Learning to Coordinate: Adaptive Learning and Equilibrium Selection in Labour Market*

Siting Estee Lu†

December 25, 2025

[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 insights on equilibrium selection. I propose two market structures, where wages are unobservable and observable to workers, and I model workers' application strategies over time using reinforcement learning and best response dynamics respectively. I show that in presence of multiple equilibria, experience-based learning generally induces workers to coordinate on a more efficient, locally asymptotically stable equilibrium in which they apply with high probability to different firms, in both static and dynamic wage-setting environments. Learning models not only highlight potential mechanisms for equilibrium selection, they also suggest process-oriented policies to improve market efficiency.

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

*Previously circulated as "Evolution of Labour Market Mismatch through Adaptive Learning with Experience".

†School of Economics, The University of Edinburgh. Contact: estee.l828@gmail.com.

I am indebted to Ed Hopkins and Axel Gottfries for their invaluable suggestions and continuous guidance. I would also like to express my gratitude to Michael Woodford for his valuable feedback and suggestions. I also hope to thank audiences at Max Planck Institute for Research on Collective Goods, Toulouse Summer School in Quantitative Social Sciences 2023, and SGPE PhD Conference 2023, for their valuable comments.

1 Introduction

Workers' job application decisions are often guided by their past experiences. While experience-based adaptive learning has been studied across strategic games, choice predictions, policy attitudes (Roth and Erev (1995); Camerer and Hua Ho (1999); Erev et al. (2010); Albarracin and Wyer Jr (2000)), job search models have largely overlooked the role of application experiences, leaving a gap in understanding how past application successes and failures shape workers' search heuristics over time and affect equilibrium outcomes. As a result, I develop an evolutionary model of job search to analyse long-run coordination in the labour market.

Empirically, job applications are often "sticky" and highly concentrated. Workers may fail to switch to better-paying jobs even with wage transparency, and they could exhibit sorting patterns not well-explained by skill differences (e.g. Archer (2016); Barbulescu and Bidwell (2013); Bamieh and Ziegler (2025)). These phenomena, which simple rational models struggle to explain, could suggest the presence of history-dependent mechanism like familiarity-based learning (Hopkins (2007)). Furthermore, theoretically, while search models may be able to identify multiple equilibria that could differ in efficiency, they are typically silent on the transition path of strategies and on equilibrium selection. Therefore, by explicitly modelling experience-based learning dynamics, I can contribute to explaining the puzzle of "sticky" search, and showing how bounded rational agents select among multiple equilibria, supported by stability analysis.

In this paper, I study a simple framework with two firms and two workers. Firms set wages, either statically or dynamically, to attract applications, and workers choose where to apply, their strategies are adaptive in response to private, learned signals from their past application experiences. I analyse two market structures, one with unobservable wages and one with observable wages. These set-ups could shed light on how experience shapes strategy transitions and equilibrium selection under different market designs.

When workers do not observe wages, they learn solely based on feedback from the jobs they have applied to. Since experimental evidence suggests that individuals often display learning pattern close to reinforcement learning (Erev and Roth (1998)), and since workers typically do not observe payoff from firms they did not apply to, I model their search behaviour using reinforcement learning, where application strategies are revised based on past personal outcomes ("self"-learning) or via knowledge diffusion from prior generations. I show that workers learn to coordinate on applying to different firms in the long run, and these pure strategy equilibria are locally asymptotically stable. I then extend the framework to two-sided reinforcement learning, allowing firms to also adjust wage in response to past realised payoffs. In this case, learning can push wages to (near-)zero, and workers' long-run search behaviour may be more random, which increases the likelihood of mismatch.

To provide a crucial point of contrast to simple reinforcement learning agents, and to capture modern job markets in which wages can be advertised and competitor's strategy is the primary source of uncertainty, I consider a second market structure in which workers choose where to apply after perfectly observing wages. I model workers' search behaviour using best response learning dynamics (Fudenberg and Levine (1998); Hopkins (1999)), in which they take into account observed wages and updated beliefs about the other worker's strategy. Given past realised actions, workers revise their choice probabilities as their beliefs about each other evolve. Firms, in turn, infer both workers' evolving beliefs from historical actions and set wages in response to how they expect workers to behave in each period. I show that, under certain conditions, multiple equilibria exist, and in presence of which, workers could converge in the long run to asymmetric equilibria of applying with higher probability to different firms that are locally asymptotically stable.

By introducing experience-based learning, I offer a behaviourally realistic account of search,

showing how workers' strategies can adapt over time under different market structures. Furthermore, experience-based learning may serve as a powerful mechanism for equilibrium selection, which actively weeds out inefficient outcomes and guides the market towards a more coordinated and efficient state, and this also opens the door to discussions of novel, process-oriented policies.

The rest of the paper is structured as follows: Section 2 studies a setting with wage opacity and introduces reinforcement learning framework. Section 3 explores a setting with perfect wage observability and introduces best response learning dynamics. Section 4 concludes by discussing market and policy implications, as well as outlining potential extensions.

Related literature. This work rides on top of extensive literature on experience-based learning to provide an evolutionary narrative for the labour market dynamics. There can be many possible learning dynamics, but Erev and Roth (1998) demonstrates that in experimental games, individuals display learning pattern that most closely resembles reinforcement learning. I therefore adopt this mechanism as the baseline approach in this paper to model job search behaviour. However, Hopkins (2002) highlights that when comparing between different learning models, this force of habit model is statistically insignificant in explaining Van Huyck et al. (1997)'s experimental results. Camerer and Hua Ho (1999) also postulates that an experience-weighted learning framework may fit individuals' learning pattern better, which would account for both actual and counterfactual payoffs if a worker select a different action. However, in a setting where workers do not observe payoffs from unchosen actions, the counterfactual component would be unavailable without additional assumptions, therefore, reinforcement learning could be the more natural baseline in this context. Nonetheless, when workers can fully observe wages ex-ante to application decisions, there could be merit in accounting for opponent's empirical frequency of choices as one can infer the expected payoff of the action not chosen. Therefore, I also explore best response dynamics (Fudenberg and Levine (1998); Hopkins (1999)) that 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 are not entirely 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 experience is experiential job search highlighted by Kanfer and Bufton (2018), where workers adapt based on past involuntary job loss. 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, a 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 use 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 as they were perceived to be more difficult to coordinate on, which is a common approach in directed search literature (Galenianos and Kircher (2009)). However, experience-based learning could suggest potential mechanisms for equilibrium selection. It not only clarifies whether workers could coordinate on applying to different firms in the long run, but also highlights how process-oriented policies may influence

the learning path and unlock efficiency gains from equilibria that might otherwise be overlooked.

This work also contributes to understanding some puzzles in real world observations. For instance, the phenomenon of workers' lack of job switch that are less affected by wages, but more by inertia such as feeling too comfortable and fear of losing identity (Archer (2016)); sorting behaviours displayed by different genders resulting from accumulation of experiences and beliefs built over the generations (Barbulescu and Bidwell (2013)), which differ from sorting due to skills (Eeckhout (2018)) or risk preferences (Fouarge et al. (2014)). These may be supported by reinforcement learning behaviour, which exhibits familiarity-based learning pattern (Hopkins (2007)). Learning models could also substantiate 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 such sorting behaviour in application choices, and provide basis for policies to tackle potential mismatch as a result of it.

Last but not least, job search models often concentrate 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 sub-market 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, showing how experience shapes micro-level strategy formation and helping to inform policies that target individual choices with potentially important macro-level implications.

2 Reinforcement Learning for "Black Box" Markets

In this section, I propose a market structure that conveys a traditional offline job search process, or search with online platforms where wage information is not explicitly revealed, thus referred to as "Black Box" markets. 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 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. Firms observe exogenous realization of productivity, denoted by $\mathbf{z} = \{z_1, z_2\}$, $\mathbf{z} \in Z$. They set wages, $\mathbf{w} = \{w_1, w_2\}$, which are fixed throughout all periods. Time is discrete, $t = \{0, 1, ..., T\}$, representing generations of workers and firms.[1] In each period:

1. While workers do not observe wages ex-ante to their selection between firm 1 and 2, they can recall from past "self" or learn from the previous generation the application strategy, the realized choice and payoff. (For example, worker 1 of $t = 1$ learns from worker 1 of $t = 0$, same applies for worker 2.)
2. They devise an application strategy based on all the information.
3. Once workers made a choice, they observe a payoff, update their choice probabilities of each firm following Erev and Roth (1998)'s reinforcement learning algorithm. They then drop out of the market, never to return.

Firms' side. In this fixed wage environment, wages are assumed to be rigid, such that firms set wages at the beginning of time and do not change them. Hereafter, suppose fixed wages satisfy $2w_1 > w_2 > \frac{w_1}{2}$, and 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 for the 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 realized payoff from the previous period (Nowé et al. (2012)). The search problem faced by the workers can be illustrated with a standard coordination game (Figure 1), where payoffs are fully revealed in the equilibrium. Workers are assumed to be homogeneous, such that they have equal probability of being 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	$\frac{w_2}{2}$, $\frac{w_2}{2}$

Figure 1: Application Game Faced by Workers

[1] For approximation of stochastic process in later parts, I consider $T \to \infty$, implying infinitely many generations.

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 $\Delta_i = \{x_t = (x_{1t}, x_{2t})\}^T$, where $\sum_{j=1}^{2} x_{jt} = 1$; and for worker $-i$, $\Delta_{-i} = \{y_t = (y_{1t}, y_{2t})\}^T$, where $\sum_{j=1}^{2} y_{jt} = 1$. 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 π_t^i, where $\pi_t^i(a_t^i, a_t^{-i})$ depends on $a_t^i \in A^i, a_t^{-i} \in A^{-i}$, which 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)). Furthermore, 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. For homogeneous workers, their probability of being hired when applying to the same firm is $\frac{1}{2}$, therefore, there is partial reinforcement of workers' choices. It can be perceived that workers' choices are 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 viewed 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) 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 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 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 reinforcement to select the same firm is discounted.

Workers only receive feedback from actions they actually take. The impact from the other worker's application strategy is implicit, where worker $-i$'s choice in period $(t-1)$ affects worker i in period t only via realized past payoffs.

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 choice probabilities** using Lemma 1 of Hopkins (2002) with the choice rule (1) and the update rule (2):

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

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

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 selecting the two firms at time t, by the two workers respectively. $R(\cdot)$ is the replicator operator, reflecting how strategies evolve based on their relative payoffs.

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

The above 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. $Q_t^i = \sum_{j=1}^{J} q_{jt}^i$ 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 last term being the error term becomes smaller, the system would approximate continuous time dynamics in the limit. This also relates to the step size, which describes the rate at which workers updates their strategies. In this model, workers have different step sizes 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. By equations (4) and (5):

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

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

Given possibility of three equilibria in this setting, workers could either coordinate on applying to different firms or adopt a mixed strategy that suggest more randomized search behaviour. Based on Hopkins and Posch (2005), Erev and Roth (1998) reinforcement learning rule would not converge to a 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 experience 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 initial conditions.

As $t \to \infty$, $Q_t^i \to \infty$ and $Q_t^{-i} \to \infty$, the expected change in strategies vanishes and the drift stabilizes, the stochastic process (x_t, y_t) can be approximated by a deterministic continuous time dynamics (Benaım and Hirsch (1999)). Based on equations (6) and (7):

$$\frac{dx_{1t}}{dt} = \lim_{Q_t^i \to \infty} Q_t^i(\mathbb{E}(x_{t+1}|q_t^i) - x_t) = R(x_t)Wy_t \tag{8}$$

$$\frac{dy_{1t}}{dt} = \lim_{Q_t^{-i} \to \infty} Q_t^{-i}(\mathbb{E}(x_{t+1}|q_t^{-i}) - y_t) = R(y_t)W^T x_t \tag{9}$$

Given payoff matrices, W and W^T (3),

$$\frac{dx_{1t}}{dt} = f(x_{1t}, y_{1t}) = x_{1t}(1 - x_{1t})[(-\frac{w_1}{2} - \frac{w_2}{2})y_{1t} + w_1 - \frac{w_2}{2}] \tag{10}$$

$$\frac{dy_{1t}}{dt} = g(x_{1t}, y_{1t}) = y_{1t}(1 - y_{1t})[(-\frac{w_1}{2} - \frac{w_2}{2})x_{1t} + w_1 - \frac{w_2}{2}] \tag{11}$$

Proposition 1 (Local Stability of Asymmetric Equilibria under Learning from Feedback)**.** *For wages satisfy $2w_1 > w_2 > \frac{w_1}{2}$, the asymmetric equilibria are locally asymptotically stable.*

Proof. Based on equations (10) and (11), the Jacobian matrix:

$$J = \begin{bmatrix} \frac{\partial f}{\partial x_{1t}} & \frac{\partial f}{\partial y_{1t}} \\ \frac{\partial g}{\partial x_{1t}} & \frac{\partial g}{\partial y_{1t}} \end{bmatrix} = \begin{bmatrix} (1 - 2x_{1t})[(-\frac{w_1}{2} - \frac{w_2}{2})y_{1t} + w_1 - \frac{w_2}{2}] & x_{1t}(1 - x_{1t})(-\frac{w_1}{2} - \frac{w_2}{2}) \\ y_{1t}(1 - y_{1t})(-\frac{w_1}{2} - \frac{w_2}{2}) & (1 - 2y_{1t})[(-\frac{w_1}{2} - \frac{w_2}{2})x_{1t} + w_1 - \frac{w_2}{2}] \end{bmatrix} \tag{12}$$

For symmetric strategy, $x_{1t} = y_{1t}$, denotes this as $\hat{xy}$, evaluating eigenvalues using $\det(J - \lambda I) = 0$:

$$\lambda = (1 - 2\hat{xy})[(-\frac{w_1}{2} - \frac{w_2}{2})\hat{xy} + w_1 - \frac{w_2}{2}] \pm \sqrt{(\hat{xy}(1 - \hat{xy})(-\frac{w_1}{2} - \frac{w_2}{2}))^2} \tag{13}$$

Given $2w_1 > w_2 > \frac{w_1}{2}$ is satisfied, at the symmetric equilibrium point, there will be a positive and a negative eigenvalue. The symmetric equilibrium is a saddle point.

In the special case of homogeneous firms, $z_1 = z_2 = z$, $w_1 = w_2 = w^*$, at $x_{1t} = y_{1t} = \frac{1}{2}$:

$$J(\frac{1}{2}, \frac{1}{2}) = \begin{bmatrix} 0 & -0.25w^* \\ -0.25w^* & 0 \end{bmatrix} \tag{14}$$

$\lambda = \pm 0.25w^*$. This is a saddle point and is unstable for any positive wages.

For asymmetric strategies, $x_{1t} = 1, y_{1t} = 0$ or $x_{1t} = 0, y_{1t} = 1$, the eigenvalues are:

$$\lambda_1 = (1 - 2x_{1t})[(-\frac{w_1}{2} - \frac{w_2}{2})y_{1t} + w_1 - \frac{w_2}{2}], \lambda_2 = (1 - 2y_{1t})[(-\frac{w_1}{2} - \frac{w_2}{2})x_{1t} + w_1 - \frac{w_2}{2}] \tag{15}$$

Given $2w_1 > w_2 > \frac{w_1}{2}$, at $x_{1t} = 1, y_{1t} = 0$, $\lambda_1 = -w_1 + \frac{w_2}{2} < 0$, $\lambda_2 = \frac{w_1}{2} - w_2 < 0$; at $x_{1t} = 0, y_{1t} = 1$, $\lambda_1 = \frac{w_1}{2} - w_2 < 0$, $\lambda_2 = -w_1 + \frac{w_2}{2} < 0$. They are locally asymptotically stable. □

Example 1. *(Unstable Mixed Strategy Equilibrium) Suppose $w_1 = 3$, $w_2 = 2$, there are 3 possible equilibria: (F1, F2), (F2, F1), and (F1, F2, $\frac{4}{5}$, $\frac{1}{5}$).*

$$\frac{dx_{1t}}{dt} = f(x_{1t}, y_{1t}) = x_{1t}(1 - x_{1t})(2 - 2.5y_{1t}) \tag{16}$$

$$\frac{dy_{1t}}{dt} = g(x_{1t}, y_{1t}) = y_{1t}(1 - y_{1t})(2 - 2.5x_{1t}) \tag{17}$$

$$J(x_{1t}, y_{1t}) = \begin{bmatrix} 0 & -0.4 \\ -0.4 & 0 \end{bmatrix} \Bigg|_{\frac{4}{5}, \frac{4}{5}} \tag{18}$$

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

Figure 2 shows the phase plot and strategy change over time for the example. Most learning trajectories demonstrate workers should end up in one of the pure strategy equilibria.

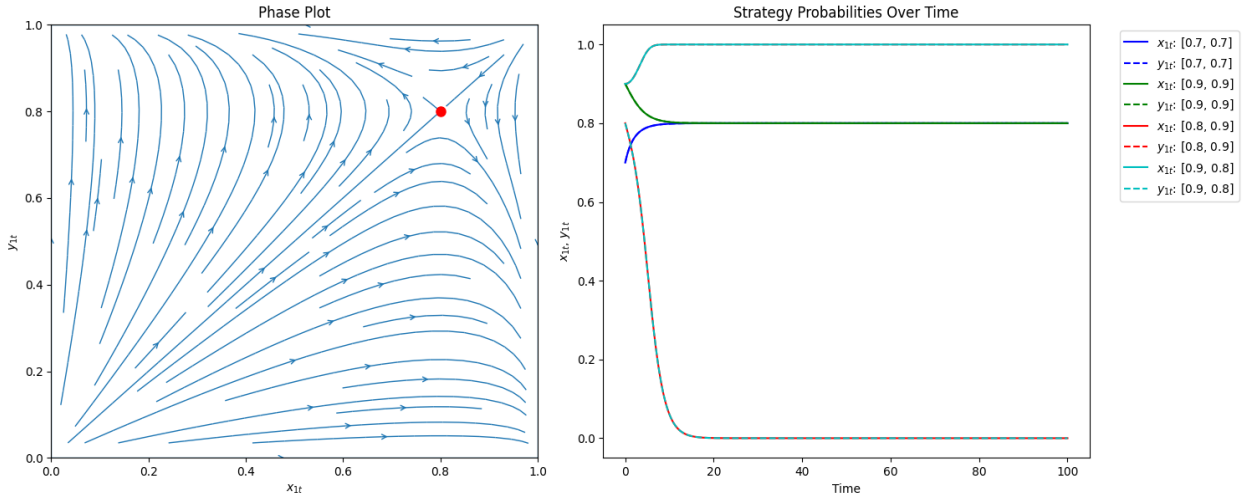


Figure 2: Illustration of Example 1

Erev and Roth (1998) highlighted a special feature of this learning mechanism, which is its heavy reliance on initial propensities and decreasing effect of recent experiences as accumulated payoffs become larger.

