakly decreasing in m for all $m \geq 1$.

Induction step: Our goal is to show that $d_m^{(N+1)}$ is weakly decreasing in m given the induction hypothesis. For convenience, define

$$\delta_m^{(N)} := V(m-1, N) - V(m, N) = -d_m^{(N)}.$$

Under the induction hypothesis, $d_m^{(N)}$ is weakly decreasing in m , and hence $\delta_m^{(N)}$ is weakly increasing in m .

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

$$\begin{aligned}
V(m, n) &= \max_{k \in [K]} F(w_k) \left((v - w_k) + V(m - 1, n - 1) \right) \\
&\quad + (1 - F(w_k)) \left(qV(m - 1, n - 1) + (1 - q)V(m, n - 1) \right).
\end{aligned} \tag{17}$$

By rearranging the terms in Eq. (17), we have

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

We define the following function:

$$H(\delta) = \max_k F(w_k)(v - w_k + (1 - q)\delta) + q\delta \tag{18}$$

such that the value function in Eq. (17) can be represented by

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

Note that $H(\delta)$ is the pointwise maximum of affine functions with slopes in $[0, 1]$. To see why, for each k , define

$$h_k(\delta) := F(w_k)(v - w_k + (1 - q)\delta) + q\delta.$$

Then $h_k(\delta)$ is affine in δ with slope

$$F(w_k)(1 - q) + q.$$

Since $F(w_k) \in [0, 1]$ and $q \in [0, 1]$, we have

$$0 \leq q + F(w_k)(1 - q) \leq 1.$$

Thus each h_k has slope in $[0, 1]$, and $H(\delta) = \max_k h_k(\delta)$ is the pointwise maximum of such affine functions.

Therefore, $H(\delta)$ is weakly increasing and 1-Lipschitz. In particular, for any $\delta_1 \leq \delta_2$,

$$0 \leq H(\delta_2) - H(\delta_1) \leq \delta_2 - \delta_1 \tag{19}$$

Applying the definition of H , we can rewrite

$$V(m, N + 1) = H(\delta_m^{(N)}) + V(m, N) \tag{20}$$

Hence,

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

By the induction hypothesis, $\delta_m^{(N)}$ is weakly increasing in m . Hence, $\delta_m^{(N)} \geq \delta_{m-1}^{(N)}$. We can then apply the inequality Eq. (19) and obtain:

$$H(\delta_m^{(N)}) - H(\delta_{m-1}^{(N)}) - \delta_m^{(N)} \leq \delta_m^{(N)} - \delta_{m-1}^{(N)} - \delta_m^{(N)} \quad (21)$$

$$= -\delta_{m-1}^{(N)} \quad (22)$$

$$= d_{m-1}^{(N)} \quad (23)$$

Therefore, we have shown that $V(m, N+1) - V(m-1, N+1) \leq d_{m-1}^{(N)}$.

Next, we show that $V(m-1, N+1) - V(m-2, N+1) \geq d_{m-1}^{(N)}$. Apply Eq. (20) again, we have

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

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

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

Combined with the previous step, we have

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

which implies

$$d_{m-1}^{(N+1)} \geq d_m^{(N+1)}$$

This concludes the proof for $d_m^{(N+1)}$ being weakly decreasing in m .

Proof of Item 2. Next, we prove that $d_m^{(n)}$ is weakly increasing in n . Again, by the definition of H , 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

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

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

By Item 1, we have

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

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

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

which concludes the proof.

Proof of Item 3. 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,

$$d_m^{(n+1)} = d_m^{(n)} + H(\delta_m^{(n)}) - H(\delta_{m-1}^{(n)}) \tag{24}$$

$$\leq d_m^{(n)} + \delta_m^{(n)} - \delta_{m-1}^{(n)} \tag{25}$$

$$= -\delta_{m-1}^{(n)} \tag{26}$$

$$= d_{m-1}^{(n)} \tag{27}$$

which concludes the proof. □

B.2 Proof of Theorem 1, Theorem 2, and Proposition 1

Proof of Theorem 1

Proof. Proof. For ease of exposition, define $\alpha_k = F(w_k)$ and $l_k = F(w_k)(v - w_k)$.

Consider a point (α_j, l_j) that lies below the convex hull. Then there exist two points (α_1, l_1) and (α_2, l_2) and a weight $\lambda \in (0, 1)$ such that

$$\alpha_j = \lambda\alpha_1 + (1 - \lambda)\alpha_2, \quad l_j < \lambda l_1 + (1 - \lambda)l_2.$$

It follows that for any scalar δ ,

$$\begin{aligned} l_j + \alpha_j\delta &< \lambda l_1 + (1 - \lambda)l_2 + (\lambda\alpha_1 + (1 - \lambda)\alpha_2)\delta \\ &= \lambda(l_1 + \alpha_1\delta) + (1 - \lambda)(l_2 + \alpha_2\delta) \\ &\leq \max\{l_1 + \alpha_1\delta, l_2 + \alpha_2\delta\}. \end{aligned}$$

Hence $l_j + \alpha_j\delta$ is strictly dominated for all δ . From the Bellman equation Eq. (8), we have

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

Therefore, evaluating at $\delta = (1 - q)\delta_m^{(n-1)}$ yields

$$l_j + \alpha_j(1 - q)\delta_m^{(n-1)} < \max_{k=1,2} \left\{ l_k + \alpha_k(1 - q)\delta_m^{(n-1)} \right\}.$$

Thus option j can never attain the maximum in the Bellman equation and cannot appear in an optimal policy. This proves the convex hull claim.

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 (α_i, l_i) on the convex hull but not on the increasing portion. Thus, there must exist (α_1, l_1) such that $\alpha_1 < \alpha_i$ and $l_1 > l_i$. Since $\delta_m^{(n-1)} \leq 0$, we must have

$$l_1 + \alpha_1(1 - q)\delta_m^{(n-1)} > l_i + \alpha_i(1 - q)\delta_m^{(n-1)}$$

which means option i is strictly dominated and cannot appear in the optimal sequence. \square

Proof of Theorem 2

Proof. Proof. We first prove Item 1 of Theorem 2 which covers the setting of $m \geq n$. We then prove Item 2 which covers the setting of $m < n$.

Proof of Item 1. We prove a stronger claim: for all $m \geq n$,

$$V(m, n) = n l_{k^*}.$$

The desired result, namely $p(m, n) = w_{k^*}$ for all $m \geq n$, then follows immediately from the Bellman equation. Proof by induction:

Base case. When $n = 1$ and $m \geq 1$, there is only one worker available. Hence,

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

which verifies the claim $V(m, 1) = 1 \cdot l_{k^*}$. Moreover, since the Bellman equation reduces to the one-shot problem, the optimal offer is $p(m, 1) = w_{k^*}$.

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

Induction step. We show that for $n = N + 1$,

$$V(m, N + 1) = (N + 1) l_{k^*} \quad \text{and} \quad p(m, N + 1) = w_{k^*}$$

for all $m \geq N + 1$.

