Evolution of Labour Market Mismatch through Adaptive Learning with Experience

Siting Estee Lu*

April 28, 2025

This is a PRELIMINARY draft.
[CLICK HERE FOR LATEST VERSION]

Abstract

Workers' past application choices can act as heuristics for future decisions. By integrating learning theory into search model, this work explores the role of experiences on workers' application choices. It provides an evolutionary perspective to the labour market dynamics, and offers some insights on equilibrium selection. I propose two market structures, where wages are unobservable and observable before workers make application decisions, and I model workers' learning behaviour using reinforcement learning and best response dynamics respectively. I show that in presence of multiple equilibria, experience-based learning generally leads to more efficient and asymptotically stable outcome of workers coordinating on applying with high probability to different firms in both static and dynamic wage settings. However, learning is nuanced as wages vary. Workers' choice probabilities are path-dependent and outcome may be more random. There can also be large fluctuations. Experiences could serve as both an accelerator and inhibitor for reaching efficient outcome. Learning models not only highlight potential mechanism towards equilibrium, they also provide novel avenue for policies to improve market efficiency by exploring experience effect and intervention on the learning trajectory.

JEL: C73, D83, J64
Keywords: Job Search, Coordination, Adaptive Learning

*School of Economics, The University of Edinburgh. Contact: s.lu@ed.ac.uk.
I am indebted to Ed Hopkins and Axel Gottfries for their invaluable suggestions and continuous guidance. I would also like to thank Michael Woodford for his comments and suggestions. The idea was previously presented at Toulouse Summer School in Quantitative Social Sciences 2023 and SGPE PhD Conference 2023.

1 Introduction

Workers' job application decisions could be guided by their past experiences, and modelling their learning behaviour offers new perspective on labour market dynamics. While experience-based adaptive learning have been studied in extensive-form games (Roth and Erev (1995)), normal-form games (Camerer and Hua Ho (1999)), choice predictions (Erev et al. (2010)), policy attitudes and future decisions (Albarracin and Wyer Jr (2000)), its application to search behaviour has not been thoroughly explored. Learning models suggest possible cognitive mechanisms underlying workers' search behaviour and shed light on several empirical puzzles: workers' apparent resistance to apply to higher wage jobs despite their ability to transit (Archer (2016)); and individuals' sorting behaviour at the application stage due to variations in perceived stereotypes (Barbulescu and Bidwell (2013)). Instead of imposing restrictions on workers' search strategies (Wright et al. (2021)), learning mechanisms could also provide insights on equilibrium selection in presence of multiple equilibria, which would have important policy implications. Therefore, this paper investigates the role of past experiences on workers' application decisions and how this affect labour market outcome.

In this paper, I explore a framework of 2 workers and 2 firms, which serves as a reduced-form representation of the labour market. I propose two type of market structures to study the impact of experiences on learning dynamics and equilibrium behaviour. These could help to shine a light on the impact of experiences for different market designs, and inform policies under different market circumstances.

In the first market structure, wages are unobservable to workers ex-ante to their application decisions, they learn solely based on feedback of jobs they have applied to. The process involves them recalling information from past "self" or learning through knowledge diffusion from prior generations, and revising their strategies via reinforcement learning (Erev and Roth (1998)). When firms set fixed wages, I found that workers eventually learn to apply to different firms and such pure strategy equilibria are stable in the long run. Initial experiences could have a persistent impact on later choices and determine which equilibrium is more likely to be selected. They may also create inertia to achieve more efficient matching, leading to longer learning trajectory. Policies could include reducing initial bias, lowering the impact from past memories to mitigate locked-in effect. When there is two-sided learning, firms are allowed to adjust wages by reinforcement learning, they would drive wages down to 0 to maximize profit as workers' strategies stabilizes in the equilibrium, and workers' equilibrium strategies are path-dependent. Given dynamically changing wages, workers could be learning different sets of Nash equilibria over time. Since experiences may have persistent impact on workers' behaviours, it could be harder to mitigate mismatch if workers engaged in application rounds with payoff structure comprises of a single equilibrium for substantial number of periods. A potential policy measure is to impose certain wage conditions, such that workers are learning to converge to the more efficient equilibria.

The second market structure consists of firms posting wages first in expectation of workers' reaction, and workers observing wages before devising their application strategies. Workers adopt best response learning dynamics (Fudenberg and Levine (1998); Hopkins (1999)) by taking into account payoffs, beliefs about the other workers' actions and their bias towards firms based on their long-term experiences or perceived stereotypes. Given past realized actions, workers adjust their strategies as their beliefs about each other evolve over time. Firms, tracking the changes in workers' beliefs, optimize wages in response to how they expect workers would behave in each period. In the long run, workers' and firms' behaviour converge to equilibria similar to the Quantal Response Equilibira (QRE). Multiple equilibria could exist when wage sensitivity and payoffs are high, and that bias from experiences are sufficiently weak. Workers are generally more likely to converge to asymmetric equilibria of applying with higher probability to different firms, but this can be affected by firms' wage-setting behaviours and workers' sensitivity to expected payoffs. The asymmetric equilibria, where workers apply more to different firms, can be asymptotically stable if certain condition on wage dynamics is fulfilled, but strong bias from experiences may destabilize the system. In view of this, policies could tackle firms' wage-setting mechanism and workers' bias from long-term experiences to induce more efficient and relatively stable outcome.

This paper shows that labour market mismatch can be a result of being on the learning trajectory. Workers' search strategies are affected by experiences, and they may eventually coordinate on applying to different firms, even when wages are dynamically changing. The lack of switching in jobs observed in reality could happen when workers do not observe wages, and they naturally sort into different jobs based on their application experiences. On the other hand, when they observe wages, such lock-in effect may be less evident if they are sensitive to payoffs. However, when bias from long-term experiences or stereotypes are strong, workers' strategies could be more resistant to changes even with wage information.

The rest of the paper is structured as follows: Section 2 introduces search with learning from feedback. Section 3 explores search with sequential learning. Section 4 discusses the implications and extensions of the work and conclude the paper.

Related literature. This work rides on top of extensive literature on experience-based learning to provide an evolutionary narrative for the labour market dynamics, which could offer novel policy insights to tackle mismatch problems. There are many possible learning dynamics (see Fudenberg and Levine (1998)), but Erev and Roth (1998) shows that individuals display learning pattern close to reinforcement in experiments, thus I adopt this learning mechanism as the baseline to model job search behaviour. However, Hopkins (2002) highlights that when comparing between different learning mechanisms, this force of habit model is statistically insignificant in explaining Van Huyck et al. (1997)'s experimental results. Camerer and Hua Ho (1999) postulates that an experience-weighted learning may fit individuals' learning pattern better. It accounts for both actual and counterfactual payoffs if the worker select a different action given the empirical frequency of opponent's choices. However, this may not be realistic in job search, particularly when wages are unobservable as counterfactual payoffs from unchosen action are not directly observable to workers. Nonetheless, there could be merit in accounting for opponent's empirical frequency of choices and changes in wage environment when workers can fully observe wages ex-ante to application decisions. Therefore, I explore best response dynamics (Fudenberg and Levine (1998); Hopkins (1999)) that are more forward-looking, which consider for expected payoffs given anticipated choice probabilities.

There are limited job search models that tracks transition path of individual firms' and workers' strategies based on application experiences. Studies have explored working experiences, which is not completely analogous to application experiences. For instance, Burdett and Mortensen (1998) proposes on-the-job experiences, which affect workers' preferences for firms, mainly due to human capital accumulation and expectation of higher wages as they climb the corporate ladder. Experiences in application stage, however, does not affect skills, yet it could also influence how one chooses firms, which constitute as possible means of directing search other than wages. Closer to application experiences is experiential job search highlighted by Kanfer and Bufton (2018), where workers adapt based on past involuntary job loss rather than feedback in the application stage. In more direct relevance, Wanberg et al. (2020) examines past application experiences on one's adjustment of search behaviours. However, the work focuses more on empirical snapshots and descriptive analysis, an unified framework in modelling workers' adaptive learning process could be beneficial in formalizing labour market dynamics.

Furthermore, a core objective of integrating learning theory into job search is to offer some insights on equilibrium selection. Despite the presence of multiple equilibria, job search literature often focus on the symmetric equilibrium, where workers adopt the same application strategy, in both incomplete information (McCall (1970)) and complete information (Wright et al. (2021)), as well as for some intermediary case of information availability (Wu (2020)). Although equilibria consist of workers applying to different firms are more efficient, even for partial information availability (Lu (2024)), these are often overlooked due to the perception that they are more difficult to coordinate on, a common approach in directed search literature (Galenianos and Kircher (2009)). However, with experience-based learning, it could provide some new insights on if workers could or could not converge to more efficient outcome of coordinating on applying to different firms, and under what conditions.

Nonetheless, this work contributes to understanding some puzzles in real world observations, which can potentially be explained by learning models. For instance, reinforcement learning exhibits familiarity-based learning pattern (Hopkins (2007)). This supports the phenomenon of workers' lack of job switch due to fundamental inertia in application decisions (Archer (2016)), as well as sorting behaviours displayed by different genders resulting from accumulation of experiences and beliefs built over the generations (Barbulescu and Bidwell (2013)) that differs from sorting due to skills (Eeckhout (2018)) or risk preferences (Fouarge et al. (2014)). It also provides the potential underlying mechanism driving the empirical evidence in Vafa et al. (2022), which shows that workers follow certain career trajectory and that past experiences are predictive of the jobs they end up in. While the predictive tool highlights observed job uptake pattern, it does not necessarily imply that workers do not attempt to apply to different jobs. Experience-based learning in the application stage formalize a possible channel behind the observed sorting behaviour. This could provide basis for policies to tackle potential mismatch due to sorting behaviour displayed in application choices.

Last but not least, job search models often concentrated on mapping from payoffs to choices, and the equilibrium choice probability distribution is often on the aggregate-level. Moen (1997) shows the probability distribution of workers choosing which submarket to apply to. Workers have deterministic choices given payoffs, choices are probabilistic only at the population-level. While Galenianos and Kircher (2009) models workers' probabilistic choices, it places restrictions like symmetry in strategies to make analysis tractable, and is more concerned with mapping payoffs to aggregate choice distributions rather than to individual strategies. Therefore, this paper investigates specifically the individual-level choice probabilities rather than market-level aggregates, effectively showing impact of experiences on strategy formulations, which can be history and path-dependent. Moreover, investigating micro-level decision-making could help inform policies influencing individual choices that can have macro implications.

2 Search with Learning from Feedback

In this section, I propose a market structure that conveys a traditional offline job search process, or search with online platforms but wage information are not explicitly revealed. In this setting, workers do not observe wages or strategies taken by other workers, who are simultaneously applying for jobs. They only learn the payoff from jobs they have applied to, and their search behaviours are solely affected by feedback received from previous periods. This learning process bear resemblance to Hopkins (2007), where consumers only receive payoff information of goods they have purchased. Using such learning mechanism, I investigate the impact of experiences on labour market dynamics.

2.1 Experience-driven Job Search with Fixed Wages

In a 2×2 set-up with 2 firms and 2 workers. Workers are indexed by $i = 1, 2$, and firms by $j = 1, 2$. Each firm offers one vacancy.

Timing and Set-up. Wage setting and application choices are made simultaneously.

1. Time is discrete, $t = \{0, 1, ..., T\}$. There are T finite generations of workers and firms.
2. At $t = 0$, firms observe exogenous realization of productivity, denoted by $\mathbf{z} = \{z_1, z_2\}$, $\mathbf{z} \in Z$, and set wages $\mathbf{w} = \{w_1, w_2\}$, which hold throughout all periods.
3. Workers do not observe the wages ex-ante to their selection between firm 1 and 2. They observe a payoff once they make a choice, and drop out of the market, never to return.
4. At $t = 1$, two new workers enter the market and firms offer two new vacancies.
5. Before making a decision, workers can recall from past "self" or learn from the previous generation. For example, worker 1 of $t = 1$ learns from worker 1 of $t = 0$, same applies for worker 2. Each worker learns the application strategy, the realized choice and the payoffs from the same indexed worker from the past period.
6. Workers in $t = 1$ then devise a new application strategy, where their choice probability of each firm are updated following Erev and Roth (1998)'s reinforcement learning algorithm.
7. This carries on for T periods.

Firms' side. In this fixed wage environment, wages are assumed to be rigid, such that firms set wages at the beginning of all application rounds and do not change them. Hereafter, suppose fixed wages satisfy $2w_1 > w_2 > \frac{w_1}{2}$, and that wages are bounded, $z_1 \geq w_1 \geq 0$, $z_2 \geq w_2 \geq 0$, such that wages are feasible and there can be multiple equilibria (as illustrated in the following workers' side) for purpose of exploring equilibrium selection.

Workers' side. Payoff structure faced by the workers is stationary, changes in workers' choice probabilities over the firms are solely based on their reward observation from the previous period. (Nowé et al. (2012)) The search problem faced by the workers can be simply illustrated as a standard coordination game, where payoffs are fully revealed in the equilibrium. Assuming no obvious heterogeneity between workers, which could influence their probability of being hired, they would have equal probability of getting hired if applying to the same firm.

		Worker 2	
		F1	F2
Worker 1	F1	$\frac{w_1}{2}$, $\frac{w_1}{2}$	w_1, w_2
	F2	w_2, w_1	$\frac{w_2}{2}$, $\frac{w_2}{2}$

Table 1: Application Game Faced by Workers

The application game could consist of three Nash Equilibria (NEs): $(F1,\ F2)$ and $(F2,\ F1)$ if $2w_1 > w_2 > \frac{w_1}{2}$; and a mixed NE, (F1, F2; $\frac{2w_1-w_2}{w_1+w_2}$, $\frac{2w_2-w_1}{w_1+w_2}$) if $2w_1 \geq w_2 \geq \frac{w_1}{2}$.

Based on Duffy and Hopkins (2005)'s market entry game, as well as van Strien (2022) and Erev and Roth (1998), I define workers' actions, strategies, rewards, choice rule and update rule:

- **Actions.** For worker i, $A^i = \{F1, F2\}$, where $F1$ represents applying to firm 1 and $F2$ to firm 2. Same applies for worker $-i$, $A^{-i} = \{F1, F2\}$.
- **Strategies.** Worker i's strategy is denoted as $\Delta_i = \{x_t = (x_{1t}, x_{2t})\}^T$, where $\sum_{j=1}^{2} x_{jt} = 1$; and worker $-i$'s is $\Delta_{-i} = \{y_t = (y_{1t}, y_{2t})\}^T$, where $\sum_{j=1}^{2} y_{jt} = 1$ (worker 2). x_{jt} and y_{jt} are the probabilities of firm j being chosen at time t. x_t is a pure strategy if $x_{jt} = 1$ for some j, similarly for y_t.
- **Rewards.** Wage matrix faced by worker i is $W = \begin{pmatrix} \frac{w_1}{2} & w_1 \\ w_2 & \frac{w_2}{2} \end{pmatrix}$. Payoff is denoted as $\pi_t^i = \pi_t^i(a_t^i, a_t^{-i})$, where $a_t^i \in A^i, a_t^{-i} \in A^{-i}$ are observed actions at the end of period t.

The reason behind this payoff structure is that I assume workers obtain positive reinforcement from both successful and unsuccessful application alike. This could arise from "good feelings" after being accepted for a job or just being interviewed, which is related to feeling validated and recognized professionally. (Briñol and Petty (2022)) When workers apply to the same firm, although one of them is not selected, they will still receive valuable information about their prospect of being hired. The positive signal of intent to hire is weighted by the probability of successful hire ($\frac{1}{2}$ for homogeneous workers), resulting in partial reinforcement of workers' choices if both applies to the same firm. Workers' choice in such case is updated based on potential payoff that encompasses the level of competition.

Workers' probability of selecting firm j at time t, x_{jt} and y_{jt}, depends on the propensities, denoted by q_{jt}^i. Each worker is endowed with an initial propensity for each action, $q_{j0}^i = \{q_{10}^i, q_{20}^i\}$. These initial propensities can be perceived as workers' respective innate preference before entering the labour market, and for subsequent periods, propensities can be referred to as accumulated payoffs obtained by selecting each firm (Beggs (2005)).

Herein, I adopt a linear **choice rule**:

$$\text{Worker } i\text{: } x_{jt} = \frac{q_{jt}^i}{\sum_{j=1}^{J} q_{jt}^i}, \text{ Worker } -i\text{: } y_{jt} = \frac{q_{jt}^{-i}}{\sum_{j=1}^{J} q_{jt}^{-i}} \tag{1}$$

Following Erev and Roth (1998) (ER) reinforcement learning mechanism, which specifies the **update rule** on how propensities are updated in each round:

$$\text{Worker } i\text{: } q_{j(t+1)}^i = q_{jt}^i + \pi_{jt}^i(a_t^i, a_t^{-i}), \text{ Worker } -i\text{: } q_{j(t+1)}^{-i} = q_{jt}^{-i} + \pi_{jt}^{-i}(a_t^i, a_t^{-i}) \tag{2}$$

When worker 1 select firm 1 in period t, if a positive feedback is received, then in the next period, the propensity of applying to firm 1 by worker 1 will increase by an increment equal to the realized payoff ($\pi_{1t}^i(F1, a_t^{-i})$) given observed actions ($a_t^i = F1,\ a_t^{-i}$).

For example, if worker 1 choose to apply to firm 1 in period 1 and worker 2 to firm 2, as payoffs (w_1, w_2) are revealed at the end of the period, action $F1$ and $F2$ are reinforced by the magnitude of w_1 and w_2 for worker 1 and 2, respectively. If both workers choose to apply to firm 1, action $F1$ will be reinforced by $\frac{w_1}{2}$ for both workers. Workers learnt the wage offered by firm 1 and how many candidates are competing for the firm. Even if they did not obtain the job, their propensity to firm 1 is positively affected because they gained some information, and if they obtained the job, the fact that another candidate was also competing for the job implies higher possibility of not getting hired, so propensity to select the same firm is slightly discounted.

Workers only receive feedback from actions they actually take. The impact of the other worker's application strategy is implicit. One do not directly observe the choices taken by the other worker in the same period, which is realistic to the labour market. But worker $-i$'s choice in period t affects worker i in the same period via realized payoffs, thus influencing worker i's choice propensities in period $t+1$.

Given payoff matrix for worker 1 (W) and worker 2 (W^T):

$$W = \begin{pmatrix} \frac{w_1}{2} & w_1 \\ w_2 & \frac{w_2}{2} \end{pmatrix}, W^T = \begin{pmatrix} \frac{w_1}{2} & w_2 \\ w_1 & \frac{w_2}{2} \end{pmatrix} \tag{3}$$

I formulate the **expected change in application strategies** using Lemma 1 of Hopkins (2002) with the choice rule (1) and the update rule (2):

$$\text{Worker } i\text{: } E(x_{t+1}|q_t^i) - x_t = \frac{R(x_t)Wy_t}{Q_t^i} + O(\frac{1}{(Q_t^i)^2}) \tag{4}$$

$$\text{Worker } -i\text{: } E(y_{t+1}|q_t^{-i}) - y_t = \frac{R(y_t)W^Tx_t}{Q_t^{-i}} + O(\frac{1}{(Q_t^{-i})^2}) \tag{5}$$

where $q_t^i = \{q_{1t}^i, q_{2t}^i\}$, $q_t^{-i} = \{q_{1t}^{-i}, q_{2t}^{-i}\}$, each comprises of the propensities for the two actions at time t, corresponding to the two workers. $R(\cdot)$ is the replicator operator, reflecting how strategies evolve based on their relative payoffs:

$$R(x_t) = \begin{pmatrix} x_{1t}(1-x_{1t}) & -x_{1t}x_{2t} \\ -x_{2t}x_{1t} & x_{2t}(1-x_{2t}) \end{pmatrix}, R(y_t) = \begin{pmatrix} y_{1t}(1-y_{1t}) & -y_{1t}y_{2t} \\ -y_{2t}y_{1t} & y_{2t}(1-y_{2t}) \end{pmatrix}.$$

This implies, for instance, as worker 1's choice probability to firm 1 (x_{1t}) increases, its growth rate is dampened, and at the same time, the choice probability to firm 2 is also reduced. This determines how quickly the probability changes and which equilibrium the system moves towards. Lastly, $Q_t^i = \sum_{j=1}^{J} q_{jt}^i$ denotes the sum of propensities, it may be interpreted as a control over the magnitude of updates. As Q_t^i grows over time, the system would approximate continuous time dynamics in the limit. This relates to the step size, which describes the rate at which workers updates their strategies. In this model, workers would have different step size determined by their payoff experiences. But as t increases, with $Q_t^i \to \infty$ and $Q_t^{-i} \to \infty$, the effective step size is of order $\frac{1}{t}$, adjustments to strategies become less significant over time.

$$E(x_{t+1}|q_t^i) - x_t \approx \frac{R(x_t)Wy_t}{Q_t^i}, E(x_{t+1}|q_t^i) - x_t \to 0 \tag{6}$$

$$E(y_{t+1}|q_t^{-i}) - y_t \approx \frac{R(y_t)W^Tx_t}{Q_t^{-i}}, E(y_{t+1}|q_t^{-i}) - y_t \to 0 \tag{7}$$

Given the possibility of three equilibria in this setting, workers could either coordinate on applying to different firms or adopting a mixed strategy that suggest more randomized search behaviour. Based on Hopkins and Posch (2005), ER learning rule would not converge to a

Nash Equilibrium (NE) linearly unstable under the replicator dynamics and it cannot converge to a rest point that is not a NE. The mixed strategy equilibrium is unstable under replicator dynamics, any perturbation would cause workers to drift towards one of the pure strategy equilibria. As a result, workers should eventually learn to coordinate on applying to different firms, and experiences would act as a natural selection mechanism for arriving at efficient outcome of one-to-one matching. However, the exact pure NE that is reached could be dependent on the initialization conditions.

Proposition 1 (Stability of Efficient Outcome)**.** *Given fixed wages that satisfy* $2w_1 > w_2 > \frac{w_1}{2}$*, workers would coordinate on applying to different firms in the long run and the outcome is stable.*

Proof. In the limit of small step size, the discrete time process can be approximated by continuous time dynamics. For payoff matrices, W and W^T (3),

$$\frac{dx_{1t}}{dt} = f(x_{1t}, y_{1t}) = x_{1t}(1 - x_{1t})[(-\frac{w_1}{2} - \frac{w_2}{2})y_{1t} + w_1 - \frac{w_2}{2}] \tag{8}$$

$$\frac{dy_{1t}}{dt} = g(x_{1t}, y_{1t}) = y_{1t}(1 - y_{1t})[(-\frac{w_1}{2} - \frac{w_2}{2})x_{1t} + w_1 - \frac{w_2}{2}] \tag{9}$$

Evaluating the Jacobian matrix:

$$J = \begin{bmatrix} \frac{\partial f}{\partial x_{1t}} & \frac{\partial f}{\partial y_{1t}} \\ \frac{\partial g}{\partial x_{1t}} & \frac{\partial g}{\partial y_{1t}} \end{bmatrix} = \begin{bmatrix} (1-2x_{1t})[(-\frac{w_1}{2} - \frac{w_2}{2})y_{1t} + w_1 - \frac{w_2}{2}] & x_{1t}(1-x_{1t})(-\frac{w_1}{2} - \frac{w_2}{2}) \\ y_{1t}(1-y_{1t})(-\frac{w_1}{2} - \frac{w_2}{2}) & (1-2y_{1t})[(-\frac{w_1}{2} - \frac{w_2}{2})x_{1t} + w_1 - \frac{w_2}{2}] \end{bmatrix} \tag{10}$$

Deriving the eigenvalue using $\det(J - \lambda I) = 0$, if λ takes both positive and negative value, the equilibrium will be a saddle point.