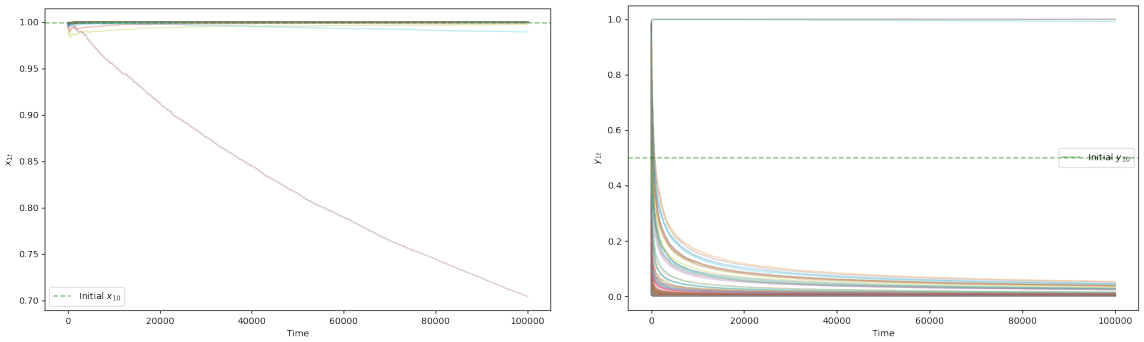
Observation 1 (Initial Propensities and Experiences on Equilibrium Selection)**.** *Equilibrium selection is influenced by initial propensities, as well as initial experiences from first few rounds of application:*

- *Strong prior preference or goodwill towards specific firm could create fundamental inertia for workers to adjust their application strategies. There could be 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. Despite this, there could be multiple periods where workers oscillate between choosing firm 1 and 2, leading to possible overcrowding for many periods. Furthermore, if both workers possess high initial propensities towards the same firm, there could be biased search for even more number of periods.

Hopkins (2007) also demonstrates that consumers could be stuck with choosing certain good that they initially prefer and undergo familiarity-based learning. The same could apply to job search context. Workers choosing to apply to firm 1 and 2 respectively in initial periods would make $(F1, F2)$ more likely to be selected than other equilibria in the long run. This may suggest that for homogeneous workers, who could start with random search, their equilibrium choice may be a result of randomness and luck, driven by positive reinforcement in the initial periods of their job applications. This also indicates that sorting behaviour can be experience-driven when skills are held constant.

Example 2. *(Heterogeneous Initial Bias) Suppose initial propensities are* $q_{10}^i = 1000, {q_{10}}^{-i} = 1$, $q_{20}^i = {q_{20}}^{-i} = 1$, *worker 1 has higher propensity to firm 1 than firm 2. Their choice probabilities to firm 1 are* $x_{10} = \frac{1000}{1001} \approx 0.999$, $y_{10} = \frac{1}{2}$. *Worker 1 applies with higher probability to firm 1 by default as compared to worker 2. As a result, as* $t \to \infty$, $(F1, F2)$ *is more likely to be reached.*



(a) Worker 1 (b) Worker 2

Figures show workers' application probability to firm 1 against number of periods 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 3: Learning Path of Worker 1 and 2

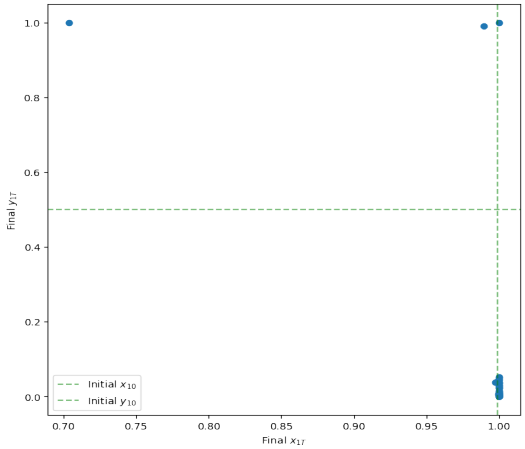


Figure 4: Workers' Choice Probability at Period $T = 100000$

Figure 3a and 3b are 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 almost immediate sorting behaviour. In this example, since worker 1 has a default strategy of applying with higher probability to firm 1, it is more likely than worker 2 in choosing firm 1 during initial rounds and thus locked-in to firm 1. However, given initial propensities are non-zero, there is always positive chance of selecting an alternative action, therefore, it remains possible for $(F2, F1)$ *to be selected in the long run, but the likelihood of converging to* $(F1, F2)$ *is much higher. In Figure 4, I show the choice probability in the last period of the simulation* $(T = 100000)$, *most of the sessions end at close to* $(F1, F2)$.

In Barbulescu and Bidwell (2013), students with the same education background 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 this model, these can be interpreted as worker heterogeneity in initial propensities, which then translate into dispersion in choice probabilities over the learning trajectory 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. For instance, 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, $\eta \in (0,1)$, in the propensity updating process, to account for the recency effect (Erev and Roth (1998)).

$$q^i_{j(t+1)} = (1-\eta)q^i_{jt} + \pi^i_{jt}(a^i_t, a^{-i}_t) \tag{19}$$

As a result, past experiences or prior knowledge could have a diminishing effect on current application decisions.

Proposition 2 (Partial Recall)**.** *Let $\eta \in (0,1)$ be the experience decay parameter for the updating process of q^i_{jt}, and π^i_{jt} be the realized payoff at time t. For T periods,*

1. *As $\eta \to 0$, the updating process converges to perfect recall.*
2. *As $\eta \to 1$, $q^i_{j(t+1)} \approx \pi^i_{jt}(a^i_t, a^{-i}_t)$, propensity could be determined solely by realized payoff.*
3. *For $0 < \eta < 1$, influence of initial propensity q^i_{j0} diminishes as $T \to \infty$.*

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

$$\lim_{\eta \to 0} q^i_{jT} = q^i_{j0} + \pi_{j0} + \pi_{j1} + ... + \pi_{jT-2} + \pi_{j(T-1)} \tag{20}$$

Same as perfect memory. And as $\eta \to 1$,

$$\lim_{\eta \to 1} q^i_{jT} = \pi_{j(T-1)} \tag{21}$$

For $0 < \eta < 1$,

$$q^i_{jT} = (1-\eta)^T q^i_{j0} + (1-\eta)^{T-1}\pi_{j0} + (1-\eta)^{T-2}\pi_{j1} + ... + (1-\eta)\pi_{jT-2} + \pi_{j(T-1)} \tag{22}$$

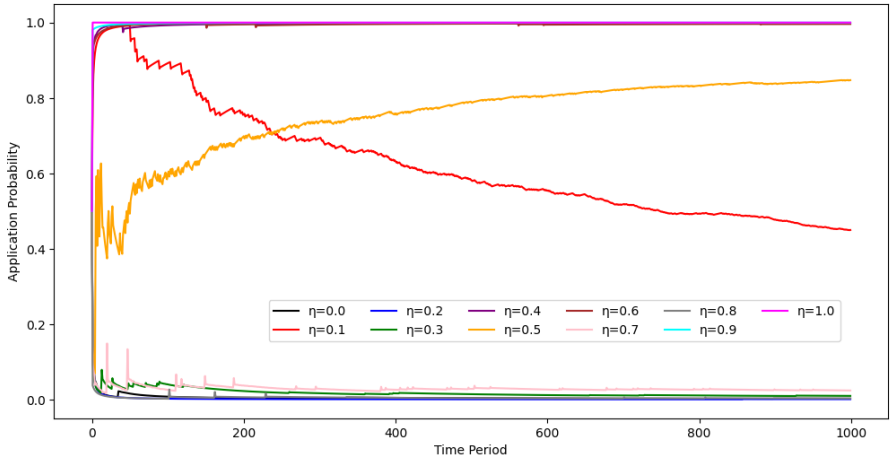
There is positive weight on both realized payoff and previous period propensities. Coefficient on q^i_{j0} is $(1-\eta)^T$, $\lim_{T\to\infty}(1-\eta)^T = 0$. □

In the extreme case of zero recall (i.e. $\eta = 1$), the probability of worker 1 selecting firm 1 depends solely on last period payoff:

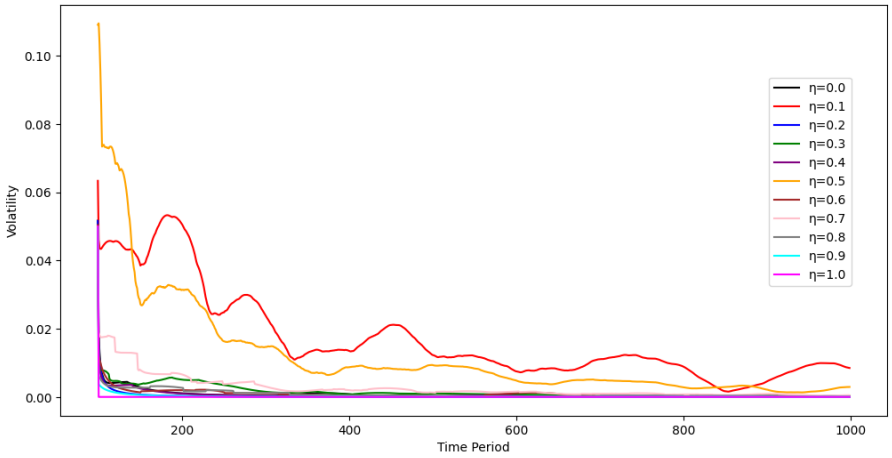
$$\lim_{\eta \to 1} x_{1T} = \frac{\pi^i_{1(T-1)}}{\pi^i_{1(T-1)} + \pi^i_{2(T-1)}}, \lim_{\eta \to 1} y_{1T} = \frac{\pi^{-i}_{1(T-1)}}{\pi^{-i}_{1(T-1)} + \pi^{-i}_{2(T-1)}} \tag{23}$$

Since workers can only select one firm in each period, this implies either $\pi_{1(T-1)}$ or $\pi_{2(T-1)}$ will be positive, therefore, for worker 1, 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, there can be slower augmentation of Q^i_t and Q^{-i}_t. While workers are expected to converge to pure NEs in the long run, limited memory could



(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, $\eta \in [0, 1]$ for the first 1000 periods ($w_1 = w_2 = 5$, $t = 100000$, ${q_{10}}^i = {q_{20}}^i = {q_{10}}^{-i} = {q_{20}}^{-i} = 1$).

Figure 5: Worker 1's Choice Probability of Firm 1 with Forgetting

contribute to longer process of reaching efficient outcome, and thus possibly more instances of mismatch.

In Figure 5a, 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, as described previously. It is possible that if both workers apply to firm 1 and receive positive feedback of $\frac{w_1}{2}$, they could end up overcrowding at firm 1. When workers retain some memories ($\eta < 1$), there is convergence to pure NEs, but choice probabilities can be noisy. In Figure 5b, I show that choice probability can be volatile[2] when $\eta < 1$. 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 necessarily imply more volatile outcomes.

Result Summary. In a fixed wage environment, where wages satisfy the existence condition for multiple equilibria, workers eventually coordinate on applying to different firms in the long run. Experience thus serves as a natural mechanism for selecting an efficient market outcome. However, coordination is not immediate, the market may experience prolonged mismatch before converging to pure NEs. When early experiences are sufficiently diverse, workers are expected to apply to different firms rapidly; but if both workers apply for the same job initially and obtain similar experiences, it could take substantial number of periods for them to adjust, leading to longer periods of mismatch.

2.2 Two-sided Reinforcement Learning with Dynamic Wages

In this section, I relax the assumption on wage rigidity, such that firms can also be adaptive learners in learning to set wages. I explore how workers react to dynamically changing wages, and if they would learn to coordinate on applying to different firms; and I also investigate how wages evolve.

In this set-up, firms and workers are learning simultaneously and both sides 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 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 chooses w_1 and receives applications from both workers and is able to produce, then the selected wage is reinforced via an update rule; if firm 1 did not receive any worker, then the selected wage 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 $\Delta_j = \{\omega_t^j = (\omega_{0t}^j, \omega_{1t}^j, ..., \omega_{zt}^j)\}^T$, 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, a^{-j} being chosen by firm 1 and 2 at time t.

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

- **Rewards.** Workers' indicator functions:

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

If at least one worker applies to firm j, firm j receives a payoff of $z_j - w_j$, otherwise 0.

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

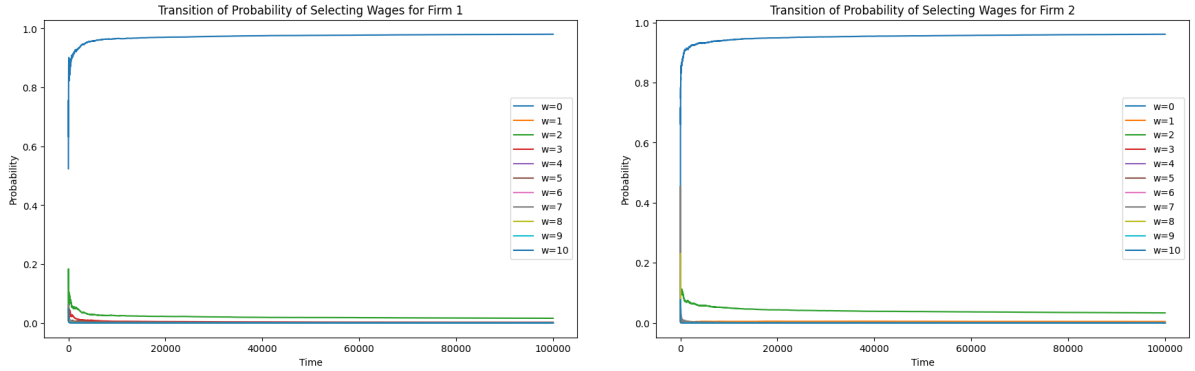
$$\text{Firm } j\text{: } \omega^j_{a_t^j} = \frac{\theta^j_{a_t^j}}{\sum_{a^j=0}^{z_j} \theta^j_{a_t^j}}, \text{Firm } -j\text{: } \omega^{-j}_{a_t^{-j}} = \frac{\theta^{-j}_{a_t^{-j}}}{\sum_{a^{-j}=0}^{z_{-j}} \theta^{-j}_{a_t^{-j}}} \tag{24}$$

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

$$\text{Firm } j\text{: } \theta^j_{a^j_{(t+1)}} = \theta^j_{a_t^j} + \pi_t^j(a_t^i, a_t^{-i}, a_t^j, a_t^{-j}), \text{Firm } -j\text{: } \theta^{-j}_{a^{-j}_{(t+1)}} = \theta^{-j}_{a_t^{-j}} + \pi_t^{-j}(a_t^i, a_t^{-i}, a_t^j, a_t^{-j}) \tag{25}$$

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 6 demonstrates growing probability of wage 0 being chosen by both firms.



(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 $t = 100000$, ${q_{j0}}^i = {q_{j0}}^{-i} = 1$, $\theta^j_{a_0^j} = \theta^{-j}_{a_0^{-j}} = 1$, $z_1 = z_2 = z = 10$.

Figure 6: Learning Path of Firm 1 and 2 in 2-sided RL

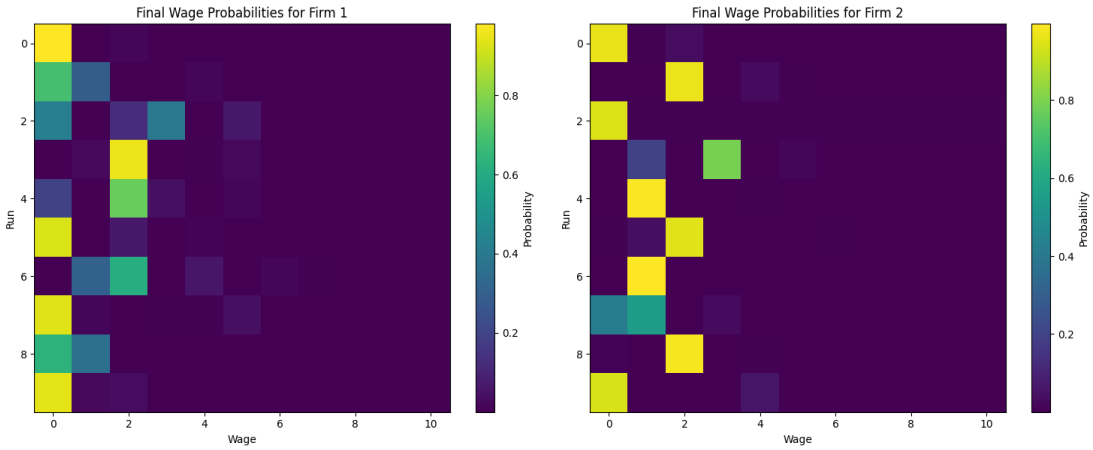


Figure 7: Firms' Probability of Selecting Each Wage at Period $T = 100000$ for 10 Sessions

In Figure 7, the choice probabilities of wages in the final period ($T = 100000$) of 10 simulated sessions (runs) were shown. The probabilities of setting low wages are higher. Both figures above suggest that wages will eventually be pushed down to 0.

For workers' side, Figure 8 shows workers' choice probabilities in the final period ($T = 100000$) of the 10 simulated sessions. There is higher likelihood of $(F1, F2)$ being selected. However, compare to the results for fixed wage environment in Figure 4, the choice probabilities are less saturated around pure NEs. There are also higher chances of workers applying to the same firm in period T, which could lead to inefficient outcome.