From the Bellman equation, we have

$$V(m, N + 1) = \max_{k \in [K]} F(w_k) \left((v - w_k) + (1 - q)(V(m - 1, N) - V(m, N)) \right) \\ + (qV(m - 1, N) + (1 - q)V(m, N)).$$

Since $m \geq N + 1$, we have $m - 1 \geq N$. By the induction hypothesis,

$$V(m - 1, N) = V(m, N) = N l_{k^*}.$$

Substituting into the Bellman equation yields

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

Moreover, since $V(m - 1, N) = V(m, N)$, the continuation term cancels, and the maximization reduces to $\max_k F(w_k)(v - w_k)$. Hence the optimal offer is $p(m, N + 1) = w_{k^*}$.

Proof of Item 2.

We now prove the monotonicity of the optimal first offer. The key observation is that the optimal offer depends on the marginal value $d_m^{(n-1)}$ through the Bellman equation. In particular, we show that the optimal offer is weakly decreasing in $d_m^{(n-1)}$. Combining this with Lemma 2 yields the desired monotonicity results.

From the Bellman equation, the optimal offer at state (m, n) satisfies

$$p(m, n) \in \arg \max_k \left\{ l_k + \alpha_k(1 - q)\delta_m^{(n-1)} \right\} = \arg \max_k \left\{ l_k - \alpha_k(1 - q)d_m^{(n-1)} \right\},$$

where $\alpha_k = F(w_k)$ and $l_k = F(w_k)(v - w_k)$.

Thus, the optimal offer corresponds to selecting the point (α_k, l_k) on the upper convex hull that maximizes

$$l_k - \alpha_k(1 - q)d_m^{(n-1)}.$$

Geometrically, this is equivalent to choosing the point on the upper convex hull that yields the highest intercept when evaluated with slope $(1 - q)d_m^{(n-1)}$. For any k_1, k_2 , the difference

$$(l_{k_1} - \alpha_{k_1}x) - (l_{k_2} - \alpha_{k_2}x) = (l_{k_1} - l_{k_2}) - (\alpha_{k_1} - \alpha_{k_2})x$$

is decreasing in x whenever $\alpha_{k_1} > \alpha_{k_2}$. Hence, as x increases, points with larger α_k become less favorable, and the maximizer shifts toward smaller α_k . This is illustrated by Fig. 16:

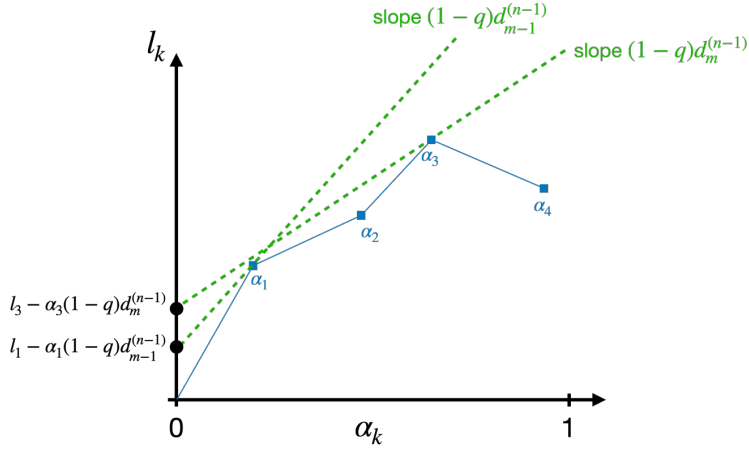


Figure 16: Graphical illustration of offer monotonicity

As the slope $(1-q)d_m^{(n-1)}$ increases, the maximizing point shifts toward smaller α_k . Since $\alpha_k = F(w_k)$ is increasing in w_k , this implies that the optimal offer $p(m, n)$ is weakly decreasing in $d_m^{(n-1)}$.

Combining this with Lemma 2, we obtain:

- Since $d_m^{(n-1)}$ is weakly decreasing in m , $p(m, n)$ is weakly increasing in m .
- Since $d_m^{(n-1)}$ is weakly increasing in n , $p(m, n)$ is weakly decreasing in n .
- Since $d_m^{(n)} \leq d_{m-1}^{(n-1)}$, we have $p(m, n+1) \geq p(m-1, n)$.

This concludes the proof. \square

\square

Proof of Proposition 1

Proof. Proof of Proposition 1. We divide the proof of Proposition 1 into two parts: (i) statements that follow directly from Theorem 2, and (ii) statements that require additional arguments. We first establish the former and then turn to the latter.

Part i: statements implied by Theorem 2. Recall that the width of the optimal pay sequence at state (m, n) is defined as the difference between the largest and smallest offers that can appear along an optimal path starting from (m, n) .

We first show that

$$b(m, n) = w_{k^*} - p(1, (n-m)^+ + 1).$$

Indeed, by Theorem 2, the optimal offer is weakly increasing in m and weakly decreasing in n . Along any feasible path from state (m, n) , n decreases by one in each stage and m may

stay the same or decrease by one (when an offer is accepted or cancelled). Hence, after each transition the supply-demand gap $n - m$ can never exceed its initial positive part $(n - m)^+$. Hence the smallest offer that can arise along an optimal path is attained at the reachable state with the smallest number of requests and the largest feasible supply-demand gap, namely

$$(1, (n - m)^+ + 1).$$

On the other hand, by Theorem 2, once the process reaches a state with demand weakly exceeding supply, the optimal offer is constant and equal to w_{k^*} . Therefore the largest offer that can appear in the optimal sequence is w_{k^*} . This proves that

$$b(m, n) = w_{k^*} - p(1, (n - m)^+ + 1).$$

The two claims in Part I now follow immediately.

First, if $m \geq n$, then $(n - m)^+ = 0$, so

$$b(m, n) = w_{k^*} - p(1, 1).$$

Since $m = 1 \geq 1$, Theorem 2 implies $p(1, 1) = w_{k^*}$. Hence

$$b(m, n) = 0.$$

Second, since $(n - m)^+$ enters the width formula only through the term

$$p(1, (n - m)^+ + 1),$$

and Theorem 2 implies that $p(1, n)$ is weakly decreasing in n , it follows that

$$b(m, n) = w_{k^*} - p(1, (n - m)^+ + 1)$$

is weakly increasing in $(n - m)^+$.

