Learning When to Say No*

David Evans
University of Oregon

George W. Evans
University of Oregon and University of St Andrews

Bruce McGough
University of Oregon

March 1st, 2021

Abstract

We consider boundedly-rational agents in McCall's model of intertemporal job search. Agents update over time their perception of the value of waiting for an additional job offer using value-function learning. A first-principles argument applied to a stationary environment demonstrates asymptotic convergence to fully optimal decision-making. In environments with actual or possible structural change our agents are assumed to discount past data. Using simulations, we consider a change in unemployment benefits, and study the effect of the associated learning dynamics on unemployment and its duration. Separately, in a calibrated exercise we show the potential of our model of bounded rationality to resolve a frictional wage dispersion puzzle.

JEL Classifications: D83; D84; E24

Key Words: Search and unemployment; Learning; Dynamic optimization; Bounded rationality; Wage dispersion

1 Introduction

We reconsider the labor-search framework in which a worker must decide whether to work at a given wage or to wait and search for a better wage. These models have been widely used

*We thank the Editor, Associate Editor, and two anonymous referees for valuable comments. Financial support from National Science Foundation Grant No. SES-1559209 is gratefully acknowledged.

in macroeconomics to study frictional unemployment and wage dispersion: see, for example, Rogerson, Shimer and Wright (2005). While these models have met with considerable success, empirical puzzles remain. For example Hornstein et al (2011) (HKV) show that a general class of labor-search environments, centered on the McCall (1970) sequential-search model, are not able to produce adequate wage dispersion. Labor search models rely heavily on optimizing agents with rational expectations (RE), and in this paper we examine what happens in these models when agents must learn over time how to optimize against unknown wage distributions.

There are a number of reasons to question the plausibility of the rational expectations hypothesis. Rational forecasting requires a full understanding of the relevant stochastic structure – a structure that is arguably out of the reach of working economists. Indeed, professional forecasters do not know and generally disagree on the correct model, and for any given specification they must estimate the model's parameters – parameters that are assumed known under RE. Separately, rational agents are also assumed to solve complex dynamic programming problems – problems that are in reality approachable only through sophisticated computational techniques.

To examine the implications of adaptive learning in labor-search models, we follow HKV and adopt the McCall model as our laboratory. The simplicity of the model allows us to obtain analytic results.[1] To model how agents learn over time, we adopt a bounded optimality approach along the lines of Evans and McGough (2018a) in which agents make decisions based on perceived trade-offs. This approach to boundedly optimal decision-making is particularly simple and natural in the context of labor search models with qualitative choices.

Our implementation of bounded optimality has a number of attractive features. Agents' behaviors are at once sophisticated and simple: they are anchored to the Bellman approach and yet only require that agents make decisions by selecting the better of two options, based on the perceived value of receiving a random wage draw. This value is revised over time as experience is gained and new data become available. Despite its simplicity, this implementation is a small deviation from fully rational decision-making in that agents can learn to optimize over time. For convenience we will refer to our framework as the bounded rationality (BR) model.

We begin by developing our version of the McCall model using the framework of Ljungqvist and Sargent (2012) who obtain a solution under rational expectations and optimal decision-making. Then, using value-function learning as suggested in Evans and McGough (2018a), we develop a framework for boundedly rational decision-making. Under very general conditions, we show directly, using the martingale convergence theorem, that agents make fully optimal decisions asymptotically. Through numerical simulations, we study the implications of learning dynamics, which are distinct from their RE counterpart. We conclude by demonstrating the potential of our approach to address the HKV puzzle.

[1] While the McCall model is partial equilibrium in the sense that the wage offer distribution is exogenous, this assumption, as pointed out by HKV, "is not at all restrictive." HKV illustrate this point using the island model of Lucas and Prescott (1974) and the matching model of Pissarides (1985).