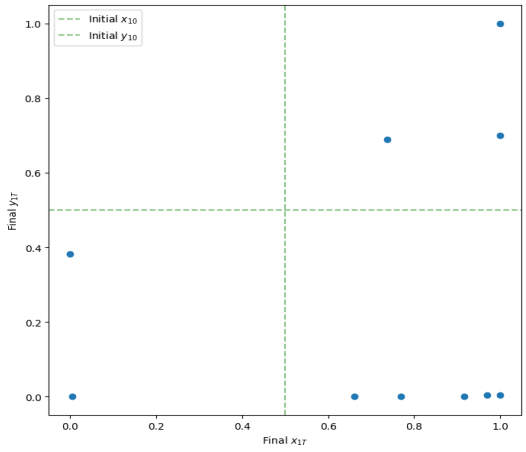


Figure 8: Workers' Choice Probability at Period $T = 100000$

Firms' expected change in probabilities of setting wages:

$$\text{Firm } j\text{: } \mathbb{E}[\omega^j_{a^j_{t+1}} | \theta^j_{a^j_t}] - \omega^j_{a^j_t} = \frac{[1-(1-x_{jt})(1-y_{jt})](z_j - a^j_t)}{S^j_t} + O(\frac{1}{S^{j\,2}_t}) \tag{26}$$

$$\text{Firm } -j\text{: } \mathbb{E}[\omega^{-j}_{a^{-j}_{t+1}} | \theta^{-j}_{a^{-j}_t}] - \omega^{-j}_{a^{-j}_t} = \frac{[1-(1-x_{(-j)t})(1-y_{(-j)t})](z_{-j} - a^{-j}_t)}{S^{-j}_t} + O(\frac{1}{S^{-j\,2}_t}) \tag{27}$$

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 sum of propensities over different actions for each firm.

As $t \to \infty$, S_t^j and S_t^{-j} grow larger, from equations (26) and (27), adjustments to strategies become less significant over time,

$$\mathbb{E}[\omega^j_{a^j_{t+1}}|\theta^j_{a^j_t}] - \omega^j_{a^j_t} \approx \frac{[1-(1-x_{jt})(1-y_{jt})](z_j - a^j_t)}{S^j_t}, \mathbb{E}[\omega^j_{a^j_{t+1}}|\theta^j_{a^j_t}] - \omega^j_{a^j_t} \to 0 \tag{28}$$

$$\mathbb{E}[\omega^{-j}_{a^{-j}_{t+1}}|\theta^{-j}_{a^{-j}_t}] - \omega^{-j}_{a^{-j}_t} \approx \frac{[1-(1-x_{(-j)t})(1-y_{(-j)t})](z_{-j} - a^{-j}_t)}{S^{-j}_t}, \mathbb{E}[\omega^{-j}_{a^{-j}_{t+1}}|\theta^{-j}_{a^{-j}_t}] - \omega^{-j}_{a^{-j}_t} \to 0 \tag{29}$$

If workers' learning processes converge, then firms' wage-setting strategies also converge.

The long run behaviour under two-sided reinforcement learning can be considerably more complex than one-sided. As workers' strategies stabilize, wages would be pushed downwards since any positive wage mechanically reduces profits. In a static, fully rational model with multiple equilibria, one would expect perfect worker coordination to be ultimately exploited in the limit, with wages driven to 0.

However, in the learning environment, the zero-wage pure strategy outcome is not "absorbing" in an economically meaningful way. When $w = 0$, the system admits a continuum of weak NEs: both the Jacobian and the Hessian collapse to 0 at these points, implying neutral stability rather than a strongly attracting steady state. As a result, equilibrium selection is inherently path-dependent, where the process ends up along this continuum depends on the direction of probability adjustments while wages are still positive.

Nonetheless, this outcome is also not stable due to asymmetric reinforcement. While it is clear on the firm side that any successful match at a lower wage is strongly reinforcing, persistently biasing wage towards 0. On the worker side, by contrast, it is less obvious. Low wages provide only weak reinforcement signals. Early in the learning process, when wages are still positive and potentially high, accumulated propensities can generate a "weight of the past", making workers' strategies sluggish and creating inertia that temporarily resembles convergence toward pure application patterns. But as wages decline, workers receive little feedback that meaningfully differentiates actions, so choice probabilities update slowly and minimally. There is no strong force pinning (x_t, y_t) to a particular pure strategy and workers could stop learning in absence of reward. However, more crucially, near-zero wages do not eliminate all dynamics. Since the framework allows firms to retain a positive, even if almost negligible, probability of occasionally posting higher wages, "accidental" high wage matches can reintroduce learning pressure and pull workers away from near-pure behaviour. This drift can in turn perturb firms' incentives to keep wages at the floor, allowing temporary wage increases and renewed randomization. As a result, the system may settle into a low wage dynamic regime, where firms remain biased toward low wages, while workers remain persistently mixed rather than converging cleanly to any single pure equilibrium. The combination of inertia and asymmetric reinforcement makes equilibrium selection path-dependent and potentially history-sensitive, the process may "lock-in" when reinforcement signals become weak, yet still drift within a neighbourhood of low wages when intermittent high-wage realizations revive learning.

Furthermore, as wages evolve endogenously, workers could effectively face a sequence of different games, characterized by the prevailing wage conditions. As the environment transitions across wage regimes, the relevant equilibrium set can change, opening the door to equilibrium switching, where agents could be learning different sets of NEs at different stages of the wage path. This shifting game landscape can further amplify persistent mixing on the workers' side and contribute to the richer, more randomized long-run patterns.

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 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 one faces NE_1 for $t \leq \tilde{t}$, 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 9, 10, 11). 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) over time.

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

Figure 9: G1: $w_{1t} > 2w_{2t}$

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

Figure 10: 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	$\frac{w_2}{2}, \frac{w_2}{2}$

Figure 11: G3: $w_{2t} > 2w_{1t}$

Based on the analysis for fixed wage environment in Section 2.1, it is expected that if wage condition for $G2$ can be sustained for longer periods of time, workers would be able to learn to converge to the pure NEs, leading to efficient outcome. However, since one may have to overcome a large inertia due to accumulated propensities from previous game plays, which could involve learning a different set of NE(s), therefore, even if equilibrium switching happens (e.g. from Figure 9 to 10), workers would experience time lag.

Suppose equilibrium switching happens at time $\tilde{t}$ due to evolving wages, workers will not immediately transition from learning NE_1 to NE_2.

When $t \leq \tilde{t}$, workers' choice probabilities (x_t, y_t) converge to NE_1 as t increases:

$$\lim_{t \to \infty, t \leq \tilde{t}} x_t, y_t \in \text{ Support of } NE_1 \tag{30}$$

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

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

$$x_t, y_t \in \text{ Support of } NE_1 \cup NE_2 \tag{31}$$

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

$$\lim_{t \to \infty, t \geq \bar{t}} x_t, y_t \in \text{ Support of } NE_2 \tag{32}$$

This learning process is illustrated by Example 3 in Appendix. Given perfect memory, time lag is infinite, the system never fully adapt to new conditions as past payoffs remain salient and never forgotten.

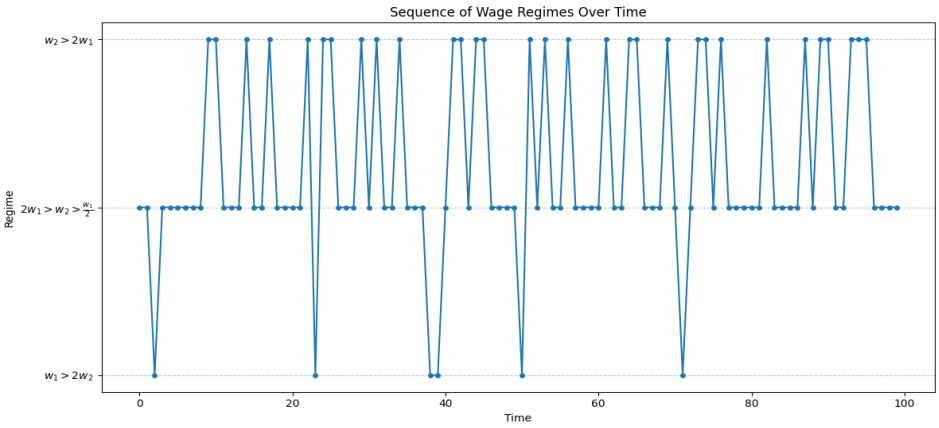


Figure shows wage transition across different regimes, characterized by wage conditions ($z_1 = z_2 = 10$, ${q_{10}}^i = {q_{20}}^i = {q_{10}}^{-i} = {q_{20}}^{-i} = 1$, $t = 100$).

Figure 12: Switching Between Wage Regimes

Over the learning trajectory, there can be multiple switching between different wage conditions. In Figure 12, I show for a simulation of 100 periods, there are jumps from one wage condition to another, and I define this to be moving across different wage regimes.

Furthermore, I demonstrate in Figure 13 the corresponding 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, workers would learn to play $(F2, F2)$, lower w_2 will be reinforced more strongly, prompting another possible regime switch to $G2$ (blue region). If workers' choice probabilities become more stabilized in $G2$, this could potentially prompt less incentive for further regime switch. However, stochasticity in workers' realized actions and equal attractiveness of $(F1, F2)$ and $(F2, F1)$ may coincidentally lead to the same firm being chosen, which could affect wages and incentivize regime switch. Therefore, there may be constant regime change.

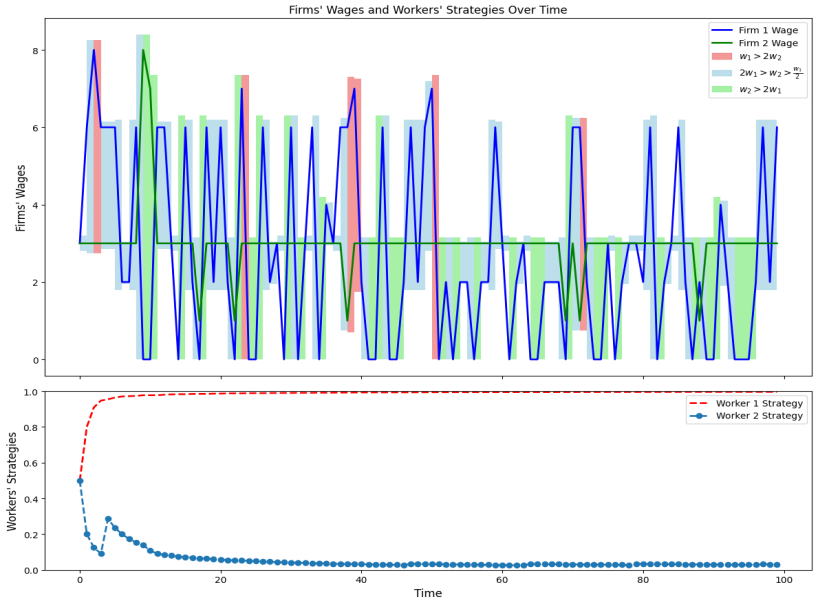


Figure 13: Changes in Wages and Workers' Choice Probabilities Over Time

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:

$$P(G_{t+1} = G_i | G_t = G_j) = \frac{n_{ij}}{n_i} \tag{33}$$

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 ($i, j \in \{1, 2, 3\}$), it also tracks the counts of staying in the same regime, $i = j$; and n_i is the counts of being in regime i. Figure 14 shows there is higher probability of switching from any regime to G_2 regime for this example. 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.

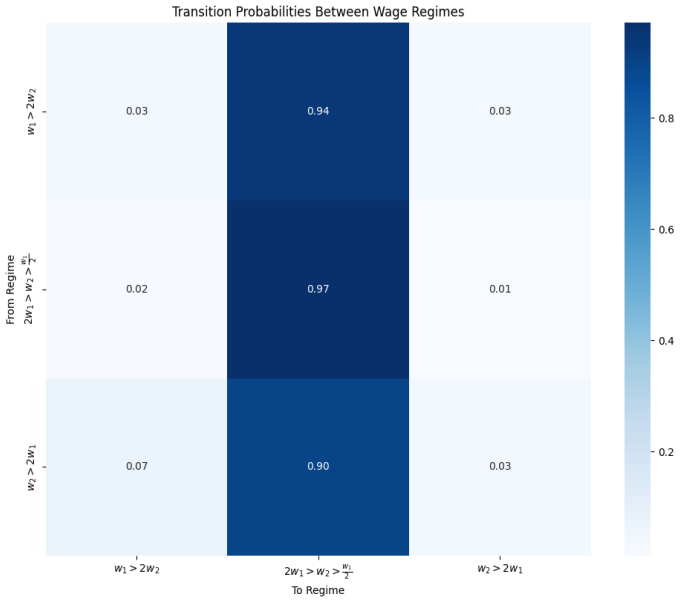


Figure 14: Transition Probility from Regime to Regime

Similar to the fixed wage environment, dynamically changing wages 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 structure shifts. This could further contributes to mismatch as workers may need to overcome inertia from previous accumulated experiences and adapt their search strategies amidst the new wage regime.

Another important question is equilibrium selection in the dynamic wage environment. Since wages are driven down to 0 in the long run, equilibrium selection is essentially path-dependent, and relies on what games were played before wages hit 0. To achieve one-to-one matching (i.e. $(F1, F2)$ or $(F2, F1)$), wage condition for $G2$ (Figure 10) needs to be maintained for substantial number of periods, 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 information by storing them in a database that can be easily maintained and retrieved, and human resources also tend to keep a record of wages offered and accepted. Therefore, I impose a memory decay factor solely on the workers' side akin to Section 2.1.1 (Equation 19).

Since workers do not have perfect recall of past experiences, it is expected there will be slower convergence towards an equilibrium. 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 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 would be more volatile.

Suppose the wage regime changes from $G1$ to $G2$ at $t = \tilde{t}$, workers would experience a time lag $(\bar{t} - \tilde{t})$ to adapt from learning NE_1 to NE_2. This transition time to "unlearn" the past and overcome previously accumulated propensities is captured by the base learning rate $(\frac{1}{\eta})$ and payoff received in $G2$:

$$(\bar{t} - \tilde{t}) \propto \frac{1}{\eta \cdot f(G2 \text{ payoff dynamics})} \tag{34}$$

Higher η, lower weight on past propensities, thus shorter time lag. (More details in Appendix A.2)

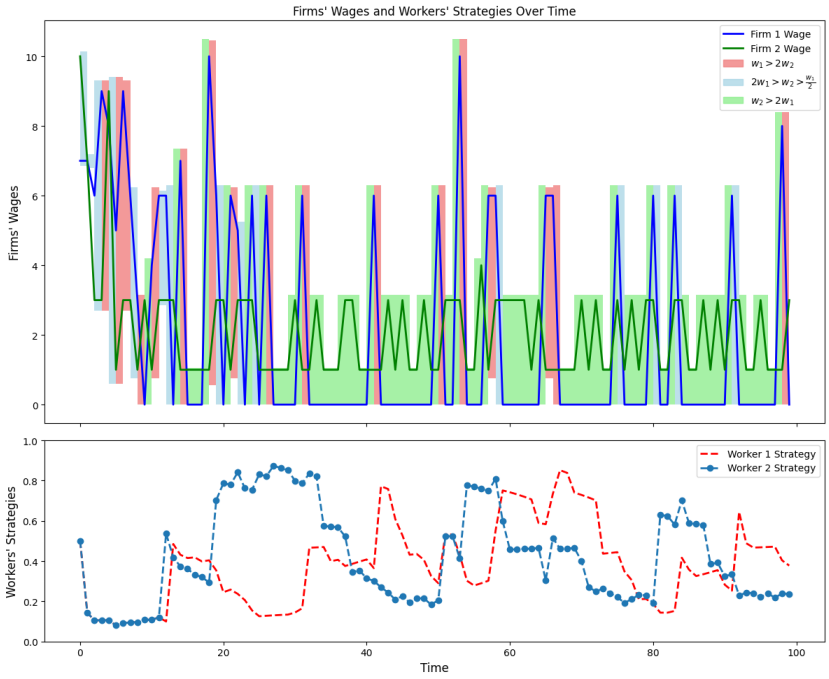


Figure 15: Changes in Wages and Workers' Choice Probabilities Over Time with $\eta = 0.8$

In Figure 15, I show a simulation of changes in wages and choice probabilities over time when there is a forgetting parameter on the workers' side (i.e. $\eta = 0.8$). Since 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, and can be perceived as being more adaptive to new market conditions. While this is beneficial in stimulating switching in job applications and inducing better coordination if workers begin with overcrowding at one of the firms. The pitfall of this is the volatility in choice probabilities. Workers could forget previous propensities and their convergence may be impeded even over an extended period of time. (More simulations in Appendix B.1)

Result Summary. Dynamic wage setting offers a more realistic environment in which firms, too, behave as adaptive learners. In this setting, long run outcomes are highly path-dependent. Rather than converging to a unique steady state, the system may settle into a low wage dynamic regime in which firms are persistently biased towards setting wages near zero, while workers could exhibit more randomized search behaviour. Furthermore, since the wage evolution process can

place workers into three distinct wage regimes, each corresponding to a different game with its own set of NE(s), workers could effectively face shifting strategic environments over time. This creates further scope for noisier or less stable strategies as workers move across regimes and learn to play different NE(s) at different parts of the wage transition path. Introducing discounting of past experiences could facilitate adaptation when the wage regime changes. However, at the same time, it may introduce a trade-off, where greater responsiveness to the current environment comes at the cost of weaker anchoring from accumulated reinforcements, potentially increasing volatility in choice probabilities as the influence of past learning fades.

3 Best Response Dynamics for "Transparent" Markets

In this section, I explore the second market structure, which resembles job search on online platforms where wages are fully revealed, thus referred to as "Transparent" markets. Workers observe firms' wage postings before choosing where to apply to. They are assumed to adopt a logit choice model and follows best-response (BR) dynamics (Fudenberg and Levine (1998); Hopkins (1999)). They consciously study the wage environment, and respond to the perceived strategy of their opponent given their experiences. This is then feedback into firms' decision problem, where they set wages knowing how workers formulate their strategies.

This model tracks two possible sources of experiences. The first constitutes of what workers observe their opponent did over time. In directed search literature, such as Wright et al. (2021), wages often have a role of directing workers, higher wage would attract higher application rate. However, when wages are the same, workers could be equally attracted to both firms, and experiences that reflect the opponent's strategy could help to determine workers' choices. Even when wages differ, this influence does not vanish. Therefore, exploring the role of experiences may shine a light on equilibrium selection in presence of multiple equilibria, and also, potentially offer an explanation for why workers hardly switch in applying for different jobs despite knowing wages beforehand.

The other source could come from workers having long-term experience that forms a static bias that is 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 are less shaped by short-term encounters. In Barbulescu and Bidwell (2013), they show 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 experiences accumulated across generations, which could affect their application choices.

Under this framework, I seek to explore how past experiences influence workers' job search behaviour given that they observe the wages, and how experience affects equilibrium selection as compared to the previous market structure.