Part ii: statements that require additional proof. The remaining items to prove are (1) $b(m, n) = 0$ when cancellation probability $q \geq q_0$. (2) $b(m, n)$ is weakly decreasing in q

We start with item (1): Fix any $m \geq 1$ and $n \geq 1$. With cancellation probability $q \in [0, 1]$, the DP recursion Eq. (17) can be written in the following form

$$V(m, n) = V(m, n - 1) + q d_m^{(n-1)} + \max_{k \in [\bar{K}]} \left\{ l_k - (1 - q) \alpha_k d_m^{(n-1)} \right\}, \quad (28)$$

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

Hence the optimal first offer at (m, n) corresponds to

$$k \in \arg \max_{k \in [\bar{K}]} \left\{ l_k - (1 - q)\alpha_k d_m^{(n-1)} \right\}.$$

By the convex-hull result, the maximizer equals k^* whenever

$$(1 - q)d_m^{(n-1)} \leq l'_{k^*} \quad \text{for all } m, n. \quad (29)$$

Therefore, it suffices to show Eq. (29).

Step 1: reduce to bounding the $m = 1$ marginal. By Lemma 2(1), for each fixed $n - 1$, the marginal $d_m^{(n-1)}$ is weakly decreasing in m . Hence, for all $m \geq 1$,

$$d_m^{(n-1)} \leq d_1^{(n-1)} = V(1, n - 1) - V(0, n - 1) = V(1, n - 1). \quad (30)$$

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

Step 2: a uniform upper bound on $V(1, n)$. When there is only one request, the value function satisfies

$$V(1, n) = \max_{k \in [\bar{K}]} \left\{ l_k + (1 - q)(1 - \alpha_k)V(1, n - 1) \right\}.$$

Since $l_k \leq l_{k^*}$ and $(1 - \alpha_k) \leq 1$ for all k , we obtain the upper bound

$$V(1, n) \leq l_{k^*} + (1 - q)V(1, n - 1).$$

Starting from $V(1, 0) = 0$, iterating this recursion yields

$$V(1, n) \leq l_{k^*} \sum_{t=0}^{n-1} (1 - q)^t \leq \frac{l_{k^*}}{q}. \quad (31)$$

Step 3: conclude the sufficient condition. Combining Eq. (30) and Eq. (31), we obtain

$$(1 - q)d_m^{(n-1)} \leq (1 - q)V(1, n - 1) \leq (1 - q)\frac{l_{k^*}}{q}.$$

Therefore, if

$$(1 - q)\frac{l_{k^*}}{q} \leq l'_{k^*}, \quad (32)$$

then Eq. (29) holds for every (m, n) , so the maximizer in Eq. (28) is always k^* . Hence the optimal offer is constant and equals \bar{w} at every state.

Finally, Eq. (32) is equivalent to

$$q \geq \frac{l_{k^*}}{l_{k^*} + l'_{k^*}},$$

which gives the desired threshold condition.

Next, we prove item (2): By Part i, we have

$$b(m, n) = w_{k^*} - p(1, (n - m)^+ + 1).$$

Since w_{k^*} does not depend on q , it suffices to show that for each fixed $r \geq 1$, the optimal offer $p(1, r)$ is weakly increasing in q .

$$V(1, r) = \max_{k \in [\bar{K}]} \left\{ \alpha_k ((v - w_k) + V(0, r - 1)) \right. \\ \left. + (1 - \alpha_k)(qV(0, r - 1) + (1 - q)V(1, r - 1)) \right\}.$$

Since $V(0, r - 1) = 0$, this simplifies to

$$V(1, r) = \max_{k \in [\bar{K}]} \left\{ l_k + (1 - q)(1 - \alpha_k)V(1, r - 1) \right\},$$

where $\alpha_k = F(w_k)$ and $l_k = F(w_k)(v - w_k)$. Define

$$\lambda_r(q) = (1 - q)V(1, r - 1).$$

Then the Bellman equation can be rewritten as

$$V(1, r) = \max_{k \in [\bar{K}]} \left\{ l_k + (1 - \alpha_k)\lambda_r(q) \right\},$$

and the optimal offer at state $(1, r)$ is characterized by

$$p(1, r) \in \arg \max_{k \in [\bar{K}]} \left\{ l_k - \alpha_k \lambda_r(q) \right\}.$$

Thus, as in the proof of Theorem 2, the optimal offer is weakly decreasing in the scalar $\lambda_r(q)$.

It therefore remains to show that $\lambda_r(q)$ is weakly decreasing in q . We prove this by induction on r .

Base case. When $r = 1$, we have $V(1, 0) = 0$, so

$$\lambda_1(q) = (1 - q)V(1, 0) = 0,$$

which is constant in q .

Induction step. Suppose $\lambda_r(q)$ is weakly decreasing in q . Then

$$\lambda_{r+1}(q) = (1 - q)V(1, r) = (1 - q) \max_{k \in [\bar{K}]} \left\{ l_k + (1 - \alpha_k)\lambda_r(q) \right\}.$$

For each fixed k , the term

$$(1 - q) \left\{ l_k + (1 - \alpha_k) \lambda_r(q) \right\}$$

is weakly decreasing in q , since both $(1 - q)$ and $\lambda_r(q)$ are weakly decreasing in q , and all coefficients are nonnegative. Taking the maximum over k preserves monotonicity, so $\lambda_{r+1}(q)$ is weakly decreasing in q .

Hence, by induction, $\lambda_r(q)$ is weakly decreasing in q for every $r \geq 1$. Since $p(1, r)$ is weakly decreasing in $\lambda_r(q)$, it follows that $p(1, r)$ is weakly increasing in q . Therefore,

$$b(m, n) = w_{k^*} - p(1, (n - m)^+ + 1)$$

is weakly decreasing in q . \square

C Proofs: Optimal Pay with Unknown Global Factor

C.1 Exact DP approach

We first prove results regarding the exact DP approach to solve the belief-augmented MDP.

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 worker 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

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

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

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

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. \square

C.2 Analysis of Direct-Commit Policy

C.2.1 Auxiliary results

This subsection develops intermediate bounds that quantify the loss from committing to a misspecified full-information policy. We proceed in two steps. First, we establish a bound on the optimal value function with respect to the per-request value (Lemma 3), which is used to bound the loss from applying π^c under $C = 0$ (Proposition 6). Second, we analyze the reverse misspecification. We first bound the sensitivity of any fixed policy to changes in the per-request value (Lemma 4), and then apply this result to bound the loss from applying π^0 under $C = c$ (Proposition 7).

Throughout this subsection, we distinguish between value functions under different global factors and policies. For $r \in \{0, c\}$, let $V^r(m, n; v)$ denote the optimal value when the global factor is $C = r$ and the per-request value is v . For any admissible policy π , let $V^r(\pi; m, n; v)$ denote the value obtained by following policy π under $C = r$ and value v . When the dependence on v is clear from context, we suppress it for notational simplicity.