A key feature of our BR approach is that, in making their decisions, workers incorporate several structural features of the economy that they know, while learning over time about a key but unknown sufficient statistic for optimal decision-making. This unknown sufficient statistic, which we denote by Q^*, measures the rational agent's expected discounted utility when they are unemployed and waiting for a random wage offer. As is well-known, optimal decision-making in this setting is characterized by a reservation wage w^* that is pinned down by Q^*. Under boundedly rational decision-making with adaptive learning, an agent uses an estimate Q of Q^* to make decisions given their knowledge of the unemployment benefit level and the probability of job separation when employed. The agent's estimate Q determines their corresponding reservation wage $\bar{w}$, and thus their boundedly optimal decisions.

The estimate Q of Q^* is updated over time based on observed wage offers. It is natural to assume that both unemployed and employed workers observe a (possibly small) sample of wage offers; an agent updates their estimate Q based on this sample. Our central theoretical result is that this procedure asymptotically yields fully optimal decision making: over time agents learn Q^*. We emphasize two distinct features of our result that are particularly attractive. First, agents do not need to have any knowledge of the distribution of wages; and second, their computations are simple as well as natural: they do not need to iterate a value function and only need to track one estimate that is updated each period.

Asymptotic optimality in stationary environments also motivates the use of closely related adaptive learning mechanisms through which agents adjust their behaviors in the presence of possible structural change. To examine the implications of possible structural change for the model's dynamics, we turn to simulations based on constant-gain learning, i.e. agents use an estimation procedure that discounts older data. We note that the use of constant gain is standard in applied work involving adaptive learning, and is particularly appropriate when the possibility of structural change is under consideration.

In Section 4 we investigate the behavior of BR agents in the presence of structural change. We consider a simple policy experiment in which unemployment benefits are unexpectedly and permanently changed. By embedding our framework in a model populated by many agents, the comparative statics and dynamics of aggregates like the mean unemployment rate and unemployment duration can be examined. We find that the impact response of an increase in benefits appears unrealistically large under RE, whereas it is much more muted under BR. This finding reflects a combination of direct effects, i.e. those effects induced by changes in structure holding fixed beliefs, and indirect effects resulting from the changes in beliefs induced by the structural change: both effects are simultaneously realized for the rational agent whereas the BR agent learns about the indirect effect only gradually over time.

In Section 5, we show the potential for the BR approach to be a simple and parsimonious way to address puzzles in the labor-search literature. Specifically, we consider the frictional-wage distribution puzzle identified by HKV, who find that calibrated search models fail to generate significant cross-sectional variation in wages. That the agents have rational expectations, and therefore fully incorporate the value of search into their decisions, is a foundational assumption of their analysis. Under rational expectations, the short unemployment durations observed in the US imply a low value to searching and, hence, require a small dispersion of wages generated by search frictions. Introducing boundedly rational decision makers is a natural way to resolve this puzzle. Indeed, using a calibrated model in line with that of HKV, we show that when agents are fully rational the spread in observed wages is minimal, while the spread is much larger when agents are boundedly rational.

Labor search models in which agents have incomplete knowledge about the wage distribution have been considered by a variety of authors. In an early contribution, Burdett and Vishwanath (1988) examine the time-series behavior of reservation wages in a modeling environment similar to ours, in which agents do not fully the understand the distribution of wage offers, and act as Bayesian decision makers that fully solve their dynamic optimization problems. While their main result is that, under certain assumptions, reservation wages are declining in the length of unemployment spells, they also find it is not trivial even to establish that a reservation-wage strategy is optimal. An advantage of our approach is that BR agents employ a reservation-wage strategy that is natural, asymptotically optimal and easy to implement.

More recently, Rotemberg (2017) uses the HKV version of the McCall model to analyze the potential for group learning to explain wage dispersion. In his setup the distribution of wage offers is unknown and is misperceived to be stationary. Conditional on their misperceptions, agents make fully optimal decisions using the observed means of their group's realized wages and the durations of their unemployment spells. He shows that if the wage-offer distribution is non-stationary in the sense that later wage offers have a higher mean, then the stationary perceptions held by agents can lead to self-confirming equilibria with increased wage-dispersion and even to equilibrium multiplicity.[2] Like Burdett and Vishwanath (1988), Rotemberg (2017) adopts the assumption that agents can fully solve their dynamic programming problem given their potentially misspecified beliefs.[3]