3.1 Experience-driven Job Search with Best Response

In a similar 2×2 set-up with 2 firms and 2 workers, where workers are indexed by $i = 1, 2$, firms by $j = 1, 2$, and each firm offering one vacancy.

Timing and Set-up. Firms observe exogenous realization of productivity, $\mathbf{z} = \{z_1, z_2\}$, $\mathbf{z} \in Z$. Time is discrete, $t = \{0, 1, ..., T\}$, representing generations of workers and firms.[3] In each period:

1. Firms infer workers' belief based on observed history of actions, and they set wages, $\mathbf{w} = \{w_{1t}, w_{2t}\}$, with respect to how workers are expected to react in the current period.
2. Workers observe wages and choose their application strategies, given beliefs about opponent's choices and any bias they adopted from the same indexed worker of the past period.[4]
3. All parties observe the payoffs at the end of the period. Workers update their beliefs based on realized actions, and they drop out of the market, never to return.

Workers' side.

[3] Same as the previous set-up, I consider $T \to \infty$, which implies infinitely many generations.

[4] Same as previous set-up, worker 1 of $t = 1$ learns from worker 1 of $t = 0$, same applies for worker 2.

- **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 form beliefs about the other worker's choice probability through realized actions. Worker 2's belief about worker 1's choice probabilities is denoted as $u_t = (u_{1t}, 1 - u_{1t})$, where $u_{1t} \in (0, 1)$; and worker 1's belief about worker 2's choice probabilities is $v_t = (v_{1t}, 1 - v_{1t})$, $v_{1t} \in (0, 1)$.
- **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:

$$\pi_t^i(a_t^i, v_t) = W_t(a_t^i, F1)v_{1t} + W_t(a_t^i, F2)(1 - v_{1t}),\ \text{where } a_t^i \in A^i \tag{35}$$

$$\pi_t^{-i}(a_t^{-i}, u_t) = W_t^T(a_t^{-i}, F1)u_{1t} + W_t^T(a_t^{-i}, F2)(1 - u_{1t}),\ \text{where } a_t^{-i} \in A^{-i} \tag{36}$$

where the payoff matrices are

$$W_t = \begin{pmatrix} \frac{w_{1t}}{2} & w_{1t} \\ w_{2t} & \frac{w_{2t}}{2} \end{pmatrix}, W_t^T = \begin{pmatrix} \frac{w_{1t}}{2} & w_{2t} \\ w_{1t} & \frac{w_{2t}}{2} \end{pmatrix} \tag{37}$$

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

  For worker 1:

$$x_t = BR_x(w_{1t}, w_{2t}, v_t) = \begin{cases} x_{1t} = \frac{\exp(\alpha_1^i + \beta\pi_t^i(F1, v_t))}{\exp(\alpha_1^i + \beta\pi_t^i(F1, v_t)) + \exp(\alpha_2^i + \beta\pi_t^i(F2, v_t))} \\ x_{2t} = \frac{\exp(\alpha_2^i + \beta\pi_t^i(F2, v_t))}{\exp(\alpha_1^i + \beta\pi_t^i(F1, v_t)) + \exp(\alpha_2^i + \beta\pi_t^i(F2, v_t))} = 1 - x_{1t} \end{cases} \tag{38}$$

  For worker 2:

$$y_t = BR_y(w_{1t}, w_{2t}, u_t) = \begin{cases} y_{1t} = \frac{\exp(\alpha_1^{-i} + \beta\pi_t^{-i}(F1, u_t))}{\exp(\alpha_1^{-i} + \beta\pi_t^{-i}(F1, u_t)) + \exp(\alpha_2^{-i} + \beta\pi_t^{-i}(F2, u_t))} \\ y_{2t} = \frac{\exp(\alpha_1^{-i} + \beta\pi_t^{-i}(F2, u_t))}{\exp(\alpha_1^{-i} + \beta\pi_t^{-i}(F1, u_t)) + \exp(\alpha_2^{-i} + \beta\pi_t^{-i}(F2, u_t))} = 1 - y_{1t} \end{cases} \tag{39}$$

- **Expected Motion.** The expected changes in choice probabilities:

$$\mathbb{E}(x_{t+1}) - x_t = BR_x(w_{1t}, w_{2t}, v_t) - x_t, \mathbb{E}(y_{t+1}) - y_t = BR_y(w_{1t}, w_{2t}, u_t) - y_t \tag{40}$$

- **Updating Rule.** Following Hopkins (2002), workers' beliefs about their opponent's choice probabilities are 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.

$$u_{t+1} = \frac{(t+1)u_t + a_t^i}{t+2}, v_{t+1} = \frac{(t+1)v_t + a_t^{-i}}{t+2} \tag{41}$$

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

$$u_t = \frac{1}{t}\sum_{k=1} a_k^i, v_t = \frac{1}{t}\sum_{k=1} a_k^{-i} \tag{42}$$

  In period $t + 1$, payoff matrices will also be adjusted based on firms' inference about workers' behaviours. Therefore, workers' expected payoffs are updated given new set of payoff matrices, (W_{t+1}, W_{t+1}^T), and beliefs, (v_{t+1}, u_{t+1}).

For the workers' side, the subgame equilibrium solution(s) resembles that of quantal response equilibrium (QRE), where workers eventually form correct beliefs about opponents' strategies (McKelvey and Palfrey (1995)). Given a 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.

To illustrate, suppose workers do not possess any bias (i.e. $\alpha^i = \alpha^{-i} = 0$), then when firms are homogenous and wages equalize (i.e. $w_1 = w_2$, $w_1 > 0$, $w_2 > 0$), Figure 16 shows workers' behaviours given variations in β. Multiple equilibria could emerge when β increases.

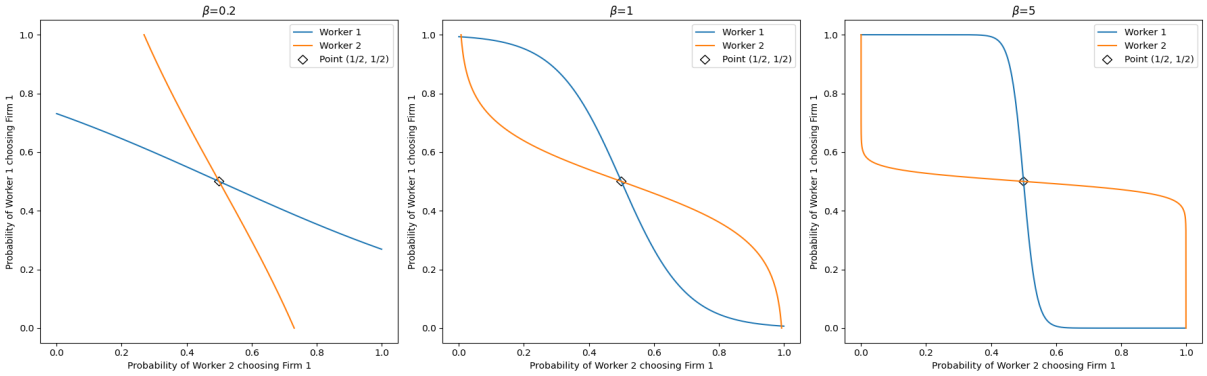


Figure 16: Workers' Response Functions to Given Set of Wages ($w_1 = w_2$)

Firms' side. Firms keeps a tally of workers' observed actions and infer workers' beliefs, they set wages in each period knowing how workers would respond. Their maximization problems:

$$\max_{w_{1t}}(1-(1-x_{1t})(1-y_{1t}))(z_1 - w_{1t}) \text{ s.t. } z_1 \geq w_{1t} \geq 0 \tag{43}$$

$$\max_{w_{2t}}(1-x_{1t}y_{1t})(z_2 - w_{2t}) \text{ s.t. } z_2 \geq w_{2t} \geq 0 \tag{44}$$

Firms' FOCs:

$$1-(1-x_{1t})(1-y_{1t}) = (z_1 - w_{1t})[(1-y_{1t})\frac{dx_{1t}}{dw_{1t}} + (1-x_{1t})\frac{dy_{1t}}{dw_{1t}}] \tag{45}$$

$$1-x_{1t}y_{1t} = -(z_2 - w_{2t})(y_{1t}\frac{dx_{1t}}{dw_{2t}} + x_{1t}\frac{dy_{1t}}{dw_{2t}}) \tag{46}$$

Workers' FOCs:

$$\frac{dx_{1t}}{dw_{1t}} = x_{1t}(1-x_{1t})(\beta\frac{1}{2}v_{1t} + \beta v_{2t}) \tag{47}$$

$$\frac{dx_{1t}}{dw_{2t}} = -x_{1t}(1-x_{1t})(\beta v_{1t} + \beta\frac{1}{2}v_{2t}) \tag{48}$$

$$\frac{dy_{1t}}{dw_{1t}} = y_{1t}(1-y_{1t})(\beta\frac{1}{2}u_{1t} + \beta u_{2t}) \tag{49}$$

$$\frac{dy_{1t}}{dw_{2t}} = -y_{1t}(1-y_{1t})(\beta u_{1t} + \beta\frac{1}{2}u_{2t}) \tag{50}$$

Combining them, the wage equations:

$$w_{1t} = \max[z_1 - \frac{1-(1-x_{1t})(1-y_{1t})}{(1-y_{1t})x_{1t}(1-x_{1t})(\beta\frac{v_{1t}}{2} + \beta(1-v_{1t})) + (1-x_{1t})y_{1t}(1-y_{1t})(\beta\frac{u_{1t}}{2} + \beta(1-u_{1t}))}, 0] \tag{51}$$

$$w_{2t} = \max[z_2 - \frac{1-x_{1t}y_{1t}}{y_{1t}x_{1t}(1-x_{1t})(\beta v_{1t} + \beta\frac{(1-v_{1t})}{2}) + x_{1t}y_{1t}(1-y_{1t})(\beta u_{1t} + \beta\frac{(1-u_{1t})}{2})}, 0] \tag{52}$$

Definition 3.1 (Sequential Learning Equilibrium)**.** *The equilibrium is a tuple* (x^*, y^*, w_1^*, w_2^*)*:*

- *(Workers' Best Response.) Workers' strategies* (x^*, y^*) *are best responses to their beliefs about their opponent* (u^*, v^*) *and firms' wages* (w_1^*, w_2^*)*.*
- *(Firms' Best Response.) Firms' wages* (w_1^*, w_2^*) *are best responses to workers' strategies* $(x^*,\ y^*)$ *and their beliefs* (u^*, v^*)*.*
- *(Consistency Condition.) Beliefs match with the actual choice probabilities,* $u^* = x^*$*,* $v^* = y^*$*.*
- *(Equilibrium Multiplicity.) Multiple equilibria could exist, corresponding to different tuples of* (x^*, y^*, w_1^*, w_2^*)*.*

In the long run, if workers and firms' behaviours converge, the system $(x_t, y_t, w_{1t}, w_{2t})$ converges to the equilibrium (x^*, y^*, w_1^*, w_2^*) described in Definition 3.1. Since multiple equilibria may exist, the system can converge to one of them, and the equilibrium selected could depend on the initial conditions and the learning path.

Based on equation (42), as $t \to \infty$, workers' beliefs (u_t, v_t) converge to long run average of observed actions:

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

Given beliefs stabilize to (u^*, v^*), from equations (38) and (39), workers' strategies (x^*, y^*):

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

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

where π^i and π^{-i} are expected payoffs depending on wages (W^*), and beliefs (u^*, v^*).

Given u^*, v^*, x^*, y^*, wages (w_1^*, w_2^*) are computed based on equations (51) and (52).

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

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

and payoff matrix can be constructed:

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

By Definition 3.1, belief consistency requires $u^* = x^*$, $v^* = y^*$. The above yield the following system of equations, which solves for $(x_1^*, y_1^*, w_1^*, w_2^*)$:

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

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

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

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

They collectively define the equilibrium, and multiple solutions may exist given their non-linearity.

Algorithm 1 Firms' Wage-setting using Grid Search Approach

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 10000 time periods **do**
    Loop for each time period
    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 (38) and (39) 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' expected payoffs using equations (43) and(44).
        Find the wage pair that lead to highest expected payoff.
      end for coarse search
      for refined search **do**
        Take the 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's 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
    for all workers **do**
      For (w_{1t}, w_{2t}), compute x_t and y_t given u_t and v_t using equations (38) and (39).
      Generate a choice of action from $(F1, F2)$ for each worker based on x_t and y_t.
    end for workers
    for reward generation and updating **do**
      Given realized workers' choices, (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}, v_{t+1}), are updated based on equation (41) 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.

In Algorithm 1, I formalize firms' wage-setting behaviour using a grid search approach, where they conduct a coarse search and then a refined search to nail down wages in each period based on potential worker reactions.

In Figure 17, I show the convergence pattern of wages (top panel), choice probabilities (middle panel) and beliefs (bottom panel) given different βs. When $\beta = 0.2$, there is a unique equilibrium, and there is convergence to it. In presence of multiple equilibria, when $\beta = 1.0$, workers converge to a strategy that is close to random search; and when $\beta = 5.0$, workers' converge to apply with high probability to different firms. Since this grid search wage-setting method emphasizes on maximizing profits, if wages differ across equilibria, firms would converge to the equilibrium with lower wages. The simulation results display some resemblance to the theoretical findings in Lu (2024), where at relatively high sensitivity to expected payoffs, wages for asymmetric equilibria tend to be lower than that of symmetric equilibrium; and vice versa for slightly lower sensitivity to expected payoffs. As a result, when firms select equilibrium with lower wages, workers converge to one of the asymmetric equilibria for $\beta = 5.0$, and to the symmetric equilibrium for $\beta = 1.0$. Whilst Figure 17c shows convergence to a single asymmetric equilibrium, 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.

An important distinction between the equilibrium achieved 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 in terms of how their current actions might influence the other worker's behaviour. 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 beliefs. Therefore, this myopic learning dynamics may converge only to a subset of QRE that are identified in forward-looking scenarios.

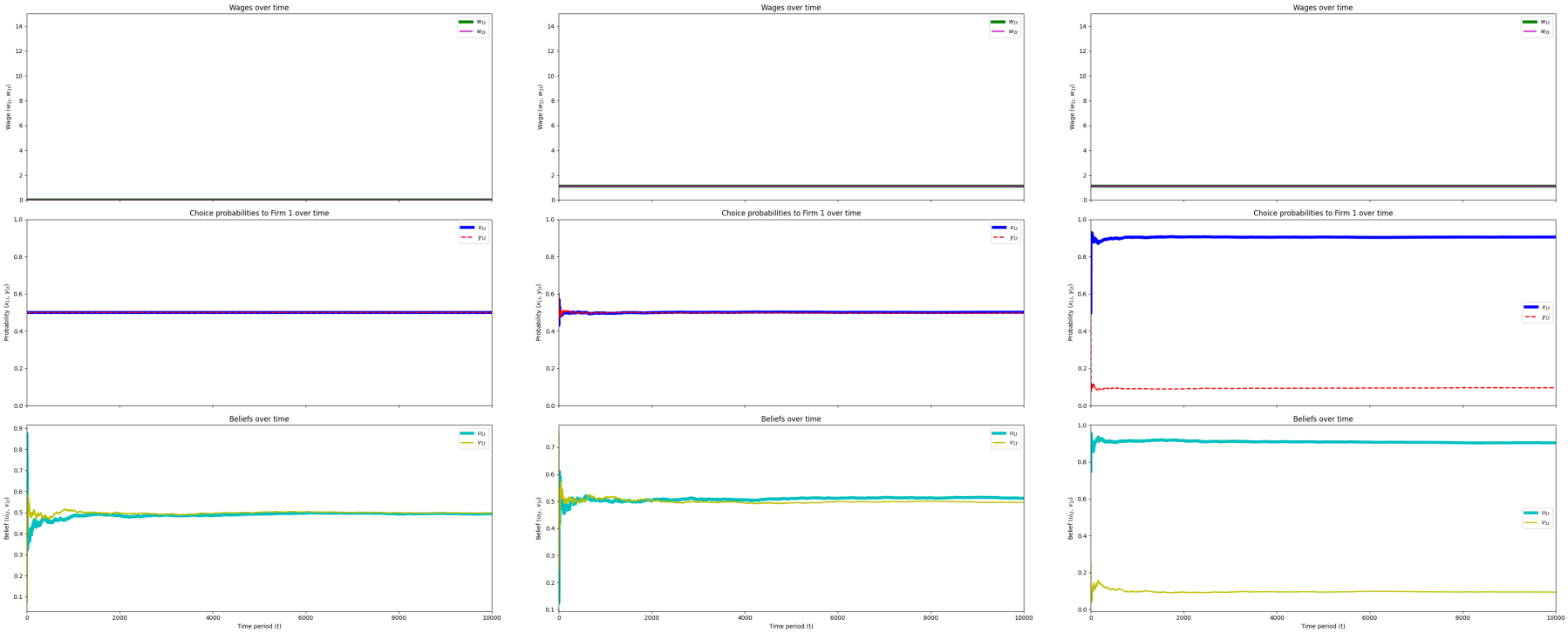
In order to gain a clearer picture of the different parts of the system, I first fix workers' beliefs to investigate how would choice probabilities and wages behave; I then fix wages to explore how might beliefs and choice probabilities evolve. In Figure 18, I show that if beliefs are fixed, workers converge to a unique set of choice probabilities. In presence of multiple equilibria, the point at which beliefs are fixed is important in determining if one converge to the asymmetric or symmetric equilibrium. If workers' beliefs are fixed at believing their opponent applies more to a different firm than them, then it is more likely for workers to converge to applying asymmetrically. Furthermore, the figure shows that if workers' behaviours stabilize, wages would also stabilize. Subsequently, in Figure 19, I fix both of the wages or one of the wages. In both cases, workers' beliefs becomes more asymmetric and they converge to applying more to different firms.

Based on the long run convergence behaviour and simulations (Figure 17), when β is low and only symmetric equilibrium exists, there is convergence to the unique equilibrium; when β is sufficiently high and multiple equilibria exist, there will be convergence to the asymmetric equilibria. This is more efficient than the symmetric case due to higher chances of one-to-one matching. However, which of the two asymmetric equilibria the workers coordinate on could depend heavily on beliefs.

Apart from grid search approach, I have also experimented other wage-setting algorithms, shown in Appendix B.2. The grid search approach could implement a cleaner selection of one of the asymmetric equilibria, but could require high computational power as firms may need to evaluate the expected payoffs for a wide range of wage combinations.