Lemma 3. *Fix $c > 0$. For each (m, n) , define the optimal value difference*

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

Then for all m, n ,

$$0 \leq \Delta^*(m, n) \leq c \min\{m, n\}. \quad (33)$$

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

```

1: function MAXN( $f_1, \dots, f_k$ )                                ▷ Upper envelope of  $k$  PWL functions on  $[0, 1]$ 
2:   if  $k = 1$  then return  $f_1$ 
3:   else if  $k = 2$  then return MAXTWO( $f_1, f_2$ )
4:   else return MAXN(MAXTWO( $f_1, f_2$ ),  $\dots$ , MAXTWO( $f_{k-1}, f_k$ ))
5:   end if
6: end function
7: function MAXTWO( $f, g$ )                                       ▷ Upper envelope of two PWL functions on  $[0, 1]$ 
8:   Input: Piecewise-linear functions  $f = (x^f, y^f)$  and  $g = (x^g, y^g)$            ▷ (breakpoints, values)
9:    $B \leftarrow [], V \leftarrow [],$                                ▷ Envelope breakpoints and values
10:   $i \leftarrow 1, j \leftarrow 1$ 
11:   $f\_larger \leftarrow (y_1^f \geq y_1^g)$ 
12:  while  $i < |x^f|$  and  $j < |y^f|$  do
13:    if  $f\_larger$  then ADDPOINT( $B, V, x_i^f, y_i^f, \varepsilon$ )
14:    else ADDPOINT( $B, V, x_j^g, y_j^g, \varepsilon$ )
15:    end if
16:     $(x_{int}, y_{int}) \leftarrow$  FINDINTERSECTION( $f, g, i, j, \varepsilon$ )   ▷ Find intersection of line segments using basic
    geometry.
17:    if  $inter \neq \text{NOTHING}$  then
18:      ADDPOINT( $B, V, x_{int}, y_{int}, \varepsilon$ )
19:       $f\_larger \leftarrow \neg f\_larger$ 
20:       $i \leftarrow i + 1, j \leftarrow j + 1$ 
21:    else
22:      if  $R_f < R_g - \varepsilon$  then  $i \leftarrow i + 1$ 
23:      else if  $R_g < R_f - \varepsilon$  then  $j \leftarrow j + 1$ 
24:      else  $i \leftarrow i + 1, j \leftarrow j + 1$ 
25:      end if
26:    end if
27:  end while
28:  Add the last point of  $f$  or  $g$  depending on which is larger. return  $(B, V)$ 
29: end function
30: function ADDPOINT( $B, V, x, y, \varepsilon$ )
31:  Only add  $x$  to  $B$  and  $y$  to  $V$  if  $x$  is sufficiently different (more than  $\varepsilon$ ) from the last  $x$  added to  $B$ .
32: end function

```

Proof. Proof of Lemma 3. 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. (33).

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

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

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

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

Moreover, define

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

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

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

Then for any fixed w ,

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

By the induction hypothesis,

$$0 \leq \Delta^*(m - 1, n - 1) \leq c \min\{m - 1, n - 1\}, \quad 0 \leq \Delta^*(m, n - 1) \leq c \min\{m, n - 1\}$$

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

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

Hence,

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

For the upper bound, it must hold that

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

Therefore,

$$\begin{aligned} \Delta^*(m, n) \leq \max_w \{ & F(w)(c + \Delta^*(m-1, n-1)) \\ & + (1 - F(w))((1 - q)\Delta^*(m, n-1) + q\Delta^*(m-1, n-1)) \} \end{aligned}$$

Applying the induction hypothesis, we have

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

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 to

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

This completes the induction. \square

\square

Next, we use Lemma 3 to bound the loss incurred when the platform applies the policy optimized for $C = c$ while the true global factor is $C = 0$.

Proposition 6. *Let π^c be the optimal policy computed under the assumption $C = c$ and per-request value v . Suppose the true global factor is $C = 0$. Then for any number of requests m and workers n ,*

$$V^0(m, n; v) - V^0(\pi^c; m, n; v) \leq c \min\{m, n\}, \quad (34)$$

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

Proof. Proof of Proposition 6.

The main idea of the proof is to relate the problem under $C = c$ and value v to the problem under $C = 0$ with value $v - c$, and show that

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

We can then apply Lemma 3 to obtain Eq. (34).

For a given pay w , let $Q(w; m, n \mid C = c, v)$ denote the Bellman objective evaluated at w , i.e., the expected value obtained by offering w in state (m, n) when the global factor is $C = c$ and the per-request value is v :

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

Define $x = w - c$, and interpret x as the offered wage in the corresponding problem with $C = 0$. Then the Bellman objective can be rewritten as

$$\begin{aligned} Q(w; m, n \mid C = c, v) &= F(x)(v - x - c) \\ &\quad + (F(x) + (1 - F(x))q)V^c(m - 1, n - 1; v) \\ &\quad + (1 - F(x))(1 - q)V^c(m, n - 1; v). \end{aligned}$$

Comparing this expression with the Bellman equation under $C = 0$ and per-request value $v - c$, we see that they coincide after identifying x as the offered wage in that problem. Therefore,

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

Moreover, the optimal policy under $C = c$ is obtained by shifting the optimal policy under $C = 0$ and value $v - c$ by c . In particular, the shifted policy $\pi^c - c$ is optimal for the problem with $C = 0$ and per-request value $v - c$.

Now consider $V^0(\pi^c; m, n; v)$. For any admissible policy π ,

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

This inequality holds because, for any fixed policy π , each offer is weakly more likely to be accepted under $C = 0$ than under $C = c$, since $F(w) \geq F(w - c)$. Since the payoff from an accepted request is the same in both environments, it follows that the expected value of following π is weakly higher under $C = 0$.

Applying this 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 3,

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

This concludes the proof. \square

\square

We now turn to the reverse setting, where the platform applies π^0 when the true global factor is $C = c$. Unlike the previous case, the analysis cannot be reduced directly to a value shift argument. Instead, we first establish a bound on the sensitivity of any fixed policy to changes in the per-request value.

Lemma 4. *Fix any admissible policy π under $C = 0$. Then for all $m, n \in \mathbb{Z}_+$ and all $c \geq 0$,*

$$V^0(\pi; m, n; v) - V^0(\pi; m, n; v - c) \leq c \min\{m, n\}. \quad (35)$$

Proof. Proof of Lemma 4. For fixed π , define

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

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

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 $\Delta^\pi(m, n) = 0 \leq c \min\{m, n\}$.

Induction step. Fix (m, n) with $m, n \geq 1$, and suppose $\Delta^\pi(m', n') \leq c \min\{m', n'\}$ 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 worker. Let w be the wage prescribed by π at state (m, n) . The fixed-policy recursion under $C = 0$ gives

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

Subtracting yields

$$\begin{aligned} \Delta^\pi(m, n) &= F(w)(c + \Delta^\pi(m - 1, n - 1)) \\ &\quad + (1 - F(w))((1 - q)\Delta^\pi(m, n - 1) + q\Delta^\pi(m - 1, n - 1)). \end{aligned}$$

By the induction hypothesis,

$$\Delta^\pi(m - 1, n - 1) \leq c \min\{m - 1, n - 1\} = c(\min\{m, n\} - 1),$$

and

$$\Delta^\pi(m, n - 1) \leq c \min\{m, n - 1\} \leq c \min\{m, n\}.$$

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

$$\begin{aligned} \Delta^\pi(m, n) &\leq F(w)(c + c(h - 1)) \\ &\quad + (1 - F(w))((1 - q)ch + qc(h - 1)) \\ &\leq ch \\ &= c \min\{m, n\}. \end{aligned}$$

This completes the induction and proves (35). \square

Next, we introduce Proposition 7 and its proof, leveraging the results from Lemma 4.