Evaluating pure strategies, Jacobian matrix shows negative eigenvalues given $2w_1 > w_2 > \frac{w_1}{2}$, implying the fixed points are asymptotically stable. In the special case of homogeneous firms, $z_1 = z_2 = z$, $w_1 = w_2 = w^*$, evaluating mixed strategy, $x_{1t} = y_{1t} = \frac{1}{2}$, $J(\frac{1}{2}, \frac{1}{2}) = \begin{bmatrix} 0 & -0.25w^* \\ -0.25w^* & 0 \end{bmatrix}$, $\lambda = \pm 0.25w^*$. For $w^* > 0$, this fixed point is a saddle point and unstable.

While trajectories very close to the stable manifold will move the system towards the mixed equilibrium, almost any perturbation should eventually lead to convergence to pure strategies, where workers coordinate on the efficient equilibria that constitute one-to-one matching. □

Example 1. *(Unstable Mixed Strategy Equilibrium) Suppose* $w_1 = 3$*,* $w_2 = 3$*, there could exist 3 possible equilibria:* $(F1, F2)$*,* $(F2, F1)$*, and* $(F1, F2, \frac{4}{5}, \frac{1}{5})$*. Assuming* $Q_t^i = Q_t^{-i} = 1$*,*

$$\frac{dx_{1t}}{dt} = f(x_{1t}, y_{1t}) = x_{1t}(1 - x_{1t})(2 - 2.5y_{1t}) \tag{11}$$

$$\frac{dy_{1t}}{dt} = g(x_{1t}, y_{1t}) = y_{1t}(1 - y_{1t})(2 - 2.5x_{1t}) \tag{12}$$

$$J(x_{1t}, y_{1t}) = \begin{bmatrix} 0 & -0.4 \\ -0.4 & 0 \end{bmatrix} \Bigg|_{\frac{4}{5}, \frac{4}{5}} \tag{13}$$

Evaluating at the point of $(x_{1t}, y_{1t}) = (\frac{4}{5}, \frac{4}{5})$*,* $\lambda = \pm 0.4$*.*

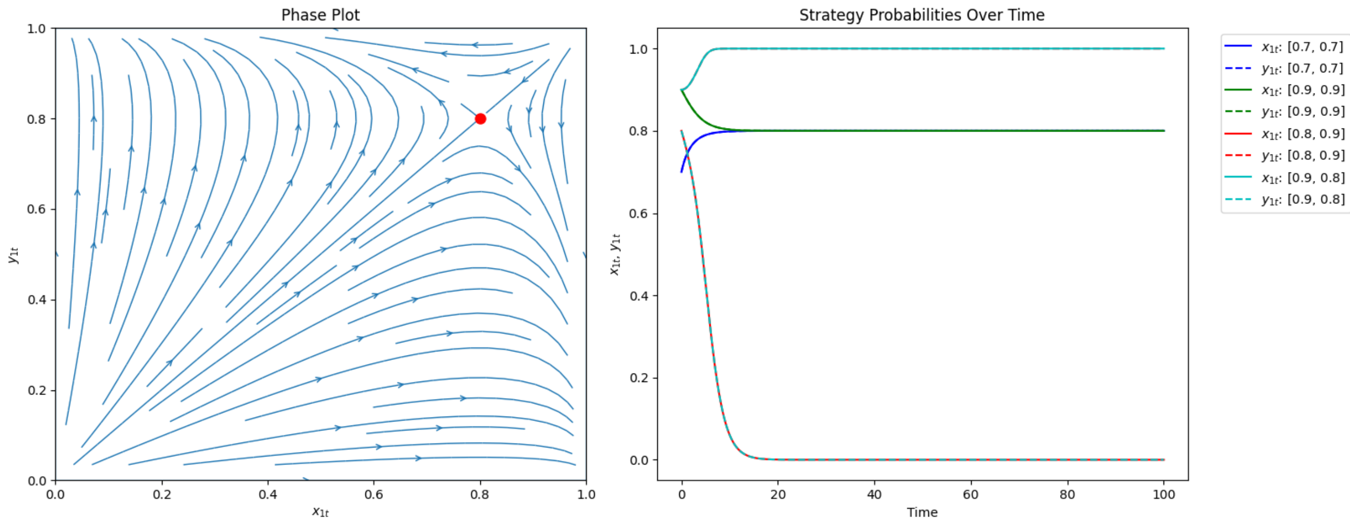


Figure 1: Illustration of Example 1

Figure 1 shows the phase plot and strategy change over time. Most learning trajectories demonstrate workers should end up in one of the pure strategy equilibria.

Erev and Roth (1998) highlighted one special feature of this learning mechanism – its heavy reliance on initial propensities and decreasing impact from recent experiences as accumulated payoffs become higher.

Observation 1 (Initial Bias and Experiences on Equilibrium selection)**.** *Equilibrium selection is influenced by initial bias, characterized by initial propensities* (q^i_{j0}, q^{-i}_{j0})*; as well as initial experiences, characterized by feedback received in the first few application rounds:*

- *Strong prior preference or goodwill towards specific firm could create fundamental inertia for workers to adjust their application strategies, contributing to immediate sorting into different firms or persistent overcrowding at a single firm.*
- *Workers could be locked-in to choices they made initially, leading to experience-based sorting.*

Given wage condition holds for multiple equilibria, $2w_1 > w_2 > \frac{w_1}{2}$, as $t \to \infty$, workers would eventually coordinate on applying to different firms.

$$\lim_{t\to\infty} x_{1t} = \bar{x} = 1 \text{ if } \lim_{q_1\to\infty} = \frac{q_1}{q_1+q_2} = 1 \tag{14}$$

$$\lim_{t\to\infty} x_{1t} = \bar{x} = 0 \text{ if } \lim_{q_2\to\infty} = \frac{q_1}{q_1+q_2} = 0 \tag{15}$$

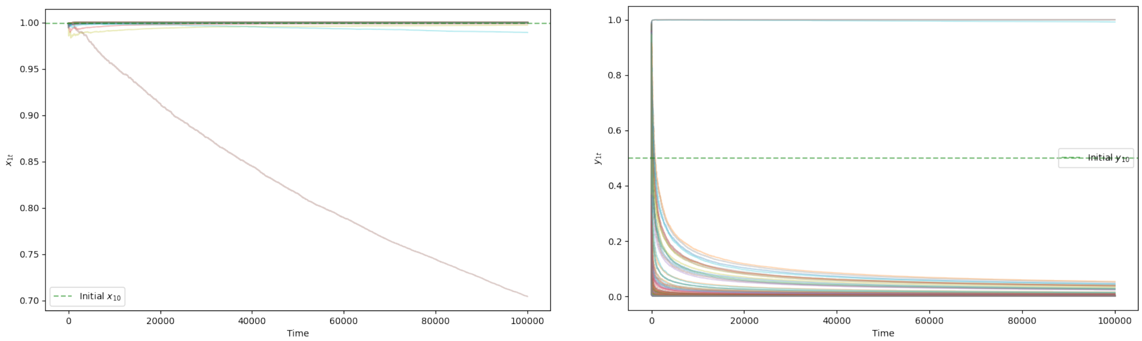
Despite this, there could be multiple periods where workers oscillate between choosing firm 1 and 2, leading to possible overcrowding for many periods over the trajectory of change in choice probabilities. Strong initial bias by both workers (i.e. high initial propensities towards the same firm) could also lead to biased search for substantial number of periods.

Hopkins (2007) also demonstrates that consumers could be locked into choosing certain good that they initially prefer and undergo familiarity-based learning. In the job search context, workers choosing to apply to firm 1 and 2 respectively in the initial periods would make $(F1, F2)$ more likely to be selected than other equilibria in the long run. This suggests that for homogeneous workers, who could start with random search, their diverse application strategies may be a result of randomness and luck, driven by positive reinforcement in initial periods of their job search. This also demonstrates that sorting behaviour can be experience-driven rather than based solely on skills.

Example 2. *(Heterogeneous Initial Bias) Suppose initialization propensities are* $q^i_{10} = 1000, {q_{10}}^{-i} = 1$, $q^i_{20} = {q_{20}}^{-i} = 1$, *worker 1 denotes higher propensity to firm 1 than firm 2. Workers' choice probabilities to firm 1:*

$$x_{10} = \frac{1000}{1001} \approx 0.999, y_{10} = \frac{1}{2} \tag{16}$$

Worker 1 applies with higher probability to firm 1 by default as compared to worker 2. As t *increases,* $(F1, F2)$ *is more likely to be reached.*



(a) Worker 1 (b) Worker 2

Figures show workers' application probability to firm 1 (y-axis) against number of periods (x-axis) for 10 simulation sessions ($w_1 = w_2 = 5$, $t = 100000$, ${q_{10}}^i = 1000, {q_{20}}^i = {q_{10}}^{-i} = {q_{20}}^{-i} = 1$).

Figure 2: Learning Path of Worker 1 and 2.

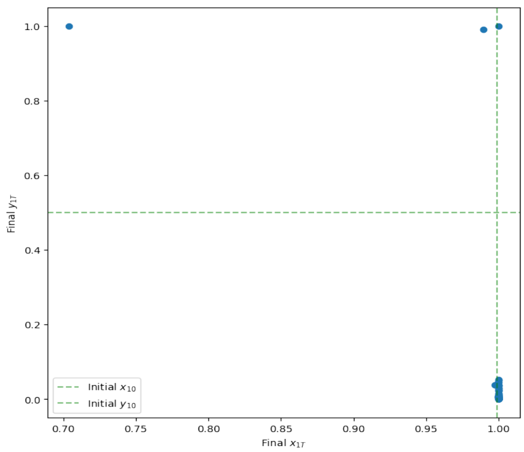


Figure 3: Workers' Final Choice Probability in Period T

Figure 2a & 2b demonstrate simulations of how workers' choice probabilities of firms evolve over time given past experiences. Worker heterogeneity, on the basis of differences in initial propensities, could lead to sorting behaviour. In this example, it is more likely for worker 1 to choose firm 1 during initial rounds, thus more likely to be lock-in to firm 1. Since oscillations are fundamental characteristic of this learning mechanism, there is always a positive chance of selecting an alternative action, thus it remain probable for $(F2, F1)$ *to be selected in the long run, but the likelihood of converging to* $(F1, F2)$ *is higher. Figure 3 shows the final choice probability in period* T*, most of the sessions stop at close to* $(F1, F2)$*.*

In Barbulescu and Bidwell (2013), students with same education backgrounds could display segregation in job applications due to gender stereotypes associated with the jobs; and Terjesen et al. (2007) also found that females, in comparison to male counterparts in universities, put greater weights on "using your degree skills", thus are more likely to go for jobs related to their degree. In the learning model, these can be interpreted as worker heterogeneity in initial propensities, which translate into dispersion in choice probabilities and affect equilibrium selection.

2.1.1 Partial Recall of Experiences

Perfect recall of experiences may be a stringent assumption, individuals are often subjected to limited memory. Cognitive load theory suggests individuals' working memory is constrained to a capacity of approximately 4 elements of information (Paas and Ayres (2014)). To implement the impact of partial recall on workers, I include a forgetting parameter, η, $\eta \in (0,1)$, to account for recency effect. As a result, past experiences or knowledge could have a diminishing effect on current application decisions (Erev and Roth (1998)).

$$q^i_{j(t+1)} = (1-\eta)q^i_{jt} + \pi^i_{jt}(a^i_t, a^{-i}_t) \tag{17}$$

Proposition 2 (Partial Recall)**.** *Let $\eta \in (0,1)$ be the experience decay parameter in the propensity updating process, for equation (17):*

1. *As $\eta \to 0$, the propensity updating process converges to perfect recall.*
2. *As $\eta \to 1$, next period propensity $q^i_{j(t+1)}$ converges to being determined solely by realized payoffs in t, $\pi^i_{jt}(a^i_t, a^{-i}_t)$.*
3. *For $0 < \eta < 1$, weight placed on previous period propensities is lower, the influence of initial propensity q^i_{j0} diminishes as $T \to \infty$.*

Proof. For T periods, as $\eta \to 0$,

$$\lim_{\eta \to 0} q^i_{jT} = q_{j0} + \pi_{j0} + \pi_{j1} + ... + \pi_{jT-2} + \pi_{j(T-1)} \tag{18}$$

As $\eta \to 1$,

$$\lim_{\eta \to 1} q^i_{jT} = \pi_{j(T-1)} \tag{19}$$

For $0 < \eta < 1$,

$$q^i_{jT} = (1-\eta)^T q_{j0} + (1-\eta)^{T-1}\pi_{j0} + (1-\eta)^{T-2}\pi_{j1} + ... + (1-\eta)\pi_{jT-2} + \pi_{j(T-1)} \tag{20}$$

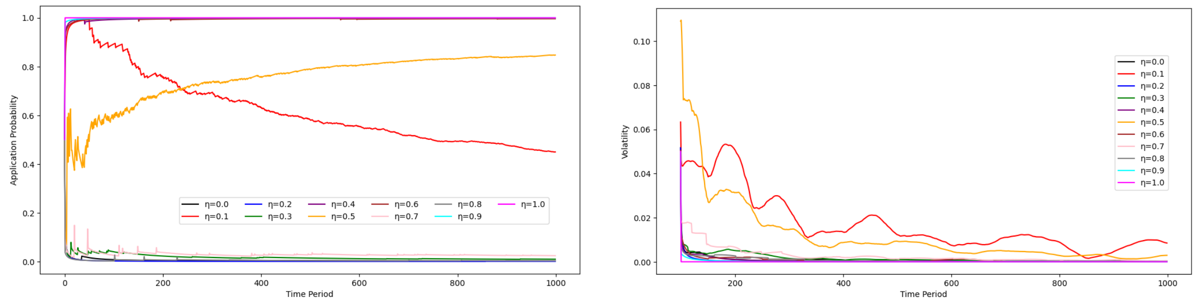
Coefficient on q^i_{j0} is $(1-\eta)^T$, $\lim_{T\to\infty}(1-\eta)^T = 0$. □

In the extreme case of complete forgetting ($\eta = 1$), the probability of worker 1 selecting firm 1 depends solely on the last period feedback:

$$\lim_{\eta \to 1} x_{1T} = \frac{\pi^i_{1(T-1)}}{\pi^i_{1(T-1)} + \pi^i_{2(T-1)}}, \lim_{\eta \to 1} y_{1T} = \frac{\pi^{-i}_{1(T-1)}}{\pi^{-i}_{1(T-1)} + \pi^{-i}_{2(T-1)}} \tag{21}$$

Since workers can only select one firm in specific period, this implies for worker 1, either $\pi_{1(T-1)}$ or $\pi_{2(T-1)}$ will be positive, therefore, either $x_{1T} \to 1$ or $x_{1T} \to 0$. There is convergence to applying with certainty to firm 1 or 2. Same applies for worker 2.

For intermediary forgetfulness ($0 < \eta < 1$), past events will have some impact on application choices. But as compared to perfect recall, past events have diminishing impact on propensities, which could lead to slower augmentation of Q_t^i and Q_t^{-i}. Therefore, it may take longer for the system to settle into equilibrium. Whilst in the fixed wage environment, workers are expected to converge to pure NEs, limited memory could contribute to longer process of reaching efficient outcomes and there may be more instances of mismatch.



(a) Learning Paths with Different η (b) Volatility in Choice Probability

Figure shows worker 1's application probability and volatility in choice probability for different η, $\eta \in [0, 1]$ across first 1000 periods. ($w_1 = w_2 = 5$, $t = 100000$, ${q_{10}}^i = {q_{20}}^i = {q_{10}}^{-i} = {q_{20}}^{-i} = 1$).

Figure 4: Worker 1's Choice Probability of Firm 1 with Forgetting

In Figure 4a, I show a simulation of worker 1's choice probability of firm 1 for different forgetting parameters. When workers do not remember any past events ($\eta = 1$), choice probability go straight to 1 or 0. It is possible that if both workers apply to firm 1 and receive positive feedback of $\frac{w_1}{2}$ end up overcrowding at firm 1. When workers retain some memories ($\eta < 1$), there is convergence to pure NEs, but choice probabilities are noisy due to positive weight placed on past propensities. For instance, Figure 4b shows high volatility[1] in choice probability when $\eta \neq 1$. While intuitively, larger forgetting parameter should imply more volatile choice probabilities, but in this setting, since propensities depend on realized payoffs, which are inherently stochastic when choice probabilities are not deterministic, so larger forgetting parameter may not imply more volatile outcomes.

2.1.2 Policy Implications

In a static wage environment, where wages satisfy the condition for multiple equilibria to exist, workers would eventually coordinate on applying to different firms in the long run. Experiences could serve as a natural mechanism paving towards an efficient market outcome. However, there could be many periods of mismatch before pure NEs are reached, thus suggesting policies could tackle the learning trajectory.

The convergence can be slower when workers have strong initial bias to the same firm, indicating policies could cater to reducing initial bias to facilitate faster coordination. Eliciting different prior preferences also have some merits as having heterogeneous initial bias could be beneficial to facilitate faster coordination on applying to different firms.

Furthermore, in this learning model, initial experiences are important for later choices and could influence the equilibrium selected. Once workers start with a job, then they are likely to settle into similar jobs in the future. If both workers apply for the same job in the initial periods and obtain similar experiences, it will take substantial number of periods for them to adjust, policies could speed up the process by lowering the memory impact. This may be done by influencing the degree of forgetting through platform recommendation system that

[1]Higher volatility implies how much application probability to firm 1 in t is higher than average choice probability of the past 100 periods (between t and $t - 99$).

control how much exposure one has to past information when searching for jobs. However, one need to determine whether to induce more or less forgetfulness. For instance, when workers have sufficiently diverse experiences, perfect recall could encourage faster convergence towards efficient outcome; but when workers' initial propensity are high and biased towards the same firm, or experiences from first few applications reinforce potential overcrowding, then there is benefits to "unlearn".

2.2 Two-sided Simultaneous Learning with Dynamic Wages

In this section, I relax the assumption on wage rigidity, such that firms can also be adaptive learners and wages are affected by workers' search behaviours. I explore how workers react to a dynamic wage environment, and if they would learn to coordinate on applying to different firms. I will also investigate how wages evolve.

In this set-up, both firms and workers are learning simultaneously and are assumed to adopt Erev and Roth (1998) reinforcement learning algorithm.

Workers' side. Workers follow the same learning pattern as Section 2.1.

Firms' side. Given exogenous realization of productivities, $\mathbf{z} = \{z_1, z_2\}$, firms select wages, $\mathbf{w} = \{w_1, w_2\}$, where $0 \leq \mathbf{w} \leq \mathbf{z}$. It is assumed that as a new firm in the labour market, it is unlikely for it to know which wage to set to attract workers at the beginning of application rounds, but over time, it would learn to choose wages based on the responses it obtained. For example, in time t, if firm 1 choose w_1 and receives both workers' application and is able to produce this period, then it is likely to select the same wage again in period $t+1$ given an update rule. However, if it receives no worker in period t, then the choice of wage value w_1 is not reinforced. I define firms' actions, strategies, rewards, choice rule and update rule below:

- **Actions.** Firm j has finite and discrete number of actions, $A^j = (0, 1, 2, ..., z_j)$. Assuming firm homogeneity, $A^j = A^{-j} = (0, 1, 2..., z)$. Each action effectively corresponds to the wage offered, $a^j = w_j$.

- **Strategies.** Firm j's strategy is denoted as $\Delta_j = \omega_t^j = (\omega_{0t}^j, \omega_{1t}^j, ..., \omega_{zt}^j)^T$, where $\sum_{a^j=0}^{z} \omega_{a^j t}^j = 1$; and firm $-j$'s is $\Delta_{-j} = \omega_t^{-j} = (\omega_{0t}^{-j}, \omega_{1t}^{-j}, ..., \omega_{zt}^{-j})^T$, where $\sum_{a^j=0}^{z} \omega_{a^j t}^{-j} = 1$. $\omega_{a^j t}^j$ and $\omega_{a^j t}^{-j}$ are the probabilities of each action a^j being chosen by firm 1 and 2 at time t.

- **Rewards.** If at least one worker applies to firm j, firm j receives a payoff of $z_j - w_j$, otherwise 0. Two possible outcomes:

$$I_i^j = \begin{cases} 1 & \text{if worker } i \text{ chooses firm } j \\ 0 & \text{otherwise} \end{cases}, \quad I_{-i}^j = \begin{cases} 1 & \text{if worker } -i \text{ chooses firm } j \\ 0 & \text{otherwise} \end{cases}$$

- **Choice Rule.** Firms' choice probabilities of selecting each action at time t, $\omega_{a_t^j}^j$ and $\omega_{a_t^{-j}}^{-j}$, depend on propensities of each action being chosen, denoted as $\theta_{a_t^j}^j$ and $\theta_{a_t^j}^{-j}$ respectively. Adopting a linear choice rule:

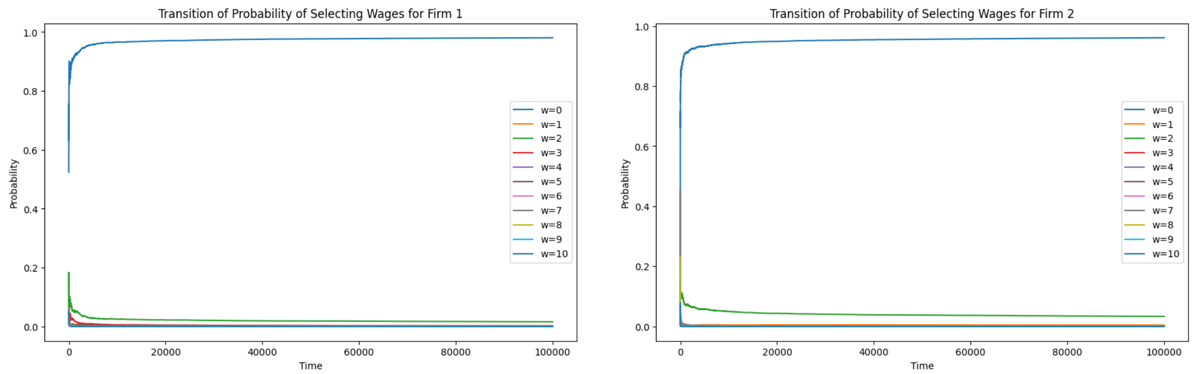
$$\text{Firm } j\text{: } \omega_{a_t^j}^j = \frac{\theta_{a_t^j}^j}{\sum_{a^j=0}^{z_j} \theta_{a_t^j}^j}, \text{Firm } -j\text{: } \omega_{a_t^{-j}}^{-j} = \frac{\theta_{a_t^{-j}}^j}{\sum_{a^{-j}=0}^{z-j} \theta_{a_t^{-j}}^j} \tag{22}$$

- **Update Rule.** Propensities are updated based on realized payoffs:

$$\text{Firm } j\text{: } \theta_{a_{(t+1)}^j}^j = \theta_{a_t^j}^j + \pi_t^j(a_t^i, a_t^{-i}, a_t^j, a_t^{-j}), \text{Firm } -j\text{: } \theta_{a_{(t+1)}^{-j}}^{-j} = \theta_{a_t^{-j}}^{-j} + \pi_t^{-j}(a_t^i, a_t^{-i}, a_t^j, a_t^{-j}) \tag{23}$$

While workers gain positive reinforcement from both successful and unsuccessful application alike. For the firms, they only receive reinforcement for the wage they set when they successfully hire a worker. This is because they do not gain any additional information about how close a worker was to choosing them if they receive no application. As a result, workers and firms have slightly different learning dynamics due to differences in their access to information.

Suppose both firms and workers are homogeneous, they start with uniform probability over their action space. Figure 5 demonstrates growing probability of wage 0 being chosen by both firms. Figure 6 shows that in the final period of 10 simulated sessions (runs), the probabilities of setting each wage are higher for low wage values. Both figures suggest that wages will eventually be pushed down towards 0.



(a) Firm 1's Choice Probabilities of Wages (b) Firm 2's Choice Probabilities of Wages

Figure shows firms' probability of choosing each discrete wage value, $\mathbf{w} \in [0, 10]$ for $t = 100000$, ${q_{j0}}^{i} = {q_{j0}}^{-i} = 1$, $\theta^{j}_{a^{j}_{0}} = \theta^{-j}_{a^{-j}_{0}} = 1$, $z_1 = z_2 = z = 10$.