Definition 3.2 (Dynamic System)**.** *The dynamic system is defined by:*

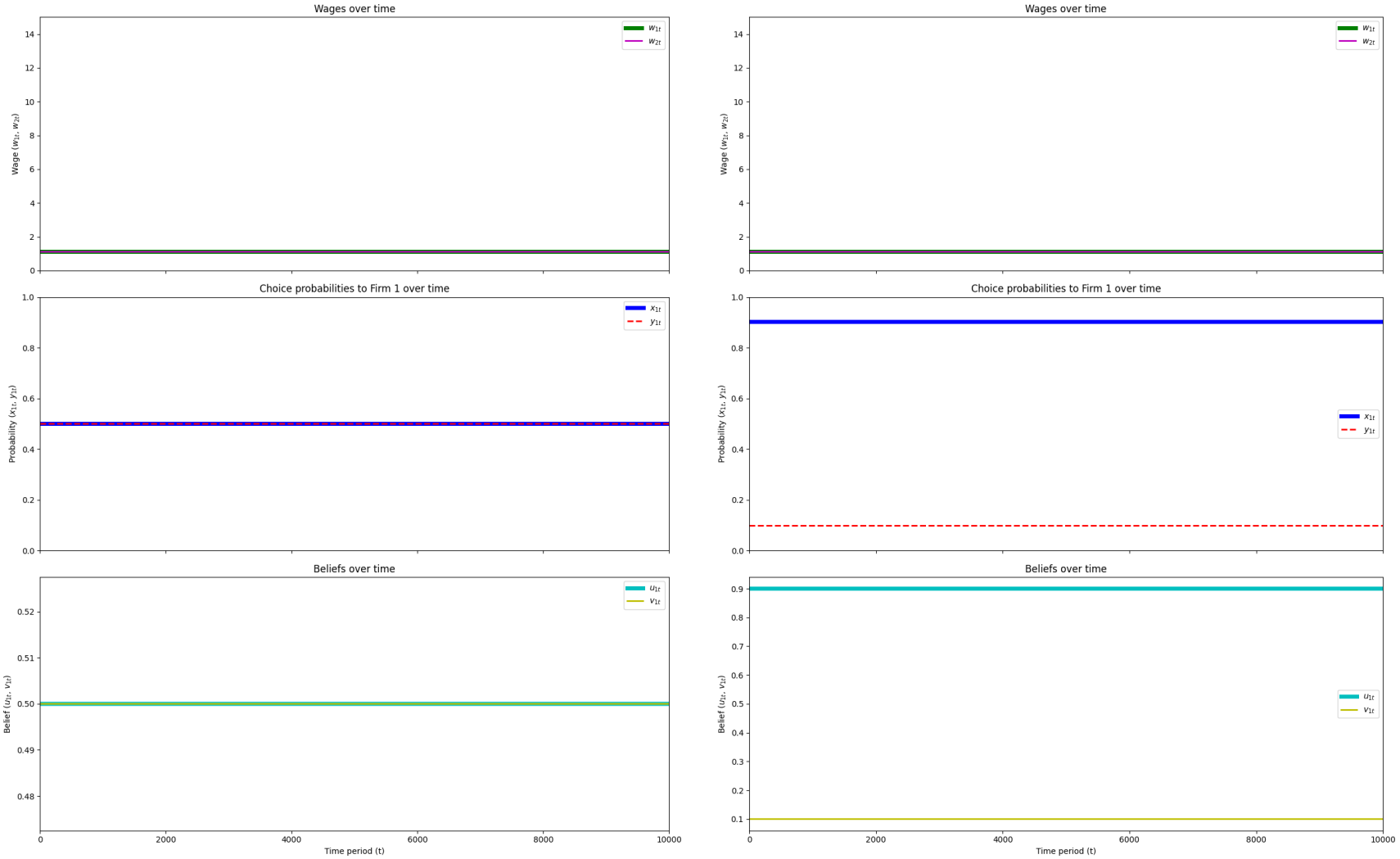
$$\dot{u}_t = x_t(v_t; w_{1t}, w_{2t}) - u_t \tag{63}$$



(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 $z_1 = z_2 = 10$, $t = 10000$, $\alpha_1^i = \alpha_1^{-i} = 0$, $u_{10} = 0.5$, $v_{10} = 0.5$, no smoothing.

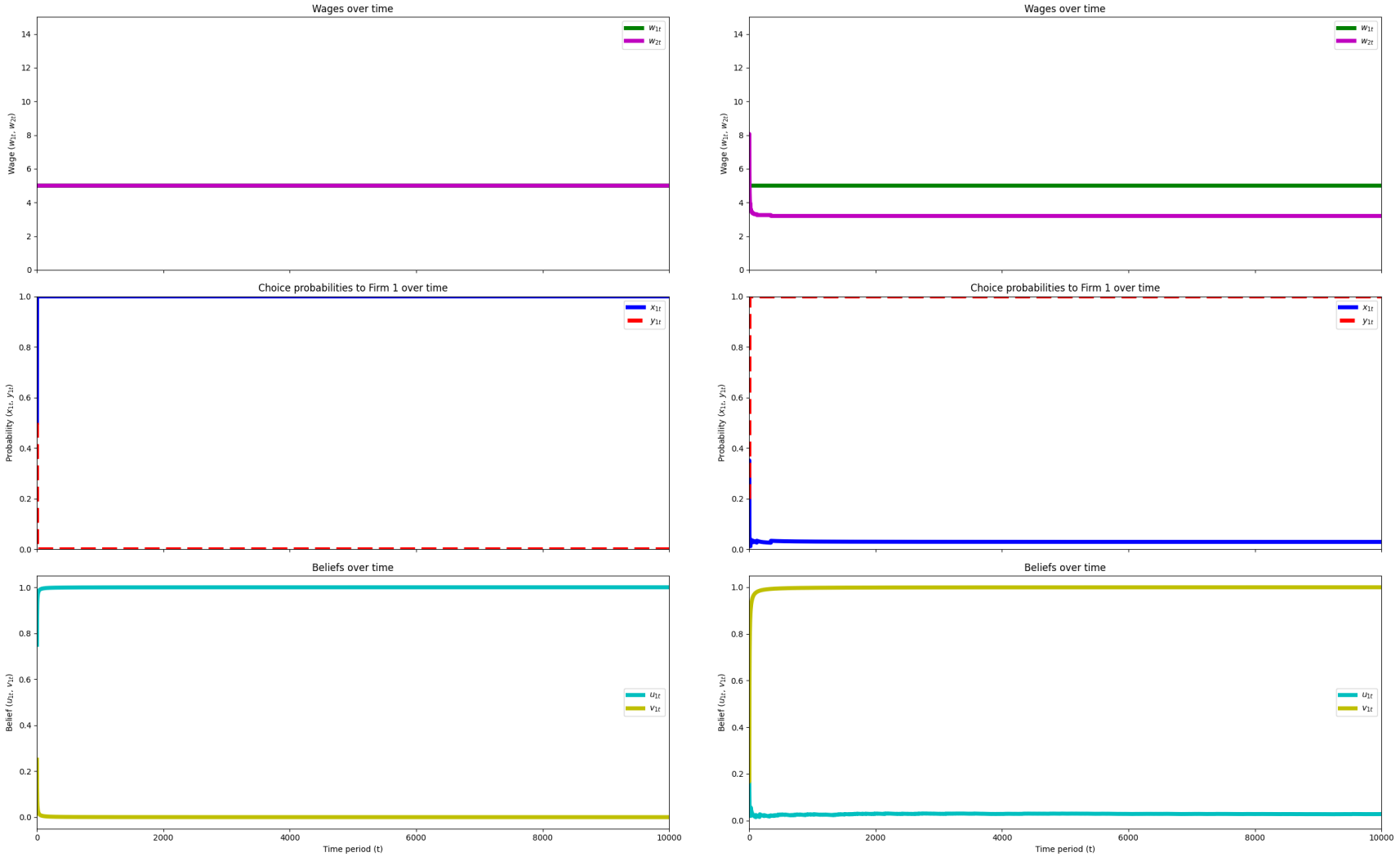
Figure 17: Changes in Wages, Workers' Choice Probabilities and Beliefs Over Time (Algorithm 1)



(a) $(u_1, v_1) = (0.5, 0.5)$ (b) $(u_1, v_1) = (0.9, 0.1)$

Figure shows wages, choice probabilities, fixing beliefs at $(u_1, v_1) = (0.5, 0.5)$ and $(0.9, 0.1)$, given $\beta = 5.0$, $z_1 = z_2 = 10$, $t = 10000$, $\alpha_1^i = \alpha_1^{-i} = 0$, and no smoothing.

Figure 18: Changes in Wages, Workers' Choice Probabilities Over Time for Fixed Beliefs (Algorithm 1)



(a) $w_1 = w_2 = 5$ (b) $w_1 = 5$

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

Figure 19: Changes in Workers' Choice Probabilities, Beliefs Over Time for Fixed Wages (Algorithm 1)

$$\dot{v}_t = y_t(u_t; w_{1t}, w_{2t}) - v_t \tag{64}$$

where (x_t, y_t) are workers' choice probabilities, (w_{1t}, w_{2t}) are firms' wages, and they are functions of workers' beliefs (u_t, v_t). In equilibrium, $\dot{u}_t = 0$, $\dot{v}_t = 0$, which implies $u^* = x^*$, $v^* = y^*$.

Proposition 3 (Local Stability of Asymmetric Equilibria under Sequential Learning)**.** *Let (x^*, y^*, w_1^*, w_2^*) be an asymmetric equilibrium of the dynamic system (Definition 3.2). The Jacobian evaluated at the equilibrium:*

$$J = J_{base} + W_1 + W_2 \tag{65}$$

where J_{base} captures direct effects of u_t and v_t on x_t and y_t, and W_1, W_2 capture indirect effects of u_t and v_t on x_t and y_t through w_{1t} and w_{2t}. In the limiting case, $(x_1^*, y_1^* \to (1,0),(0,1))$, such equilibria are locally asymptotically stable.

Proof. To analyse stability, I restate the following equations for clearer view.

Beliefs:

$$u_{t+1} = \frac{(t+1)u_t + a_t^i}{t+2}, v_{t+1} = \frac{(t+1)v_t + a_t^{-i}}{t+2} \tag{66}$$

Workers' choice probabilities:

$$x_{1t} = \frac{1}{1+\exp(-[(\alpha_1^i - \alpha_2^i) + \beta(w_{1t} - \frac{w_{2t}}{2} - (\frac{w_{1t}}{2} + \frac{w_2}{2})v_{1t}])} \tag{67}$$

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

Wages:

$$w_{1t} = \max[z_1 - \frac{1-(1-x_{1t})(1-y_{1t})}{(1-y_{1t})x_{1t}(1-x_{1t})(\beta\frac{v_{1t}}{2} + \beta(1-v_{1t})) + (1-x_{1t})y_{1t}(1-y_{1t})(\beta\frac{u_{1t}}{2} + \beta(1-u_{1t}))}, 0] \tag{69}$$

$$w_{2t} = \max[z_2 - \frac{1-x_{1t}y_{1t}}{y_{1t}x_{1t}(1-x_{1t})(\beta v_{1t} + \beta\frac{(1-v_{1t})}{2}) + x_{1t}y_{1t}(1-y_{1t})(\beta u_{1t} + \beta\frac{(1-u_{1t})}{2})}, 0] \tag{70}$$

For stability analysis, by Definition 3.2, where beliefs are the only dynamical values, the choice probabilities and wages are instantaneous functions of u_t and v_t. Based on equations (63) and (64), find the Jacobian:

$$J = \begin{pmatrix} \frac{\partial \dot{u}_t}{\partial u_t} & \frac{\partial \dot{u}_t}{\partial v_t} \\ \frac{\partial \dot{v}_t}{\partial u_t} & \frac{\partial \dot{v}_t}{\partial v_t} \end{pmatrix} \tag{71}$$

where

$$\frac{\partial \dot{u}_t}{\partial u_t} = \frac{\partial x_t}{\partial w_{1t}}\frac{\partial w_{1t}}{\partial u_t} + \frac{\partial x_t}{\partial w_{2t}}\frac{\partial w_{2t}}{\partial u_t} - 1 \tag{72}$$

$$\frac{\partial \dot{u}_t}{\partial v_t} = \frac{\partial x_t}{\partial v_t} + \frac{\partial x_t}{\partial w_{1t}}\frac{\partial w_{1t}}{\partial v_t} + \frac{\partial x_t}{\partial w_{2t}}\frac{\partial w_{2t}}{\partial v_t} \tag{73}$$

$$\frac{\partial \dot{v}_t}{\partial u_t} = \frac{\partial y_t}{\partial u_t} + \frac{\partial y_t}{\partial w_{1t}}\frac{\partial w_{1t}}{\partial u_t} + \frac{\partial y_t}{\partial w_{2t}}\frac{\partial w_{2t}}{\partial u_t} \tag{74}$$

$$\frac{\partial \dot{v}_t}{\partial v_t} = \frac{\partial y_t}{\partial w_{1t}}\frac{\partial w_{1t}}{\partial v_t} + \frac{\partial y_t}{\partial w_{2t}}\frac{\partial w_{2t}}{\partial v_t} - 1 \tag{75}$$

The full form and the corresponding condensed version can be written as:

$$J = \underbrace{\begin{pmatrix} 0 & \frac{\partial x_t}{\partial v_t} \\ \frac{\partial y_t}{\partial u_t} & 0 \end{pmatrix} - I}_{J_{\text{base}}} + \underbrace{\begin{pmatrix} \frac{\partial x_t}{\partial w_{1t}}\frac{\partial w_{1t}}{\partial u_t} & \frac{\partial x_t}{\partial w_{1t}}\frac{\partial w_{1t}}{\partial v_t} \\ \frac{\partial y_t}{\partial w_{1t}}\frac{\partial w_{1t}}{\partial u_t} & \frac{\partial y_t}{\partial w_{1t}}\frac{\partial w_{1t}}{\partial v_t} \end{pmatrix}}_{W_1} + \underbrace{\begin{pmatrix} \frac{\partial x_t}{\partial w_{2t}}\frac{\partial w_{2t}}{\partial u_t} & \frac{\partial x_t}{\partial w_{2t}}\frac{\partial w_{2t}}{\partial v_t} \\ \frac{\partial y_t}{\partial w_{2t}}\frac{\partial w_{2t}}{\partial u_t} & \frac{\partial y_t}{\partial w_{2t}}\frac{\partial w_{2t}}{\partial v_t} \end{pmatrix}}_{W_2} \tag{76}$$

The Jacobian determinant:

$$\det\begin{pmatrix} -1-\lambda+\underbrace{\frac{\partial x_t}{\partial w_{1t}}\frac{\partial w_{1t}}{\partial u_t}+\frac{\partial x_t}{\partial w_{2t}}\frac{\partial w_{2t}}{\partial u_t}}_{a} & \underbrace{\frac{\partial x_t}{\partial v_t}+\frac{\partial x_t}{\partial w_{1t}}\frac{\partial w_{1t}}{\partial v_t}+\frac{\partial x_t}{\partial w_{2t}}\frac{\partial w_{2t}}{\partial v_t}}_{c} \\ \underbrace{\frac{\partial y_t}{\partial u_t}+\frac{\partial y_t}{\partial w_{1t}}\frac{\partial w_{1t}}{\partial u_t}+\frac{\partial y_t}{\partial w_{2t}}\frac{\partial w_{2t}}{\partial u_t}}_{d} & -1-\lambda+\underbrace{\frac{\partial y_t}{\partial w_{1t}}\frac{\partial w_{1t}}{\partial v_t}+\frac{\partial y_t}{\partial w_{2t}}\frac{\partial w_{2t}}{\partial v_t}}_{b} \end{pmatrix}=0 \tag{77}$$

$$\lambda=-1-\frac{(a+b)\pm\sqrt{(a+b)^2+4(cd-ab)}}{2} \tag{78}$$

Breaking down based on workers' equations (67), (68) and firms' equations (69), (70):

$$\frac{\partial x_{1t}}{\partial w_{1t}}=\beta x_{1t}(1-x_{1t})(1-\frac{v_{1t}}{2}) \tag{79}$$

$$\frac{\partial x_{1t}}{\partial w_{2t}}=-\beta x_{1t}(1-x_{1t})(\frac{1}{2}+\frac{v_{1t}}{2}) \tag{80}$$

$$\frac{\partial y_{1t}}{\partial w_{1t}}=\beta y_{1t}(1-y_{1t})(1-\frac{u_{1t}}{2}) \tag{81}$$

$$\frac{\partial y_{1t}}{\partial w_{2t}}=-\beta y_{1t}(1-y_{1t})(\frac{1}{2}+\frac{u_{1t}}{2}) \tag{82}$$

$$\frac{\partial x_{1t}}{\partial v_{1t}}=-\beta x_{1t}(1-x_{1t})(\frac{w_{1t}}{2}+\frac{w_{2t}}{2}) \tag{83}$$

$$\frac{\partial y_{1t}}{\partial u_{1t}}=-\beta y_{1t}(1-y_{1t})(\frac{w_{1t}}{2}+\frac{w_{2t}}{2}) \tag{84}$$

$$\frac{\partial w_{1t}}{\partial u_{1t}}=-\frac{1}{2}\beta(1-x_{1t})y_{1t}(1-y_{1t})\frac{1-(1-x_{1t})(1-y_{1t})}{[(1-y_{1t})x_{1t}(1-x_{1t})\beta(1-\frac{v_{1t}}{2})+(1-x_{1t})y_{1t}(1-y_{1t})\beta(1-\frac{u_{1t}}{2})]^2} \tag{85}$$

$$\frac{\partial w_{1t}}{\partial v_{1t}}=-\frac{1}{2}\beta(1-y_{1t})x_{1t}(1-x_{1t})\frac{1-(1-x_{1t})(1-y_{1t})}{[(1-y_{1t})x_{1t}(1-x_{1t})\beta(1-\frac{v_{1t}}{2})+(1-x_{1t})y_{1t}(1-y_{1t})\beta(1-\frac{u_{1t}}{2})]^2} \tag{86}$$

$$\frac{\partial w_{2t}}{\partial u_{1t}}=\frac{1}{2}\beta x_{1t}y_{1t}(1-y_{1t})\frac{1-x_{1t}y_{1t}}{[y_{1t}x_{1t}(1-x_{1t})\beta\frac{(1+v_{1t})}{2}+x_{1t}y_{1t}(1-y_{1t})\beta\frac{(1+u_{1t})}{2}]^2} \tag{87}$$

$$\frac{\partial w_{2t}}{\partial v_{1t}}=\frac{1}{2}\beta y_{1t}x_{1t}(1-x_{1t})\frac{1-x_{1t}y_{1t}}{[y_{1t}x_{1t}(1-x_{1t})\beta\frac{(1+v_{1t})}{2}+x_{1t}y_{1t}(1-y_{1t})\beta\frac{(1+u_{1t})}{2}]^2} \tag{88}$$

If $W_1+W_2\approx 0$, $J\approx J_{\text{base}}$, the eigenvalues for J_{base}:

$$\lambda=-1\pm\sqrt{\beta^2x_{1t}(1-x_{1t})y_{1t}(1-y_{1t})(\frac{w_{1t}}{2}+\frac{w_{2t}}{2})^2} \tag{89}$$

By equations (67) and (68), higher β corresponds to more extreme x_{1t} and y_{1t} (i.e. $(x_{1t},y_{1t})\to(1,0)$ or $(0,1)$). In this limiting case, $x_{1t}(1-x_{1t})y_{1t}(1-y_{1t})\to 0$. Hence, $\beta(\frac{w_1^*}{2}+\frac{w_2^*}{2})\sqrt{x_1^*(1-x_1^*)y_1^*(1-y_1^*)}<1$, eigenvalues are negative. The equilibria are locally asymptotically stable. For smaller β, $\beta(\frac{w_1^*}{2}+\frac{w_2^*}{2})\sqrt{x_1^*(1-x_1^*)y_1^*(1-y_1^*)}<1$ may still be satisfied due to direct dependence on β.