The approach presented in this paper is related to several lines of inquiry in the macroeconomics and machine-learning literature. Like the adaptive learning approach in macroeconomics, e.g. Bray and Savin (1986), Marcet and Sargent (1989) and Evans and Honkapohja (2001), which focuses on least-squares learning, we consider decision-making procedures that, while not fully rational, have the potential to converge to rational expectations. Like Marimon, McGrattan and Sargent (1990), Preston (2005) and Cogley and Sargent (2008), our framework has long-lived agents that must solve a challenging dynamic stochastic optimization problem. In these settings two issues are of concern: (i) there are parameters that govern the state dynamics that may not be known; and (ii) the assumption that agents know how to solve dynamic stochastic programming problems is implausibly strong.

Cogley and Sargent (2008) examine the first issue in a permanent-income model in which

[2]We note that Rotemberg (2017) is focused on equilibrium outcomes, and in particular does not explicitly include consideration of learning dynamics.

[3]Using a different labor-search model that includes Nash bargaining, Damdinsuren and Zaharieva (2018) assume boundedly rational agents use (misspecified) linear forecasting models based on economy-wide data to form expectations of future wages. The dynamics of the model with adaptive learning are studied numerically in an agent-based version of the model. Among their findings is that higher bargaining power is associated with higher wages and larger wage dispersion.

income follows a two-state Markov process with unknown transition probabilities, taking the agent's problem outside the usual dynamic programming framework. Two approaches to decision making are considered. Bayesian decision-makers follow a fully optimal decision rule within an expanded state space, which requires considerable sophistication and expertise for the agent. The second approach employs the "anticipated utility" model of Kreps (1998), in which agents make decisions each period by solving their dynamic programing problem going forward based on current estimates of the transition probabilities. This procedure is boundedly rational, since agents ignore that their estimates will be revised in the future, but is computationally simpler and performed almost as well as optimal decision-making.

Bayesian decision-making has also been used in boundedly-rational settings. Adam, Marcet and Beutel (2017) have shown how to implement this approach in an asset-pricing environment. In their set-up agents are "internally rational," in the sense that they have a *prior* over variables exogenous to their decision-making, which they update over time using Bayes Law. These beliefs may not be externally rational in the sense of fully agreeing with the actual law of motion for these variables. By imposing simple natural forms of beliefs, it is possible to solve the agents' dynamic programming problem.

The anticipated utility framework has also been employed in adaptive learning set-ups in which long-lived agents use least-squares learning. Preston (2005) developed an approach in which agents estimate and update over time the forecasting models for relevant variables exogenous to the agents' decision-making. For given forecasts of these variables over an infinite horizon, agents make decisions based on the solution to their dynamic optimization problem. Again, these decisions are boundedly optimal in the sense that the procedure does not take account of the fact that their estimates of the parameters will change over time.

The approach adopted in the current paper is closest to the general bounded-optimality framework of Evans and McGough (2018a). In their approach infinitely-lived agents optimize by solving a two-period problem in which a suitable variable in the second period encodes benefits for the entire future. While their primary focus is on shadow-price learning, in which the key second-period variables are shadow prices for the endogenous state vector, they also show how a value function learning approach can equivalently be employed in a setting with continuously measured state variables. Within a linear-quadratic (LQ) framework, Evans and McGough (2018a) use the anticipated utility approach and obtain conditions under which an agent can learn to optimize over time.[4]

The current paper applies a version of value-function learning in a discrete choice setting in which a worker must choose whether or not to take a wage offer. While this a natural setting for value-function learning, the discrete choice and non-LQ features of the model take the agent's decision problem outside the theoretical framework developed in Evans and McGough (2018a). In our McCall-type set-up, the single sufficient statistic needed is the value associated with the dynamic optimization problem when the agent is unemployed and facing a random wage draw. We show how, given an estimate of this value, an agent can make boundedly optimal decisions under the anticipated utility assumption, and we demonstrate

[4]Evans and McGough (2018b) show how to apply shadow-price learning to a wide range of DSGE macroeconomic frameworks using computational techniques.

that when agents in addition use a natural adaptive-learning scheme for updating their estimates over time, they will asymptotically learn with probability one how to make optimal decisions within a stationary environment.[5]