Proposition 7. *Let π^0 be the optimal policy computed under the assumption $C = 0$ and per-request value v . Suppose the true global factor is $C = c$. Define $w_{\min} = \min_k w_k$ and $\bar{v} = v - w_{\min}$. Then for any number of requests m and workers n ,*

$$V^c(m, n; v) - V^c(\pi^0; m, n; v) \leq \beta \min\{m, n\} \quad (36)$$

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

Proof. Proof of Proposition 7. We first note that

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

Indeed, under $C = c$, the lowest possible acceptable pay is $w_{\min} + c$, so the platform's profit from any accepted request is at most

$$(v - (w_{\min} + c))^+ = (\bar{v} - c)^+.$$

Moreover, at most $\min\{m, n\}$ requests can be served. Therefore,

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

which implies the claimed bound.

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

Step 1. Write

$$\begin{aligned} V^c(m, n; v) - V^c(\pi^0; m, n; v) &= (V^c(m, n; v) - V^0(m, n; v)) \\ &\quad + (V^0(m, n; v) - V^c(\pi^0; m, n; v)). \end{aligned}$$

By the shift identity,

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

so the first term becomes

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

where the inequality follows from monotonicity of $V^0(m, n; v)$ in v . Therefore,

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

Step 2. Define a transformed policy $\tilde{\pi}^0$ for the baseline environment $C = 0$ as follows: whenever π^0 prescribes wage w at a given state, $\tilde{\pi}^0$ prescribes wage $w - c$.

We claim that

$$V^c(\pi^0; m, n; v) = V^0(\tilde{\pi}^0; m, n; v - c). \quad (37)$$

To see this, note that under $C = c$, offering wage w results in acceptance probability $F(w - c)$. Under $C = 0$, offering wage $w - c$ yields the same acceptance probability $F(w - c)$. Moreover, conditional on acceptance, the instantaneous reward under $(C = c, v)$ with wage w equals $v - w$, while under $(C = 0, v - c)$ with wage $w - c$ it equals

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

Since the cancellation probability q is identical in both environments, it follows that the two systems induce the same distribution over transitions and the same reward along every realized path. This establishes (37).

Therefore,

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

Adding and subtracting $V^0(\tilde{\pi}^0; m, n; v)$ yields

$$\begin{aligned} V^0(m, n; v) - V^c(\pi^0; m, n; v) &= (V^0(m, n; v) - V^0(\tilde{\pi}^0; m, n; v)) \\ &\quad + (V^0(\tilde{\pi}^0; m, n; v) - V^0(\tilde{\pi}^0; m, n; v - c)). \end{aligned}$$

The first term is nonpositive because $V^0(m, n; v)$ is the optimal value under $(C = 0, v)$, so for any admissible policy π ,

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

and in particular for $\pi = \tilde{\pi}^0$. Hence,

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

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

Step 3. Applying Lemma 4 with $\pi = \tilde{\pi}^0$ yields

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

Combining this bound with the inequality established in Step 2 gives

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

as claimed. \square

\square

C.2.2 Proof of Theorem 3

Having established the bounds in Propositions 6 and 7, we are now ready to bound the performance of the Direct-Commit policy.

Proof. Proof of Theorem 3. 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

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

The first term is 0 since π^0 is optimal under $C = 0$, and by Proposition 7, the second term is at most

$$\mu_0 \beta \min\{m, n\}.$$

Hence, the regret from committing to π^0 is at most

$$\mu_0 \beta \min\{m, n\}.$$

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

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

The second term is 0 since π^c is optimal under $C = c$, and by Proposition 6, the first term is at most

$$(1 - \mu_0) c \min\{m, n\}.$$

Hence, the regret from committing to π^c is at most

$$(1 - \mu_0) c \min\{m, n\}.$$

By Definition 2, π^{DC} selects π^0 when $\mu_0 \leq \frac{c}{c+\beta}$ and π^c otherwise. Equivalently, it selects the policy corresponding to the smaller of the two bounds above. Therefore,

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

which proves the result. \square

\square

C.3 Analysis for the Probe-and-Commit Policy

We now provide proofs of results characterizing the performance of the probe-and-commit policy.

Proof. Proof of Theorem 4. We first prove the regret bound for a fixed probe value p . Then we show that minimizing the continuation term is equivalent to maximizing the function $G(p)$ when $\beta\mu_0 \geq c(1 - \mu_0)$ and constant otherwise.

Part I: Regret bound. Fix (m, n) , prior belief μ_0 , and probe value p . Let $V_p^{\text{probe}}(m, n)$ denote the expected value of a policy that is forced to offer p to the first worker and thereafter behaves clairvoyantly, i.e., optimally conditional on the realized global factor $C \in \{0, c\}$.

We decompose the regret of Probe-and-Commit as

$$V^{\text{clair}}(m, n) - V^{\pi^{\text{PC}}(p)}(m, n \mid \mu_0) = \left(V^{\text{clair}}(m, n) - V_p^{\text{probe}}(m, n) \right) + \left(V_p^{\text{probe}}(m, n) - V^{\pi^{\text{PC}}(p)}(m, n \mid \mu_0) \right). \quad (38)$$

Step 1: Immediate loss. Under the clairvoyant benchmark, the first worker can contribute at most \bar{v} in expected profit, since $\bar{v} = v - w_{\min}$ is an upper bound on the profit from any accepted request. Under the probe policy, the first-period expected profit is

$$P_{\text{accept}}(p, \mu_0)(v - p).$$

Therefore,

$$V^{\text{clair}}(m, n) - V_p^{\text{probe}}(m, n) \leq \bar{v} - P_{\text{accept}}(p, \mu_0)(v - p). \quad (39)$$

Step 2: Continuation loss. Let $Y \in \{A, R\}$ denote the probe outcome, and let (m_Y, n_Y) denote the remaining state after the probe and the possible cancellation realization. Thus,

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

Conditional on the realized outcome, both the forced-probe clairvoyant policy and Probe-and-Commit face the same continuation subproblem, characterized by the state (m_Y, n_Y) and posterior belief $\mu'_Y(p)$. The former continues with the clairvoyant-optimal policy, while the latter applies Direct-Commit.