$W_1+W_2\approx 0$ could hold at asymmetric equilibria when wage effect vanishes. Based on equations (79) to (82), as $x_{1t},y_{1t}\to(1,0)$ or $(0,1)$, $\frac{\partial x_{1t}}{\partial w_{1t}}$, $\frac{\partial x_{1t}}{\partial w_{2t}}$, $\frac{\partial y_{1t}}{\partial w_{1t}}$, $\frac{\partial y_{1t}}{\partial w_{2t}}\approx 0$. If the terms $\frac{\partial w_{1t}}{\partial u_{1t}}$, $\frac{\partial w_{1t}}{\partial v_{1t}}$, $\frac{\partial w_{2t}}{\partial u_{1t}}$, $\frac{\partial w_{2t}}{\partial v_{1t}}$ do not grow faster than the derivatives of x_t, y_t with respect to wages, then $W_1+W_2\approx 0$ and $J\approx J_{\text{base}}$; otherwise, the $W_1+W_2\approx 0$ may not hold.

Deriving the full expression of W_1 and W_2, for each element in W_1:

$$\frac{\partial x_t}{\partial w_{1t}}\frac{\partial w_{1t}}{\partial u_t} = -\frac{1}{2}x_{1t}(1-\frac{v_{1t}}{2})y_{1t}\frac{1-(1-x_{1t})(1-y_{1t})}{(1-y_{1t})[x_{1t}(1-\frac{v_{1t}}{2})+y_{1t}(1-\frac{u_{1t}}{2})]^2} \tag{90}$$

$$\frac{\partial x_t}{\partial w_{1t}}\frac{\partial w_{1t}}{\partial v_t} = -\frac{1}{2}x_{1t}(1-\frac{v_{1t}}{2})x_{1t}\frac{1-(1-x_{1t})(1-y_{1t})}{(1-y_{1t})[x_{1t}(1-\frac{v_{1t}}{2})+y_{1t}(1-\frac{u_{1t}}{2})]^2} \tag{91}$$

$$\frac{\partial y_t}{\partial w_{1t}}\frac{\partial w_{1t}}{\partial u_t} = -\frac{1}{2}y_{1t}(1-\frac{u_{1t}}{2})y_{1t}\frac{1-(1-x_{1t})(1-y_{1t})}{(1-x_{1t})[x_{1t}(1-\frac{v_{1t}}{2})+y_{1t}(1-\frac{u_{1t}}{2})]^2} \tag{92}$$

$$\frac{\partial y_t}{\partial w_{1t}}\frac{\partial w_{1t}}{\partial v_t} = -\frac{1}{2}y_{1t}(1-\frac{u_{1t}}{2})x_{1t}\frac{1-(1-x_{1t})(1-y_{1t})}{(1-x_{1t})[x_{1t}(1-\frac{v_{1t}}{2})+y_{1t}(1-\frac{u_{1t}}{2})]^2} \tag{93}$$

$$W_1 = \begin{pmatrix} 0 & -\frac{1}{2} \\ 0 & 0 \end{pmatrix} \text{ for } x_{1t}=1, y_{1t}=0; W_1 = \begin{pmatrix} 0 & 0 \\ -\frac{1}{2} & 0 \end{pmatrix} \text{ for } x_{1t}=0, y_{1t}=1 \tag{94}$$

The same can be done for W_2:

$$\frac{\partial x_t}{\partial w_{2t}}\frac{\partial w_{2t}}{\partial u_t} = -\frac{1}{2}(1-x_{1t})(\frac{1}{2}+\frac{v_{1t}}{2})(1-y_{1t})\frac{1-x_{1t}y_{1t}}{y_{1t}[(1-x_{1t})\frac{(1+v_{1t})}{2}+(1-y_{1t})\frac{(1+u_{1t})}{2}]^2} \tag{95}$$

$$\frac{\partial x_t}{\partial w_{2t}}\frac{\partial w_{2t}}{\partial v_t} = -\frac{1}{2}(1-x_{1t})(\frac{1}{2}+\frac{v_{1t}}{2})(1-x_{1t})\frac{1-x_{1t}y_{1t}}{y_{1t}[(1-x_{1t})\frac{(1+v_{1t})}{2}+(1-y_{1t})\frac{(1+u_{1t})}{2}]^2} \tag{96}$$

$$\frac{\partial y_t}{\partial w_{2t}}\frac{\partial w_{2t}}{\partial u_t} = -\frac{1}{2}(1-y_{1t})(\frac{1}{2}+\frac{u_{1t}}{2})(1-y_{1t})\frac{1-x_{1t}y_{1t}}{x_{1t}[(1-x_{1t})\frac{(1+v_{1t})}{2}+(1-y_{1t})\frac{(1+u_{1t})}{2}]^2} \tag{97}$$

$$\frac{\partial y_t}{\partial w_{2t}}\frac{\partial w_{2t}}{\partial v_t} = -\frac{1}{2}(1-y_{1t})(\frac{1}{2}+\frac{u_{1t}}{2})(1-x_{1t})\frac{1-x_{1t}y_{1t}}{x_{1t}[(1-x_{1t})\frac{(1+v_{1t})}{2}+(1-y_{1t})\frac{(1+u_{1t})}{2}]^2} \tag{98}$$

$$W_2 = \begin{pmatrix} 0 & 0 \\ -\frac{1}{2} & 0 \end{pmatrix} \text{ for } x_{1t}=1, y_{1t}=0; W_2 = \begin{pmatrix} 0 & -\frac{1}{2} \\ 0 & 0 \end{pmatrix} \text{ for } x_{1t}=0, y_{1t}=1 \tag{99}$$

The full expression shows some elements of W_1+W_2 can depart from 0. Based on equations (94) and (99), in the limit of $x_{1t}, y_{1t} \to (1,0)$ or $(0,1)$, the Jacobian determinant (77) ensures that eigenvalues are negative (i.e. $\lambda_1 = -\frac{3}{2}$, $\lambda_2 = -\frac{1}{2}$), and the asymmetric equilibria are locally asymptotically stable.

Although $W_1+W_2 \not\approx 0$, but if W_1+W_2 is small, J_{base} could determine the stability. Equations (90) to (98) show that W_1+W_2 are independent of β. Increasing β only affects J_{base}, which grows relative to W_1+W_2, thus stability can be predominantly determined by J_{base}. As $\beta \to \infty$, x_{1t} and y_{1t} tend to limiting case, then evaluating Jacobian shows asymmetric equilibria to be locally asymptotically stable. □

Corollary 3.0.1 (Stability with Approximately Constant Wages)**.** *Suppose w_1, w_2 are approximately constant ($\frac{\partial w_{1t}}{\partial u_{1t}}$, $\frac{\partial w_{1t}}{\partial v_{1t}}$, $\frac{\partial w_{2t}}{\partial v_{1t}}$, $\frac{\partial w_{2t}}{\partial v_{1t}} \approx 0$), then $W_1+W_2 \approx 0$. By Proposition 3, J_{base} determines the stability. The eigenvalues are:*

$$\lambda = -1 \pm \sqrt{\frac{\partial x_t}{\partial v_t}\frac{\partial y_t}{\partial u_t}} \tag{100}$$

If $\sqrt{\frac{\partial x_t}{\partial v_t}\frac{\partial y_t}{\partial u_t}} < 1$, all eigenvalues have negative real parts, the system is locally asymptotically stable. In the limit of $x_{1t}, y_{1t} \to (1,0)$ or $(0,1)$, this condition holds.

Corollary 3.0.2 (High Sensitivity to Wages)**.** *As β increases, workers become more sensitive to payoffs. J_{base} grows with β, while W_1+W_2 is independent of β. By Proposition 3, as $\beta \to \infty$, $x_{1t}, y_{1t} \to (1,0)$ or $(0,1)$, the system is locally asymptotically stable.*

These corollaries follow immediately from proof of Proposition 3.

As for local stability of symmetric equilibrium, if $W_1 + W_2 \approx 0$, then based on equation (89), symmetric equilibrium is a saddle point when β is large (one positive and one negative eigenvalue) and locally asymptotically stable if β is small (both eigenvalues are negative). However, $W_1 + W_2 \approx 0$ may not hold near the equilibrium point as wage effect depart from 0 (i.e. $\frac{\partial x_{1t}}{\partial w_{1t}}$, $\frac{\partial x_{1t}}{\partial w_{2t}}$, $\frac{\partial y_{1t}}{\partial w_{1t}}$, $\frac{\partial y_{1t}}{\partial w_{2t}} \not\approx 0$), but if wages are approximately constant, then this analysis would ensue.

If W_1+W_2 is non-negligible, by equation (78), only if $\frac{(a+b)\pm\sqrt{(a+b)^2+4(cd-ab)}}{2} < 1$, then equilibrium is locally asymptotically stable. Since J_{base} is dependent on β and W_1+W_2 is not, as β increases, stability can be predominantly determined by J_{base}. As $\beta \to \infty$, based on equation (89), the equilibrium is a saddle point and unstable. On the other hand, for $\beta \to 0$, there could be a unique symmetric equilibrium, which can be locally asymptotically stable. As a result, when workers are sensitive to wages, symmetric equilibrium is more susceptible to instability.

3.2 Anchoring Bias from Long-term Experiences

The next step is to consider non-zero α^i and α^{-i}. Based on equations (38) and (39):

$$x_{1t} = \frac{1}{1+\exp(-[(\alpha_1^i - \alpha_2^i) + \beta(\pi_t^i(F1, v_t) - \pi_t^i(F2, v_t))])} \tag{101}$$

$$y_{1t} = \frac{1}{1+\exp(-[(\alpha_1^{-i} - \alpha_2^{-i}) + \beta(\pi_t^{-i}(F1, u_t) - \pi_t^{-i}(F2, u_t))]} \tag{102}$$

These imply that relative bias matters more than the absolute bias, and introducing some bias towards one firm over the other could affect choice probabilities. (see Appendix B.3)

From equation (101) and (102), let $\Delta\pi^i = \pi_t^i(F1, v_t) - \pi_t^i(F2, v_t)$, $\Delta\pi^{-i} = \pi_t^{-i}(F1, v_t) - \pi_t^{-i}(F2, v_t)$; $\Delta\alpha^i = \alpha_1^i - \alpha_2^i$, $\Delta\alpha^{-i} = \alpha_1^{-i} - \alpha_2^{-i}$. The magnitude of relative bias towards a firm ($\Delta\alpha^i$) as compared to the expected payoff difference between the two options ($\Delta\pi^i$) could affect equilibrium multiplicity and selection.

If $|\Delta\alpha^i| > \beta|\Delta\pi^i|$, experience bias is strong. This implies that relative bias would have a stronger influence on x_{1t} and y_{1t}, and beliefs about opponents would have less impact, there can be convergence to a unique equilibrium. For instance, assuming $\beta|\Delta\pi^i| \to 0$,

$$x_{1t} \approx \frac{1}{1+\exp(-(\alpha_1^i - \alpha_2^i))}, y_{1t} \approx \frac{1}{1+\exp(-(\alpha_1^{-i} - \alpha_2^{-i}))} \tag{103}$$

x_{1t} and y_{1t} converge to constant values, and by equations (51) and (52), w_{1t} and w_{2t} also converge.

If $|\Delta\alpha^i| < \beta|\Delta\pi^i|$, experience bias is weak. This implies that relative expected payoffs would have a larger influence on x_{1t} and y_{1t} than relative bias. Multiple equilibria could exist and there is convergence to equilibrium defined by Definition 3.1, but selection would depend on belief updating. For instance, assuming $|\Delta\alpha^i| \to 0$,

$$x_{1t} \approx \frac{1}{1+\exp(-[\beta(\pi_t^i(F1, v_t) - \pi_t^i(F2, v_t))])}, y_{1t} \approx \frac{1}{1+\exp(-[\beta(\pi_t^{-i}(F1, u_t) - \pi_t^{-i}(F2, u_t))]} \tag{104}$$

In equilibrium, the selection depends on u^*, v^*.

Figure 20 shows simulations for workers having experience bias towards firm 1 relative to firm 2. There is clear convergence to one set of workers' strategies for both βs. At $\beta = 0.2$, experience bias is relatively strong, workers' strategies are heavily influenced by it, therefore, they both

apply with higher probability to firm 1. The probabilities can be computed with equation (103). At $\beta = 5.0$, experience bias is relatively weak, and there are multiple equilibria. Convergence to one of the asymmetric equilibria is observed as beliefs stabilize.

From equations (67) and (68),

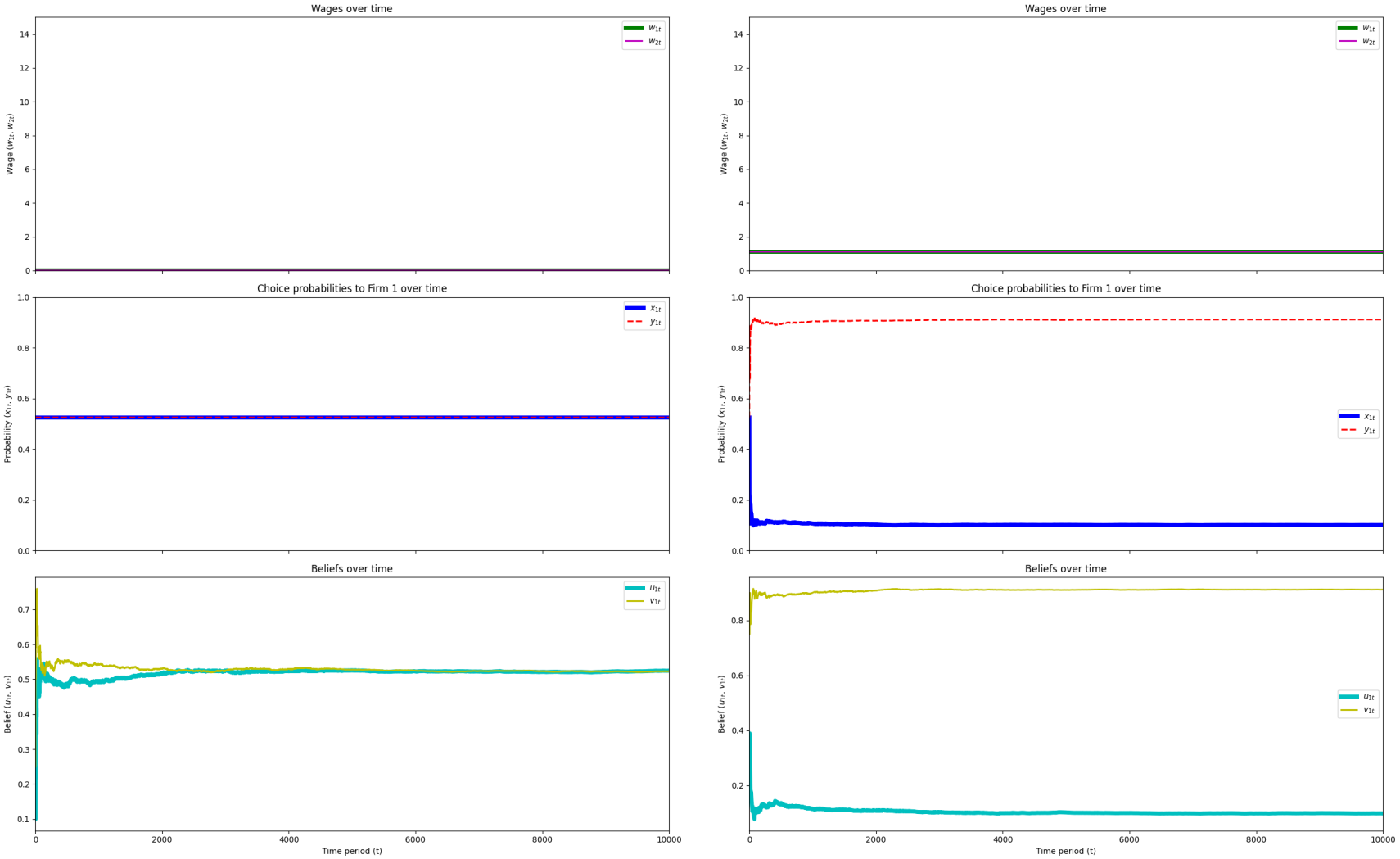
$$\frac{dx_{1t}}{d\Delta\alpha^i} = x_{1t}(1 - x_{1t}) > 0, \frac{dy_{1t}}{d\Delta\alpha^{-i}} = y_{1t}(1 - y_{1t}) > 0 \tag{105}$$

$\Delta\alpha^i$ and $\Delta\alpha^{-i}$ affect x_{1t} and y_{1t} positively, implying stronger bias towards firm 1 would increase application probability to firm 1.

Result Summary. In this section, I explore the market structure where workers observe the wages before formulating their application strategies. In the long run, workers' and firms' behaviour would converge to equilibria similar to QRE. Using Algorithm 1, I show that as beliefs stabilize, there is clear convergence to a single equilibrium, however, which equilibrium is selected depends on whether multiple equilibria are feasible for the system.