Our framework is also related to the "Q-learning" approach developed originally by Watkins (1989) and Watkins and Dayan (1992) as well its extensions to temporal difference learning from the computer science literature.[6] In these approaches agents make decisions based on estimates of quality-action pairs, with the quality function updated over time. As in the current paper the Q-learning approach is motivated by the Bellman equation, but it is typically and most effectively implemented in set-ups in which the state as well as action spaces are finite. In our set-up agents make decisions facing a continuously-valued wage distribution, where the distribution is unknown to the agents; furthermore, when making their boundedly optimal choices, our agents are able to incorporate features of the transition dynamics, such as benefit levels, that are known to the agents.

Our paper proceeds as follows. Section 2 outlines the environment. Section 3 presents our model of boundedly optimal decision-making. Section 4 discusses structural change and learning dynamics in the context of an unexpected benefits change. Section 5 covers the application to the HKV puzzle. Section 6 concludes.

2 The Model and Optimal Decision-making

We consider an infinitely-lived agent who receives utility from consumption via the instantaneous utility function U. Time is discrete, wages are paid in perishable goods, and there is no storage technology. At the beginning of a given period the agent receives a wage offer, and decides whether or not to accept it. The wage offer is drawn from a distribution that depends on whether the agent was employed or unemployed at the end of the previous period. If the agent was employed, her wage in the previous period constitutes her wage offer in the current period. If the agent was unemployed at the end of the previous period, she receives a wage offer w drawn from a time-invariant exogenous distribution F (density dF). In either case, the agent must decide whether or not to accept the offer.

If the wage offer is not accepted the agent is unemployed in the current period, and receives an unemployment benefit $b > 0$; and, because she is unemployed at the end of the current period, she will receive a wage offer drawn from F at the beginning of the next period. If the offer is accepted then the agent receives the wage w in the current period. We assume exogenous job destruction parameterized by α. At the end of the period, with probability $1 - \alpha$ the match with the firm is preserved and, because she is employed at the end of the current period, she will receive the same wage offer in the next period. With probability α the match is destroyed, the worker becomes unemployed at the end of the period, and at the beginning of the next period she receives a wage offer drawn from F. We remark that, under

[5] By deriving our results from first principles we are able to obtain global stability, which is also in contrast to Evans and McGough (2018a).

[6] See Sutton and Barto (2011) for a detailed introduction to reinforcement learning and in particular temporal difference learning.

full rationality, an agent employed in the preceding period will always accept her wage offer in the current period; however, under bounded rationality, previously employed agents may decide to enter unemployment as their understanding of the world evolves.

We make the following assumptions to ensure the that the worker's problem is well behaved, which we set out for future reference:

Assumption A:

1. U is twice continuously differentiable, with $U' > 0$ and $U'' \leq 0$.
2. F has support $[w_{\min}, w_{\max}]$, where $0 < w_{\min} < w_{\max}$.
3. All wage draws are independent over time and across agents.
4. $0 < \alpha < 1$.

The first two items ensure the existence and continuity of the worker's value function, while the third item guarantees that the worker's optimal value of search does not depend on additional state variables.

It remains to specify how agents decide whether or not to accept the wage offer. In this Section we adopt the conventional assumption that agents are fully rational and we characterize the corresponding optimal behavior. Section 3 takes up the case of boundedly rational agents, and shows, under suitable assumptions, that boundedly optimal decision-making converges to fully optimal behavior.

Assumption A implies that the fully optimal agent makes decisions by solving the following dynamic programing problem. Let $V^*(w)$ be the expected present value of utility of a fully rational worker entering the period with wage offer w. It follows that

$$\begin{aligned} V^*(w) &= \max_{a\in\{0,1\}} U\left(c(a,w)\right) + \beta E\left(V^*(w')|a,w\right) \\ & \quad w' = g(w,a,\hat{w},s), \end{aligned} \tag{1}$$