Figure 5: Learning Path of Firm 1 and 2 in 2-sided RL

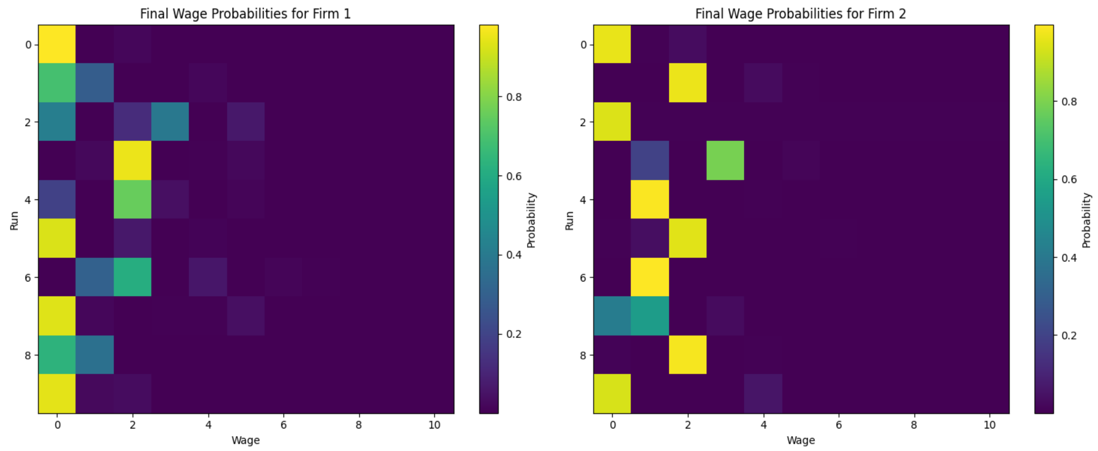


Figure 6: Firms' Probability of Selecting Each Wage in Period T Across 10 Sessions

For workers' side, Figure 7 shows the choice probabilities in the final period of 10 simulated sessions. There is higher likelihood of $(F1, F2)$ being selected. However, when compare to the static wage environment in Figure 3, the results are less saturated around pure NEs. There are also higher chances of workers applying to the same firm after substantial number of periods at end of period T, which are inefficient outcomes.

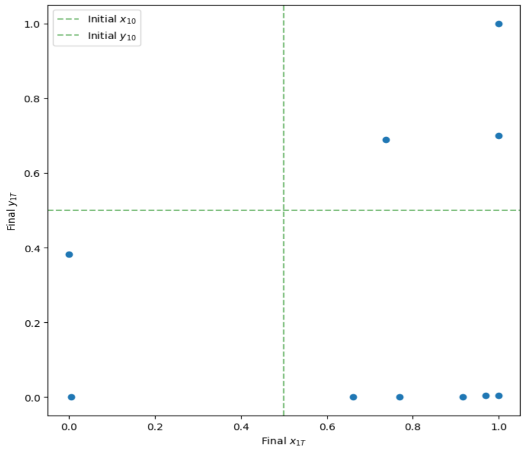


Figure 7: Workers' Final Choice Probability in Period T

Proposition 3 (Convergence in Dynamic Wage Environment)**.** *Given workers' strategies, x_t and y_t, and firms' strategies, $\omega^j_{a^j_t}$ and $\omega^{-j}_{a^{-j}_t}$, at time t. The system converges to a fixed point, $(x^*, y^*, {\omega^j_{a^j_t}}^*, {\omega^{-j}_{a^{-j}_t}}^*)$, as $t \to \infty$.*

Proof. For workers, I show in Section 2.1 that as $t \to \infty$, by equations 6 and 7, $E(x_{t+1}|q^i_t) - x_t \to 0$ and $E(y_{t+1}|q^{-i}_t) - y_t \to 0$.

For firms, their expected change in strategies are:

$$\text{Firm } j\text{: } E[\omega^j_{a^j_{t+1}}|\theta^j_{a^j_t}] - \omega^j_{a^j_t} = \frac{[1-(1-x_{jt})(1-y_{jt})](z_j - a^j_t)}{S^j_t} + O(\frac{1}{{S^j_t}^2}) \tag{24}$$

$$\text{Firm } -j\text{: } E[\omega^{-j}_{a^{-j}_{t+1}}|\theta^{-j}_{a^{-j}_t}] - \omega^{-j}_{a^{-j}_t} = \frac{[1-(1-x_{(-j)t})(1-y_{(-j)t})](z_{-j} - a^{-j}_t)}{S^{-j}_t} + O(\frac{1}{{S^{-j}_t}^2}) \tag{25}$$

where $[1-(1-x_{jt})(1-y_{jt})]$ reflects the probability that at least one worker applies to the firm; $S^j_t = \sum_{a^j=0}^{z_j} \theta^j_{a^j_t}$, $S^{-j}_t = \sum_{a^{-j}=0}^{z_{-j}} \theta^{-j}_{a^{-j}_t}$ are the sum of propensities over different actions in period t for each firm.

As $t \to \infty$, S^j_t and S^{-j}_t grow larger. From equations 24 and 25,

$$E[\omega^j_{a^j_{t+1}}|\theta^j_{a^j_t}] - \omega^j_{a^j_t} \approx \frac{[1-(1-x_{jt})(1-y_{jt})](z_j - a^j_t)}{S^j_t}, E[\omega^j_{a^j_{t+1}}|\theta^j_{a^j_t}] - \omega^j_{a^j_t} \to 0 \tag{26}$$

$$E[\omega^{-j}_{a^{-j}_{t+1}}|\theta^{-j}_{a^{-j}_t}] - \omega^{-j}_{a^{-j}_t} \approx \frac{[1-(1-x_{(-j)t})(1-y_{(-j)t})](z_{-j} - a^{-j}_t)}{S^{-j}_t}, E[\omega^{-j}_{a^{-j}_{t+1}}|\theta^{-j}_{a^{-j}_t}] - \omega^{-j}_{a^{-j}_t} \to 0 \tag{27}$$

Both workers' and firms' expected change in strategies converge to 0. □

As workers' strategies stabilize, wages will be driven down to 0 as positive wages reduce profits. At the point where wages are exactly 0, there will be a continuum of equilibria. Both Jacobian and Hessian matrix evaluated at the equilibrium points become 0, which means all equilibrium points are neutrally stable and are weak NEs. Based on this logic, the equilibrium selected depends on the direction of strategy adjustment at the time when wages are still positive. Once wages hit 0, strategies are "locked-in". Such outcome is seemingly "uninteresting" as both sides simply stop learning because there are no rewards and no adaptive pressure, this does not imply strategies are optimal. Despite the outcome, it may be interesting to explore the learning trajectory since equilibrium selection is path-dependent, efficient outcomes of one-to-one matching may be reached depending on the learning dynamics.

Along the learning path, since wages are dynamically changing, workers could be facing different "games" depending on the wage conditions. There could be equilibrium switching as one moves from one "game" to another.

Definition 2.1. *(Equilibrium Switching) Consider the game faced by workers at time t to be G_t, the set of Nash Equilibria (NEs) associated with the specific game is defined to be $NE(G_t)$. Equilibrium switching is defined to occur if:*

1. *There exist two distinct equilibrium sets NE_1 and NE_2, such that $NE(G_t) = NE_1$ for $t \leq \tilde{t}$; $NE(G_t) = NE_2$ for $\tilde{t} < t \leq T$, $\tilde{t}$ is the critical time of game change, T being total number of periods.*

2. *There can be multiple equilibrium switching over workers' learning trajectory.*

Workers could effectively be facing three possible games for some given wage conditions (see Figure 8, 9, 10). They encompass different sets of NE(s). Equilibrium switching (Definition 2.1) could occur as payoffs evolve, and workers are learning different sets of NE(s) along the learning trajectory.

		Worker 2	
		F1	F2
Worker 1	F1	$\boxed{\frac{w_1}{2}, \frac{w_1}{2}}$	w_1, w_2
	F2	w_2, w_1	$\frac{w_2}{2}, \frac{w_2}{2}$

Figure 8: G1: $w_{1t} > 2w_{2t}$

		Worker 2	
		F1	F2
Worker 1	F1	$\frac{w_1}{2}, \frac{w_1}{2}$	$\boxed{w_1, w_2}$
	F2	$\boxed{w_2, w_1}$	$\frac{w_2}{2}, \frac{w_2}{2}$

Figure 9: G2: $2w_{1t} > w_{2t} > \frac{w_{1t}}{2}$

		Worker 2	
		F1	F2
Worker 1	F1	$\frac{w_1}{2}, \frac{w_1}{2}$	w_1, w_2
	F2	w_2, w_1	$\boxed{\frac{w_2}{2}, \frac{w_2}{2}}$

Figure 10: G3: $w_{2t} > 2w_{1t}$

Based on the analysis for static wage environment in Section 2.1, it is expected that if wage condition for $G2$ can be sustained for a long period of time, workers would be able to learn to converge to the pure NEs, leading to efficient outcomes. However, since one may have to overcome a large inertia due to accumulated propensities from previous game play, which could involve learning a different set of NE(s), therefore, even if equilibrium switching happens (e.g. from Figure 8 to 9), workers' actual behaviour from converging to one set of NE(s) to another happens with a time lag.

Proposition 4. *(Delayed Adaptation in Equilibrium Switching) Following equilibrium switching at time $\tilde{t}$ due to evolving wages, workers will not immediately transition from learning NE_1 to NE_2. There exists a time lag in adapting their strategies:*

For $t \leq \tilde{t}$, workers converge to NE_1:

$$\lim_{t \to \infty, t \leq \tilde{t}} x_t, y_t \in \textit{ Support of } NE_1 \tag{28}$$

At $t = \tilde{t}$, game changes with NE_2 becomes the set of new equilibria.

For $\tilde{t} < t < \bar{t}$, workers can be described to be in the transitional state that are still influenced by NE_1, but beginning to adapt to NE_2:

$$x_t, y_t \in \textit{ Support of } NE_1 \cup NE_2 \tag{29}$$

For $t \geq \bar{t}$, workers fully converge to NE_2:

$$\lim_{\bar{t} \to \infty, t \geq \bar{t}} x_t, y_t \in \textit{ Support of } NE_2 \tag{30}$$

where x_t, y_t are set of choice probability over firms in time t for worker 1 and 2, respectively.

Proof. Shown by Example 3. □

Given perfect memory, time lag is infinite, the system never fully adapt to new conditions due to accumulated past payoffs. In Figure 11, I show for a single simulation session with 100 periods, there is switching between different wage regimes, corresponding to the different games ($G1$, $G2$, $G3$).

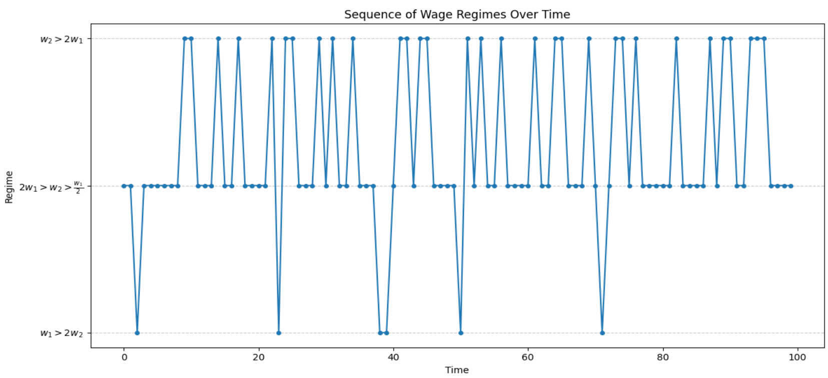


Figure shows wage transition across different regimes, characterized by wage conditions for ($z_1 = z_2 = 10$, ${q_{10}}^{i} = {q_{20}}^{i} = {q_{10}}^{-i} = {q_{20}}^{-i} = 1$), $t = 100$.

Figure 11: Switching Between Wage Regimes

In Figure 12, I illustrate the evolution of wages and choice probabilities. Suppose workers and firms are in regime 1 ($G1$) (red region), workers are converging towards $(F1, F1)$, firm 1's wage setting behaviour will be reinforced, and lower w_1 value will receive stronger reinforcement as profit $(z_1 - w_1)$ received is higher. Decreasing w_1 could lead to regime switch. If a switch to regime 3 ($G3$) (green region) ensues, as workers learn to play $(F2, F2)$, lower w_2 will be reinforced more strongly, prompting another possible regime switch to $G2$. While workers' choice probabilities stabilizing in $G2$ (blue region) potentially prompt less incentive for further regime switch, stochasticity in workers' realized actions and equal attractiveness of $(F1, F2)$ and $(F2, F1)$ may lead to same action, such as both choosing firm 1, being realized, which could affect wages and incentivize regime switch. As a result, there may be constant regime change until workers' behaviour stabilizes and wages are driven down to 0.

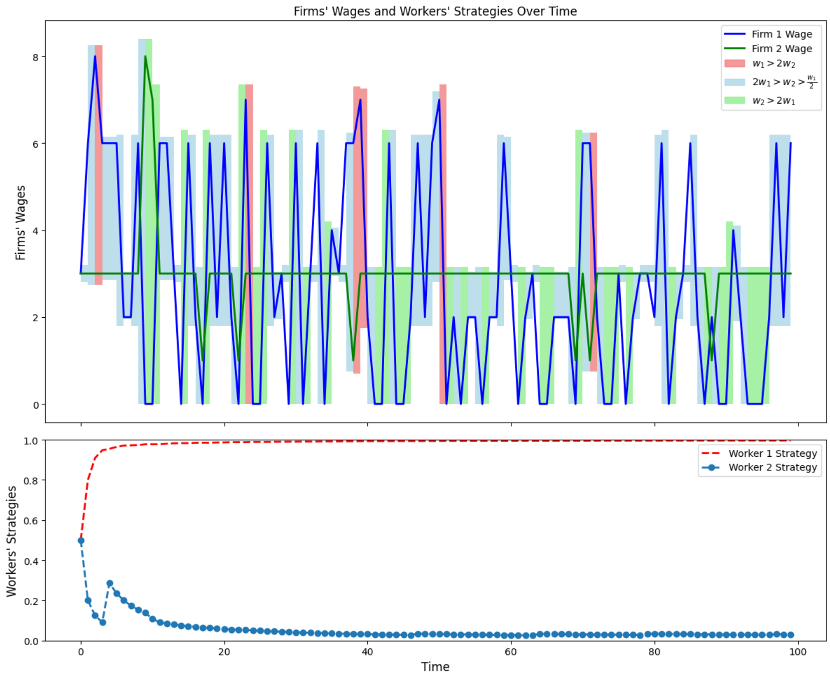


Figure 12: Changes in Wages and Workers' Choice Probabilities Over Time

Suppose I run the session for 10000 periods, and compute the conditional probability of switching from one regime to another based on the empirical frequency:

$$P(G_{t+1} = G_i | G_t = G_j) = \frac{n_{ij}}{n_i} \tag{31}$$

where G_t refers to the game played in period t, reflective of the regime workers are in; n_{ij} is the counts of regime change from i to j ($i, j \in \{1, 2, 3\}$), and n_i is the counts of being in regime i. Figure 13 shows there is higher probability of switching from any regime to the G_2 regime for this simulation. However, given the stochasticity of action realizations by the workers and the firms due to their probabilistic behaviours, the transition patterns could vary across different simulation sessions.

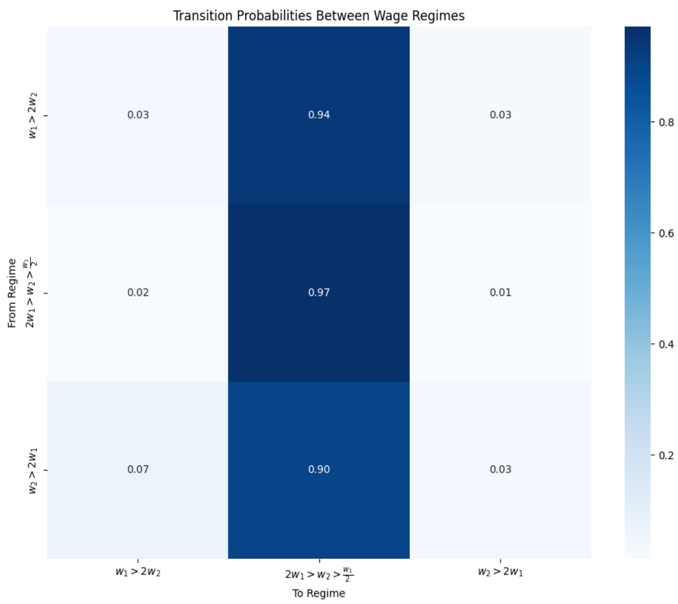


Figure 13: Transition Probility from Regime to Regime

Similar to static wage environment, dynamic wage environment could lead to prolonged periods of mismatch as workers need time to learn the NE(s). However, given wages vary over time, workers could be learning different sets of NE(s) as payoff structures shift over time. This could further contributes to mismatch as workers may need to overcome inertia from previous accumulated experiences as they adapt their search strategies amidst the new wage condition.

Another important question is equilibrium selection in the dynamic wage environment. Given that wages are pushed down to 0 in the long run, workers will be indifferent between selecting firm 1 or 2, and there will be a continuum of equilibria. Equilibrium selection is essentially path-dependent, and relies on what game is being played before wages hit 0. For one-to-one matching (i.e. $(F1, F2)$ or $(F2, F1)$) to be more likely, wage condition for $G2$ (Figure 9) needs to be maintained for substantial periods of time, such that pure NEs are possible outcomes where workers can learn to coordinate on.

2.2.1 Partial Recall of Experiences

Workers may not have perfect recall of past experiences, but firms are likely able to keep track of all past feedback by storing hiring information in a database that can be easily maintained and retrieved, and human resources tend to keep a record of wages offered and accepted. Therefore, I impose a memory decay factor on workers' side akin to Section 2.1.1 (Equation 17).

Since workers do not have perfect recall of past experiences, it is expected there will be slower convergence towards an equilibrium point. While longer learning trajectory could generate more instances of mismatch, the ability to forget "misaligned" market states as payoff structure shifts may also be an asset. As regime changes, the time lag for workers to start adapting to learn new set of NE(s) would be shorter under partial recall than perfect memory. However, discounting past experiences can also result in more fluctuations as workers are less locked-in and wages in response would be more volatile.

Proposition 5. *(Time to "Unlearn" the Past with Partial Memory) Upon change in wage regime at time $t = \tilde{t}$ from $G1$ to $G2$, there exists a time lag $(\bar{t} - \tilde{t})$ for workers to adapt from learning NE_1 to NE_2. The transition time between games in overcoming initial propensities can be captured by the base learning rate $(\frac{1}{\eta})$ and payoff dynamics in playing $G2$:*

$$(\bar{t} - \tilde{t}) \propto \frac{1}{\eta \cdot f(G2 \text{ payoff dynamics})} \tag{32}$$

Higher η, lower weight on past propensities, thus shorter time lag.

Proof. See Appendix A.2. □

In Figure 14, I show a simulation of changes in wages and choice probabilities over time when there exist a forgetting parameter (e.g. $\eta = 0.8$). As compared to perfect recall, such learning trajectory suggests that workers are less locked-in by past experiences and place greater weight on recent payoffs. They switch faster to playing the set of NE(s) supported by the wage regime they are in, thus can be perceived as more adaptive to new market conditions. While this is beneficial in stimulating switching in job applications and inducing coordination among workers when they could begin the application rounds with overcrowding at one of the firms. The pitfall of this is the volatility in choice probabilities. There could be many periods of mismatch as workers could forget previous propensities and they may not exhibit convergence to applying more to different firms even over a finite and extended period of time. (More simulation examples in Appendix B.1).

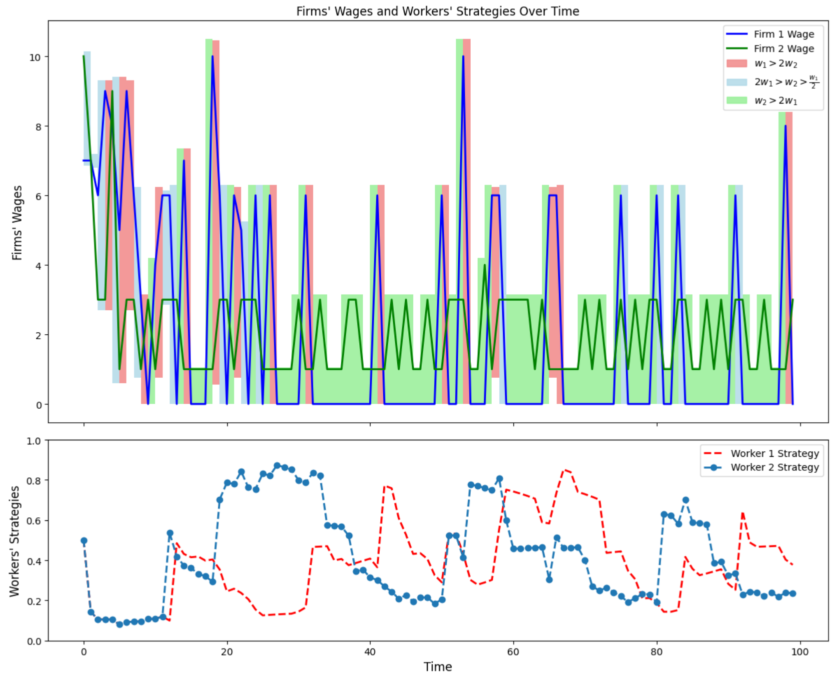


Figure 14: Changes in Wages and Workers' Choice Probabilities Over Time with $\eta = 0.8$

2.2.2 Policy Implications

Building on top of static wage environment, the dynamic wage setting portrays a slightly more realistic scenario where firms are also adaptive learners. As workers settle into more stable strategies and firms drive wages down to 0, there is a continuum of equilibria, and equilibrium selection would depend on the learning path. However, this long run behaviour is less "interesting" as workers and firms simply stop learning due to 0 reward, it is less informative of whether the eventual strategy is in fact optimal. The merit of this learning mechanism is it allows me to explore the evolution of workers' and firms' behaviours when wages remain positive and dynamically changing.

Since wages could change over time, workers may face 3 possible games, consisting of different sets of NE(s). Workers' choice probabilities could be noisier as they maneuver within different wage regimes. In order to nudge workers towards a more efficient outcome, policies could focus on maintaining a wage environment with pure NEs, where workers could converge to applying to different firms eventually. Furthermore, since initial experiences could create inertia for workers to switch to learning a different set of NE(s), if workers persistently overcrowd at firm 1 for a substantial number of periods, this will influence future behaviours of workers and affect their convergence rate. This could also disproportionately benefit firm that happens to be setting higher wages at the initial application rounds. Policy makers looking to decrease probability of overcrowding could consider measures for workers to "unlearn" past experiences. While forgetting can happen naturally, it is also possible to de-emphasize past instances by re-framing each application rounds as less analogous to one another and highlighting that market conditions has changed drastically from a distant past. Imposing higher discounting could improve transition to learning a new set of NE(s) that may be more efficient, but there is a trade-off between higher adaptivity to new situations and increased volatility of choice probabilities.

3 Search with Sequential Learning

In this section, I explore the second market structure, which resembles job search on online platforms where the wages are fully revealed. Firms post wages first, follow by workers. Workers are assumed to adopt a logit choice model and follows best-response (BR) dynamics highlighted by Fudenberg and Levine (1998) and Hopkins (1999). They consciously study the wage environment, and respond to the perceived strategy of their opponent given their own past experiences. This is feedback into firms' decision problem, where they set wages knowing how workers would formulate their strategies.

Under this framework, I seek to explore how past experiences influence workers' job search behaviour given that they observe the wages before deciding on their application strategies, and how this affects equilibrium selection as compared to the previous market structure, where workers only observe past feedback. In directed search literature, such as Wright et al. (2021), wages have a role of directing workers, higher wage would attract higher application rate, and workers' past experiences may dampen this effect. Experiences may also affect equilibrium selection in presence of multiple equilibria. Furthermore, this model can potentially shine a light on why workers hardly switch in applying for different jobs despite knowing wages beforehand.