For workers to coordinate on applying to different firms, the presence of multiple equilibria is crucial. This can arise when workers' sensitivity (β) to expected payoffs is sufficiently high. In such case, workers would converge to applying with higher probabilities to different firms, leading to more efficient outcome, and in the limiting case, these asymmetric equilibria are locally asymptotically stable. However, equilibrium selection among asymmetric outcomes remains indeterminate, it would be contingent on the belief updating process and the point at which beliefs stabilize.

I also show that exogenous bias from long-term experiences could affect both the existence of multiple equilibria and long run outcome. When experience bias is strong relative to expected payoffs, workers could rely solely on their bias to inform application strategies, leading to a single equilibrium for the system. Having strong and diverse bias towards different firms could direct workers naturally to different firms. On the other hand, when experience bias is less prominent as compared to expected payoffs, multiple equilibria could emerge and previous analysis about convergence and selection apply accordingly.



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

Figure shows wages, choice probabilities and belief evolution, 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 20: Changes in Wages, Workers' Choice Probabilities and Beliefs Over Time with Bias (Algorithm 1)

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. 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

These learning models offer an evolutionary account in which the economy may experience long episodes of mismatch prior to reaching equilibrium. However, when the search problem admits multiple equilibria, this in turn calls into question which equilibrium will be selected and whether workers can ultimately learn to coordinate and 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, they would converge to coordinate on applying to different firms in the long run when wages are fixed and multiple equilibria exist. 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" effect and is illuminating 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, both workers and firms are adaptive learners, wages are driven down to 0 in the long run, learning ceases in the absence of rewards, thus the eventual equilibrium outcome simply relies on the learning path. Workers could end up at an equilibrium strategy that constitute more randomized choice probabilities.

In the second market structure, where workers observe the wages before making an application decision, their choice probabilities depend on wages and beliefs about their opponent's choices. Firms, on the other hand, are setting wages based on their inference about workers' beliefs given historical realized actions. In presence of multiple equilibria, when workers are highly sensitive to expected payoffs, they converge to the asymmetric equilibria, where they apply with high probability to different firms. For relatively low wage sensitivity, however, there could be convergence to the symmetric equilibrium. The presence of long-term experiences as an exogenous bias could affect equilibrium multiplicity. The equilibrium selection, as a result, would also be influenced depending on the extent of relative bias as compared to expected payoffs. When facing strong experience bias, there could be a unique equilibrium where workers converge to.

In general, workers coordinating on applying with high probability to different firms can happen with experience-based learning without intervention. However, it may not be clear which pure NE or asymmetric equilibria will be selected, which can depend on initial conditions and the learning path.

4.2 Wage Transparency

The market structures and learning dynamics that I have adopted in this paper also provide direct basis for analysing the efficiency of wage transparency.

When wages are not observable and workers are learning via reinforcement of past successes, they are facing a substantially more complex learning problem. They must simultaneously infer the underlying wage distribution and learn how to coordinate on application strategies. In such setting, feedback is limited to a realised payoff from their own application, which provides a noisy reinforcement signal that does not disentangle wage payoffs and choices made by opponent. The adjustment process could be slow and the market can remain trapped in inefficient, mismatched configurations for extended periods of time. By contrast, when wages are observable and workers are adopting best response dynamics, the learning is markedly simplified. Since wage structure is common knowledge, workers only need to learn the coordination problem and how to best respond to beliefs about their opponent's strategies. The information contained in observed past wages and application choices generates a much richer and more informative feedback signal. This facilitates faster convergence to the efficient, coordinated allocation.

As a result, by comparing the two learning mechanism, it is possible to infer that wage transparency operates as a coordination-enhancing device. It simplifies the learning problem, accelerates the selection of efficient equilibria, and overall reduces the cumulative welfare losses associated with transitional mismatch.

4.3 Policy Implications

The model's dynamic, learning-based framework delivers policy implications that go well beyond simple information provision. Because wages effectively define the game that workers are playing, managing the wage environment is itself a key policy lever. The wage structure shapes the strategic possibilities of the coordination problem, it determines whether the more efficient coordinated equilibria exist. Policies could therefore aim to preserve wage conditions under which multiple equilibria could emerge and coordination on efficient outcomes is possible.

A second implication concerns the timing of interventions. The strong path dependence in the model implies that the effectiveness of policies aimed at fostering labour mobility and rapid lock-in is highly time-dependent. There is a critical window before experience effect kicks in, and during which interventions are particularly powerful. As a result, policies such as internships, enhanced career counselling and rotational programmes should be front-loaded. This can help to either diversify initial experiences, allowing more exploration in the early stage of the career, or induce a rapid lock-in to matches that pushes the market toward one of the coordinated equilibrium.

Finally, for more experienced workers, the model highlights the risk of becoming trapped in inefficient learning cycles. Such traps can arise from strong initial biases that concentrate applications on one of the firms, or from a sequence of bad luck that generates a long history of outcomes, which are difficult to overturn. In these cases, marginal adjustments to information provision are unlikely to be sufficient, and instead, interventions could aim to deliver informational shocks that are strong enough to counteract the weight of past experiences. For example, job search platforms can actively push "out-of-the-box" recommendations to the job seekers, the platform designs can also deliberately devalue or down-weight old search history, and retraining programmes can be offered to help the workers to "unlearn" entrenched search heuristics.

4.4 Potential Extensions

One area of extension is to incorporate learning with imprecision to portray a more realistic job search scenario. While the learning rules in the paper already contain some aspects of constrained decision-making, such as limited feedback under reinforcement learning when individuals do not explore enough; and the inability to observe opponent's strategy under learning with best response, hence relying on belief updating from past realized actions, it would be valuable to model additional sources of imprecision in learning the wage environment and in recalling past experiences, diving deeper into cognitively-founded learning dynamics. For instance, the sequential learning environment provides a natural basis for studying inattention. A logit choice model could encompass noise from both inattention (Matějka and McKay (2015); Mattsson and Weibull (2002)) and experience updating, postulating a novel form of learning rule that links information frictions and dynamic adjustments. Furthermore, even though I considered decay in memories, it may be more realistic to allow for noisy memories (Azeredo da Silveira et al. (2024); Aridor et al. (2024)), which could generate richer adjustment paths and help explain fluctuations or temporary reversals in matching outcomes.

A second extension is to move beyond the two types of learning models postulated in the paper by allowing mixed learning models. In practice, agents may combine reinforcement from realized outcomes with belief-based adjustment, or switch between different heuristics over time. The two learning models analysed could provide the baseline, starting points for studying hybrid learning dynamics.

Last but not least, this paper naturally motivates scaling up the framework beyond the 2×2 environment to multi-player markets. In an $N \times N$, or more generally $I \times J$ setting, the core insights are likely to remain robust, where inefficient symmetric outcomes are expected to remain unstable. However, equilibrium selection may involve many possible one-to-one matchings, up to $N!$ perfect matchings or fuzzy, near-perfect matchings. The increased complexity could imply slower convergence and longer transitory mismatch. Nonetheless, this opens up possibility for large-scale agent-based simulations grounded in the learning framework developed in this paper, enabling a systematic study of convergence rates, mismatch durations and equilibrium selection patterns as market size grows.

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.

Azeredo da Silveira, R., Sung, Y., and Woodford, M. (2024). Optimally imprecise memory and biased forecasts. *American Economic Review*, 114(10):3075–3118.

Bamieh, O. and Ziegler, L. (2025). Can wage transparency alleviate gender sorting in the labour market? *Economic Policy*, 40(122):401–426.

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.

Benaım, M. and Hirsch, M. W. (1999). Mixed equilibria and dynamical systems arising from fictitious play in perturbed games. *Games and Economic Behavior*, 29(1-2):36–72.

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.

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: December 27, 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 Example 3

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

Example 3. *Period 1: Initial wages are randomly picked. Suppose workers start off from* $G1$ *(Figure 9), 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 continues to experiment wages randomly, assuming uniform probability distribution 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 continues 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 10). Choosing* $F1$ *will lead to lower reinforcement as compared to choosing* $F2$*, propensities could thus be updated, and slowly, there will be convergence towards* $(F1, F2)$ *or* $(F2, F1)$*. 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 Time to "Unlearn" the Past with Partial Recall

When there is a change from $G1$ to $G2$ at $t = \tilde{t}$, workers start experiencing decay for propensities accumulated in $G1$:

$$\text{Worker } i\text{: } q^i_{j\tilde{t}} = (1-\eta)q^i_{j(\tilde{t}-1)}, \text{Worker } -i\text{: } q^{-i}_{j\tilde{t}} = (1-\eta)q^{-i}_{j(\tilde{t}-1)} \tag{106}$$

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

$$\text{Worker } i\text{: } q^i_{j(\tilde{t}+1)} = (1-\eta)q^i_{j(\tilde{t})} + \pi^i_{\tilde{t}}(a^i_{\tilde{t}}, a^{-i}_{\tilde{t}}, a^j_{\tilde{t}}, a^{-j}_{\tilde{t}}) \tag{107}$$

$$\text{Worker } -i\text{: } q^{-i}_{j(\tilde{t}+1)} = (1-\eta)q^{-i}_{j(\tilde{t})} + \pi^{-i}_{\tilde{t}}(a^i_{\tilde{t}}, a^{-i}_{\tilde{t}}, a^j_{\tilde{t}}, a^{-j}_{\tilde{t}}) \tag{108}$$

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

$$q^{i*}_j = \frac{\pi^i_j}{\eta} \tag{109}$$

At $t = \tilde{t}$, one starts new accumulation of propensity according to $G2$, thus for $t > \tilde{t}$,

$$q^i_{jt} = (1-\eta)^{t-\tilde{t}} q^i_{j\tilde{t}} + \sum_{k=0}^{t-\tilde{t}-1} (1-\eta)^k \pi^i_j \tag{110}$$

$$q^i_{jt} = (1-\eta)^{t-\tilde{t}} q^i_{j\tilde{t}} + \frac{1-(1-\eta)^{t-\tilde{t}}}{\eta} \pi^i_j \tag{111}$$

$$q^i_{jt} = (1-\eta)^{t-\tilde{t}} (q^i_{j\tilde{t}} - q^{i*}_j) + q^*_j \tag{112}$$

Set $q^i_{j\tilde{t}} = 0$, the new accumulated propensities for $t > \tilde{t}$:

$$q^i_{jt} = (1-\eta)^{t-\tilde{t}} (-q^{i*}_j) + q^*_j \tag{113}$$

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

$$(1-\eta)^{t-\tilde{t}} (-q^{i*}_j) + q^*_j \geq q^i_{j\tilde{t}} (1-\eta)^{t-\tilde{t}} \tag{114}$$

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

From the inequality, I obtain the time lag $(t - \tilde{t})$:

$$t - \tilde{t} \leq \frac{\ln(\frac{q^*_j}{q_{j\tilde{t}} + q^*_j})}{\ln(1-\eta)} \tag{115}$$

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

$$t - \tilde{t} \propto \frac{1}{\eta} \tag{116}$$

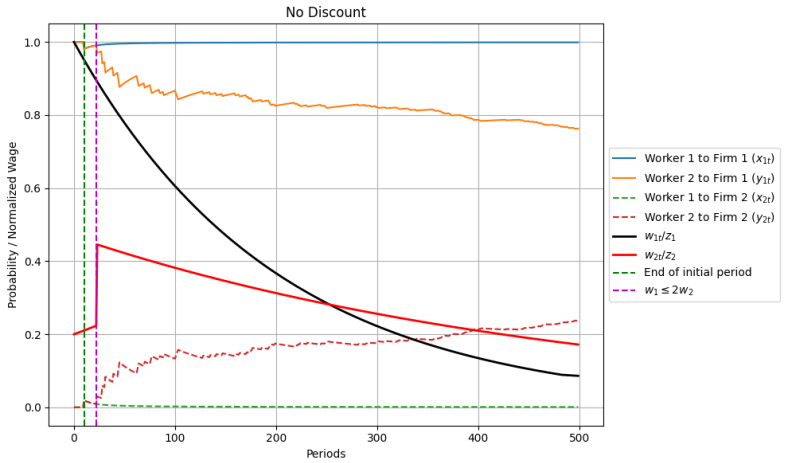
But since $G2$ payoffs are likely to be non-constant as payoffs evolve through reinforcement learning on the firms' side, the dependency of time lag on payoff changes can be more generically written as dependency on η and $f(G2$ payoff dynamics), shown in equation (34), which captures the learning process.

Return to Section 2.2.1.

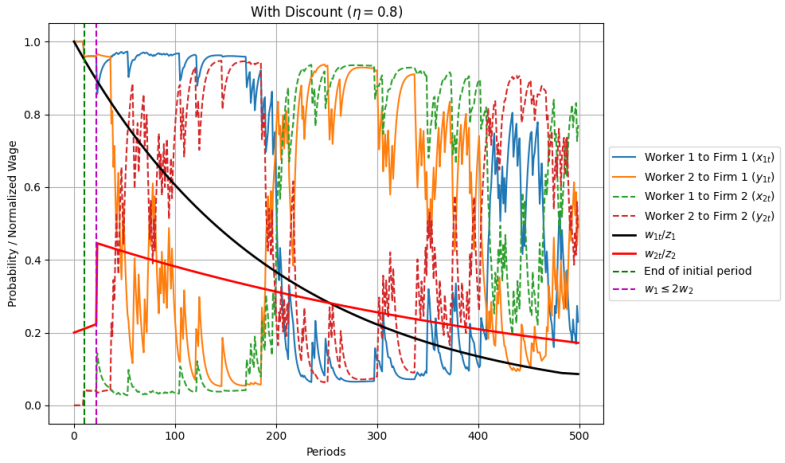
B Simulations

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

Figure 21a and 21b show simulations 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 choose firm 1 for 10 periods, during which, w_1 is arbitrarily adjusted downwards and w_2 upwards. Given exogenous wage changes, there can be a shift from G1 to G2 (i.e. $2w_1 > w_2 > \frac{w_1}{2}$), at the point of pink dotted line, leading to a switch in equilibrium to learn from $(F1, F1)$ to the set $(F1, F2)$ and $(F2, F1)$.



(a) Perfect Recall



(b) Partial Recall

Figures show for $z_1 = z_2 = 10$, $w_{10} = 10, w_{20} = 2$, all initial propensities are fixed at 1 ($x_{10} = y_{10} = 1$, $\theta^j_{a^j_0} = \theta^{-j}_{a^j_0} = 1$):

1. Assume $w_1 \geq 2w_2$ for initial 10 periods, workers are programmed to choose $(F1, F1)$, and there is exogenous adjustment in w_1 downward by 0.5% and w_2 upward 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 even though both workers

start applying to firm 1, over time, one of the worker would adjust his/her strategy and redirect search towards firm 2. In the long run, if the wage condition is sustained, it is expected that workers will apply 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 a different set of NE in G1.

In Figure 21b, when there is a forgetting parameter, workers' choice probabilities change more rapidly. There is a rapid 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. When 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, as workers' strategies are more influenced by recent events, their choices could exhibit greater stochasticity. As a result, their may not converge to applying more to different firms even over an extended time period.

Return to Section 2.2.1.

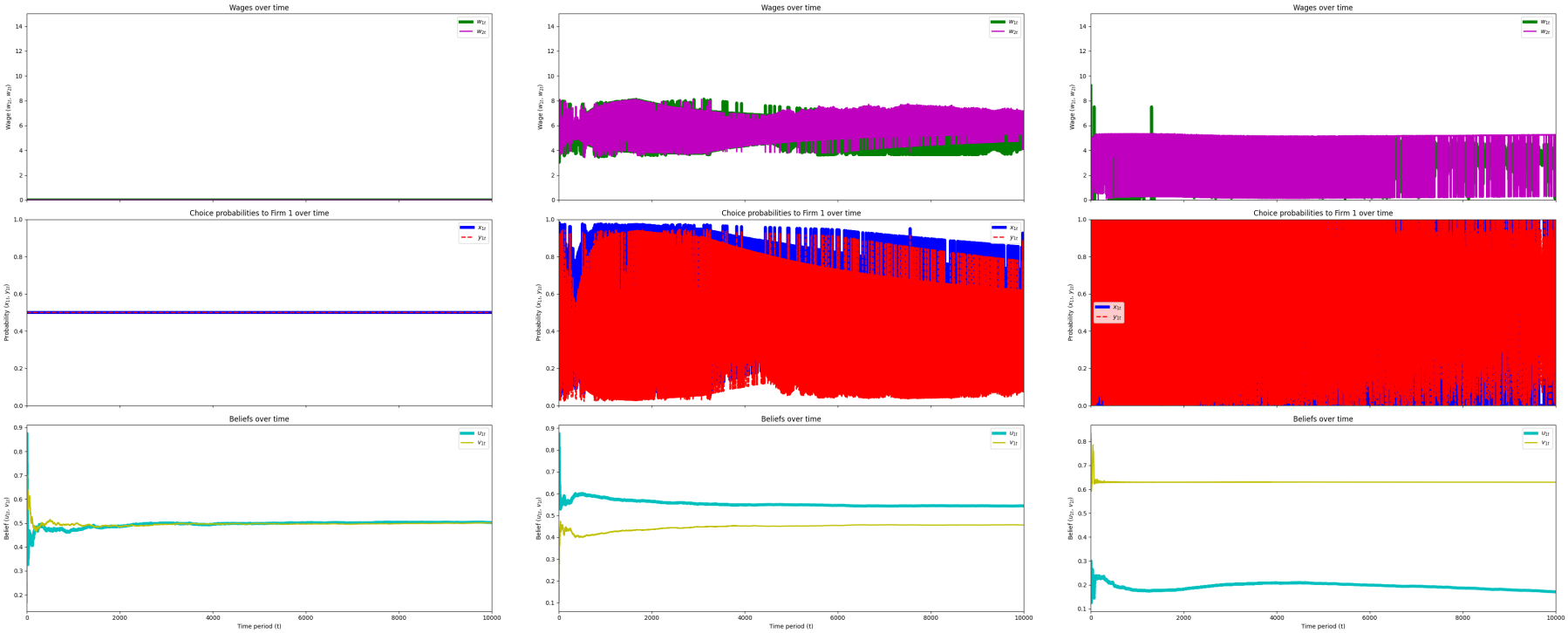
B.2 Different Algorithms for Sequential Learning

In this section, I experimented with other algorithms, which mainly differ on the firms' side.