with the expectation E taken over random variables $\hat{w}$ and s. Here $a \in \{0,1\}$ is a control variable identifying whether the job is accepted ($a = 1$) or rejected ($a = 0$), $\hat{w}$ is an i.i.d. random variable drawn from F, and $s \in \{0,1\}$ is an i.i.d. random variable taking the value 1 with probability α, thus capturing job destruction. Finally, functions $c(a,w)$ and $g(w,a,\hat{w},s)$ are defined as follows

$$c(a,w) = \begin{cases} w & \text{if } a = 1 \\ b & \text{if } a = 0 \end{cases} \text{ and } g(w,a,\hat{w},s) = \begin{cases} w & \text{if } a = 1 \text{ and } s = 0 \\ \hat{w} & \text{otherwise} \end{cases}$$

and codify how the consumption and the availability of future wage offers depend on the worker's choice of accepting or rejecting the wage offer.

The optimal value $V^*(w)$ of having a wage offer w in hand allows us to define

$$Q^* = E\left(V^*(\hat{w})\right) \equiv \int_{w_{\min}}^{w_{\max}} V^*(\hat{w})dF(\hat{w}),$$

where we note that Q^* is the value, under optimal decision-making, associated with being unemployed at the start of the period, i.e. before $\hat{w}$ is realized. Moreover, as we will see in our introduction of bounded optimality, Q^* encapsulates all of the complicated features of this problem: that the wage offer distribution may not be known and that, even conditional on knowing the wage offer distribution, making optimal decisions requires solving a complicated fixed point problem.

3 Boundedly Optimal Decision-making

In this Section we specify how boundedly optimal agents make decisions, which requires allowing for an explicit dependence of the value function on beliefs. First in Section 3.1 we show how boundedly optimal decision-making can be formulated in terms of an agent's perception of the expected discounted utility of receiving a random wage draw, a value we denote by Q. We note that only unemployed agents receive random wage draws; thus, Q may be interpreted as the value associated with being unemployed. In Section 3.2 we demonstrate that optimal behavior can be viewed as a special case, i.e. $Q = Q^*$. In Section 3.3 we show that under a natural updating rule the agent's perceptions Q converge over time to Q^*, i.e. agents learn over time to make optimal decisions. Finally, Section 3.4 discusses the implications of altering the learning rule to discount older data.

3.1 Decision-making under subjective beliefs

Denote by Q the agent's current perceived (i.e. subjective) value of receiving a random wage offer drawn from F. Let $V(w, Q)$ denote the perceived value of a wage offer w. With this notation we assume that boundedly optimal agents with beliefs Q make decisions by solving the following optimization problem

$$V(w, Q) = \max\left\{U(b) + \beta Q, U(w) + \beta(1-\alpha)V(w, Q) + \beta\alpha Q\right\}. \tag{2}$$

The agent accepts the wage offer w if

$$U(b) + \beta Q < U(w) + \beta(1-\alpha)V(w, Q) + \beta\alpha Q \tag{3}$$

and otherwise rejects the offer.[7] Now observe that if (3) holds then

$$V(w, Q) = U(w) + \beta(1-\alpha)V(w, Q) + \beta\alpha Q, \tag{4}$$

which implies

$$V(w, Q) = \phi U(w) + \beta\alpha\phi Q, \tag{5}$$

where $\phi = (1 - \beta(1-\alpha))^{-1}$, and we note that $0 < \alpha\phi < 1$.

[7] In the probability zero case that the agent is indifferent between accepting the job or remaining unemployed we assume that the agent rejects the offer.

We think of the optimal belief Q^* as difficult to determine, requiring as it does, a complete understanding of the wage distribution as well as the ability to compute fixed points. In contrast, given Q, the determination of $V(w,Q)$ is relatively straightforward: if (3) holds then $V(w,Q)$ is given by (5). The intuition for this equation can be given by rearranging (4) as

$$V(w,Q) = U(w) + \beta\left(V(w,Q) + \alpha\left(Q - V(w,Q)\right)\right).$$

This says that if accepting a job at w is optimal then its value its equal to $U(w)$ plus the discounted expected value in the coming period, which is again $V(w,Q)$ if employment continues, but must be adjusted for the "capital loss" $Q - V(w,Q)$ in value that arises if the agent becomes unemployed, which occurs with probability α.