3.1 Anchoring Bias from Long-term Experiences

Suppose each worker has a static bias from accumulated experiences exogenous to the current learning problem. This can be perceived as an anchoring bias based on historical events (Lieder et al. (2018)), or as social and cultural stereotypes (Langenhove and Harré (1994)) that is less shaped by short-term encounters but built more by long-term circumstances. In Barbulescu and Bidwell (2013), they shows that similarly qualified students in MBA program were found to display gender segregation in job applications, where women are less likely to apply for traditionally masculine jobs than men due to gender role stereotypes. Therefore, workers can be perceived to possess bias from long-term experiences, which affects their application choices.

Timing and Set-up.

1. Time is discrete, $t = \{0, 1, ..., T\}$. There are T finite generations of workers and firms.

2. At $t = 0$, two workers enter the market and two firms are offering two vacancies.

3. Firms observe exogenous realization of productivity, $\mathbf{z} = \{z_1, z_2\}$, $\mathbf{z} \in Z$, and post wages, $\mathbf{w} = \{w_1, w_2\}$. Firms set wages optimally period-by-period based on how workers are expected to react in each period.

4. Workers observe the wages, and decide their application strategies based on given wages, their beliefs about opponent's choices, and their own long-term experiences.

5. All parties observe a payoff at the end of the period, and then drop out of the market, never to return.

6. At $t = 1$, two new workers enter the market and firms offer two new vacancies.

7. Workers at $t = 1$ observed new set of wages, and based on realized actions at $t = 0$, they update their beliefs about opponent's choices based on empirical frequency of realized actions. They then update their choice probabilities.

8. This carries on for T periods.

Workers' side.

- **Actions.** Both workers are choosing between firm 1 and 2, $A^i = A^{-i} = \{F1, F2\}$.
- **Strategies.** Let $x = (x_1, x_2)$, $y = (y_1, y_2)$, where (x_1, x_2) and (y_1, y_2) denote the probability distribution over the two actions for worker 1 and 2 respectively. At time t, their choice probabilities are: $x_t = (x_{1t}, x_{2t})$ and $y_t = (y_{1t}, y_{2t})$, where $\sum x_t = 1$, $\sum y_t = 1$.
- **Beliefs.** Workers do not observe the exact strategies of their opponent, they can update their beliefs about the other worker's choice probability through realized actions. Let worker 2's belief about worker 1's choice probabilities be $u_t = (u_{1t}, u_{2t})$, where $u_{1t} + u_{2t} = 1$, and worker 1's belief about worker 2's choice probabilities be $v_t = (v_{1t}, v_{2t})$, $v_{1t} + v_{2t} = 1$.
- **Bias.** Worker 1's bias is $\alpha^i = (\alpha_1^i, \alpha_2^i)$, where α_1^i is the bias towards selecting firm 1, and α_2^i is the bias towards selecting firm 2. Correspondingly, worker 2's bias is $\alpha^{-i} = (\alpha_1^{-i}, \alpha_2^{-i})$.
- **Expected Payoffs.** Given beliefs about the other worker's choice probability, the expected payoffs for worker 1 and 2 when selecting an action at time t:

$$\pi_t^i(a_t^i, v_t) = W_t(a_t^i, F1)v_{1t} + W_t(a_t^i, F2)v_{2t}, \text{ where } a_t^i \in A^i \tag{33}$$

$$\pi_t^{-i}(a_t^{-i}, u_t) = W_t^T(a_t^{-i}, F1)u_{1t} + W_t^T(a_t^{-i}, F2)u_{2t}, \text{ where } a_t^{-i} \in A^i \tag{34}$$

and

$$W_t = \begin{pmatrix} \frac{w_{1t}}{2} & w_{1t} \\ w_{2t} & \frac{w_{2t}}{2} \end{pmatrix}, W_t^T = \begin{pmatrix} \frac{w_{1t}}{2} & w_{2t} \\ w_{1t} & \frac{w_{2t}}{2} \end{pmatrix} \tag{35}$$

- **Choice Rule.** Workers are reacting to wages, their beliefs about the other worker's strategy and their own bias. β is a rationality or sensitivity parameter to the expected payoffs. For worker 1:

$$x_t = BR_x(w_{1t}, w_{2t}, v_t) = \begin{cases} x_{1t} = \frac{\exp(\alpha_1^i + \beta\pi_t^i(F1, v_t))}{\exp(\alpha_1^i + \beta\pi_t^i(F1, v_t)) + \exp(\alpha_2^i + \beta\pi_t^i(F2, v_t))} \\ x_{2t} = \frac{\exp(\alpha_2^i + \beta\pi_t^i(F2, v_t))}{\exp(\alpha_1^i + \beta\pi_t^i(F1, v_t)) + \exp(\alpha_2^i + \beta\pi_t^i(F2, v_t))} = 1 - x_{1t} \end{cases} \tag{36}$$

For worker 2:

$$y_t = BR_y(w_{1t}, w_{2t}, u_t) = \begin{cases} y_{1t} = \frac{\exp(\alpha_1^{-i} + \beta\pi_t^{-i}(F1, u_t))}{\exp(\alpha_1^{-i} + \beta\pi_t^{-i}(F1, u_t)) + \exp(\alpha_2^{-i} + \beta\pi_t^{-i}(F2, u_t))} \\ y_{2t} = \frac{\exp(\alpha_1^{-i} + \beta\pi_t^{-i}(F2, u_t))}{\exp(\alpha_1^{-i} + \beta\pi_t^{-i}(F1, u_t)) + \exp(\alpha_2^{-i} + \beta\pi_t^{-i}(F2, u_t))} = 1 - y_{1t} \end{cases} \tag{37}$$

- **Expected Motion.** Learning dynamics are captured by changes in choice probabilities over time:

$$E(x_{t+1}) - x_t = BR_x(w_{1t}, w_{2t}, v_t) - x_t, E(y_{t+1}) - y_t = BR_y(w_{1t}, w_{2t}, u_t) - y_t \tag{38}$$

- **Updating Rule.** Following Hopkins (2002), workers' beliefs about their opponent's strategy is updated in each period after observing the realized actions by both workers in the previous period, and the weight attributed to initial beliefs is assumed to be 1.

$$u_{t+1} = \frac{(t+1)u_t + a_t^i}{t+2}, v_{t+1} = \frac{(t+1)v_t + a_t^{-i}}{t+2} \tag{39}$$

This can be expressed as running average of actions chosen in each period:

$$u_t = \frac{1}{t}\sum_{k=1} a_k^i, v_t = \frac{1}{t}\sum_{k=1} a_k^{-i} \tag{40}$$

In period $t+1$, payoff matrix will also be adjusted as firms learn. Therefore, workers' expected payoffs are updated given new payoffs, (W_{t+1}, W_{t+1}^T), and beliefs, (v_{t+1}, u_{t+1}).

This updating rule effectively encapsulates the aggregate effect of opponent's bias and their responses to wage changes. The simplicity of updating based on what one observes avoid explicit modelling of opponent's internal decision-making process and could be more practical.

On workers' side, the subgame equilibrium solution(s) resemble that of quantal response equilibrium (QRE), where workers eventually form correct beliefs about opponents' strategies (McKelvey and Palfrey (1995)). For a give set of wages (w_1, w_2), the presence of multiple equilibria depends on the sensitivity parameter, β. Higher β implies one is more sensitive to changes in expected payoffs and there is less noise in strategies, thus close to pure strategies could exist; whereas as β tends to 0, a unique mixed equilibrium emerge. Suppose workers do not possess bias from long-term experiences, $\alpha^i = \alpha^{-i} = 0$, when firms are homogenous and wages equalize, $w_1 = w_2$, where $w_1 > 0$, $w_2 > 0$, Figure 22 illustrates workers' behaviours given variations in β, where there can be multiple equilibria:

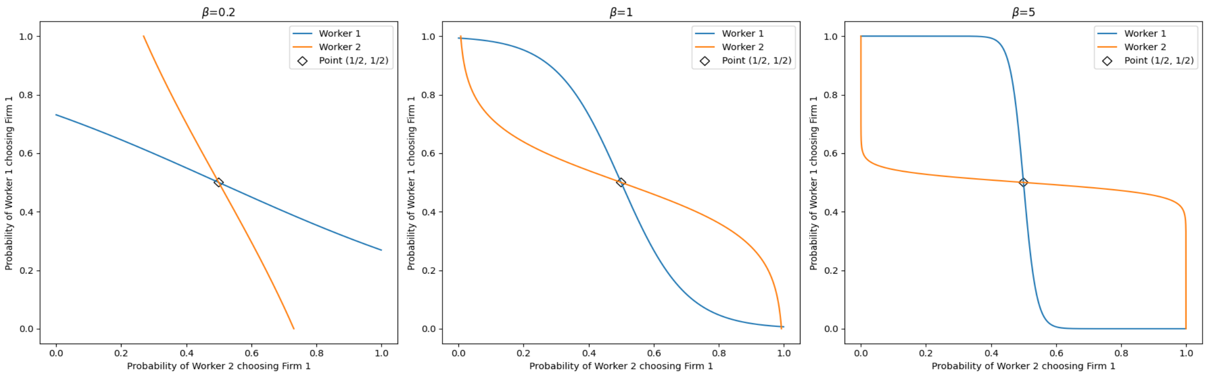


Figure 15: Workers' Response Functions to Given Set of Wages

Firms' side. Instead of using reinforcement learning approach, firms are setting wages optimally in expectation of workers' reaction. They essentially keeps a tally of workers' observed actions, and update workers' belief about each other, and set wages in each period knowing how workers would respond. As a result, they face maximization problems:

$$\max_{w_{1t}} (1 - (1 - x_{1t})(1 - y_{1t}))(z_1 - w_{1t}) \text{ s.t. } z_1 \geq w_{1t} \geq 0 \tag{41}$$

$$\max_{w_{2t}} (1 - x_{1t}y_{1t})(z_2 - w_{2t}) \text{ s.t. } z_2 \geq w_{2t} \geq 0 \tag{42}$$

Firms' FOCs:

$$1 - (1 - x_{1t})(1 - y_{1t}) = (z_1 - w_{1t})[(1 - y_{1t})\frac{dx_{1t}}{dw_{1t}} + (1 - x_{1t})\frac{dy_{1t}}{dw_{1t}}] \tag{43}$$

$$1 - x_{1t}y_{1t} = (z_2 - w_{2t})[y_{1t}\frac{dx_{1t}}{dw_{2t}} + x_{1t}\frac{dy_{1t}}{dw_{2t}}] \tag{44}$$

Workers' FOCs:

$$\frac{dx_{1t}}{dw_{1t}} = x_{1t}(1 - x_{1t})(\beta\frac{1}{2}v_{1t} + \beta v_{2t}) \tag{45}$$

$$\frac{dx_{1t}}{dw_{2t}} = x_{1t}(1 - x_{1t})(\beta v_{1t} + \beta\frac{1}{2}v_{2t}) \tag{46}$$

$$\frac{dy_{1t}}{dw_{1t}} = y_{1t}(1 - y_{1t})(\beta\frac{1}{2}u_{1t} + \beta u_{2t}) \tag{47}$$

$$\frac{dy_{1t}}{dw_{2t}} = y_{1t}(1 - y_{1t})(\beta u_{1t} + \beta\frac{1}{2}u_{2t}) \tag{48}$$

Combining them, the wage equations are:

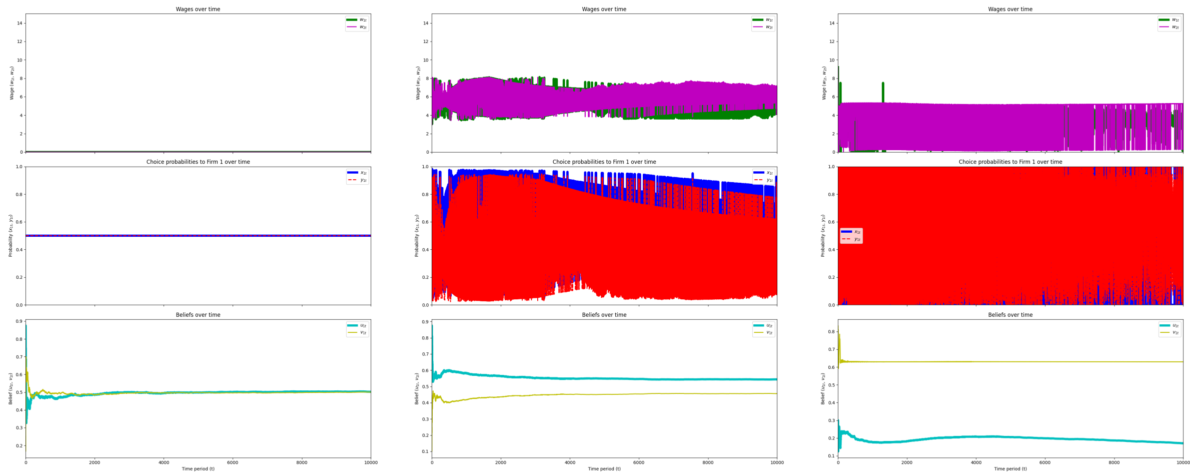
$$w_{1t} = \max[z_1 - \frac{1-(1-x_{1t})(1-y_{1t})}{(1-y_{1t})x_{1t}(1-x_{1t})(\beta\frac{v_{1t}}{2}+\beta(1-v_{1t})) + (1-x_{1t})y_{1t}(1-y_{1t})(\beta\frac{u_{1t}}{2}+\beta(1-u_{1t}))}, 0] \tag{49}$$

$$w_{2t} = \max[z_2 - \frac{1-x_{1t}y_{1t}}{y_{1t}x_{1t}(1-x_{1t})(\beta v_{1t}+\beta\frac{(1-v_{1t})}{2}) + x_{1t}y_{1t}(1-y_{1t})(\beta u_{1t}+\beta\frac{(1-u_{1t})}{2})}, 0] \tag{50}$$

Algorithm 1 Sequential Learning Algorithm

Initialize for $t = 0$, set α_1^i, α_1^{-i}, u_0^i, v_0^i, x_0, y_0; Compute w_{10}, w_{20}.
for one session **do**
  Loop the following
  for 100000 time periods **do**
    Loop for each time period
    for all firms **do**
      Loop for 500 iterations
      for one iteration **do**
        Guess a set of wages, $(w_{1t}^{\text{Guess}}), (w_{2t}^{\text{Guess}})$.
        Compute workers' reaction based on equations (36) and (37) to obtain $(x_{1t}^{\text{Potential}}, y_{1t}^{\text{Potential}})$.
        Based on workers' potential application rates, compute wages using equations (49) and (50) to obtain $(w_{1t}^{\text{Potential}}, w_{2t}^{\text{Potential}})$.
        Compare $(w_{1t}^{\text{Potential}}, w_{2t}^{\text{Potential}})$ with $(w_{1t}^{\text{Guess}}), (w_{2t}^{\text{Guess}})$.
        If wage guess differs from the wage that would be set optimally given workers' response, the guess is incorrect, reiterate the process and start with a different guess.
        Make a smaller adjustment from previous guess if it was close, otherwise, make a bigger adjustment.
      end for one iteration
      Set wages (w_{1t}, w_{2t}) to be when the difference between the guessed wages and potential wages is negligible, such that these are the optimal wages in each period given workers' reaction.
    end for firms
    for all workers **do**
      Observe (w_{1t}, w_{2t}), compute x_t and y_t given u_t and v_t using equations (36) and (37).
      Generate a choice of action from $(F1, F2)$ for each worker based on x_t and y_t, which will be the realized workers' choice in period t.
    end for workers
    for reward generation and updating **do**
      Given realized workers' actions (a_t^i, a_t^{-i}) and firms' wages (w_{1t}, w_{2t}), compute the rewards for all agents.
      Workers' beliefs about each other (u_{t+1} and v_{t+1}) are updated based on equation (39) for use in the following period.
    end for one period
  end for all periods
end for all sessions
return results $(x_{1t}, x_{2t}, y_{1t}, y_{2t}, u_{1t}, u_{2t}, v_{1t}, v_{2t}, w_{1t}, w_{2t})$ for all periods.

Based on Algorithm 1, I show in Figure 16 one simulation session, which records changes in wages over time (top panel), workers' choice probability of firm 1 (middle panel) and beliefs of other worker's choice probability to firm 1 (bottom panel). At low β (Figure 16a), workers' choice probabilities stays at 0.5 and beliefs converge to 0.5, thus wages are 0 based on equation (49) and (50). For intermediary β (Figure 16b), wages are positive. While beliefs about opponents are seemingly stable and converge to a point slightly different from random, wages and workers' choice probabilities fluctuate rapidly. In the final case of high β (Figure 16c), wages are also positive. There is convergence of beliefs to opponents applying with higher probability to different firms, but more drastic fluctuations in firms' wage setting and workers' choice probabilities are observed. (Multiple Sessions in Appendix B.2)



(a) $\beta = 0.2$ (b) $\beta = 1.0$ (c) $\beta = 5.0$

Figure shows wages, choice probabilities and belief evolution for different $\beta = \{0.2, 1.0, 5.0\}$, given $z_1 = z_2 = 10$, $t = 10000$, $\alpha_1^i = 0$, $\alpha_1^{-i} = 0$, $u_{10} = 0.5$, $v_{10} = 0.5$, and no smoothing.

Figure 16: Changes in Wages and Workers' Choice Probabilities Over Time

In presence of multiple equilibria when β is sufficiently large, tiny shifts in beliefs could push the system across boundaries between different basins of attractions, thereby contributing to large fluctuations in firms' and workers' behaviours. This could be the main reason for the apparent jumps across different convergence paths. However, since Algorithm 1 start with random guesses of wages each period (Line 9), this may also contribute to large fluctuations in wages and workers' choice probabilities as values in different basin of attractions may be selected each time. This method effectively explore local optima points, where optimal wages are found near the neighbourhood of initial guesses in each period. Therefore, even with same belief inputs (u_t, v_t), it is not guaranteed that the same convergence path is selected.

In Algorithm 2, I modify the firms' behaviours (Line 6-17 in Algorithm 1) slightly to put some structure on wage guesses to check if fluctuations could be mitigated slightly if next period wage guesses are made on the basis of previous period optimal wages. When there are no smoothing, the simulation results are similar to before, where small changes in beliefs could lead to large fluctuations in choice probabilities and wages. Smoothing the variables would allow a clearer illustration of the underlying trend. Therefore, in Figure 17, suppose a moving average for wages, choice probabilities and beliefs are taken. This does not make any difference for small β, where there is convergence to the unique equilibrium, but it does affect moderate β.

Algorithm 2 Firms' Wage-setting with Small Adjustment from Previous Period

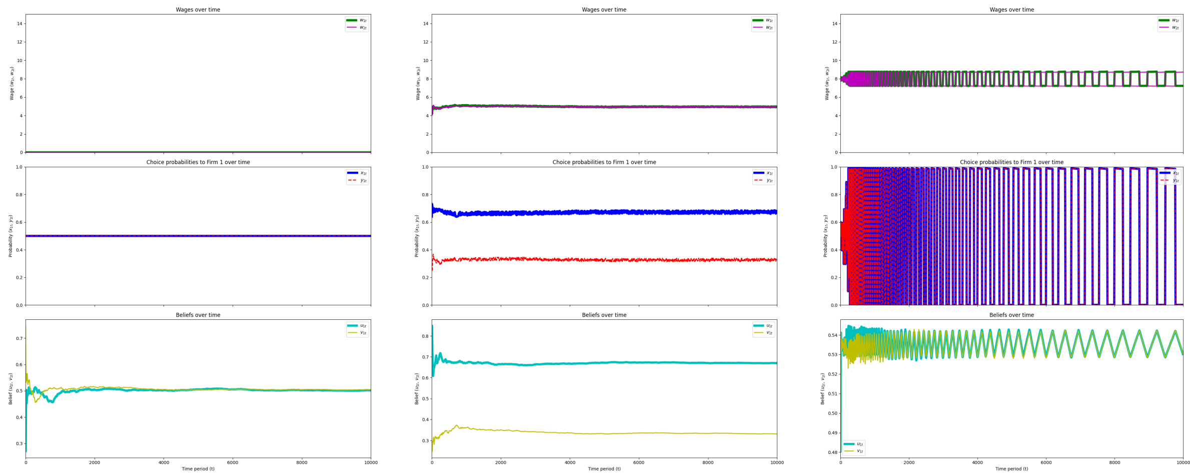
```
for all firms do
    Check if previous wages are still optimal given updated workers' beliefs based on past period realized actions.
    for a set of (w_{1(t-1)}, w_{2(t-1)}) do
        Compute workers' (x_{1t}^{Potential}, y_{1t}^{Potential}) based on equations (36) and (37).
        Based on workers' potential application rates, compute corresponding wages, (w_{1t}^{Potential}, w_{2t}^{Potential}), using equations (49) and (50).
        Compare (w_{1t}^{Potential}, w_{2t}^{Potential}) with (w_{1(t-1)}), (w_{2(t-1)}).
    end for checking previous wages
    If previous wages is optimal, set (w_{1t}, w_{2t}) = (w_{1(t-1)}), (w_{2(t-1)}); adjust otherwise.
    for previous wages no longer optimal do
        Loop for 500 iterations
        for first iteration do
            Compute difference between (w_{1t}^{Potential}, w_{2t}^{Potential}) and (w_{1(t-1)}), (w_{2(t-1)}).
            Make new wages guesses in small increment if wage difference is small, otherwise make a bigger adjustment:
```

$$w_{1t}^{\text{Guess}} = w_{1(t-1)} + \text{increment} * \text{wage difference} \tag{51}$$

$$w_{2t}^{\text{Guess}} = w_{2(t-1)} + \text{increment} * \text{wage difference} \tag{52}$$

```
            Given wage guesses, compute workers' reaction, (x_{1t}^{Potential}, y_{1t}^{Potential}), based on potential reaction, compute potential wages, (w_{1t}^{Potential}, w_{2t}^{Potential}).
            Compare (w_{1t}^{Potential}, w_{2t}^{Potential}) and (w_{1t}^{Guess}, w_{2t}^{Guess}), if far apart, the guess is incorrect, start a new guess in subsequent iteration based on their differences.
        end for first iteration
        Subsequent iterations make small increments from previous guesses. The possible bound for adjustment is smaller for first 5 iterations, bigger for the next 15, and even larger afterwards. This helps to ensure convergence within 500 iterations.
    end for all iterations
    Set wages (w_{1t}, w_{2t}) to be when the difference between guessed wages and potential wages is negligible.
end for firms
```

In Figure 17b, wages (top panel), workers' choice probabilities (middle panel) and beliefs (bottom panel) show more stabilized values. Workers are predominantly in one of the basins of attraction, where worker 1 applies more to firm 1 and worker 2 to firm 2. However, as workers become more sensitive to payoffs at higher β (Figure 17c, perturbations in beliefs again land workers and firms in different basin of attractions, leading to fluctuations in wages and workers' choice probabilities. When sensitivity to payoff is high, the system spends little time transitioning between equilibria, so the oscillation is rapid; and when sensitivity to payoff is moderate, the system may still spend time transitioning between different basin of attractions, but it could have spent more time in one than the other on average. Theoretically, if the beliefs over time stabilizes and firmly fall within the basin of attraction for one equilibrium, then it is expected that the system would stay in that equilibrium, but this may not be witnessed in the simulation.



(a) $\beta = 0.2$ (b) $\beta = 1.0$ (c) $\beta = 5.0$

Figure shows moving average of wages, choice probabilities and belief evolution over last 10 periods for different $\beta = \{0.2, 1.0, 5.0\}$, given $z_1 = z_2 = 10$, $t = 10000$, $\alpha_1^i = 0$, $\alpha_1^{-i} = 0$, $u_{10} = 0.5$, $v_{10} = 0.5$.