Applying Theorem 3 to this continuation subproblem yields, for each realization,

$$V^{\text{clair}}(m_Y, n_Y \mid \mu'_Y(p)) - V^{\pi^{\text{DC}}}(m_Y, n_Y \mid \mu'_Y(p)) \leq L(\mu'_Y(p); c) \min\{m_Y, n_Y\}.$$

Since the above inequality holds for every realized continuation outcome, taking expectations over the probe outcome and the cancellation realization preserves the inequality. Hence,

$$V_p^{\text{probe}}(m, n) - V^{\pi^{\text{PC}}(p)}(m, n \mid \mu_0) \leq \mathbb{E}[L(\mu'_Y(p); c) \min\{m_Y, n_Y\}].$$

Finally, since $m_Y \leq m$ and $n_Y = n - 1$ in all cases, we have

$$\min\{m_Y, n_Y\} \leq \min\{m, n - 1\},$$

which implies

$$V_p^{\text{probe}}(m, n) - V^{\pi^{\text{PC}}(p)}(m, n \mid \mu_0) \leq \min\{m, n - 1\} \mathbb{E}[L(\mu'_Y(p); c)]. \quad (40)$$

Step 3: Combine. Combining (38), (39), and (40) gives

$$\text{Regret}(\pi^{\text{PC}}(p)) \leq (\bar{v} - P_{\text{accept}}(p, \mu_0)(v - p)) + \min\{m, n - 1\} \mathbb{E}[L(\mu'_Y(p); c)],$$

which proves Eq. (15). This concludes the proof for the regret bound.

Part II: minimizing continuation value. We next show the connection between the continuation term in (15) and the function $G(p)$. By the law of total probability,

$$\mathbb{E}[L(\mu'_Y(p); c)] = P_{\text{accept}}(p, \mu_0) L(\mu'_A(p); c) + (1 - P_{\text{accept}}(p, \mu_0)) L(\mu'_R(p); c).$$

Substituting $L(\mu; c) = \min\{\mu\beta, (1 - \mu)c\}$ and using the identity $a \min\{x, y\} = \min\{ax, ay\}$, we obtain

$$\begin{aligned} \mathbb{E}[L(\mu'_Y(p); c)] &= \min\{P_{\text{accept}}(p, \mu_0)\mu'_A(p)\beta, P_{\text{accept}}(p, \mu_0)(1 - \mu'_A(p))c\} \\ &\quad + \min\{(1 - P_{\text{accept}}(p, \mu_0))\mu'_R(p)\beta, (1 - P_{\text{accept}}(p, \mu_0))(1 - \mu'_R(p))c\}. \end{aligned}$$

By Bayes' rule,

$$P_{\text{accept}}(p, \mu_0)\mu'_A(p) = \mu_0 F(p - c), \quad P_{\text{accept}}(p, \mu_0)(1 - \mu'_A(p)) = (1 - \mu_0)F(p),$$

and similarly for rejection. Substituting yields

$$\mathbb{E}[L(\mu'_Y(p); c)] = \min\{\beta\mu_0 F(p - c), c(1 - \mu_0)F(p)\} + \min\{\beta\mu_0(1 - F(p - c)), c(1 - \mu_0)(1 - F(p))\}.$$

Let

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

From the derivation above,

$$\mathbb{E}[L(\mu'_Y(p); c)] = \min\{aF(p - c), bF(p)\} + \min\{a(1 - F(p - c)), b(1 - F(p))\}.$$

We next show that the continuation loss depends on p only through $G(p)$.

Let

$$a \triangleq \beta\mu_0, \quad b \triangleq c(1 - \mu_0), \quad G(p) \triangleq bF(p) - aF(p - c).$$

Starting from

$$\mathbb{E}[L(\mu'_Y(p); c)] = \min\{aF(p-c), bF(p)\} + \min\{a(1-F(p-c)), b(1-F(p))\},$$

we rewrite the first term as

$$\min\{aF(p-c), bF(p)\} = aF(p-c) + \min\{0, bF(p) - aF(p-c)\} = aF(p-c) + \min\{0, G(p)\}.$$

Similarly, for the second term,

$$\begin{aligned} \min\{a(1-F(p-c)), b(1-F(p))\} &= \min\{a - aF(p-c), b - bF(p)\} \\ &= -bF(p) + \min\{a + bF(p) - aF(p-c), b\} \\ &= -bF(p) + \min\{a + G(p), b\}. \end{aligned}$$

Combining the two expressions and using $aF(p-c) - bF(p) = -G(p)$, we obtain

$$\mathbb{E}[L(\mu'_Y(p); c)] = -G(p) + \min\{0, G(p)\} + \min\{a + G(p), b\}.$$

Equivalently,

$$\mathbb{E}[L(\mu'_Y(p); c)] = \min\{0, G(p)\} + \min\{a, b - G(p)\}.$$

We now analyze this expression by cases.

First consider the case $b \geq a$. Since $F(p) \geq F(p-c)$, we have

$$G(p) = bF(p) - aF(p-c) = (b-a)F(p) + a(F(p) - F(p-c)) \geq 0.$$

Hence

$$\min\{0, G(p)\} = 0,$$

and therefore

$$\mathbb{E}[L(\mu'_Y(p); c)] = \min\{a, b - G(p)\}.$$

It follows that the continuation loss is weakly decreasing in $G(p)$.

Next consider the case $a > b$. Then $b - a < 0$. We distinguish three subcases.

If $G(p) < b - a$, then $G(p) < 0$, so

$$\min\{0, G(p)\} = G(p).$$

Moreover,

$$b - G(p) > b - (b - a) = a,$$

so

$$\min\{a, b - G(p)\} = a.$$

Thus,

$$\mathbb{E}[L(\mu'_Y(p); c)] = a + G(p).$$

If $b - a \leq G(p) \leq 0$, then again

$$\min\{0, G(p)\} = G(p),$$

while

$$b - G(p) \leq b - (b - a) = a,$$

so

$$\min\{a, b - G(p)\} = b - G(p).$$

Hence,

$$\mathbb{E}[L(\mu'_Y(p); c)] = G(p) + b - G(p) = b.$$

Finally, if $G(p) > 0$, then

$$\min\{0, G(p)\} = 0,$$

and since $b - G(p) < b < a$, we have

$$\min\{a, b - G(p)\} = b - G(p).$$

Therefore,

$$\mathbb{E}[L(\mu'_Y(p); c)] = b - G(p).$$

Putting the three subcases together, when $a > b$,

$$\mathbb{E}[L(\mu'_Y(p); c)] = \begin{cases} a + G(p), & G(p) < b - a, \\ b, & b - a \leq G(p) \leq 0, \\ b - G(p), & G(p) > 0. \end{cases}$$

However, the region $G(p) < b - a$ cannot arise. Indeed,

$$G(p) = bF(p) - aF(p - c) \geq bF(p) - aF(p) = -(a - b)F(p) \geq b - a,$$

where the first inequality uses $F(p - c) \leq F(p)$. Hence only the latter two cases are feasible, and thus

$$\mathbb{E}[L(\mu'_Y(p); c)] = \begin{cases} b, & b - a \leq G(p) \leq 0, \\ b - G(p), & G(p) > 0. \end{cases}$$

Equivalently,

$$\mathbb{E}[L(\mu'_Y(p); c)] = b - (G(p))^+.$$

We have therefore shown that in both cases, $\mathbb{E}[L(\mu'_Y(p); c)]$ depends on p only through $G(p)$, and is weakly decreasing in $G(p)$. \square

D Proofs: Comparing Pay Strategies

Proof of Proposition 3

Proof. Proof. Under the no-heterogeneity assumption, each worker’s reservation wage is either 0 when $C = 0$ or c when $C = c$. Thus, under full information, the optimal payment is 0 if $C = 0$ and c if $C = c$.

