A Markovian Employment Model with Hostility, Severance, and Surplus

Marcus Lisman

April 1, 2025

We develop a multi-period employment model in which a firm repeatedly observes a two-state productivity process ($\alpha_t \in \{H, L\}$) governed by Markov parameter p. The worker's outside utility is u_0. The firm must pay a wage $w \geq u_0$. If the firm fires the worker, it pays an exogenous severance $\bar{s}$. If $\bar{s} < k$ (where k is the worker's firing cost), the worker becomes *hostile*, imposing an additional cost $h = (k - \bar{s})$ *on the firm*. We present a multi-period (and infinite-horizon) profit function, as well as a social-surplus function using a discount factor δ.

1 Introduction

In many labor markets, a firm is forced to consider whether to keep or fire a worker whose productivity fluctuates. Workers may suffer significant personal costs upon dismissal, but it is the *firm* that must absorb any hostility cost h if severance $\bar{s}$ is insufficient to cover the worker's cost k. Consequently, the firm's decision to fire a low-productivity worker is tempered by the possibility of paying hostility. We show how to write out a multi-period or infinite-horizon formulation, define the firm's profit function, and measure total surplus with discounting.

2 Model Setup

2.1 States and Markov Productivity

We have an infinite set of discrete periods $t = 0, 1, 2, \ldots$ (or we can truncate at T for a finite-horizon scenario). Each period:

$$\alpha_t \in \{H, L\},$$

with a Markov transition parameter $p \in [0.5, 1]$:

$$\Pr(\alpha_{t+1} = \alpha_t) = p, \quad \Pr(\alpha_{t+1} \neq \alpha_t) = 1 - p.$$

Hence α_t is High or Low. We allow $p = 0.5$ to represent an i.i.d. process, and p close to 1 for a highly persistent process.

2.2 Wage, Outside Utility, and Worker's Firing Cost k

- $\mathbf{w} \geq u_0$: The worker's wage. We assume the worker's outside option is u_0, so the firm must pay at least u_0 for the worker to accept the job.
- $\mathbf{k}$: The worker's cost if fired. For simplicity, we treat k as known; or we imagine a fraction q of workers have a higher k. We can also make k unknown to the firm that it will have to learn over time.

2.3 Exogenous Severance $\bar{s}$ and Hostility Cost to the Firm

If the firm fires the worker and $\bar{s} < k$, the worker's net shortfall is $(k - \bar{s})$. We interpret that as the worker's hostility, which directly translates into a cost to the *firm*, denoted

$$h = (k - \bar{s}) \quad (\text{if } \bar{s} < k),$$

and $h = 0$ if $\bar{s} \geq k$. That is, if severance meets or exceeds the worker's cost of job loss, no hostility arises. If severance is insufficient, the firm pays the price through hostility.

If $\bar{s} \geq k$, firing yields zero cost (other than the forgoing of production). If $\bar{s} < k$, the firm pays the difference $\big(k - \bar{s}\big)$. This hostility is an intangible friction inflicted on the firm.

2.4 Per-Period Payoffs and Discounting

We let $\delta \in (0, 1)$ denote the discount factor for future payoffs. In each period:

$$\text{Firm payoff if keep} = \alpha_t - w, \quad \text{Worker payoff if keep} = w.$$

$$\text{Firm payoff if fire} = 0 \; - \; \underbrace{(k - \bar{s})}_{\text{if } \bar{s}<k} \; = \; - \max\{0,\; k - \bar{s}\}, \quad \text{Worker payoff if fire} = \; \underbrace{\bar{s} - k}_{\leq 0 \text{ if } \bar{s}<k} \; .$$

For simplicity, we do not add a hiring cost or firing cost to the firm beyond hostility.