Figure 17: Changes in Wages and Workers' Choice Probabilities Over Time (Algorithm 2)

To formally define the equilibrium that the system is converging to:

Proposition 6 (Stabilization of the System)**.** *In the long run, workers and firms' behaviours converge to form equilibria similar to the Quantal Response Equilibria (QRE).*

- *(Workers' Beliefs.) Workers' beliefs of opponents' choices converge to long run average of observed actions.*

$$u^* = \lim_{t\to\infty} u_t = \mathbb{E}(a^i), v^* = \lim_{t\to\infty} v_t = \mathbb{E}(a^{-i}) \tag{53}$$

- *(Workers' Strategies.) Workers' strategies converge to (x^*, y^*) for stabilized beliefs.*

$$x^* = \begin{cases} x_1 = \frac{\exp(\alpha_1^i + \beta\pi^i(F1, v^*))}{\exp(\alpha_1^i + \beta\pi^i(F1, v^*)) + \exp(\alpha_2^i + \beta\pi^i(F2, v^*))} \\ x_2 = 1 - x_1 \end{cases} \tag{54}$$

$$y^* = \begin{cases} y_1 = \frac{\exp(\alpha_1^{-i} + \beta\pi_t^{-i}(F1, u^*))}{\exp(\alpha_1^{-i} + \beta\pi^{-i}(F1, u^*)) + \exp(\alpha_2^{-i} + \beta\pi^{-i}(F2, u^*))} \\ y_2 = 1 - y_1 \end{cases} \tag{55}$$

where π^i and π^{-i} are expected payoffs computed using W^*, u^* and v^*.

- *(Firms' Wages.) Wages (w_1^*, w_2^*) are determined given (x^*, y^*, u^*, v^*).*

$$w_1^* = \max[z_1 - \frac{1 - (1 - x_1^*)(1 - y_1^*)}{(1 - y_1^*)x_1^*(1 - x_1^*)(\beta\frac{v_1^*}{2} + \beta(1 - v_1^*)) + (1 - x_1^*)y_1^*(1 - y_1^*)(\beta\frac{u_1^*}{2} + \beta(1 - u_1^*))}, 0] \tag{56}$$

$$w_2^* = \max[z_2 - \frac{1 - x_1^* y_1^*}{y_1^* x_1^*(1 - x_1^*)(\beta v_1^* + \beta\frac{(1 - v_1^*)}{2}) + x_1^* y_1^*(1 - y_1^*)(\beta u_1^* + \beta\frac{(1 - u_1^*)}{2})}, 0] \tag{57}$$

and payoff matrix:

$$W^* = \begin{pmatrix} \frac{w_1^*}{2} & w_1^* \\ w_2^* & \frac{w_2^*}{2} \end{pmatrix} \tag{58}$$

- *(Consistency Condition.) In the equilibrium, $u^* = x^*$, $v^* = y^*$, beliefs matches the actual choice probabilities.*

- *(Equilibrium Multiplicity.) For given set of parameters, multiple equilibria could exist, defined by different sets of (x^*, y^*, w_1^*, w_2^*).*

- *(Long Run Fluctuations.) In presence of multiple equilibria, different local optima may be achieved. There could be jumps between different basins of attractions before stabilization.*

Proof. Based on equation (40), as $t \to \infty$, workers' beliefs of opponents' choices converges to the running average of observed actions (shown by 53). Given u^* and v^*, workers' strategies (x^*, y^*) are computed based on equations (54) and (55). Firms' equilibrium wage-setting (w_1^*, w_2^*) given workers' response is determined by equations (49) and (50).

In the equilibrium, belief consistency needs to be achieved, such that their beliefs match the choice probabilities. Given the learning dynamics, the system would converge to equilibrium defined by $(x_1^*, y_1^*, w_1^*, w_2^*)$ as $t \to \infty$:

$$x_1^* = \frac{\exp(\alpha_1^i + \beta\pi^i(F1, y^*))}{\exp(\alpha_1^i + \beta\pi^i(F1, y^*)) + \exp(\alpha_2^i + \beta\pi_t^i(F2, y^*))} \tag{59}$$

$$y_1^* = \frac{\exp(\alpha_1^{-i} + \beta\pi_t^{-i}(F1, x^*))}{\exp(\alpha_1^{-i} + \beta\pi_t^{-i}(F1, x^*)) + \exp(\alpha_2^{-i} + \beta\pi_t^{-i}(F2, x^*))} \tag{60}$$

$$w_1^* = \max[z_1 - \frac{1-(1-x_1^*)(1-y_1^*)}{(1-y_1^*)x_1^*(1-x_1^*)(\beta\frac{y_1^*}{2} + \beta(1-y_1^*)) + (1-x_1^*)y_1^*(1-y_1^*)(\beta\frac{x_1^*}{2} + \beta(1-x_1^*))}, 0] \tag{61}$$

$$w_2^* = \max[z_2 - \frac{1-x_1^* y_1^*}{y_1^* x_1^*(1-x_1^*)(\beta y_1^* + \beta\frac{(1-y_1^*)}{2}) + x_1^* y_1^*(1-y_1^*)(\beta x_1^* + \beta\frac{(1-x_1^*)}{2})}, 0] \tag{62}$$

These equations collectively define the equilibrium. Multiple solutions could exist. □

An important distinction between the equilibrium achieved via this learning dynamics as compared to traditional QRE is that workers and firms are not forward-looking. Workers are myopic in a sense that they only update their beliefs based on past observations of opponents' actions. They do not strategize given how their current actions might influence how the other worker would react. Similarly, for the firms, they also have limited foresight. While they take into account workers' reaction to the wages posted in the current period and optimize based on potential response, they do not account for how wages might affect belief evolution over time. Therefore, it is entirely plausible that this myopic learning dynamics converge only to a subset of QRE that can be identified in forward-looking scenarios.

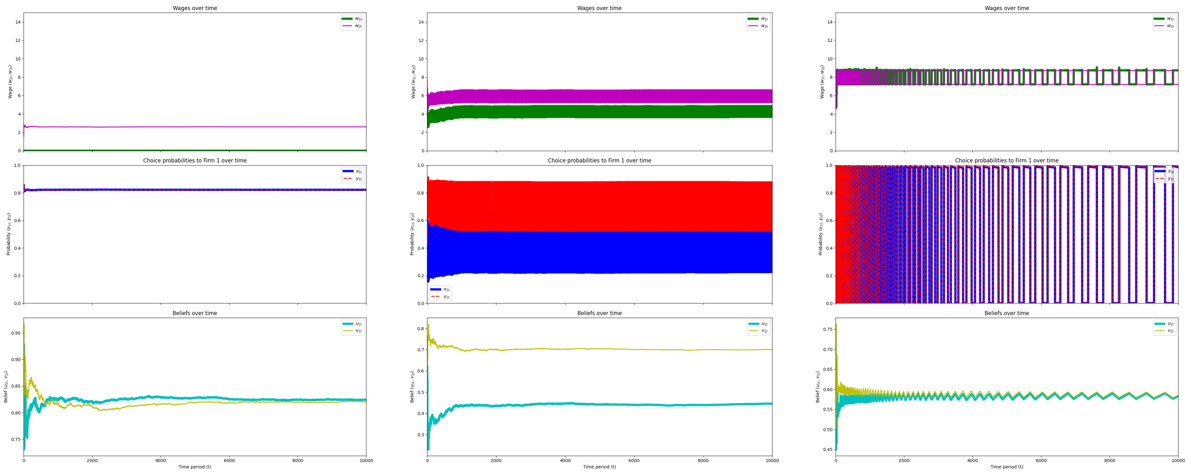
While it is straightforward that there is convergence to unique equilibrium when β is low, and to the asymmetric equilibria when β is high. In the succeeding case, it remain hard to distinguish exactly which of the two asymmetric equilibria will the workers coordinate on.

Therefore, the next step is to investigate the impact of bias from long-term experiences (α^i, α^{-i}). Rewriting workers' choice probabilities to firm 1:

$$x_{1t} = \frac{1}{1 + \exp(-[(\alpha_1^i - \alpha_2^i) + \beta(\pi_t^i(F1, v_t) - \pi_t^i(F2, v_t))]])} \tag{63}$$

$$y_{1t} = \frac{1}{1 + \exp(-[(\alpha_1^{-i} - \alpha_2^{-i}) + \beta(\pi_t^{-i}(F1, u_t) - \pi_t^{-i}(F2, u_t))]]} \tag{64}$$

Based on the equations, relative bias matters more than the absolute bias. Introducing some bias towards one firm over the other could affect choice probabilities. In the following simulation using Algorithm 2 with no smoothing, I showcase an example when both workers have long-term experiences that create bias towards firm 1 (i.e. $\alpha_1^i - \alpha_2^i > 0$).



(a) $\beta = 0.2$ (b) $\beta = 1.0$ (c) $\beta = 5.0$

Figure shows wages, choice probabilities and belief evolution for different $\beta = \{0.2, 1.0, 5.0\}$, given $z_1 = z_2 = 10$, $t = 10000$, $\alpha_1^i = 2$, $\alpha_1^{-i} = 2$, $u_{10} = 0.5$, $v_{10} = 0.5$, no smoothing.

Figure 18: Changes in Wages and Workers' Choice Probabilities Over Time with Bias

In Figure 18, at low β, workers converge to a mixed strategy of applying with higher probability to firm 1; At intermediary β, there is convergence to workers applying with different probability to different firms; Lastly, at high β, there is more fluctuations in choice probability and no apparent convergence over time. Since both workers are slightly biased to firm 1 by default, their search strategies can be more skewed towards firm 1. This can be beneficial for equilibrium selection in presence of multiple equilibria, particularly when workers are only moderately sensitive to expected payoffs. As shown in Figure 18b, as beliefs stabilize more, there is an underlying trend of workers applying with higher probabilities to different firms. However, if workers are highly sensitive to changes to expected payoffs (Figure 18c), then small experience bias may not have much effect on assisting convergence to a single equilibrium, perturbations in beliefs could still push the system into different basins of attraction, leading to fluctuations between different convergence paths. (Workers' side illustration in Appendix B.3)

Proposition 7 (Impact of Long-term Experiences on Equilibrium Selection)**.** *The larger the relative bias towards a firm (i.e. $|\alpha_1^i - \alpha_2^i| > 0$) than expected payoff difference between the two options (i.e. $\Delta\pi^i = |\pi^i(F1) - \pi^i(F2)| > 0$), the more likely long-term experiences play a significant role in facilitating equilibrium selection.*

- *Strong Bias from Experience ($|\alpha_1^i - \alpha_2^i| > \beta|\Delta\bar{\pi}^i|$): Convergence to unique equilibrium.*
- *Moderate Bias from Experience ($\beta|\Delta\bar{\pi}^i| > |\alpha_1^i - \alpha_2^i| > \beta|\Delta\underline{\pi}^i|$): Likely to sink into a basin of attraction and converge to one of the equilibria given equilibrium multiplicity.*
- *Weak Experience Bias ($\beta|\Delta\underline{\pi}^i| > |\alpha_1^i - \alpha_2^i|$): Likely to fluctuate between different convergence paths given equilibrium multiplicity.*

Proof. From equation (63) and (64), let $\Delta\pi^i = \pi_t^i(F1, v_t) - \pi_t^i(F2, v_t)$, $\Delta\pi^{-i} = \pi_t^{-i}(F1, v_t) - \pi_t^{-i}(F2, v_t)$. If $|\alpha_1^i - \alpha_2^i| > \beta|\Delta\bar{\pi}^i|$, relative bias from experience would have a larger influence on x_{1t} and y_{1t}, and beliefs about opponents would have less impact, such that

$$x_{1t} \approx \frac{1}{1+\exp(-(\alpha_1^i - \alpha_2^i))}, y_{1t} \approx \frac{1}{1+\exp(-(\alpha_1^{-i} - \alpha_2^{-i}))} \tag{65}$$

In the equilibrium, x^* and y^* converge to constant values, similarly for w_1^* and w_2^*, following equations (61) and (62).

If $\beta|\Delta\bar{\pi}^i| > |\alpha_1^i - \alpha_2^i| > \beta|\Delta\underline{\pi}^i|$, x_{1t} and y_{1t} are influenced by both relative bias from experience and expected payoff difference. While it still may be difficult to determine which equilibrium the system will converge to in presence of multiple equilibria, the relative bias could dampen the fluctuations across different equilibria. Since workers' sensitivities to expected payoff, including beliefs about opponents, are moderate, the small perturbations in beliefs could still allow x_{1t} and y_{1t} to quickly sink into a basin of attraction, resulting in convergence to one of the equilibria.

Lastly, if $\beta|\Delta\underline{\pi}^i| > |\alpha_1^i - \alpha_2^i|$, relative expected payoff, including belief perturbations, would have a larger influence on x_{1t} and y_{1t} than relative bias, such that

$$x_{1t} \approx \frac{1}{1+\exp(-[\beta(\pi_t^i(F1, v_t) - \pi_t^i(F2, v_t))])}, y_{1t} \approx \frac{1}{1+\exp(-[\beta(\pi_t^{-i}(F1, u_t) - \pi_t^{-i}(F2, u_t))]} \tag{66}$$

It is expected there will be more fluctuations as there can be jumps between different convergence paths. □

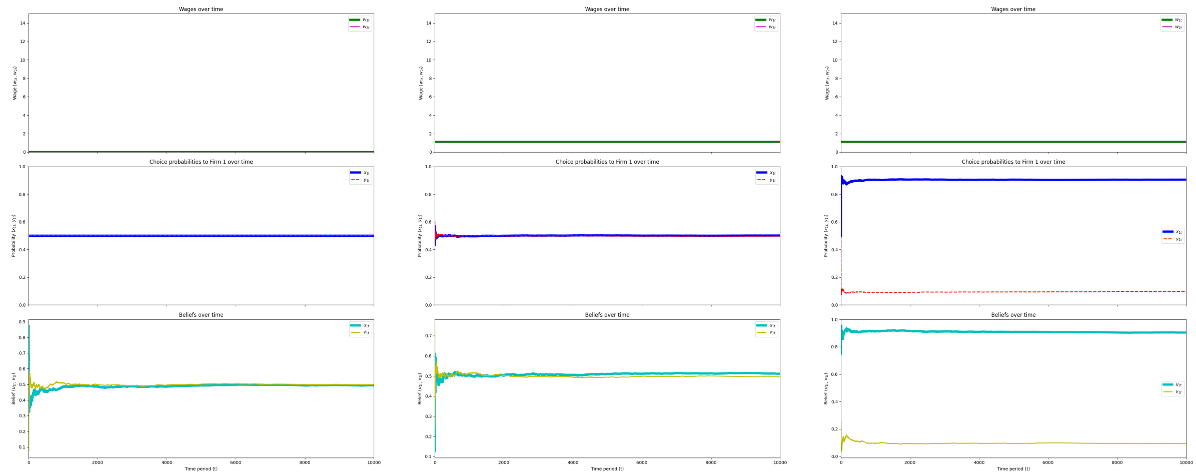
A possible question that could arise is that the above wage-setting approach focuses on local optimization, where firms adjust wages based on past periods, creating some path dependency, could it be possible for firms to consider all possible wage and workers' reaction combinations and select an optimal wage?

Algorithm 3 Firms' Wage-setting using Grid Search Approach

for all firms **do**
    Conduct coarse search, followed by more refined search
    for coarse search **do**
        Set two arrays of wages bounded by $[0, z_j]$, divide the range into 10 evenly spaced values, such that there can be finite pairs of $(w_{1t}^{\text{Coarse}}), (w_{2t}^{\text{Coarse}})$.
        Compute workers' reaction based on equations (36) and (37) to obtain $(x_{1t}^{\text{Potential}}, y_{1t}^{\text{Potential}})$ to each pair of $(w_{1t}^{\text{Coarse}}), (w_{2t}^{\text{Coarse}})$.
        Based on workers' potential application rates, compute firms' payoffs using equations (41) and (42).
        Find the wage pair that lead to highest profit.
    end for coarse search
    for refined search **do**
        Take the best pair of wages previously identified, $(w_{1t}^{\text{Coarse}}, w_{2t}^{\text{Coarse}})$, and create a finer search window (i.e. ± 0.5), dividing this range into 20 equal units.
        Search all combination of wages in this range to find the pair that maximizes individual firm profit, $(w_{1t}^{\text{Refined}}, w_{2t}^{\text{Refined}})$.
    end for refined search
    Set the wages $(w_{1t}, w_{2t}) = (w_{1t}^{\text{Refined}}, w_{2t}^{\text{Refined}})$.
end for firms

Algorithm 3 portrays a global maximization approach, which effectively ask the firms to pick the optimal wage across the full range of wage possibilities and potential worker reactions in each period.

When $\beta = 0.2$, there is a unique equilibrium, therefore, all three algorithms would lead to the same convergence results. The grid search approach could lead to more substantial difference when there exist multiple equilibria, such as in Figure 23b and 23c. Firstly, it is noticeable that the wages selected can be lower than that under the previous local optimization approach, implying possibility of higher monopsony power. It also shows that for $\beta = 1.0$, workers converge to a strategy that is close to random search; and for $\beta = 5.0$, workers' converge to almost pure coordination on applying with high probability to different firms. This draws some resemblance to the results in Lu (2024): For relatively high sensitivity to expected payoff, workers are more likely to coordinate on applying to different firms, resulting in lower wages and higher profits for the firms; whereas for workers having lower sensitivity to wages, there is higher risk of firms not meeting at least one worker, thus asymmetric strategies correspond to higher wages as compared to symmetric strategies. As a result, when firms maximize their profits by setting lower wages, workers would converge to asymmetric strategies for high β, and to symmetric strategies for low β, given that multiple equilibria exist. Whilst simulation results in Figure 23c show convergence to a single asymmetric equilibrium under this approach, it is necessary to note that the two asymmetric equilibria are equally attractive, and which one is selected would be contingent on the learning path.



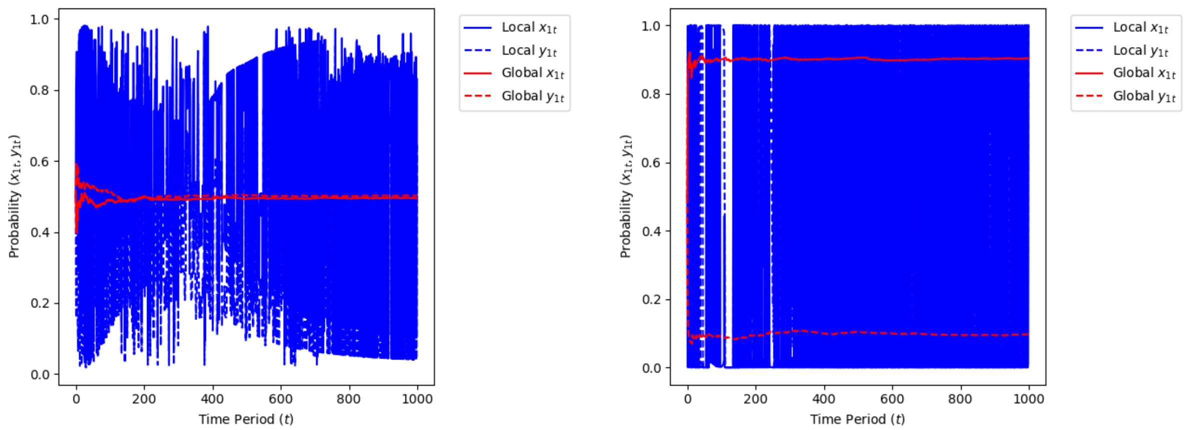
(a) $\beta = 0.2$ (b) $\beta = 1.0$ (c) $\beta = 5.0$

Figure shows wages, choice probabilities and belief evolution for different $\beta = \{0.2, 1.0, 5.0\}$, given $z_1 = z_2 = 10$, $t = 10000$, $\alpha_1^i = 0$, $\alpha_1^{-i} = 0$, $u_{10} = 0.5$, $v_{10} = 0.5$, no smoothing.

Figure 19: Changes in Wages and Workers' Choice Probabilities Over Time (Algorithm 3)

Compare the evolution of workers' choice probabilities to firm 1 using Algorithm 2 and 3 for the cases with multiple equilibria. In Figure 20, I show that there could be much more fluctuations in workers' strategies under the local optimization approach, but workers can seemingly converge to the asymmetric equilibria; and under the global optimization approach, there is less fluctuations, but workers are more likely to converge to the symmetric equilibrium when workers' sensitivity to expected payoff is moderate, and asymmetric equilibria when sensitivity is high. The results are heavily reliant on firms' behaviours. The second approach can implement a clearer selection of one of the asymmetric equilibria, but it would require firms to have perfect information of the profit landscape for any wage combinations, and the computational power required to evaluate the expected payoffs. If long-term experiences are involved, the evaluation of $|\alpha_1^i - \alpha_2^i|$ against $\beta|\Delta\pi^i|$ remains the same as Proposition 7. However, the difference is that there could be clearer convergence to one set of workers' strategies for all three cases listed. (Example in Appendix B.4)

The global maximization approach is more likely to propel convergence towards a single equilibrium



(a) $\beta = 1.0$ (b) $\beta = 5.0$

Figure shows comparison of choice probabilities evolution for the two wage-setting approach, denoted as local and global respectively. Given $\beta = \{1.0, 5.0\}$, $z_1 = z_2 = 10$, $t = 1000$, $\alpha_1^i = 0$, $\alpha_1^{-i} = 0$, $u_{10} = 0.5$, $v_{10} = 0.5$, no smoothing.

Figure 20: Comparing Workers' Strategies For the Two Wage-setting Approach

when there are multiple. More efficient outcome where workers converge to asymmetric strategies are more likely to be achieved when β is high and long-term experiences have small influence. However, wages tend to be low, implying high monopsony power.

In order to analyze the stability of the equilibrium outcome(s), I transform the the system from discrete to continuous time.