We first show that $\pi^{PC}(0)$ is optimal. Consider using the first offer as a probe at $p = 0$. If the offer is accepted, then necessarily $C = 0$, since under $C = c$ a payment of 0 would be rejected. If the offer is rejected, then necessarily $C = c$. Hence the first worker’s response fully reveals the regime. Since $n > m$, one rejected probe does not create a capacity shortage: after at most one probe, the platform still has enough workers to implement the full-information optimal policy for all m requests. Therefore, Probe-and-Commit with probe $p = 0$ achieves the clairvoyant benchmark, and

$$\text{Regret}(\pi^{PC}(0)) = 0.$$

We next analyze DP-based Thompson Sampling. Under this policy, at each step the platform samples a regime from its current posterior and offers the corresponding full-information optimal payment: 0 if it samples $C = 0$, and c if it samples $C = c$.

In this setting, offering $p = 0$ is always informative: acceptance reveals $C = 0$, while rejection reveals $C = c$. In contrast, offering $p = c$ always leads to acceptance regardless of the true state, and is therefore uninformative. Thus, conditional on the posterior still being unresolved, the only way to remain in the same informational state in the next period is to sample $C = c$ and offer $p = c$, which occurs with probability μ_0 .

An uninformative acceptance is costly only when the true state is $C = 0$. This event occurs with probability

$$p_\epsilon = \mu_0(1 - \mu_0),$$

and incurs regret c , since the clairvoyant benchmark would have paid 0 instead of c .

Let R_k denote the expected regret of Thompson Sampling when k requests remain and the posterior is still equal to the prior. By the discussion above, with probability p_ϵ the platform incurs immediate regret c , and with probability μ_0 the posterior remains unresolved so the platform continues with value R_{k-1} . Therefore,

$$R_k = c p_\epsilon + \mu_0 R_{k-1}, \quad R_0 = 0.$$

Iterating this recursion yields

$$R_m = c p_\epsilon \sum_{t=0}^{m-1} \mu_0^t = c p_\epsilon \cdot \frac{1 - \mu_0^m}{1 - \mu_0}.$$

Hence,

$$\text{Regret}(\pi^{TS}) = c p_\epsilon \cdot \frac{1 - \mu_0^m}{1 - \mu_0}.$$

This completes the proof. \square

Proof of Proposition 4

Proof. Proof. Suppose that $\mu_0 \leq \frac{c}{c+\beta}$, i.e., $\pi^{\text{DC}} = \pi^0$. Then we observe

$$\text{Regret}(\pi^{\text{DC}}) = \mu_0 [V^c(m, n) - V^c(\pi^0; m, n)].$$

The maximum pay offer under π^0 is w_K which is less than $w_1 + c$ since $c > w_K - w_1$. Therefore, applying π^0 under $C = c$ means no offer is accepted and $V^c(\pi^0; m, n) = 0$. Now let $\alpha = \max_{k \in [K]} F(w_k)(v - c - w_k)$. We can lower bound $V^c(m, n)$ as follows: if $m \geq n$, then the optimal pay policy is myopic, and $V^c(m, n) = n \times \max_{k \in [K]} F(w_k)(v - c - w_k) = n\alpha$. If $m < n$, we can use value function monotonicity to say $V^c(m, n) \geq V^c(m, m) = m\alpha$. Thus, in this case

$$\text{Regret}(\pi^{\text{DC}}) \geq \mu_0 \min\{m, n\}\alpha.$$

Now suppose that $\mu_0 > \frac{c}{c+\beta}$, i.e., $\pi^{\text{DC}} = \pi^c$. Then we can write

$$\text{Regret}(\pi^{\text{DC}}) = (1 - \mu_0) [V^0(m, n) - V^0(\pi^c; m, n)].$$

The optimal value function when $C = 0$ must exceed the value obtained from offering every worker w_K , i.e., $V^0(m, n) \geq (v - w_K) \min\{m, n\}$. Meanwhile, applying π^c when $C = 0$ means every pay offer is at least $w_1 + c$ (and is accepted). Thus, we can write $V^0(\pi^c; m, n) \leq (v - w_1 - c) \min\{m, n\}$. Thus, we observe

$$\text{Regret}(\pi^{\text{DC}}) \geq (1 - \mu_0) \min\{m, n\}(v - w_K - (v - w_1 - c)) = (1 - \mu_0) \min\{m, n\}(c - (w_K - w_1)).$$

Thus, by Corollary 1, if $\min\{m, n\} > \frac{\bar{v} - (1 - \mu_0)(v - w_K)}{\delta(\mu_0)}$, where

$$\delta(\mu_0) = \begin{cases} \mu_0 \alpha & , \text{ if } \mu_0 \leq \frac{c}{c+\beta} \\ (1 - \mu_0)(c - (w_K - w_1)) & , \text{ if } \mu_0 > \frac{c}{c+\beta}, \end{cases}$$

then $\text{Regret}(\pi^{\text{DC}}) > \text{Regret}(\pi^{\text{PC}})$ \square

E Proofs: Value of Flexibility

E.1 Optimal Static Pay

We first present proofs of the results regarding the optimal static pay policy.

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 $V_k(\mu_0, m, n)$ designate the value of this fixed-pay policy for a particular m , n , and initial belief μ_0 . Clearly, we can write

$$V_k(\mu_0, m, n) = (1 - \mu_0)V_k(0, m, n) + \mu_0 V_k(1, m, n),$$

since with probability μ_0 , the global factor verifies $C = c$ and we will obtain value $V_k(1, m, n)$, and with probability $(1 - \mu_0)$, the global factor verifies $C = 0$ and we will obtain value $V_k(0, m, n)$. Both cases correspond to a known global factor value. Thus, for the rest of this proof we study $V_k(m, n) = V_k(0, m, n)$, with the understanding that the same results apply to $V_k(1, m, n)$.