3 Multi-Period Profit Function

We define the firm's problem as a dynamic choice each period: *keep* or *fire*. If keep, the firm's immediate payoff is $\alpha_t - w$. If fire, the immediate payoff is $-\max\{0,\; k - \bar{s}\}$ and a new hire might come next period with productivity drawn from the same Markov chain (or you can consider continuing the same chain). For an infinite horizon, let

$$V(\alpha) = \max\big\{(\alpha - w) + \delta\, \mathbb{E}[V(\alpha') \mid \alpha], \;\; - \max\{0, k - \bar{s}\} + \delta\, \mathbb{E}[V(\alpha_{\text{new}})]\big\},$$

where:

$$\mathbb{E}[V(\alpha') \mid \alpha] = \Big\{ p\, V(\alpha) + (1 - p)\, V(\text{other state}), \quad \text{since } \alpha' \text{ can stay or flip},$$

and α_{new} is the new worker's state if we assume re-entry at some baseline distribution. The key is that if the firm fires, the immediate payoff is $\underbrace{0}_{\text{no output}} - \max\{0, k - \bar{s}\}$.

Hence, the firm's infinite-horizon expected profit from state α is $V(\alpha)$. One can solve this system by standard dynamic programming for $\{V(H), V(L)\}$. For instance:

$$V(H) = \max\Big\{(H - w) + \delta\big[p\,V(H) + (1-p)\,V(L)\big], -\max\{0, k - \bar{s}\} + \delta\,\mathbb{E}[V(\alpha_{\text{new}})]\Big\},$$
$$V(L) = \max\Big\{(L - w) + \delta\big[(1-p)\,V(H) + p\,V(L)\big], -\max\{0, k - \bar{s}\} + \delta\,\mathbb{E}[V(\alpha_{\text{new}})]\Big\}.$$

We see that if $\bar{s} \geq k$, the fire branch has no hostility cost ($= 0$), making it cheaper to separate. If $\bar{s} < k$, the firm must pay $(k - \bar{s})$ in hostility whenever it fires. The wage w is typically pinned down by requiring $w \geq u_0$ so that the worker accepts ex ante.

4 Conditional Profit Function $\pi(\alpha \mid u_0, k, q, p)$

We can also denote the *firm's single-period* profit or the entire multi-period value as

$$\pi(\alpha \mid u_0, k, q, p),$$

emphasizing that k, q, p, u_0 shape the environment. Typically:

- p sets how likely we stay in α or switch.
- u_0 sets a lower bound on w (the wage).
- k might be the cost to the worker if fired, but it becomes a cost $k - \bar{s}$ to the firm if $\bar{s} < k$.
- q is relevant if we have multiple classes of workers or want an expected payoff across k distributions.

Hence we might interpret $\pi(\cdots)$ as an expected or realized payoff in a given scenario.

5 Social Surplus: Discounted Welfare Computation

We define the social surplus (or total welfare) as the discounted sum of:

(Output) − (Worker's outside alternative) − (hostility cost paid by firm if any).

One approach is:

$$\text{Social Surplus} = \sum_{t=0}^{\infty} \delta^t \big[\alpha_t - u_0 - H_t\big],$$

where H_t is the hostility cost *the firm* must pay in period t if it fires insufficiently. Specifically:

$$H_t = \begin{cases} k - \bar{s}, & \text{if fired at } t \text{ and } \bar{s} < k, \\ 0, & \text{otherwise.} \end{cases}$$

We subtract the worker's fallback u_0 each period because from a societal viewpoint, $\alpha_t - u_0$ is the net benefit of the job relative to the worker's next-best option. (Alternatively, one might define surplus as simply α_t, or incorporate more detailed outside-option references. The main point is that if the worker is not employed, their utility is u_0, so the net gain from employing them is output minus that fallback.)