For workers' side, discrete time equations are reinstated below:

$$u_{t+1} = \frac{(t+1)u_t + a_t^i}{t+2}, v_{t+1} = \frac{(t+1)v_t + a_t^{-i}}{t+2} \tag{67}$$

$$x_{1t} = \frac{1}{1 + \exp(-[(\alpha_1^i - \alpha_2^i) + \beta(w_{1t} - \frac{w_{2t}}{2} - (\frac{w_{1t}}{2} + \frac{w_2}{2})v_{1t}])} \tag{68}$$

$$y_{1t} = \frac{1}{1 + \exp(-[(\alpha_1^{-i} - \alpha_2^{-i}) + \beta(w_{1t} - \frac{w_{2t}}{2} - (\frac{w_{1t}}{2} + \frac{w_{2t}}{2})u_{1t})]} \tag{69}$$

In continuous time:

$$\dot{u}_t = a_t^i - u_t, \dot{v}_t = a_t^{-i} - v_t \tag{70}$$

$$\dot{x}_{1t} = \frac{dx_{1t}}{dw_{1t}}\dot{w}_{1t} + \frac{dx_{1t}}{dw_{2t}}\dot{w}_{2t} + \frac{dx_{1t}}{dv_{1t}}\dot{v}_{1t} \tag{71}$$

$$\dot{y}_{1t} = \frac{dy_{1t}}{dw_{1t}}\dot{w}_{1t} + \frac{dy_{1t}}{dw_{2t}}\dot{w}_{2t} + \frac{dy_{1t}}{dv_{1t}}\dot{v}_{1t} \tag{72}$$

From equations (71) and (72):

$$\dot{x}_{1t} = x_{1t}(1 - x_{1t})\beta[(1 - \frac{v_{1t}}{2})\dot{w}_{1t} - (\frac{1}{2} + \frac{v_{1t}}{2})\dot{w}_{2t} - (\frac{w_{1t}}{2} + \frac{w_{2t}}{2})\dot{v}_{1t}] \tag{73}$$

$$\dot{y}_{1t} = y_{1t}(1 - y_{1t})\beta[(1 - \frac{u_{1t}}{2})\dot{w}_{1t} - (\frac{1}{2} + \frac{u_{1t}}{2})\dot{w}_{2t} - (\frac{w_{1t}}{2} + \frac{w_{2t}}{2})\dot{u}_{1t}] \tag{74}$$

Based on equation (49) and (50), wages in discrete time:

$$w_{1t} = \max[z_1 - \frac{1 - (1 - x_{1t})(1 - y_{1t})}{(1 - y_{1t})x_{1t}(1 - x_{1t})(\beta\frac{v_{1t}}{2} + \beta(1 - v_{1t})) + (1 - x_{1t})y_{1t}(1 - y_{1t})(\beta\frac{u_{1t}}{2} + \beta(1 - u_{1t}))}, 0] \tag{75}$$

$$w_{2t} = \max[z_2 - \frac{1 - x_{1t}y_{1t}}{y_{1t}x_{1t}(1-x_{1t})(\beta v_{1t} + \beta\frac{(1-v_{1t})}{2}) + x_{1t}y_{1t}(1-y_{1t})(\beta u_{1t} + \beta\frac{(1-u_{1t})}{2})}, 0] \tag{76}$$

In continuous time:

$$\dot{w}_{1t} = \frac{dw_{1t}}{dx_{1t}}\dot{x}_{1t} + \frac{dw_{1t}}{dy_{1t}}\dot{y}_{1t} + \frac{dw_{1t}}{du_{1t}}\dot{u}_{1t} + \frac{dw_{1t}}{dv_{1t}}\dot{v}_{1t} \tag{77}$$

$$\dot{w}_{2t} = \frac{dw_{2t}}{dx_{1t}}\dot{x}_{1t} + \frac{dw_{2t}}{dy_{2t}}\dot{y}_{1t} + \frac{dw_{2t}}{du_{1t}}\dot{u}_{1t} + \frac{dw_{2t}}{dv_{1t}}\dot{v}_{1t} \tag{78}$$

From equations (77) and (78):

$$\begin{aligned}
\dot{w}_{1t} = \{ & \frac{1-(1-x_{1t})(1-y_{1t})}{[(1-y_{1t})x_{1t}(1-x_{1t})\beta(1-\frac{v_{1t}}{2}) + (1-x_{1t})y_{1t}(1-y_{1t})\beta(1-\frac{u_{1t}}{2})]^2}[(1-y_{1t})(1-2x_{1t})\beta(1-\frac{v_{1t}}{2}) \\
& - y_{1t}(1-y_{1t})\beta(1-\frac{u_{1t}}{2})] - \frac{1-y_{1t}}{(1-y_{1t})x_{1t}(1-x_{1t})\beta(1-\frac{v_{1t}}{2}) + (1-x_{1t})y_{1t}(1-y_{1t})\beta(1-\frac{u_{1t}}{2})}\}\dot{x}_{1t} \\
& + \{\frac{1-(1-x_{1t})(1-y_{1t})}{[(1-y_{1t})x_{1t}(1-x_{1t})\beta(1-\frac{v_{1t}}{2}) + (1-x_{1t})y_{1t}(1-y_{1t})\beta(1-\frac{u_{1t}}{2})]^2}[-x_{1t}(1-x_{1t})\beta(1-\frac{v_{1t}}{2}) \\
& + (1-x_{1t})(1-2y_{1t})\beta(1-\frac{u_{1t}}{2})] - \frac{1-x_{1t}}{(1-y_{1t})x_{1t}(1-x_{1t})\beta(1-\frac{v_{1t}}{2}) + (1-x_{1t})y_{1t}(1-y_{1t})\beta(1-\frac{u_{1t}}{2})}\}\dot{y}_{1t} \\
& + \{\frac{1-(1-x_{1t})(1-y_{1t})}{[(1-y_{1t})x_{1t}(1-x_{1t})\beta(1-\frac{v_{1t}}{2}) + (1-x_{1t})y_{1t}(1-y_{1t})\beta(1-\frac{u_{1t}}{2})]^2}(1-x_{1t})y_{1t}(1-y_{1t})\beta(-\frac{1}{2})\dot{u}_{1t} \\
& + \{\frac{1-(1-x_{1t})(1-y_{1t})}{[(1-y_{1t})x_{1t}(1-x_{1t})\beta(1-\frac{v_{1t}}{2}) + (1-x_{1t})y_{1t}(1-y_{1t})\beta(1-\frac{u_{1t}}{2})]^2}(1-y_{1t})x_{1t}(1-x_{1t})\beta(-\frac{1}{2})\dot{v}_{1t}
\end{aligned} \tag{79}$$

$$\begin{aligned}
\dot{w}_{2t} = \{ & \frac{1-x_{1t}y_{1t}}{[y_{1t}x_{1t}(1-x_{1t})\beta\frac{(1+v_{1t})}{2} + x_{1t}y_{1t}(1-y_{1t})\beta\frac{(1+u_{1t})}{2}]^2}[y_{1t}(1-2x_{1t})\beta\frac{(1+v_{1t})}{2} \\
& + y_{1t}(1-y_{1t})\beta\frac{(1+u_{1t})}{2}] + \frac{y_{1t}}{y_{1t}x_{1t}(1-x_{1t})\beta\frac{(1+v_{1t})}{2} + x_{1t}y_{1t}(1-y_{1t})\beta\frac{(1+u_{1t})}{2}}\}\dot{x}_{1t} \\
& + \{\frac{1-x_{1t}y_{1t}}{[y_{1t}x_{1t}(1-x_{1t})\beta\frac{(1+v_{1t})}{2} + x_{1t}y_{1t}(1-y_{1t})\beta\frac{(1+u_{1t})}{2}]^2}[x_{1t}(1-x_{1t})\beta\frac{(1+v_{1t})}{2} \\
& + x_{1t}(1-2y_{1t})\beta\frac{(1+u_{1t})}{2}] + \frac{x_{1t}}{y_{1t}x_{1t}(1-x_{1t})\beta\frac{(1+v_{1t})}{2} + x_{1t}y_{1t}(1-y_{1t})\beta\frac{(1+u_{1t})}{2}}\}\dot{y}_{1t} \\
& + \frac{1-x_{1t}y_{1t}}{[y_{1t}x_{1t}(1-x_{1t})\beta\frac{(1+v_{1t})}{2} + x_{1t}y_{1t}(1-y_{1t})\beta\frac{(1+u_{1t})}{2}]^2}x_{1t}y_{1t}(1-y_{1t})\beta\frac{1}{2}\dot{u}_{1t} \\
& + \frac{1-x_{1t}y_{1t}}{[y_{1t}x_{1t}(1-x_{1t})\beta\frac{(1+v_{1t})}{2} + x_{1t}y_{1t}(1-y_{1t})\beta\frac{(1+u_{1t})}{2}]^2}y_{1t}x_{1t}(1-x_{1t})\beta\frac{1}{2}\dot{v}_{1t}
\end{aligned} \tag{80}$$

For $\dot{u}_{1t} = 0$, $\dot{v}_{1t} = 0$, $\dot{x}_{1t} = 0$, $\dot{y}_{1t} = 0$, $\dot{w}_{1t} = 0$, $\dot{w}_{2t} = 0$. As belief converges to 0 or 1 depending on the realized action, equilibrium points can be solved with equations (68), (69), (75), (76).

The stability analysis can be intricate, to simplify the process, I suppose slower dynamics in the $w_1 - w_2$ subsystem, and explore the $x - y$ subsystem, and then the reverse.

Proposition 8 (Stability of the (x_1, y_1, w_1, w_2) system)**.** *Let $(x_1^*, y_1^*, w_1^*, w_2^*)$ be an equilibrium point in the firm-worker matching problem.*

- *Given belief updating rule, belief stabilizes in continuous time ($u_{1t} \to u_1^*$, $v_{1t} \to v_1^*$).*

- *Workers' side $x - y$ subsystem can be asymptotically stable when dynamics of w_1 and w_2 are assumed to be slow and fulfill the condition*

$$2\dot{w}_{2t} > \dot{w}_{1t} > \frac{1}{2}\dot{w}_{2t} \tag{81}$$

in the limit of $x_1, y_1 \to (0,1), (1,0)$.

- *The $x - y$ subsystem is unstable when workers behave symmetrically, $x_1 = y_1$.*

- *Firms' side $w_1 - w_2$ subsystem is asymptotically stable when dynamics of x_1 and y_1 are assumed to be slow and workers behave asymmetrically, close to limit $(x_1, y_1 \to (0,1), (1,0))$.*

- *Under timescale separation and sufficiently small perturbations around equilibrium point, the full system can be locally stable.*

Proof. On the workers' side, I compute the Jacobian matrix and evaluate the eigenvalues. On the firms' side, since the Jacobian of $\dot{w}_{1t}$ and $\dot{w}_{2t}$ with respect to x_{1t} and y_{1t} blows up at the limit $x_{1t}, y_{1t} \to (1,0), (0,1)$ due to division by 0, I adopt Lyapunov analysis.

Full details see Appendix A.3. □

Let $\Delta\alpha^i = \alpha_1^i - \alpha_2^i$ and $\Delta\alpha^{-i} = \alpha_1^{-i} - \alpha_2^{-i}$, from equations (68) and (69),

$$\frac{dx_{1t}}{d\Delta\alpha^i} = x_{1t}(1 - x_{1t}), \frac{dy_{1t}}{d\Delta\alpha^{-i}} = y_{1t}(1 - y_{1t}) \tag{82}$$

$\Delta\alpha^i$ and $\Delta\alpha^{-i}$ affect x_{1t} and y_{1t} positively, implying stronger bias towards firm 1 would increase application rate towards firm 1, and they also influence wages indirectly through x_{1t}, y_{1t}.

Proposition 9 (Impact of Experience Bias on Workers' Learning Dynamics.)**.** *Given belief stabilizes and wage dynamics condition, $2\dot{w}_{2t} > \dot{w}_{1t} > \frac{\dot{w}_{2t}}{2}$, is satisfied.*

- *For $v_{1t} = u_{1t}$ and workers behaving symmetrically ($x_{1t} = y_{1t}$): Higher relative bias ($|\Delta\alpha^i|$, $|\Delta\alpha^{-i}| > 0$) could slow down learning dynamics (i.e. $\frac{d\dot{x}_{1t}}{d\Delta\alpha^i} < 0$, $\frac{d\dot{y}_{1t}}{d\alpha^{-i}} < 0$).*

- *For $v_{1t} \neq u_{1t}$ and workers behaving asymmetrically ($x_{1t}, y_{1t} \to (0,1), (1,0)$): Higher relative bias could contribute to asymmetric learning dynamics. (i.e. $\frac{d\dot{x}_{1t}}{d\Delta\alpha^i} < 0$, $\frac{d\dot{y}_{1t}}{d\alpha^{-i}} > 0$).*

Proof. See Appendix A.4. □

For convergence to symmetric equilibrium, higher relative bias and slower learning process does not imply greater stability, it indicates that the system could appear settle in the neighbourhood of the equilibrium. Workers would display slower adaptation as they are less sensitive to changes in bias or payoffs. As a result, they would appear "locked-in" into symmetric strategies and crowding at a single firm. For convergence to asymmetric equilibria, the asymmetric learning dynamics caused by relative bias implies that one worker could stay near the equilibrium, but the other moves away more rapidly. This imbalance in sensitivity generated by relative bias could destabilize workers' $x - y$ subsystem.

As for the firms, even though workers' bias from experience do not directly affect wages, they have an indirect effect via x_{1t} and y_{1t}. Therefore, relative bias would also influence wage dynamics, $(\dot{w}_{1t}, \dot{w}_{2t})$ in equations (101) and (102), and could lead to destabilization in firms' $w_1 - w_2$ subsystem.

Result Summary. In this section, I explore sequential learning where workers observe the wages before formulating their application strategies. Their strategies are subjected not only to payoffs but also to their beliefs about the other workers' choices and their static bias from long-term experiences. Firms, on the other hand, optimize wages given how they expect workers to behave in each period. Workers' and firms' behaviour converge to equilibria similar to the QRE in the long run. When workers possess sufficiently high sensitivity to expected payoffs (β), multiple equilibria could exist.

The different learning algorithms help to formalize firms' wage-setting and workers' learning behaviours, and they mainly differs in how firms optimize wages in each period. In presence of multiple equilibria, the local optimization method (Algorithm 1 and 2) could lead to fluctuations in both wages and workers' choice probabilities as the system jumps between different convergence paths. This is mainly contributed by perturbations in beliefs, which may push the system into different basins of attractions. But in general, there is convergence to asymmetric equilibria. For the global optimization method (Algorithm 3), it is less prone to belief fluctuations. As beliefs settle into one of the basin of attractions, firms search through a wide range of possible wages to nail down the optimal wages instead of being restricted to the neighbourhood of initial wage guesses or previous period optimal wages, thus there could be clearer convergence pattern to a single equilibrium. However, workers and firms would converge to the symmetric equilibria when workers have moderate sensitivity to expected payoffs, and only converge to the asymmetric equilibria when sensitivity are sufficiently high.

I also transform the problem into continuous time to analyze equilibrium stability via timescale separation. I found that workers' subsystem is asymptotically stable when they behave asymmetrically for certain wage dynamics condition, and given workers' search behaviour, firms' subsystem would also be asymptotically stable. Therefore, under timescale separation and small perturbations around the equilibrium points, the full system can be locally stable in the asymmetric case.

The presence of bias from long-term experiences could affect equilibrium selection depending on the weighing between sensitivity to expected payoffs and extent of bias. When experience bias is strong, workers would rely on their bias to inform application strategies, and if the extent of bias is the same, they could converge to a unique, symmetric equilibrium. Experiences also affect workers' learning dynamics. Stronger bias towards a firm could lead to workers being seemingly "locked-in" to the neighbourhood of symmetric strategies, where they display slower adaptivity. They could also contribute to asymmetric learning dynamics at the asymmetric equilibria points, and this may destabilize the system.

3.2 Policy Implications

This market structure involves workers observing the wages before making their application decisions. Given presence of multiple equilibria and in the absence of workers' bias from long-term experiences, equilibrium selection could be heavily affected by workers' beliefs and also by how firms set wages. While the former depends on the realized actions and may be harder to intervene, policies could potentially influence the later. When workers' sensitivity to expected payoff is high, both local optimization and global optimization method on the firms' side could lead to more efficient outcome, where workers apply with higher probability to different firms, albeit the former could lead to more noisy results in choice probabilities. However, when sensitivity is only moderate, prompting firms to adjust wages slightly based on past periods instead of searching through the full range of wages may be beneficial in achieving asymmetric equilibria as there is greater tendency for firms to choose local optima points period-by-period, leading to jumps between convergence paths towards asymmetric equilibria.

When workers possess bias from long-term experiences, the extent of bias against expected payoffs is essential to determine the existence of multiple equilibria. If workers' possess large bias towards different firm from one another, and overriding the expected payoff difference between selecting each firm, then there may be a single unique equilibrium. In this case, workers with diverse experience could be directed naturally and apply with higher probability to different firms, thus leading to higher likelihood of one-to-one matching. Potential measures to implement this could be early exposure programmes, such as internships, that could prime workers to hold different bias based on early experiences. This could also support development of identity connotations associated with different firm types, which may appeal to workers with different cultural and social backgrounds. However, when workers are biased similarly, the presence of multiple equilibria is crucial to make it possible for workers to be able to coordinate on applying to different firms. In which case, long-term experiences have to be less prominent as compared to expected payoffs. Nonetheless, even when workers are converging to the asymmetric equilibria, high experience bias may destabilize workers' subsystem, making asymmetric equilibria less stable. Therefore, weakening such bias could also be one of the policy objectives to achieve greater stability. Policymakers may consider measures such as information provision to reduce cultural stereotypes associated with different jobs.

4 Discussions and Conclusion

In this paper, I explore the role of experiences on workers' adaptive learning behaviour in application choices. By integrating experience-based learning into search, I am able to track the evolution of firms' wages and workers' choice probabilities over time, which provides some insights on labour market dynamics. It also answers to the question of equilibrium selection and whether workers learn to apply to jobs more efficiently, defined as higher likelihood of one-to-one matching, in both static wage environment and when wages are dynamically changing. Furthermore, this could also help to explain the apparent puzzling phenomenon of workers' lack of switching in applying to higher wage jobs even when they are able to transit (Archer (2016)), and provide an argument for sorting at application stage that is less related to skills (Barbulescu and Bidwell (2013)).

4.1 Equilibrium Selection

The presence of equilibrium multiplicity often call into question regarding which equilibrium workers would coordinate on, and if in the long run, they would learn to apply efficiently to different firms.

For the first market structure, where workers do not observe wages and simply learn to apply based on past feedback. When wages are fixed and there exist multiple equilibria, workers would converge to coordinate on applying to different firms in the long run. Such learning mechanism emphasizes heavily on initial propensities and initial experiences, which can have prominent impact in determining the equilibrium chosen. As a result, if workers start off with bias towards different firms, these create a natural "lock-in" and is illuminate of the eventual equilibrium outcome. Similarly, positive reinforcements in the initial application rounds could propel workers into different directions and they would be "stuck" applying to the firm they are more familiar with. On the other hand, in an environment where wages are dynamically changing, wages are pushed down to 0 in the long run, and a continuum of equilibria emerge. Both workers and firms stop learning in the absence of rewards, thus the eventual equilibrium outcome simply depend on the learning path. There could be convergence to more randomized choice probabilities for the workers, and that efficient outcome of one-to-one matching may not be achieved.

In the second market structure, where workers observe the wages before making an application decision, their choice probabilities over the firms would depend on wages, beliefs about their opponent's choices and their own bias from long-term experiences. Since firms are setting wages optimally while tracking workers' belief updates and taking into consideration their reactions to wage changes, this can also constitute as a dynamic wage setting. In the long run, the system converges to equilibria similar to that of QRE. Given multiple equilibria, without experience bias, workers could converge to the asymmetric equilibria when workers are highly sensitive to expected payoffs. For moderate wage sensitivity, firms choosing optimal wages through global optimization method would lead to convergence to symmetric equilibria. The same applies for weak experience bias. With strong experience bias, efficient outcome may also emerge if workers have diverse experiences that bias them to different firms, the asymmetric equilibrium would also be the unique equilibrium.

In general, convergence of workers' strategies to coordinating on applying to different firms can happen with experience-based learning without intervention. However, the mechanism is not very enlightening on which asymmetric equilibria will be selected. Furthermore, experiences may behave both as an accelerator and inhibitor for reaching efficient outcome under dynamic wages. For instance, when workers learn from feedback, experiences could hinder learning as workers move from one game type to another, and the eventual choices could be more randomized as a result; and when workers are able to observe wages before making decisions, having strong bias in the same direction could result in eventual overcrowding at one of the firms.

4.2 Mismatch Problems

Even though workers generally converge to applying to different firms, these learning models portrays an evolutionary narrative which implies prolonged periods of mismatch before reaching equilibrium.

In the first market structure with fixed wages, not only that workers have to learn to coordinate, before which, there will be many periods of mismatch, if workers have similar experiences, this process will be longer. Policies could cater to reducing impact of past experiences by inducing forgetfulness. For example, filling in past working experiences and reusing them for all future applications could put less emphasis on past experiences. As for dynamic wage environment, workers could be playing different games depending on the evolution of payoffs. It could be harder to mitigate mismatch if workers play games with a single NE, where they apply to the same firm, for too long. Policies could similarly induce forgetfulness or focus on maintaining a wage environment with pure NEs of workers applying more to different firms, such that workers could eventually learn to coordinate.

For the second market structure, strong experience bias towards the same firm would induce prolonged periods of mismatch, even in the equilibrium. For weaker experiences relative to expected payoffs, in presence of multiple equilibria, there could be fluctuations between convergence paths towards different equilibrium points, thereby contributing to mismatch on top of being on the learning trajectory towards equilibrium. Policy-makers could reduce bias from long-term experiences or create different anchoring bias for each worker through influencing cultural and social stereotypes.

4.3 Extensions

One area of extension is to have an even more realistic portrayal of learning process with imprecision. While the learning rules in the paper already contain some aspects of constrained decision-making, such as the absence of information if one does not explore options enough in reinforcement learning; and in the case of sequential learning, the inability to observe opponent strategies and has to rely on belief updating based on past realized actions. The prospect of diving deeper into imprecision in learning the wage environment as well as recalling could spark more research in cognitively-founded learning dynamics. For instance, the dynamic wage environment with sequential learning provides basis for incorporation of inattention. The choice of logit model could encompass both rational inattention and experience updating, postulating a novel form of learning rule. (Matějka and McKay (2015), Mattsson and Weibull (2002)) Furthermore, even though I considered decay in memories, it could be more realistic to account for noisy memories (Da Silveira et al. (2020), Aridor et al. (2024)), which could lead to more nuanced learning mechanism and even longer periods of mismatch.

Last but not least, this paper provides a baseline for incorporating adaptive learning behaviour into labour market dynamics. It facilitates understanding of pairwise learning and potential equilibrium outcomes under different learning dynamics. An obvious extension is to scale up the model to include more firms and workers, such it turns to a $I \times J$ game. Analyzing learning behaviour when more interactions are involved may encompass richer results.

References

Albarracin, D. and Wyer Jr, R. S. (2000). The cognitive impact of past behavior: influences on beliefs, attitudes, and future behavioral decisions. *Journal of personality and social psychology*, 79(1):5.

Archer, S. . T. G. (31 May 2016). Why we don't change jobs enough – and why we should. https://www.theguardian.com/careers/2016/may/31/change-jobs-risks-dead-end-take-career-happiness. Accessed: 15 Jun 2024.

Aridor, G., da Silveira, R. A., and Woodford, M. (2024). Information-constrained coordination of economic behavior. Technical report, National Bureau of Economic Research.

Barbulescu, R. and Bidwell, M. (2013). Do women choose different jobs from men? mechanisms of application segregation in the market for managerial workers. *Organization Science*, 24(3):737–756.

Beggs, A. W. (2005). On the convergence of reinforcement learning. *Journal of economic theory*, 122(1):1–36.

Briñol, P. and Petty, R. E. (2022). Self-validation theory: An integrative framework for understanding when thoughts become consequential. *Psychological Review*, 129(2):340.

Burdett, K. and Mortensen, D. T. (1998). Wage differentials, employer size, and unemployment. *International economic review*, pages 257–273.

Camerer, C. and Hua Ho, T. (1999). Experience-weighted attraction learning in normal form games. *Econometrica*, 67(4):827–874.

Da Silveira, R. A., Sung, Y., and Woodford, M. (2020). Optimally imprecise memory and biased forecasts. Technical report, National Bureau of Economic Research.

Duffy, J. and Hopkins, E. (2005). Learning, information, and sorting in market entry games: theory and evidence. *Games and Economic behavior*, 51(1):31–62.

Eeckhout, J. (2018). Sorting in the labor market. *Annual Review of Economics*, 10(1):1–29.

Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S. M., Hau, R., Hertwig, R., Stewart, T., West, R., and Lebiere, C. (2010). A choice prediction competition: Choices from experience and from description. *Journal of Behavioral Decision Making*, 23(1):15–47.

Erev, I. and Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American economic review*, pages 848–881.

Fouarge, D., Kriechel, B., and Dohmen, T. (2014). Occupational sorting of school graduates: The role of economic preferences. *Journal of Economic Behavior & Organization*, 106:335–351.

Fudenberg, D. and Levine, D. K. (1998). *The theory of learning in games*, volume 2. MIT press.

Galenianos, M. and Kircher, P. (2009). Directed search with multiple job applications. *Journal of economic theory*, 144(2):445–471.

Hopkins, E. (1999). A note on best response dynamics. *Games and Economic Behavior*, 29(1-2):138–150.

Hopkins, E. (2002). Two competing models of how people learn in games. *Econometrica*, 70(6):2141–2166.

Hopkins, E. (2007). Adaptive learning models of consumer behavior. *Journal of economic behavior & organization*, 64(3-4):348–368.

Hopkins, E. and Posch, M. (2005). Attainability of boundary points under reinforcement learning. *Games and Economic Behavior*, 53(1):110–125.