If instead (3) does not hold, the wage offer is rejected and the agent's present value of utility is simply $U(b) + \beta Q$. We conclude that

$$V(w,Q) = \max\left\{U(b) + \beta Q, \phi U(w) + \beta\alpha\phi Q\right\}. \tag{6}$$

Thus, given perceived Q, decision-making is straightforward based on (6). We now obtain results that characterize the properties of boundedly optimal decision-making based on Q, and in the next Section we relate these results to fully optimal decision-making.

Our first result establishes the existence of a "reservation wage" $\bar{w}$ that depend on beliefs Q. Because this dependency is piece-wise it is useful to define

$$Q_\star = \frac{\phi U(w_\star) - U(b)}{\beta(1-\alpha\phi)}, \text{ where } \star \in \{\min, \max\}.$$

Lemma 1. *There is a continuous, non-decreasing function* $\bar{w}: \mathbb{R} \to [w_{\min}, w_{\max}]$*, which is differentiable on* $(Q_{\min}, Q_{\max})$*, such that* $\bar{w}(Q_\star) = w_\star$ *for* $\star \in \{\min, \max\}$*, and such that*

$$V(w,Q) = \begin{cases} U(b) + \beta Q & \text{if } Q > Q_{\max} \text{ or if } Q \in [Q_{\min}, Q_{\max}] \text{ and } w \leq \bar{w}(Q) \\ \phi U(w) + \beta\alpha\phi Q & \text{if } Q < Q_{\min} \text{ or if } Q \in [Q_{\min}, Q_{\max}] \text{ and } w > \bar{w}(Q) \end{cases}. \tag{7}$$

The proof of this and all results in this Section are in Appendix A. This lemma immediately implies the following proposition characterizing boundedly optimal behavior.

Proposition 1. ***(Boundedly optimal behavior)*** *Given beliefs* Q*, there exists* $\bar{w}(Q) \geq w_{\min}$ *such that the policy* $a = 1$ *if and only if* $w > \bar{w}$ *solves the boundedly optimal agent's problem (2).*

Noting from Lemma 1 that $\bar{w}$ depends on Q and b, we conclude this section with simple comparative statics results with respect to these variables that will be useful in Section 4. Provided that $w_{\min} < \bar{w}(Q,b) < w_{\max}$, $\bar{w}$ is implicitly defined by

$$\phi U\left(\bar{w}(Q,b)\right) + \beta\alpha\phi Q = U(b) + \beta Q. \tag{8}$$

From Assumption A we have that u is C^1 and thus

$$\frac{\partial \bar{w}}{\partial Q} = \frac{\beta(1-\alpha\phi)}{\phi U'\left(\bar{w}(Q,b)\right)} \text{ and } \frac{\partial \bar{w}}{\partial b} = \frac{U'(b)}{\phi U'\left(\bar{w}(Q,b)\right)}, \tag{9}$$

which are both positive provided $U' > 0$.

Below we drop the explicit dependence of $\bar{w}$ on b except when considering cases in which b is changed.

3.2 Optimal beliefs

We now establish a link between optimal decision-making and decisions under subjective beliefs. To this end we define a map $T : \mathbb{R} \to \mathbb{R}$ by

$$T(Q) = E\left(V(\hat{w}, Q)\right) = \int_{w_{\min}}^{w_{\max}} V(\hat{w}, Q) dF(\hat{w}). \tag{10}$$

We interpret $T(Q)$ as the expected value today, induced by beliefs Q and the boundedly optimal reservation wage $\bar{w}(Q)$ from Proposition 1. Lemma A.2 in Appendix A establishes that T is continuous, and is differentiable except at finite number of points, with a positive derivative strictly less than one. As one would expect there is a tight link between the fixed point of this T map and optimal decision making by the agent.