Algorithm 2 Firms' Wage-setting using Random Initial Guesses

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 10000 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 (38) and (39) to obtain $(x_{1t}^{\text{Potential}}, y_{1t}^{\text{Potential}})$.
        Based on workers' potential application rates, compute wages using equations (51) and (52) 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**
      For (w_{1t}, w_{2t}), compute x_t and y_t given u_t and v_t using equations (38) and (39).
      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 (41) 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.

With Algorithm 2, I show in Figure 22 one simulation example, 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 $\beta = 0.2$, workers' choice probabilities



(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 22: Changes in Wages and Workers' Choice Probabilities and Beliefs Over Time (Algorithm 2)

stays at 0.5, beliefs converge to 0.5, and wages are 0 based on equation (51) and (52). At $\beta = 1.0$, wages are positive. While beliefs about opponents are seemingly stable and converge to around 0.5, wages and workers' choice probabilities fluctuate rapidly. In the final case of $\beta = 5.0$, wages are also positive. There seems to be convergence of beliefs to opponents applying with higher probability to different firms, but there is more drastic fluctuations in firms' wage setting and workers' choice probabilities.

In Table 1, 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.

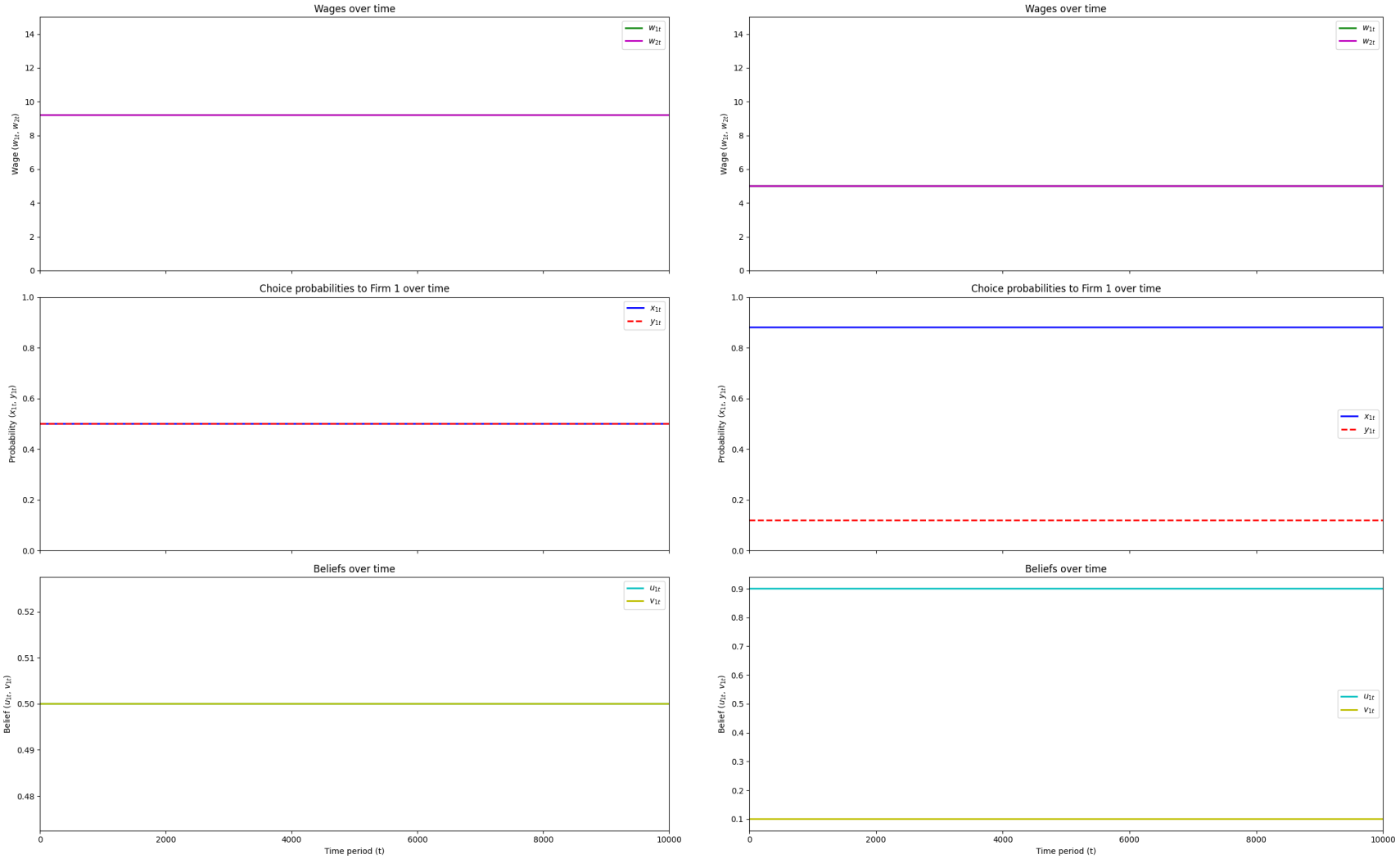
First and last session	w_1	w_2	x_{1t}	y_{1t}
Low β (0.2) [First]	0.000000	0.000000	0.500000	0.500000
Low β (0.2) [Last]	0.000000	0.000000	0.500000	0.500000
Medium β (1.0) [First]	4.331617	4.332060	0.499997	0.499995
Medium β (1.0) [Last]	4.331806	4.331872	0.499994	0.500004
High β (5.0) [First]	4.599850	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 1: Average Values of Last 20% of 10000 Periods for 20 Simulation Sessions

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 2 starts with random guesses of wages in 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. Algorithm 2 is effectively a local optimization method, it explores local optima points, where wages are searched near the neighbourhood of initial guesses in each period.

In order to nail down the reason behind the fluctuations in choice probabilities and wages, (1) I first fix workers' beliefs to investigate if workers' choice probabilities and firms' wage-setting behaviour stabilize when beliefs stabilize; and (2) I then fix both wages or only one of the wages to verify if firms' wage-setting behaviour in this algorithm is the triggering factor for the fluctuations.

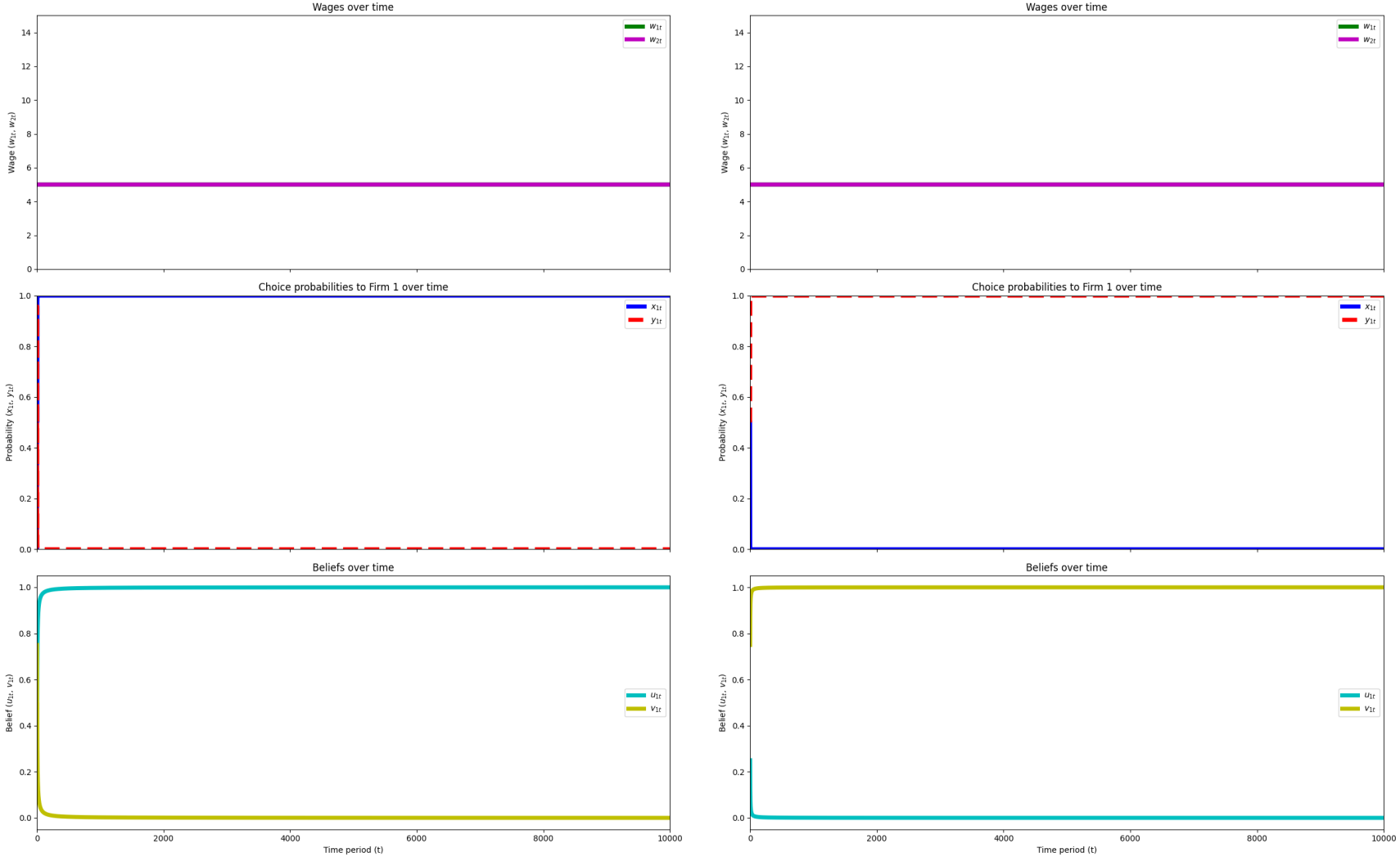
Figures 23, 24 imply that if beliefs stabilize, workers and firms' behaviour would stabilize. Similarly, fixing wages would lead to convergence to a unique set of beliefs and choice probabilities. As a result, the observed fluctuations in the full system is due to small shifts in beliefs, which move workers' choice probabilities, in turn affect wages significantly, and further feedback into choice probabilities, leading to large fluctuations in the feedback loop ($u_t, v_t \to x_t, y_t \leftrightarrow w_{1t}, w_{2t}$).



(a) $(u_1, v_1) = (0.5, 0.5)$ (b) $(u_1, v_1) = (0.9, 0.1)$

Figure shows wages, choice probabilities, fixing beliefs at $(u_1, v_1) = (0.5, 0.5)$ and $(0.9, 0.1)$, given $\beta = 5.0$, $z_1 = z_2 = 10$, $t = 10000$, $\alpha_1^i = 0$, $\alpha_1^{-i} = 0$, and no smoothing.

Figure 23: Changes in Wages, Workers' Choice Probabilities Over Time for Fixed Beliefs (Algorithm 2)



(a) $w_1 = w_2 = 5$ (b) $w_1 = 5$

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

Figure 24: Changes in Workers' Choice Probabilities, Beliefs Over Time for Fixed Wages (Algorithm 2)

Algorithm 3 Firms' Wage-setting with Small Adjustment from Previous Period

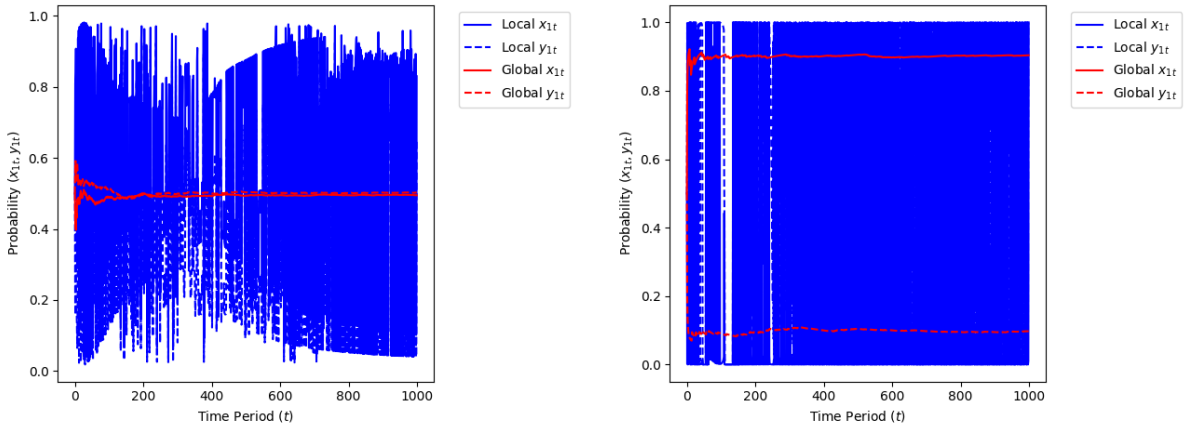
1: **for** all firms **do**
2: Check if previous wages are still optimal given updated workers' beliefs based on past period realized actions.
3: **for** a set of $(w_{1(t-1)}, w_{2(t-1)})$ **do**
4: Compute workers' $(x_{1t}^{\text{Potential}}, y_{1t}^{\text{Potential}})$ based on equations (38) and (39).
5: Based on workers' potential application rates, compute corresponding wages, $(w_{1t}^{\text{Potential}}, w_{2t}^{\text{Potential}})$, using equations (51) and (52).
6: Compare $(w_{1t}^{\text{Potential}}, w_{2t}^{\text{Potential}})$ with $(w_{1(t-1)}), (w_{2(t-1)})$.
7: **end for** checking previous wages
8: If previous wages is optimal, set $(w_{1t}, w_{2t}) = (w_{1(t-1)}), (w_{2(t-1)})$; adjust otherwise.
9: **for** previous wages no longer optimal **do**
10: Loop for 500 iterations
11: **for** first iteration **do**
12: Compute difference between $(w_{1t}^{\text{Potential}}, w_{2t}^{\text{Potential}})$ and $(w_{1(t-1)}), (w_{2(t-1)})$.
13: Make new wages guesses in small increment if wage difference is small, otherwise make a bigger adjustment:

$$w_{1t}^{\text{Guess}} = w_{1(t-1)} + \text{increment} * \text{wage difference} \tag{117}$$

$$w_{2t}^{\text{Guess}} = w_{2(t-1)} + \text{increment} * \text{wage difference} \tag{118}$$

14: Given wage guesses, compute workers' reaction, $(x_{1t}^{\text{Potential}}, y_{1t}^{\text{Potential}})$, based on potential reaction, compute potential wages, $(w_{1t}^{\text{Potential}}, w_{2t}^{\text{Potential}})$.
15: Compare $(w_{1t}^{\text{Potential}}, w_{2t}^{\text{Potential}})$ and $(w_{1t}^{\text{Guess}}, w_{2t}^{\text{Guess}})$, if far apart, the guess is incorrect, start a new guess in subsequent iteration based on their differences.
16: **end for** first iteration
17: 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.
18: **end for** all iterations
19: Set wages (w_{1t}, w_{2t}) to be the set for which the difference between guessed wages and potential wages is negligible.
20: **end for** firms

I experiment with another algorithm. Algorithm 3 shows the firms' side, the rest is the same as before. In this algorithm, I modify the firms' behaviours (Line 6-17) slightly to put some structure on wage guesses, such that next period wage guesses are made in the vicinity of previous period wages. The simulation results are similar to Algorithm 2, again implying that small shifts in beliefs can contribute to fluctuations in choice probabilities and wages when there exist multiple equilibria.



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

Figure shows comparison of choice probabilities evolution for two wage-setting approach, denoted as local (Algorithm 3) and global (Algorithm 1) respectively ($\beta = \{1.0, 5.0\}$, $z_1 = z_2 = 10$, $t = 1000$, $\alpha_1^i = \alpha_1^{-i} = 0$, $u_{10} = 0.5$, $v_{10} = 0.5$, no smoothing)

Figure 25: Comparing Workers' Strategies For Local and Global Wage-setting Approach

However, firms' wage-setting behaviour also plays a role in the fluctuations of workers' choice probabilities. Comparing the evolution of workers' choice probabilities to firm 1 using Algorithm 1 and 3 for the cases with multiple equilibria in Figure 25, I show that there is a clearer selection of one equilibrium in the former case, and drastic fluctuations in choice probabilities in the later. The potential reason behind this is the difference between global and local optimization approach. The local optimization method by Algorithm 3 can lead to circling around a single equilibrium as wages react to small adjustment in beliefs when they move from previous period values, and there can also be jumps across different equilibria due to large effect from the feedback loop. On the other hand, global optimization method by Algorithm 1 scans the entire wage landscape for each set of beliefs, making it more robust to small shifts in beliefs and avoiding such fluctuations.

Return to Section 3.1.

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

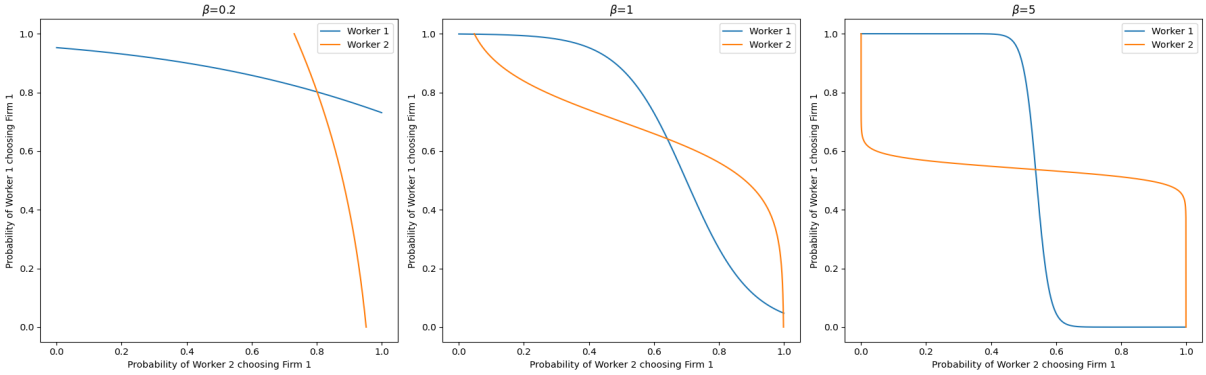


Figure shows workers' choice probability for $\alpha_1^i = \alpha_1^{-i} = 2$, $\alpha_2^i = \alpha_2^{-i} = 0$, when fixing the wages to $w_1 = w_2 = 10$.

Figure 26: Workers' Response Functions with Bias

Return to Section 3.2