Kanfer, R. and Bufton, G. M. (2018). Job loss and job search: A social-cognitive and self-regulation perspective. *The Oxford handbook of job loss and job search*, pages 143–158.

Langenhove, L. v. and Harré, R. (1994). Cultural stereotypes and positioning theory. *Journal for the Theory of Social Behaviour*, 24(4):359–372.

Lieder, F., Griffiths, T. L., M. Huys, Q. J., and Goodman, N. D. (2018). The anchoring bias reflects rational use of cognitive resources. *Psychonomic bulletin & review*, 25:322–349.

Lu, S. (2024). Attention as a scarce resource: Costly job search with inattentive workers. *Working Paper (Under Revision).*

Matějka, F. and McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–298.

Mattsson, L.-G. and Weibull, J. W. (2002). Probabilistic choice and procedurally bounded rationality. *Games and Economic Behavior*, 41(1):61–78.

McCall, J. J. (1970). Economics of information and job search. *The Quarterly Journal of Economics*, 84(1):113–126.

McKelvey, R. D. and Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and economic behavior*, 10(1):6–38.

Moen, E. R. (1997). Competitive search equilibrium. *Journal of political Economy*, 105(2):385–411.

Nowé, A., Vrancx, P., and De Hauwere, Y.-M. (2012). Game theory and multi-agent reinforcement learning. *Reinforcement Learning: State-of-the-Art*, pages 441–470.

Paas, F. and Ayres, P. (2014). Cognitive load theory: A broader view on the role of memory in learning and education. *Educational Psychology Review*, 26:191–195.

Roth, A. E. and Erev, I. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and economic behavior*, 8(1):164–212.

Terjesen, S., Vinnicombe, S., and Freeman, C. (2007). Attracting generation y graduates: Organisational attributes, likelihood to apply and sex differences. *Career development international*, 12(6):504–522.

Vafa, K., Palikot, E., Du, T., Kanodia, A., Athey, S., and Blei, D. M. (2022). Career: A foundation model for labor sequence data. *arXiv preprint arXiv:2202.08370.*

Van Huyck, J. B., Battalio, R. C., and Rankin, F. W. (1997). On the origin of convention: Evidence from coordination games. *The Economic Journal*, 107(442):576–596.

van Strien, S. (2022). Dynamics of learning and iterated games lecture notes, math60007/70007/97069. Accessed: April 28, 2025.

Wanberg, C. R., Ali, A. A., and Csillag, B. (2020). Job seeking: The process and experience of looking for a job. *Annual Review of Organizational Psychology and Organizational Behavior*, 7(1):315–337.

Wright, R., Kircher, P., Julien, B., and Guerrieri, V. (2021). Directed search and competitive search equilibrium: A guided tour. *Journal of Economic Literature*, 59(1):90–148.

Wu, L. (2020). *Partially Directed Search in the Labor Market.* PhD thesis, The University of Chicago.

A Proofs

A.1 Proposition 4

To show delayed adaptation in equilibrium switching, I show the following exemplary learning mechanism for 3 periods:

Example 3. *Period 1: Initial wages are randomly picked. Suppose workers start off from* $G1$ *(Figure 8), where* $w_{10} > 2w_{20}$ *and chose* $(F1, F1)$*:*

$$\textit{Workers' propensities:} \quad q^i_{11} = q^i_{10} + \frac{w_{10}}{2}, \quad q^{-i}_{11} = q^{-i}_{10} + \frac{w_{10}}{2}$$

$$\textit{Firms' propensities:} \quad \theta^j_{(w_{11})1} = \theta^j_{(w_{10})0} + (z_1 - w_{10}), \quad \theta^{-j}_{(w_{21})1} = \theta^{-j}_{(w_{20})0}$$

Workers' choice of firm 1 is reinforced; Firm 1's choice of w_{10} *is reinforced, and firm 2 continue to experiment wages randomly, assuming uniform probability over its action space.*

Period 2: Suppose workers continue to choose $(F1, F1)$ *and firm 1 picked a new wage that is lower than the previous period,* $w_{11} < w_{10}$*, but the relationship* $w_{11} > 2w_{21}$ *still hold, then:*

$$q^i_{12} = q^i_{11} + \frac{w_{11}}{2}, q^{-i}_{12} = q^{-i}_{11} + \frac{w_{11}}{2}$$

$$\theta^j_{(w_{11})2} = \theta^j_{(w_{11})1} + (z_1 - w_{11}), \theta^{-j}_{(w_{21})2} = \theta^{-j}_{(w_{21})1}$$

Workers' choice of firm 1 is again reinforced; Firm 1's choice of w_{11} *is reinforced, and the strength of reinforcement is higher than that of a wage value equals to* w_{10}*. This logic implies there will be higher chance of picking lower wage values as more iterations occur. Since firm 2 was not selected in round 2, it continues to experiment within its action space randomly in the next period.*

Period 3: Suppose firm 1 picked an even lower wage than the previous period, $w_{12} < w_{11}$*, and the wage condition becomes* $2w_{12} > w_{22} > \frac{w_{12}}{2}$*. There is a switch in the game played. Given previous reinforcement, there is higher probability of selecting* $F1$*, if* $(F1, F1)$ *is chosen again:*

$$q^i_{13} = q^i_{12} + \frac{w_{12}}{2}, q^{-i}_{13} = q^{-i}_{12} + \frac{w_{12}}{2}$$

$$\theta^j_{(w_{12})3} = \theta^j_{(w_{12})2} + (z_1 - w_{12}), \theta^{-j}_{(w_{22})3} = \theta^{-j}_{(w_{22})2}$$

Workers' propensities to firm 1 are positively reinforced, but with even less strength than before; firm 1's choice of lower wage is reinforced, while firm 2 continue to sample the action space randomly.

This shows that as w_{1t} *is driven downwards,* $w_{1t} > 2w_{2t}$ *could break down, and* $2w_{1t} > w_{2t} > \frac{w_{1t}}{2}$ *may arise, leading to a game change from* $G1$ *to* $G2$ *(Figure 9). Choosing* $F1$ *will lead to lower reinforcement as compared to choosing* $F2$*, propensities could thus be updated, and slowly, there will be convergence towards* $(F1, F2)$ *and* $(F2, F1)$*. Nonetheless, it is possible to have multiple switching (i.e. from* $G1$ *to* $G2$ *to* $G3$*, etc.) depending on the changes in wage conditions.*

Return to Section 2.2.

A.2 Proposition 5

When wage regime change from $G1$ to $G2$ at $t = \tilde{t}$, workers' start experiencing decay for propensities accumulated in $G1$:

$$\text{Worker } i\text{: } q^i_{j\tilde{t}} = (1-\eta)q^i_{j(\tilde{t}-1)}, \text{Worker } -i\text{: } q^{-i}_{j\tilde{t}} = (1-\eta)q^{-i}_{j(\tilde{t}-1)} \tag{83}$$

And future propensities will be accumulated based on $G2$ payoffs:

$$\text{Worker } i\text{: } q^i_{j(\tilde{t}+1)} = (1-\eta)q^i_{j(\tilde{t})} + \pi^i_{\tilde{t}}(a^i_{\tilde{t}}, a^{-i}_{\tilde{t}}, a^j_{\tilde{t}}, a^{-j}_{\tilde{t}}) \tag{84}$$

$$\text{Worker } -i\text{: } q^{-i}_{j(\tilde{t}+1)} = (1-\eta)q^{-i}_{j(\tilde{t})} + \pi^{-i}_{\tilde{t}}(a^i_{\tilde{t}}, a^{-i}_{\tilde{t}}, a^j_{\tilde{t}}, a^{-j}_{\tilde{t}}) \tag{85}$$

Suppose payoffs in $G2$ are constant for worker i, π^{i*}_t, steady state propensities would be:

$$q^{i*}_j = \frac{\pi^i_j}{\eta} \tag{86}$$

The propensities at time $t > \tilde{t}$, where at $\tilde{t}$, one start new accumulation of propensity according to $G2$:

$$q^i_{jt} = (1-\eta)^{t-\tilde{t}} q^i_{j\tilde{t}} + \sum_{k=0}^{t-\tilde{t}-1} (1-\eta)^k \pi^i_j \tag{87}$$

$$q^i_{jt} = (1-\eta)^{t-\tilde{t}} q^i_{j\tilde{t}} + \frac{1-(1-\eta)^{t-\tilde{t}}}{\eta}\pi^i_j \tag{88}$$

$$q^i_{jt} = (1-\eta)^{t-\tilde{t}} (q^i_{j\tilde{t}} - q^{i*}_j) + q^*_j \tag{89}$$

For $t > \tilde{t}$, computing propensities starting from 0, set $q^i_{j\tilde{t}} = 0$:

$$q^i_{jt} = (1-\eta)^{t-\tilde{t}} (-q^{i*}_j) + q^*_j \tag{90}$$

The time taken for propensity after $\tilde{t}$ to exceed the ones before can be found by:

$$(1-\eta)^{t-\tilde{t}} (-q^{i*}_j) + q^*_j \geq q^i_{j\tilde{t}}(1-\eta)^{t-\tilde{t}} \tag{91}$$

where $q^i_{j\tilde{t}} = (1-\eta)^{\tilde{t}} q^i_{j0} + \sum_{k=0}^{\tilde{t}-1}(1-\eta)^{\tilde{t}-k-1}\pi^i_k$.

I am interested in $t - \tilde{t}$:

$$t - \tilde{t} \leq \frac{\ln(\frac{q^*_j}{q_{j\tilde{t}} + q^*_j})}{\ln(1-\eta)} \tag{92}$$

For small η, $\eta \in (0,1)$, $\ln(1-\eta) \approx -\eta$,

$$t - \tilde{t} \propto \frac{1}{\eta} \tag{93}$$

But since $G2$ payoffs are likely to be non-constant as payoffs evolve through reinforcement learning on the firms' side, the dependency of time lag on payoff changes can be more generically written as dependency on η and $f(G2 \text{ payoff dynamics})$, which captures the learning process. It could lead to longer or shorter transition process.

Return to Section 2.2.1.

A.3 Proposition 8

Workers' side. The Jacobian I need to solve for this system is a 4×4 matrix:

$$J = \begin{pmatrix} \frac{\partial \dot{x}_{1t}}{\partial x_{1t}} & \frac{\partial \dot{x}_{1t}}{\partial y_{1t}} & \frac{\partial \dot{x}_{1t}}{\partial u_{1t}} & \frac{\partial \dot{x}_{1t}}{\partial v_{1t}} \\ \frac{\partial \dot{y}_{1t}}{\partial x_{1t}} & \frac{\partial \dot{y}_{1t}}{\partial y_{1t}} & \frac{\partial \dot{y}_{1t}}{\partial u_{1t}} & \frac{\partial \dot{y}_{1t}}{\partial v_{1t}} \\ \frac{\partial \dot{u}_{1t}}{\partial x_{1t}} & \frac{\partial \dot{u}_{1t}}{\partial y_{1t}} & \frac{\partial \dot{u}_{1t}}{\partial u_{1t}} & \frac{\partial \dot{u}_{1t}}{\partial v_{1t}} \\ \frac{\partial \dot{v}_{1t}}{\partial x_{1t}} & \frac{\partial \dot{v}_{1t}}{\partial y_{1t}} & \frac{\partial \dot{v}_{1t}}{\partial u_{1t}} & \frac{\partial \dot{v}_{1t}}{\partial v_{1t}} \end{pmatrix}$$

$$= \begin{pmatrix} \begin{matrix}(1-2x_{1t})\beta[(1-\frac{v_{1t}}{2})\dot{w}_{1t} \\ -(\frac{1}{2}+\frac{v_{1t}}{2})\dot{w}_{2t}-(\frac{w_{1t}}{2}+\frac{w_{2t}}{2})\dot{v}_{1t}]\end{matrix} & 0 & 0 & \begin{matrix}x_{1t}(1-x_{1t}) \\ \beta(-\frac{1}{2}\dot{w}_{1t}-\frac{1}{2}\dot{w}_{2t})\end{matrix} \\ 0 & \begin{matrix}(1-2y_{1t})\beta[(1-\frac{u_{1t}}{2})\dot{w}_{1t} \\ -(\frac{1}{2}+\frac{u_{1t}}{2})\dot{w}_{2t}-(\frac{w_{1t}}{2}+\frac{w_{2t}}{2})\dot{u}_{1t}]\end{matrix} & \begin{matrix}y_{1t}(1-y_{1t}) \\ \beta(-\frac{1}{2}\dot{w}_{1t}-\frac{1}{2}\dot{w}_{2t})\end{matrix} & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{94}$$

Since beliefs converge to realized actions and become deterministic eventually, I simplify the analysis further to focus solely on the $x - y$ interactions, reducing the Jacobian matrix:

$$J = \begin{pmatrix} (1-2x_{1t})\beta[(1-\frac{v_{1t}}{2})\dot{w}_{1t}-(\frac{1}{2}+\frac{v_{1t}}{2})\dot{w}_{2t}-(\frac{w_{1t}}{2}+\frac{w_{2t}}{2})\dot{v}_{1t}] & x_{1t}(1-x_{1t})\beta(-\frac{1}{2}\dot{w}_{1t}-\frac{1}{2}\dot{w}_{2t}) \\ (1-2y_{1t})\beta[(1-\frac{u_{1t}}{2})\dot{w}_{1t}-(\frac{1}{2}+\frac{u_{1t}}{2})\dot{w}_{2t}-(\frac{w_{1t}}{2}+\frac{w_{2t}}{2})\dot{u}_{1t}] & y_{1t}(1-y_{1t})\beta(-\frac{1}{2}\dot{w}_{1t}-\frac{1}{2}\dot{w}_{2t}) \end{pmatrix} \tag{95}$$

For $\dot{v}_{1t} = 0$, $\dot{u}_{1t} = 0$, find eigenvalues:

$$\lambda = \frac{\begin{matrix}(1-2x_{1t})\beta[(1-\frac{v_{1t}}{2})\dot{w}_{1t}-(\frac{1}{2}+\frac{v_{1t}}{2})\dot{w}_{2t}]+y_{1t}(1-y_{1t})\beta(-\frac{1}{2}\dot{w}_{1t}-\frac{1}{2}\dot{w}_{2t}) \\ \pm\sqrt{\begin{matrix}[(1-2x_{1t})\beta[(1-\frac{v_{1t}}{2})\dot{w}_{1t}-(\frac{1}{2}+\frac{v_{1t}}{2})\dot{w}_{2t}]+y_{1t}(1-y_{1t})\beta(-\frac{1}{2}\dot{w}_{1t}-\frac{1}{2}\dot{w}_{2t})]^2 \\ -4[(1-2x_{1t})y_{1t}(1-y_{1t})\beta^2[(1-\frac{v_{1t}}{2})\dot{w}_{1t}-(\frac{1}{2}+\frac{v_{1t}}{2})\dot{w}_{2t}](-\frac{1}{2}\dot{w}_{1t}-\frac{1}{2}\dot{w}_{2t}) \\ -x_{1t}(1-x_{1t})(1-2y_{1t})\beta^2(-\frac{1}{2}\dot{w}_{1t}-\frac{1}{2}\dot{w}_{2t})[(1-\frac{u_{1t}}{2})\dot{w}_{1t}-(\frac{1}{2}+\frac{u_{1t}}{2})\dot{w}_{2t}]]\end{matrix}}\end{matrix}}{2} \tag{96}$$

For symmetric strategies ($x_{1t} = y_{1t}$), beliefs should be symmetric as well, $v_{1t} = u_{1t}$, one λ is 0 and depending the wage dynamics, the other would be either positive or negative. For example, for $v_{1t} = u_{1t} = 1$, $x_{1t} = y_{1t} = a$,

$$\lambda = \frac{\begin{matrix}(1-2a)\beta(\frac{1}{2}\dot{w}_{1t}-\dot{w}_{2t})+a(1-a)\beta(-\frac{1}{2}\dot{w}_{1t}-\frac{1}{2}\dot{w}_{2t}) \\ \pm\sqrt{[(1-2a)\beta(\frac{1}{2}\dot{w}_{1t}-\dot{w}_{2t})+a(1-a)\beta(-\frac{1}{2}\dot{w}_{1t}-\frac{1}{2}\dot{w}_{2t})]^2}\end{matrix}}{2} \tag{97}$$

Case 1: $\lambda_1 = 0$, $\lambda_2 > 0$ if $(1-2a)\beta(\frac{1}{2}\dot{w}_{1t}-\dot{w}_{2t})+a(1-a)\beta(-\frac{1}{2}\dot{w}_{1t}-\frac{1}{2}\dot{w}_{2t}) > 0$, which implies that symmetric strategies is unstable; Case 2: $\lambda_1 = 0$, $\lambda_2 < 0$ if $(1-2a)\beta(\frac{1}{2}\dot{w}_{1t}-\dot{w}_{2t})+a(1-a)\beta(-\frac{1}{2}\dot{w}_{1t}-\frac{1}{2}\dot{w}_{2t}) < 0$, symmetric strategies maybe stable but not asymptotically stable. For $\dot{w}_{1t} > 0$, $\dot{w}_{2t} > 0$, it is more likely for case 1 to prevail as a tends to be smaller than 0.5 (application rate to firm 1 is lower than half) when one believes that opponent is applying with higher probability to firm 1.

For completeness, suppose $v_{1t} = u_{1t} = 0$, $x_{1t} = y_{1t} = a$,

$$\lambda = \frac{\begin{matrix}(1-2a)\beta(\dot{w}_{1t}-\frac{1}{2}\dot{w}_{2t})+a(1-a)\beta(-\frac{1}{2}\dot{w}_{1t}-\frac{1}{2}\dot{w}_{2t}) \\ \pm\sqrt{[(1-2a)\beta(\dot{w}_{1t}-\frac{1}{2}\dot{w}_{2t})+a(1-a)\beta(-\frac{1}{2}\dot{w}_{1t}-\frac{1}{2}\dot{w}_{2t})]^2}\end{matrix}}{2} \tag{98}$$

For $\lambda_1 = 0$, $\lambda_2 < 0$, then $(1-2a)\beta(\dot{w}_{1t}-\frac{1}{2}\dot{w}_{2t})+a(1-a)\beta(-\frac{1}{2}\dot{w}_{1t}-\frac{1}{2}\dot{w}_{2t}) < 0$ needs to be satisfied. Either way, symmetric strategies are not asymptotically stable no matter the wage dynamics.

For asymmetric strategies ($x_{1t} \neq y_{1t}$, and $x_{1t}, y_{1t} \to (0,1), (1,0)$), beliefs should be asymmetric as well, λs can be both negative or one positive, one negative. For example, in the limit, $x_{1t} \to 1$,

$y_{1t} \to 0$ and $v_{1t} = 0$, $u_{1t} = 1$ (i.e. believing the other worker applies to firm 1 would increase the probability of application to firm 2);

$$\lambda = \frac{-\beta(\dot{w}_{1t} - \frac{1}{2}\dot{w}_{2t}) \pm \sqrt{[-\beta(\dot{w}_{1t} - \frac{1}{2}\dot{w}_{2t})]^2}}{2} \tag{99}$$

In the limit of x_{1t} and y_{1t} – Case 1: $\lambda_1 = 0$, $\lambda_2 > 0$ if $(\dot{w}_{1t} - \frac{1}{2}\dot{w}_{2t}) < 0$; Case 2: $\lambda_1 = 0$, $\lambda_2 < 0$ if $(\dot{w}_{1t} - \frac{1}{2}\dot{w}_{2t}) > 0$.

Since x_{1t} and y_{1t} are close $(1, 0)$ and not exactly there, the system can be asymptotically stable. For instance, let $f = (1-2x_{1t})y_{1t}(1-y_{1t})\beta^2[(1-\frac{v_{1t}}{2})\dot{w}_{1t} - (\frac{1}{2}+\frac{v_{1t}}{2})\dot{w}_{2t}](-\frac{1}{2}\dot{w}_{1t} - \frac{1}{2}\dot{w}_{2t}) - x_{1t}(1-x_{1t})(1-2y_{1t})\beta^2(-\frac{1}{2}\dot{w}_{1t} - \frac{1}{2}\dot{w}_{2t})[(1-\frac{u_{1t}}{2})\dot{w}_{1t} - (\frac{1}{2}+\frac{u_{1t}}{2})\dot{w}_{2t}]$. If $(\dot{w}_{1t} - \frac{1}{2}\dot{w}_{2t}) > 0$, and $f > 0$, both eigenvalues will be negative.

On the other hand, if $(\dot{w}_{1t} - \frac{1}{2}\dot{w}_{2t}) < 0$, and $f > 0$, both eigenvalues will be positive; and if $f < 0$, one eigenvalue will be negative and the other be positive. Either case, the system would be unstable.

For completeness, suppose in the limit, $x_{1t} \to 0$, $y_{1t} \to 1$ and $v_{1t} = 1$, $u_{1t} = 0$;

$$\lambda = \frac{\beta(\frac{1}{2}\dot{w}_{1t} - \dot{w}_{2t}) \pm \sqrt{[\beta(\frac{1}{2}\dot{w}_{1t} - \dot{w}_{2t})]^2}}{2} \tag{100}$$

In the limit of x_{1t} and y_{1t}, $\lambda_1 = 0$, $\lambda_2 < 0$ if $(\frac{1}{2}\dot{w}_{1t} - \dot{w}_{2t}) < 0$. In general, for $2\dot{w}_{2t} > \dot{w}_{1t} > \frac{1}{2}\dot{w}_{2t}$, then in the limit, the system is marginally stable. It can be asymptotically stable when x_{1t} and y_{1t} are close to the limit but not there.

Overall, on the workers' side, in the $x-y$ system, the system would not be asymptotically stable when workers behave symmetrically. However, the system may be asymptotically stable in the asymmetric case for some wage dynamics conditions.