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:

$$V_k(m, n) = \begin{cases} l_k \left[m \frac{1 - a_k^n}{1 - a_k} + \sum_{i=1}^{m-1} b_i(m, n) a_k^{n-i} \right] & \text{if } m < n \\ n l_k & \text{if } m \geq n \end{cases} \quad (41)$$

where

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

We begin by proving the second case by induction over n . First, observe that if $n = 1$, for any $m \geq 1$ 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 $V_k(m, n) = n l_k$ for some n and all $m \geq n$. Then let $m \geq n + 1$ (meaning $m - 1 \geq n$) 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 (41). 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

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

which matches (41) 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

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

which proves the base case $n = 3$. Now assume the claim is true for $n \geq 2$. 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 (41) holds for every (m_0, n) with $m_0 \leq m - 1$. We first check the base case $m = n$:

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

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 $n + 1$:

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

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 (41). 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:

$$V_k(m, \infty) := \lim_{n \rightarrow \infty} V_k(m, n) = m l_k \frac{1}{1 - a_k} = \frac{m F(w_k)(v - w_k)}{1 - (1 - q)(1 - F(w_k))} = m V_k(1, \infty).$$

Due to the discrete worker types, the only values of p that can maximize the equation in Lemma 1 are $\cup_{k \in [K]} \{w_k, w_k + c\}$. We denote these allowed pay levels by $\{\tilde{w}_k\}_{k \in [2K]}$, where $\tilde{w}_k = w_k$ for $k \in [K]$ and $\tilde{w}_k = w_{k-K} + c$ for $k \in \{K + 1, \dots, 2K\}$.

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

$$(1 - \mu_0) \frac{F(\tilde{w}_{k_s})(v - \tilde{w}_{k_s})}{1 - (1 - q)(1 - F(\tilde{w}_{k_s}))} + \mu_0 \frac{F(\tilde{w}_{k_s} - c)(v - \tilde{w}_{k_s})}{1 - (1 - q)(1 - F(\tilde{w}_{k_s} - c))} \\ > (1 - \mu_0) \frac{F(\tilde{w}_k)(v - \tilde{w}_k)}{1 - (1 - q)(1 - F(\tilde{w}_k))} + \mu_0 \frac{F(\tilde{w}_k - c)(v - \tilde{w}_k)}{1 - (1 - q)(1 - F(\tilde{w}_k - c))} + \varepsilon$$

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

$$\left| m \frac{F(w_k)(v - w_k)}{1 - (1 - q)(1 - F(w_k))} - V_k(m, n) \right| < \frac{\varepsilon}{2},$$

and similarly for $V_k(1, m, n)$, which completes the proof. \square

Proof of Corollary 2

Proof. Proof. The first result follows immediately from setting $q = 1$ in Lemma 1. For the second, observe that when $q = 0$, the right-hand side of equation 16 becomes

$$\max_p (1 - \mu_0)(v - p) + \mu_0(v - p) \mathbb{1}[F(p - c) > 0],$$

or equivalently

$$\max_p (1 - \mu_0)(v - p) + \mu_0(v - p) \mathbb{1}[p \geq w_1 + c].$$

There are two possible maximizers, $p = w_1$, and $p = w_1 + c$, so the optimal profit is either $(1 - \mu_0)(v - w_1)$ or $(1 - \mu_0)(v - w_1 - c) + \mu_0(v - w_1 - c) = v - w_1 - c$. Therefore,

$$p = w_1 \Leftrightarrow (1 - \mu_0)(v - w_1) > v - w_1 - c \Leftrightarrow \mu_0(v - w_1) < c \Leftrightarrow \mu_0 < \frac{c}{v - w_1}.$$

\square

Proof of Proposition 5

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

$$\frac{F(w_{k_s})(v - w_{k_s})}{1 - (1 - q)(1 - F(w_{k_s}))} > \frac{F(w_k)(v - w_k)}{1 - (1 - q)(1 - F(w_k))}, \quad (42)$$

but

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

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 (42) to lower bound the right-hand side we obtain

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

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

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

which violates the monotonicity of the optimal value function. \square

\square

E.2 Value of Flexibility

Proof of Theorem 5

Proof. **Statement 1.** The first statement in the proposition follows immediately from the definition of the single-pay and optimal policies.

Statement 2, part 1. 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:

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

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

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

which completes the proof of the second statement.

Statement 2, part 2. 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 worker, then only offers w_{k^*} if the first request does not cancel, otherwise offers the optimal pay sequence with $m - 1$ requests and $n_0 - 1$ workers. 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 $\text{VoF}(m, n_0) > 0$.

Statement 3. 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 workers 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.,

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

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

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

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

$$\begin{aligned} \frac{PI(m, n)}{m} - \frac{S(m, n)}{m} \geq \min & \left(\mu_0 \left(\frac{F(w_c - c)(v - w_c)}{1 - (1 - q)(1 - F(w_c - c))} \right), \right. \\ & \left. (1 - \mu_0) \left(\frac{F(w_0)(v - w_0)}{1 - (1 - q)(1 - F(w_0))} - (v - w_c) \right) \right). \end{aligned} \quad (43)$$

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

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

which for n_0 large enough and $n \geq n_0$, we know from Theorem 5 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

$$\begin{aligned}
PI(m, n) - V^*(m, n) &\leq (1 - \mu_0) \left[\frac{F(w_0)(v - w_0)}{1 - (1 - q)(1 - F(w_0))} - (m - 1)\varepsilon - (v - w_K) \right] \\
&\quad + \mu_0 \left[q \frac{F(w_c - c)(v - w_c)}{1 - (1 - q)(1 - F(w_c - c))} - (m - q)\varepsilon \right] \\
&\leq (1 - \mu_0) \left[\frac{F(w_0)(v - w_0)}{1 - (1 - q)(1 - F(w_0))} - (v - w_K) \right] \\
&\quad + \mu_0 \left[q \frac{F(w_c - c)(v - w_c)}{1 - (1 - q)(1 - F(w_c - c))} \right]. \quad (44)
\end{aligned}$$

Putting (43) and (44) together, we obtain

$$\begin{aligned}
V^*(m, n) - S(m, n) &\geq m \cdot \min \left(\mu_0 \left[\frac{F(w_c - c)(v - w_c)}{1 - (1 - q)(1 - F(w_c - c))} \right], \right. \\
&\quad \left. (1 - \mu_0) \left[\frac{F(w_0)(v - w_0)}{1 - (1 - q)(1 - F(w_0))} - v + w_c \right] \right) \\
&\quad - (1 - \mu_0) \left[\frac{F(w_0)(v - w_0)}{1 - (1 - q)(1 - F(w_0))} - (v - w_K) \right] - \mu_0 q \frac{F(w_c - c)(v - w_c)}{1 - (1 - q)(1 - F(w_c - c))}. \quad (45)
\end{aligned}$$

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. \square

Proof of Corollary 3

Proof. Proof. Let $q = 0$. Then assuming $F(w_0 = 0) > 0$, then $w_0 = 0$, $w_c = c$, and Equation (45) simplifies to

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

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

$$\begin{aligned} \frac{m\mu_0(v-c)}{\mu_0(v-c) + (1-\mu_0)c} &> 1 \\ m\mu_0(v-c) - \mu_0(v-c) + \mu_0c &> c \\ \mu_0 &> \frac{c}{c + (m-1)(v-c)}. \end{aligned}$$

□