Hence, if the firm never fires, the stream of surplus is $\sum_{t=0}^{\infty} \delta^t[\alpha_t - u_0]$. If the firm does fire at time τ, we add a penalty $\delta^\tau(k - \bar{s})$ if $\bar{s} < k$. We can do a dynamic programming approach on surplus as well:

$$W(\alpha) = \max\Big\{[\alpha - u_0] + \delta\,\mathbb{E}[W(\alpha') \mid \alpha], -\mathbf{1}_{\{\bar{s}<k\}}(k - \bar{s}) + \delta\,\mathbb{E}[W(\alpha_{\text{new}})]\Big\}.$$

We can compare the firm's private decision rule to the surplus-maximizing rule to see if there is any divergence, especially if $\bar{s} < k$ is large.

6 Role of Hostility in the Model

In this section, we go into the role of hostility:

6.1 1) Incentive to Avoid Firing if $\bar{s} < k$

Hostility arises exactly when $\bar{s} < k$. Thus, if the worker is in a low state $\alpha = L$, the firm might consider separation, but it realizes that by firing it must pay $\max\{0, k - \bar{s}\}$. If $\bar{s} < k$, the firm's short-run payoff from firing is negative. This can deter the firm from firing immediately, especially if there is a reasonable chance the worker returns to $\alpha = H$ soon (p is low or moderate). Conversely, if p is near 1 and the worker is stuck in low state L, the firm may choose to fire anyway, absorbing the hostility cost once.

6.2 2) Worker's Firing Cost vs. Actual Implementation

While the worker's personal cost is k, in this model, the cost is *passed on to the firm* if severance is insufficient. That is, $\bar{s} < k \implies$ the firm pays $(k - \bar{s})$ in intangible or direct losses from a hostile departure. This arrangement can be interpreted as:

- Legal or reputational damage to the firm if it fires at a severance below the worker's legitimate cost k.

Hence hostility is a penalty on the firm for failing to meet the worker's firing cost in severance.

6.3 3) Effect on Profit and Surplus

1. Profit: If the firm is near indifferent about continuing a low-productivity worker, a large hostility cost (big $k - \bar{s}$) might tilt the balance to keep. The firm's optimal policy might be forbear firing unless the expected productivity is so poor that it's worth incurring hostility now.

2. Social Surplus: Because hostility is purely a cost to the firm, but not necessarily recouped by the worker, it's effectively a net efficiency loss. The difference $(k - \bar{s})$ is lost. This reduces total welfare if the firm chooses to fire with insufficient severance.

6.4 4) Parameter Sweeps in Hostility Context

When you vary $(p, k, \bar{s}, u_0)$:

- If $\bar{s}$ is always above k, no hostility arises. The firm can separate at will, making the model akin to frictionless firing.
- If $\bar{s} < k$, hostility triggers whenever the firm dismisses the worker. We can examine how large $k - \bar{s}$ must be before the firm simply keeps the worker in L for many periods (especially if p is small).

7 Q-Learning Implementation with Hostility

The hostility cost factor changes the *reward* in an RL or Q-learning algorithm for the firm. Specifically, if the firm chooses fire when $\bar{s} < k$:

$$r_t = -\big(k - \bar{s}\big).$$

Thus, the Q-values incorporate that negative payoff. Over repeated episodes, the firm's policy emerges as balancing potential future productivity gains vs. immediate hostility hits.

8 Q-Learning Framework

We next show how Q-learning might be applied:

8.1 1) Firm Learning Across Many Worker Types

Suppose the *firm* is the single learner (RL agent). The worker's parameters (k, u_0) might vary across episodes, or we have a fraction q with big k. In each episode:

(The state is $\alpha_t \in \{H, L\}$.

(The action is {keep, fire}.

(The reward (to the firm) is:

$$r_t = \begin{cases} \alpha_t - w, & \text{if keep,} \\ -\max[0, k - \bar{s}], & \text{if fire.} \end{cases}$$

(The next state α_{t+1} is drawn from a Markov chain with parameter p.