Firms' side. I now turn to the $w_1 - w_2$ system. I will focus on the case where $x - y$ system may be asymptotically stable (i.e. when workers behave asymmetrically). Suppose workers have slower dynamics of x_{1t} and y_{1t}. Based on equations 79 and 80, when beliefs stabilizes:

$$\begin{aligned}
\dot{w}_{1t} = \{&\frac{1-(1-x_{1t})(1-y_{1t})}{[(1-y_{1t})x_{1t}(1-x_{1t})\beta(1-\frac{v_{1t}}{2}) + (1-x_{1t})y_{1t}(1-y_{1t})\beta(1-\frac{u_{1t}}{2})]^2}[(1-y_{1t})(1-2x_{1t})\beta(1-\frac{v_{1t}}{2}) \\
&- y_{1t}(1-y_{1t})\beta(1-\frac{u_{1t}}{2})] - \frac{1-y_{1t}}{(1-y_{1t})x_{1t}(1-x_{1t})\beta(1-\frac{v_{1t}}{2}) + (1-x_{1t})y_{1t}(1-y_{1t})\beta(1-\frac{u_{1t}}{2})}\}\dot{x}_{1t} \\
&+ \{\frac{1-(1-x_{1t})(1-y_{1t})}{[(1-y_{1t})x_{1t}(1-x_{1t})\beta(1-\frac{v_{1t}}{2}) + (1-x_{1t})y_{1t}(1-y_{1t})\beta(1-\frac{u_{1t}}{2})]^2}[-x_{1t}(1-x_{1t})\beta(1-\frac{v_{1t}}{2}) \\
&+(1-x_{1t})(1-2y_{1t})\beta(1-\frac{u_{1t}}{2})] - \frac{1-x_{1t}}{(1-y_{1t})x_{1t}(1-x_{1t})\beta(1-\frac{v_{1t}}{2}) + (1-x_{1t})y_{1t}(1-y_{1t})\beta(1-\frac{u_{1t}}{2})}\}\dot{y}_{1t}
\end{aligned} \tag{101}$$

$$\dot{w}_{2t}=\{\frac{1-x_{1t}y_{1t}}{[y_{1t}x_{1t}(1-x_{1t})\beta\frac{(1+v_{1t})}{2}+x_{1t}y_{1t}(1-y_{1t})\beta\frac{(1+u_{1t})}{2}]^2}[y_{1t}(1-2x_{1t})\beta\frac{(1+v_{1t})}{2}$$
$$+y_{1t}(1-y_{1t})\beta\frac{(1+u_{1t})}{2}]+\frac{y_{1t}}{y_{1t}x_{1t}(1-x_{1t})\beta\frac{(1+v_{1t})}{2}+x_{1t}y_{1t}(1-y_{1t})\beta\frac{(1+u_{1t})}{2}}\}\dot{x}_{1t}$$
$$+\{\frac{1-x_{1t}y_{1t}}{[y_{1t}x_{1t}(1-x_{1t})\beta\frac{(1+v_{1t})}{2}+x_{1t}y_{1t}(1-y_{1t})\beta\frac{(1+u_{1t})}{2}]^2}[x_{1t}(1-x_{1t})\beta\frac{(1+v_{1t})}{2}$$
$$+x_{1t}(1-2y_{1t})\beta\frac{(1+u_{1t})}{2}]+\frac{x_{1t}}{y_{1t}x_{1t}(1-x_{1t})\beta\frac{(1+v_{1t})}{2}+x_{1t}y_{1t}(1-y_{1t})\beta\frac{(1+u_{1t})}{2}}\}\dot{y}_{1t} \quad (102)$$

In the limit of $x_{1t}, y_{1t} \to (1,0),(0,1)$, the Jacobian of $\dot{w}_{1t}$ and $\dot{w}_{2t}$ with respect to x_{1t} and y_{1t} blows up at the limit due to division by 0. As a result, I adopt Lyapunov analysis.

$$\dot{V}=w_{1t}\dot{w}_{1t}+w_{2t}\dot{w}_{2t}$$

$$\dot{V}=w_{1t}\{\{\frac{1-(1-x_{1t})(1-y_{1t})}{[(1-y_{1t})x_{1t}(1-x_{1t})\beta(1-\frac{v_{1t}}{2})+(1-x_{1t})y_{1t}(1-y_{1t})\beta(1-\frac{u_{1t}}{2})]^2}[(1-y_{1t})(1-2x_{1t})\beta(1-\frac{v_{1t}}{2})$$
$$-y_{1t}(1-y_{1t})\beta(1-\frac{u_{1t}}{2})]-\frac{1-y_{1t}}{(1-y_{1t})x_{1t}(1-x_{1t})\beta(1-\frac{v_{1t}}{2})+(1-x_{1t})y_{1t}(1-y_{1t})\beta(1-\frac{u_{1t}}{2})}\}\dot{x}_{1t}$$
$$+\{\frac{1-(1-x_{1t})(1-y_{1t})}{[(1-y_{1t})x_{1t}(1-x_{1t})\beta(1-\frac{v_{1t}}{2})+(1-x_{1t})y_{1t}(1-y_{1t})\beta(1-\frac{u_{1t}}{2})]^2}[-x_{1t}(1-x_{1t})\beta(1-\frac{v_{1t}}{2})$$
$$+(1-x_{1t})(1-2y_{1t})\beta(1-\frac{u_{1t}}{2})]-\frac{1-x_{1t}}{(1-y_{1t})x_{1t}(1-x_{1t})\beta(1-\frac{v_{1t}}{2})+(1-x_{1t})y_{1t}(1-y_{1t})\beta(1-\frac{u_{1t}}{2})}\}\dot{y}_{1t}\}$$
$$+w_{2t}\{\{\frac{1-x_{1t}y_{1t}}{[y_{1t}x_{1t}(1-x_{1t})\beta\frac{(1+v_{1t})}{2}+x_{1t}y_{1t}(1-y_{1t})\beta\frac{(1+u_{1t})}{2}]^2}[y_{1t}(1-2x_{1t})\beta\frac{(1+v_{1t})}{2}$$
$$+y_{1t}(1-y_{1t})\beta\frac{(1+u_{1t})}{2}]+\frac{y_{1t}}{y_{1t}x_{1t}(1-x_{1t})\beta\frac{(1+v_{1t})}{2}+x_{1t}y_{1t}(1-y_{1t})\beta\frac{(1+u_{1t})}{2}}\}\dot{x}_{1t}$$
$$+\{\frac{1-x_{1t}y_{1t}}{[y_{1t}x_{1t}(1-x_{1t})\beta\frac{(1+v_{1t})}{2}+x_{1t}y_{1t}(1-y_{1t})\beta\frac{(1+u_{1t})}{2}]^2}[x_{1t}(1-x_{1t})\beta\frac{(1+v_{1t})}{2}$$
$$+x_{1t}(1-2y_{1t})\beta\frac{(1+u_{1t})}{2}]+\frac{x_{1t}}{y_{1t}x_{1t}(1-x_{1t})\beta\frac{(1+v_{1t})}{2}+x_{1t}y_{1t}(1-y_{1t})\beta\frac{(1+u_{1t})}{2}}\}\dot{y}_{1t}\} \quad (103)$$

Suppose $v_{1t}=0$, $u_{1t}=1$, and $x_{1t}\to 1$, $y_{1t}\to 0$. While it is analytically complex to evaluate $\dot{V}$, numerically, if I plug in values of x_{1t} and y_{1t} close to the limit but not exactly there, for example, $x_{1t}=0.99$, $y_{1t}=0.01$:

$$
\begin{aligned}
\dot{V} = w_{1t}\{\{&\frac{1-(1-0.99)(1-0.01)}{[(1-0.01)0.99(1-0.99)\beta+(1-0.99)0.01(1-0.01)\beta\frac{1}{2}]^2}[(1-0.01)(1-2*0.99)\beta \\
&-0.01(1-0.01)\beta\frac{1}{2}]-\frac{1-0.01}{(1-0.01)0.99(1-0.99)\beta+(1-0.99)0.01(1-0.01)\beta\frac{1}{2}}\}\dot{x}_{1t} \\
&+\{\frac{1-(1-0.99)(1-0.01)}{[(1-0.01)0.99(1-0.99)\beta+(1-0.99)0.01(1-0.01)\beta\frac{1}{2}]^2}[-0.99(1-0.99)\beta \\
&+(1-0.99)(1-2*0.01)\beta\frac{1}{2}]-\frac{1-0.99}{(1-0.01)0.99(1-0.99)\beta+(1-0.99)0.01(1-0.01)\beta\frac{1}{2}}\}\dot{y}_{1t}\} \\
&+w_{2t}\{\{\frac{1-0.99*0.01}{[0.01*0.99(1-0.99)\beta\frac{1}{2}+0.99*0.01(1-0.01)\beta]^2}[0.01(1-2*0.99)\beta\frac{1}{2} \\
&+0.01(1-0.01)\beta]+\frac{0.01}{0.01*0.99(1-0.99)\beta\frac{1}{2}+0.99*0.01(1-0.01)\beta\frac{(1+1)}{2}}\}\dot{x}_{1t} \\
&+\{\frac{1-0.99*0.01}{[0.01*0.99(1-0.99)\beta\frac{1}{2}+0.99*0.01(1-0.01)\beta]^2}[0.99(1-0.99)\beta\frac{1}{2} \\
&+0.99(1-2*0.01)\beta]+\frac{0.99}{0.01*0.99(1-0.99)\beta\frac{1}{2}+0.99*0.01(1-0.01)\beta}\}\dot{y}_{1t}\} \\
&\approx \frac{1}{\beta}[w_{1t}(-10050\dot{x}_{1t}-52\dot{y}_{1t})+w_{2t}(509\dot{x}_{1t}+10050\dot{y}_{1t}) \qquad (104)
\end{aligned}
$$

Since $\dot{x}_{1t} > 0$, $\dot{y}_{1t} < 0$, $\dot{V} < 0$. For the equilibrium point where $x_{1t} \to 1$, $y_{1t} \to 0$, it is asymptotically stable in the $w_1 - w_2$ system.

Return to Section 3.1.

A.4 Proposition 9

Relative bias could affect strategy dynamics, $(\dot{x}_{1t}, \dot{y}_{1t})$, via x_{1t} and y_{1t}. Based on equations (73) and (74):

$$\frac{d\dot{x}_{1t}}{d\Delta\alpha^i} = x_{1t}(1-x_{1t})(1-2x_{1t})\beta[(1-\frac{v_{1t}}{2})\dot{w}_{1t}-(\frac{1}{2}+\frac{v_{1t}}{2})\dot{w}_{2t}-(\frac{w_{1t}}{2}+\frac{w_{2t}}{2})\dot{v}_{1t}] \tag{105}$$

$$\frac{d\dot{y}_{1t}}{d\Delta\alpha^{-i}} = y_{1t}(1-y_{1t})(1-2y_{1t})\beta[(1-\frac{u_{1t}}{2})\dot{w}_{1t}-(\frac{1}{2}+\frac{u_{1t}}{2})\dot{w}_{2t}-(\frac{w_{1t}}{2}+\frac{w_{2t}}{2})\dot{u}_{1t}] \tag{106}$$

Given stabilized beliefs, and suppose $v_{1t} = u_{1t} = 1$:

$$\frac{d\dot{x}_{1t}}{d\Delta\alpha^i} = x_{1t}(1-x_{1t})(1-2x_{1t})\beta(\frac{1}{2}\dot{w}_{1t}-\dot{w}_{2t}) \tag{107}$$

$$\frac{d\dot{y}_{1t}}{d\Delta\alpha^{-i}} = y_{1t}(1-y_{1t})(1-2y_{1t})\beta(\frac{1}{2}\dot{w}_{1t}-\dot{w}_{2t}) \tag{108}$$

Since both workers believe their opponent will apply with certainty to firm 1, such that $x_{1t} < \frac{1}{2}$, $y_{1t} < \frac{1}{2}$, and workers behave symmetrically, $x_{1t} = y_{1t}$. If $2\dot{w}_{2t} > \dot{w}_{1t}$, then $\frac{d\dot{x}_{1t}}{d\Delta\alpha^i} < 0$ and $\frac{d\dot{y}_{1t}}{d\alpha^{-i}} < 0$. Stronger relative bias to firm 1 would slow down the learning process. As for $v_{1t} = u_{1t} = 0$ and $2\dot{w}_{1t} > \dot{w}_{2t}$, stronger relative bias to firm 2 (i.e. $\Delta\alpha^i$, $\Delta\alpha^{-i} < 0$) would have the same impact.

For $v_{1t} = 0$, $u_{1t} = 1$, such that $x_{1t} \to 1$, $y_{1t} \to 0$,

$$\frac{d\dot{x}_{1t}}{d\Delta\alpha^i} = x_{1t}(1-x_{1t})(1-2x_{1t})\beta(\dot{w}_{1t}-\frac{1}{2}\dot{w}_{2t}) \tag{109}$$

$$\frac{d\dot{y}_{1t}}{d\Delta\alpha^{-i}} = y_{1t}(1-y_{1t})(1-2y_{1t})\beta(\frac{1}{2}\dot{w}_{1t} - \dot{w}_{2t}) \tag{110}$$

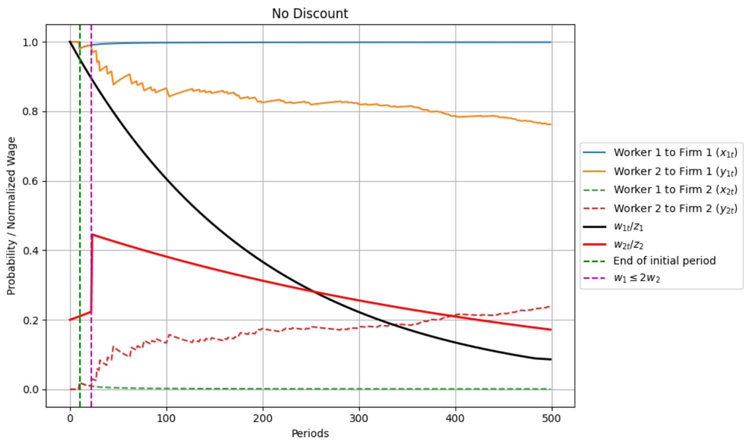
For $2\dot{w}_{2t} > \dot{w}_{1t} > \frac{1}{2}\dot{w}_{2t}$, $\frac{d\dot{x}_{1t}}{d\Delta\alpha^{i}} < 0$ and $\frac{d\dot{y}_{1t}}{d\alpha^{-i}} > 0$. As bias to firm 1 grows, there is diminishing sensitivity for worker 1. Since x_{1t} is already close to 1, the capacity for change is low. There is faster dynamics for worker 2, and the capacity for change remains high as y_{1t} is close to 0. Relative bias contribute to this asymmetric learning dynamics, where one side's learning curve is flattened and the other is more reactive. (Worker 1 stay near the equilibrium and worker 2 moves away.) This could lead to greater destabilization of workers' subsystem.

Return to Section 3.1.

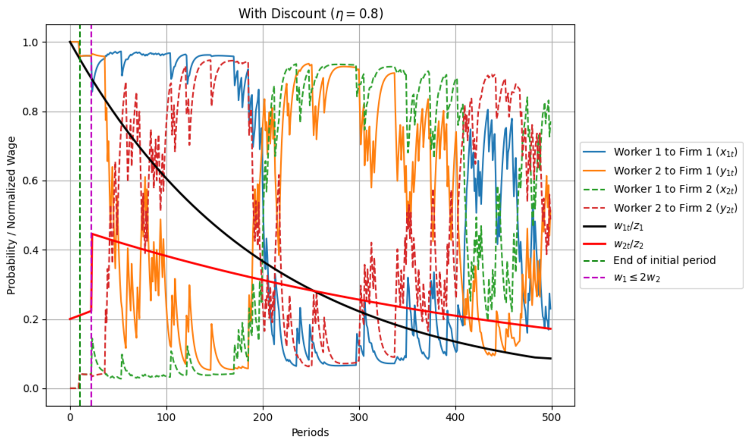
B Simulations

B.1 Example for Partial Recall of Experiences in 2-sided RL

Figure 21a and 21b show simulation of potential learning trajectory for workers with perfect and partial recall when wages are fixed to be some exogenous values that decline at a constant rate. Suppose I start with $G1$ (i.e. $w_1 \geq 2w_2$), and workers are fixed to be choosing firm 1 for 10 periods, during which, w_1 is arbitrarily adjusted downwards and w_2 upwards. Along the trajectory of exogeneous wage changes, wage condition could vary, thus shifting the game from G1 to G2 (i.e. $2w_1 > w_2 > \frac{w_1}{2}$), leading to a switch in equilibrium to learn from $(F1, F1)$ to the set $(F1, F2)$ and $(F2, F1)$ (the switching point is indicated by the pink dotted line). In this example, further shift from $G2$ to $G3$ is possible as w_1 and w_2 continues to decrease at the current rate, but not shown in the figures.



(a) Perfect Recall



(b) Partial Recall

For $z_1 = z_2 = 10$, $w_{10} = 10, w_{20} = 2$, all initial propensities are fixed at 1 ($x_{10} = y_{10} = 1$, $\theta^j_{a_0^j} = \theta^{-j}_{a_0^j} = 1$):

1. Assume $w_1 \geq 2w_2$ for initial 10 periods, workers are programmed to choose $(F1, F1)$, and w_1 adjust downwards by 0.5% and w_2 increases by 0.5% in each period. Over time, wage condition could reverse, and the instance where $w_1 \leq 2w_2$ is marked.
2. After 10 periods, workers are no longer fixed to choose firm 1. They can freely choose based on updated propensities and choice probabilities.
3. Once workers choose to apply to different firms, both w_1 and w_2 are programmed to decrease steadily by 0.5% per period.

Figure 21: Possible Learning Trajectory for Workers

When workers remember past events perfectly, Figure 21a shows that one of the worker would adjust his/her strategy and redirecting search towards a different firm from the other worker over time. In the long run, if the wage condition is sustained, it is expected that workers will apply with higher probability to different firms. However, the process of learning to play new equilibria can be long. There is a combined effect from learning the new set of NEs in G2, and also to overcome the initial learning experiences which direct workers to different set of NE in G1.

In Figure 21b, when there is a discount factor on past payoffs, the application strategies change more rapidly. There is a faster switch to applying more to different firms after the equilibrium switching point. But since workers are less locked-in by past experiences and put greater weight on recent payoffs, they are also more responsive to short-term positive feedback, resulting in greater fluctuations in choice probabilities. For such learning trajectory, where workers initially overcrowd at firm 1, the partial recall set-up could be beneficial in stimulating switching of job application and inducing coordination among workers at a faster pace. Workers are quicker to forget the initial experiences, thus are more adaptive to new market conditions. However, the pitfall of this is the volatility in application strategies. This could lead to a higher likelihood of mismatch, as workers' strategies are more influenced by recent events and exhibit greater stochasticity. As a result, their behaviour does not converge to applying more to different firms, even over an extended time period.

Return to Section 2.2.1.

B.2 Further Illustrations of Multiple Sessions (Algorithm 1)

In Table 2, I run the simulation 20 times to obtain values for 20 sessions. I then compute the average values for the last 20% of the periods for each session and listed the first and last session. I show that wages are positive when workers are more responsive to them (i.e. higher β). It is also more likely for workers to apply with higher probability to different firms when β is higher. Even though by averaging across the 20 sessions, I found for all β, the average choice probabilities concentrated on 0.5, which implies that workers are equally likely to select either of the asymmetric strategies in presence of multiple equilibria.

First and last session	w_1	w_2	x_{1t}	y_{1t}
Low β (0.2) [First]	0.000000	0.000000	0.500000	0.500000
Low β (0.2) [Last]	0.000000	0.000000	0.500000	0.500000
Medium β (1.0) [First]	4.331617	4.332060	0.499997	0.499995
Medium β (1.0) [Last]	4.331806	4.331872	0.499994	0.500004
High β (5.0) [First]	4.599850	4.600150	0.749995	0.250005
High β (5.0) [Last]	4.599290	4.600707	0.250003	0.749996
Average across sessions				
Low β (0.2)	0.000000	0.000000	0.500000	0.500000
Medium β (1.0)	4.331838	4.331839	0.500000	0.500000
High β (5.0)	5.870716	5.169464	0.462453	0.589455

Table 2: Average Values of Last 20% of 10000 Periods for 20 Simulation Sessions

Return to Section 3.1

B.3 Workers' Side Illustration of Long-Term Experience Bias

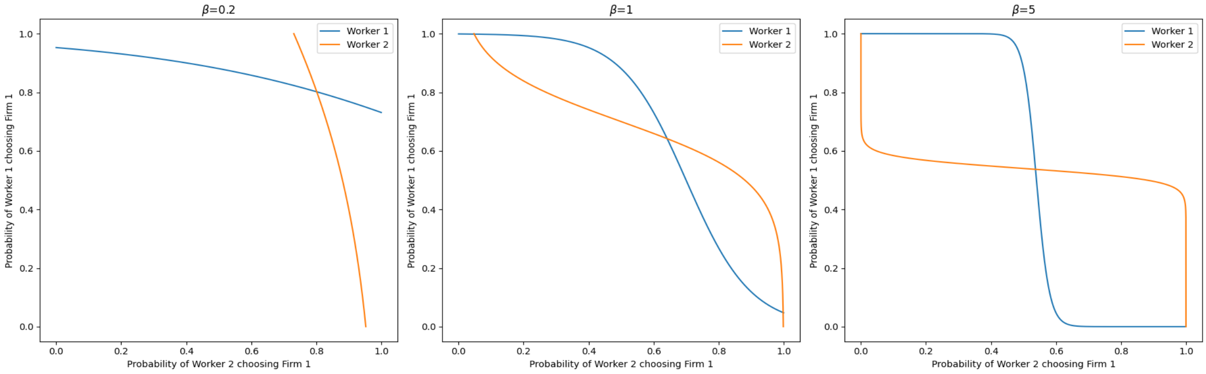


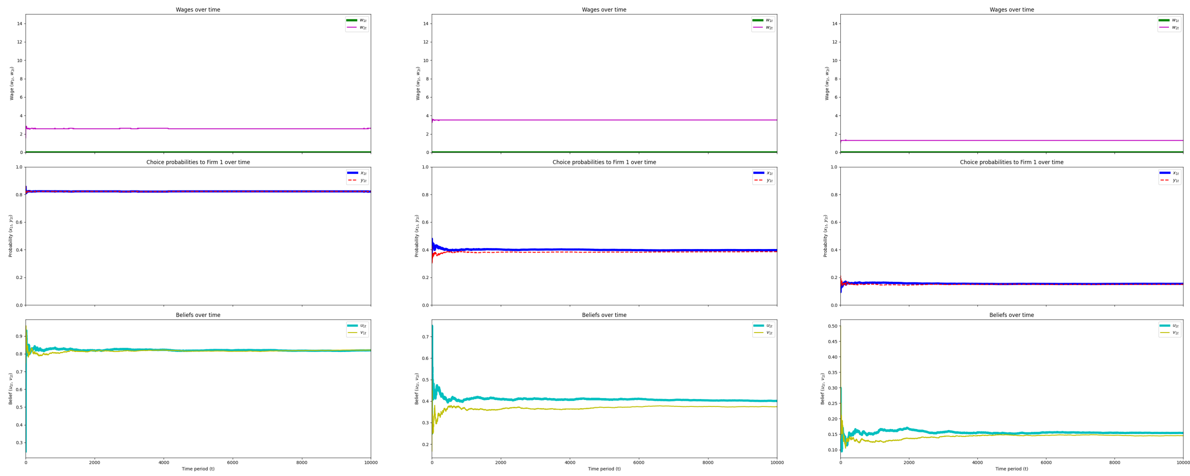
Figure show workers' choice probability for $\alpha_1^i = \alpha_1^{-i} = 2$ when fixing the wages to $w_1 = w_2 = 10$.

Figure 22: Workers' Response Functions with Long-Term Experiences

Return to Section 3.1

B.4 Changes in Wages and Workers' Choice Probabilities with Long-Term Experience Bias (Algorithm 3)

Figure 23 shows simulation examples when workers have experience bias towards firm 1 relative to firm 2. There is clear convergence to one set of workers' strategies for all β. At low β, workers' strategies are heavily influenced by their own bias, therefore, they apply with higher probability to firm 1. For high β, workers are more sensitive to expected payoff, their beliefs about each other adjusted accordingly, thus they could end up converging to apply with higher probability firm 2. Since expected payoffs are relatively low as compared to relative experiences, it is possible that the system only has a unique, symmetric equilibrium.



(a) $\beta = 0.2$ (b) $\beta = 1.0$ (c) $\beta = 5.0$

Figure shows wages, choice probabilities and belief evolution for different $\beta = \{0.2, 1.0, 5.0\}$, given $z_1 = z_2 = 10$, $t = 10000$, $\alpha_1^i = 2$, $\alpha_1^{-i} = 2$, $u_{10} = 0.5$, $v_{10} = 0.5$, no smoothing.

Figure 23: Changes in Wages and Workers' Choice Probabilities with Bias (Algorithm 3)

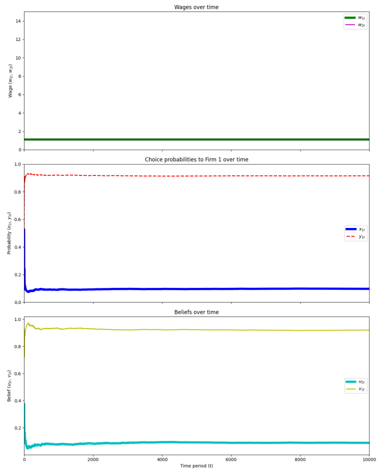


Figure shows wages, choice probabilities and belief evolution for different $\beta = \{5.0\}$, given $z_1 = z_2 = 10$, $t = 10000$, $\alpha_1^i = 0.1$, $\alpha_1^{-i} = 0.1$, $u_{10} = 0.5$, $v_{10} = 0.5$, no smoothing.

Figure 24: Changes in Wages and Workers' Choice Probabilities Over Time with Low Experience Bias (Algorithm 3)

For low experience bias, such as in Figure 24, there will still be multiple equilibria and convergence to one of the asymmetric equilibria will be observed as beliefs stabilize.

Return to Section 3.1.