We use Q-learning updates

$$Q_{\text{new}}(\alpha, a) = (1 - \alpha_Q)\, Q_{\text{old}}(\alpha, a) + \alpha_Q \big[r_t + \delta \max_{a'} Q_{\text{old}}(\alpha_{t+1}, a')\big].$$

By repeatedly simulating episodes (where the fraction q of episodes has large k), the firm learns an approximate keep/fire policy for each α that maximizes its expected discounted payoff, capturing the hostility cost if $\bar{s} < k$.

8.2 Parameter Sweeps

Having set up Q-learning, we can sweep across:

- **k**: The firing cost to the worker, e.g. from 0 up to some k_{max}.
- **q**: The fraction of high-k workers.
- **p**: The Markov parameter from 0.5 (i.i.d.) to near 1 (very persistent).
- $\mathbf{u_0}$: The outside utility; if it is large, the wage is higher.
- $\bar{\mathbf{s}}$: The severance level. If $\bar{s} < k$, hostility cost arises.

After each parameter combination, we can run Q-learning for many episodes and measure:

- How often the firm chooses to fire in Low states.
- The fraction of episodes with hostility cost.
- The firm's average discounted profit (approx. $V(H)$ or $V(L)$).
- The social surplus, comparing actual keep/fire decisions vs. the notional welfare-maximizing policy.

Such sweeps highlight the interplay among $\{k, \bar{s}, p, q, u_0\}$. For instance:

- If p is small (0.5), the chain is nearly i.i.d., so even if the worker is in L, we might keep them hoping for H next period.
- If $\bar{s} < k$, frequent hostility arises upon firing. Does that deter the firm from firing when L persists?

9 An Experiment with Unknown k, q

A further extension is to assume the firm *does not know* (a) the workers cost k, or (b) the fraction q of high-cost workers. Then:

- Motivation: In real labor markets, an employer rarely knows each workers exact cost of unemployment. Some portion of the workforce might be extremely costly to fire, but the firm cannot observe it ex ante.
- Learning Setup:
  1. The environment draws a workers k from a distribution: with prob. q, $k > \bar{s}$ (high-cost), else $k \leq \bar{s}$ (low-cost).
  2. The firm sees only the productivity $\{H, L\}$ each period; it chooses keep or fire.
  3. If it fires, it *discovers* whether $k > \bar{s}$ or not *ex post*, from whether hostility cost arises.

4. Over many repeated episodes (or a single long run with repeated new hires), the firm updates its Q-values about how often and how large a penalty $k - \bar{s}$ might be.

- Value: This addresses partial ignorance. The firms policy emerges from trial and error. Early on, the firm may test firing in Low states more often, occasionally incurring big hostility. Over time, if hostility is frequent, the Q-learning algorithm shifts to more retention. If hostility is rare, it learns that $\bar{s} \geq k$ for most workers and fires freely.

- Experiment: We can run:

  1. **Parameter sweeps** in (q, p), the fraction of high-cost workers and the Markov stickiness.
  2. **Monitor how quickly** the Q-learner converges to a stable keep/fire policy.
  3. **Evaluate profit and the frequency of hostility** for each combination.

  This experiment highlights the advantage of model-free RL in uncertain firing-cost environments.

Hence, letting k, q be unknown ex ante adds a realistic dimension to the model, ensuring the firm must discover how often $\bar{s} < k$ occurs by actually risking hostility. The resulting learned policy can deviate from a frictionless approach or from the case where the firm had full information from the start.

10 An Experiment with Unknown p (Maybe?)

10.1 Motivation

In many real-world labor scenarios, the firm does not precisely know how persistent the workers productivity state is. It only sees, period to period, whether productivity stays the same or changes. If p^* is near 1, once a worker hits L, they remain L with high probability, encouraging early firing if hostility cost is not too large. But if p^* is near 0.5, L might flip to H next period quite often, so the firm might tolerate L episodes. Without direct knowledge of p^*, the firm must learn from observations.

10.2 RL Implementation

In a Q-learning approach:

- The **environment** draws α_{t+1} from the *true* p^* each step, but the firm does *not* know p^*.

- The **firms RL** states are still $\{H, L\}$ plus possibly an internal dimension for the firms current guess $\hat{p}$. Alternatively, the firm can keep a separate tracker for p, or just let the Q-table converge from repeated trial and error.

- The **reward** each period:

$$r_t = \begin{cases} \alpha_t - w, & \text{if keep,} \\ -\max\{0, k - \bar{s}\}, & \text{if fire.} \end{cases}$$

- Over many episodes, the firm sees how often $\alpha_{t+1} = \alpha_t$ or not, refining $\hat{p}$ implicitly. If it sees that Low states remain Low an unexpectedly large fraction of the time, it learns (by negative or forgone payoffs) that p must be high.

Hence, the keep/fire policy emerges from the RL updates without the firm ever seeing p directly. This is a scenario of *partial ignorance* about the Markov transition.

10.3 Parameter Sweeps and Observed Behavior

One can run repeated simulations for various *true* $p^* \in \{0.5, \ldots, 1\}$ and see:

- How quickly the firms Q-values adapt to the correct keep/fire threshold.
- Whether early episodes feature too many firings or too few, depending on the mismatch between the firms initial guess and the real p^*.
- The frequency of hostility if $\bar{s} < k$ is also relevant: if the firm tries firing early and learns p^* is large, it might still keep the worker next time to avoid repeated hostility.

Thus, unknown p fosters more realistic learning by observing transitions. As the agent updates its policy, it effectively discovers the environments persistence level.

11 Conclusion

We have formulated the model to ensure:

- **Hostility cost h is paid by the firm**, whenever $\bar{s} < k$. This means the firm directly suffers an additional cost $(k - \bar{s})$ if it dismisses a worker who has a cost k above the severance level.
- **Multi-Period and Infinite Profit Functions**: We spelled out the infinite-horizon Bellman equations:

  $$V(\alpha) = \max\Big\{(\alpha - w) + \delta\,\mathbb{E}[V(\alpha')],\ -\max[0, k - \bar{s}] + \delta\,\mathbb{E}[V(\alpha_{\text{new}})]\Big\},$$

  showing how the firm decides keep vs. fire each period.
- **Social Surplus with Discounting**:

  $$W(\alpha) = \max\Big\{[\alpha - u_0] + \delta\,\mathbb{E}[W(\alpha')],\ -\mathbf{1}_{\{\bar{s}<k\}}\,(\,k - \bar{s}\,) + \delta\,\mathbb{E}[W(\alpha_{\text{new}})]\Big\},$$

  or an equivalent summation $\sum_{t=0}^{\infty} \delta^t [\,\alpha_t - u_0 - H_t\,]$.

Such a framework illustrates how high persistence ($p \approx 1$) encourages firing in the Low state (since $\alpha = L$ will likely remain L), while with i.i.d. ($p = 0.5$), the firm might keep the worker hoping to flip to High next period. Likewise, if $\bar{s}$ is small and k is large, the firm triggers hostility cost upon firing, which may or may not be worth it.

This model opens the door to analyzing:

- When does the firm's private keep/fire decision align with surplus maximization?
- How does the fraction q of high-cost workers shape equilibrium turnover rates?
- Does raising $\bar{s}$ reduce hostility but also cause other inefficiencies?
- Is a certain level of hostility necessary for optimized total efficiency?

The model also:

- Discourages rash separations in low-productivity states if the hostility cost is sufficiently large.
- Creates an explicit difference between the firms firing decisions and the frictionless case where $\bar{s} \geq k$.
- Affects both firms private profit and overall social surplus (since $(k - \bar{s})$ is lost upon separation).

Additionally, we could have a section in my thesis comparing RL to the analytical dynamic programming solution if we assume full knowledge. Then we can run Q-learning with no knowledge or partial knowledge, and compare the final policy to the DP optimum. This quantifies how ignorance plus explorationexploitation constraints might cause differences from the fully informed solution and tie back into asymmetric information that the Cary paper discusses.