Theorem 1. ***(Optimal Behavior)*** *The expected discounted utility under optimal decision-making of receiving a random wage draw, $Q^* = E\left(V^*\left(\hat{w}\right)\right)$, is the unique fixed point of the T-map (10). The policy $a = 1$ if and only if $w > \bar{w}(Q^*) \equiv w^*$ solves the optimal agent's problem (1).*

This is the standard "reservation wage" result of the McCall search model. However, Theorem 1 comes with the additional interpretation that there exists a belief Q^* about the value of being unemployed such that a boundedly rational agent with beliefs Q^* behaves optimally. The explicit connection between Q^* and the agent's problem (1) arises from the observation $V^*(w) = V(w, Q^*)$, which is established in the proof of Theorem 1. This observation may then be coupled with Proposition 1 to show the equivalence of problems (1) and (2) when $Q = Q^*$.

Finally, it is convenient to adopt assumptions that result in non-trivial optimal decision-making, i.e. in which some wage offers are rejected and other wage offers are accepted: $w_{\min} < w^* < w_{\max}$. The following Proposition characterizes the parameter restrictions consistent with this assumption.

Proposition 2. ***(Non-trivial decision-making)*** *If*

$$\phi\left(U(w_{\min}) - \beta(1-\alpha)\int_{w_{min}}^{w_{max}} U(\hat{w})dF(\hat{w})\right) < U(b) < \phi(1-\beta)(1-\alpha)U(w_{\max}) \tag{11}$$

then $Q_{\min} < Q^* < Q_{\max}$, i.e. $w_{\min} < w^* < w_{\max}$.

We omit the straightforward proof. We remark that when condition (11) holds, the comparative statics result (9) applies to Q^*. Henceforth we assume the following:

Assumption B: $U, b, w_{\min}, w_{\max}, \alpha, \beta$ and F are such that Condition (11) holds.

3.3 Learning When to Say No

We now return to considerations involving boundedly rational agents. Recall that Proposition 1 presents a reservation-wage decision rule that is optimal for given beliefs Q. For agents to learn over time in order to improve their decision-making behavior, it is necessary to update their beliefs as new data become available.[8] We adopt the "anticipated utility" perspective introduced by Kreps (1998), and frequently employed in the adaptive learning literature, in which agents make decisions based on their current beliefs Q, while ignoring the fact that these beliefs will evolve over time.[9]

As just discussed, agents update their beliefs over time as new data become available; however, we observe that if a given agent learned only from their own experience then they would update their beliefs only when they were unemployed. Because this is an implausibly extreme assumption, we introduce a social component to the adaptive learning process: we assume that in each period each agent observes a sample of wage offers received by unemployed workers and uses this sample to revise the perceived value from being unemployed.[10] We denote by $\hat{w}_t^N = \{\hat{w}_t(k)\}_{k=1}^N$ the random sample of N wage realizations. For simplicity we assume that unemployed and employed agents use the same sample size.

Let Q_t be the value, perceived at the start of period t, of being unemployed. Note that Q_t measures the agent's perception of the value of receiving a random wage draw.[11] To update this perception the agent computes the sample mean of $V(\cdot, Q_t)$ based on his sample of wage draws. Since Q_t encodes the information from all previous wage draws, the agent updates his estimate of Q using a weighted average of Q_t with this sample mean. Formally let

$$\hat{T}(\hat{w}_t^N, Q_t) = N^{-1} \sum_{k=1}^{N} V\left(\hat{w}_t(k), Q_t\right) \tag{12}$$

denote the sample mean of $V(\cdot, Q_t)$ based on the sample $\hat{w}_t^N$. The agent is then assumed to update his beliefs at the end of period t according to the algorithm

$$Q_{t+1} = Q_t + \gamma_{t+1}\left(\hat{T}(\hat{w}_t^N, Q_t) - Q_t\right), \tag{13}$$

where $0 < \gamma_{t+1} < 1$ is specified below. Thus the revised estimate of the value of being unemployed Q_{t+1}, which is carried by the agent into the next period, adjusts the previous estimate Q_t to reflect information obtained during period t.[12]

---

[8] Proposition 1 provides optimal decision-making given beliefs under the assumption that the separation rate is known. It would be straightforward to require our agents to also estimate this separation rate. We explore this further in Section 5. Under the assumptions of Theorem 2 asymptotic rationality still obtains.

[9] See, for example Sargent (1999), Preston (2005), Cogley and Sargent (2008).

[10] An alternative approach is through the information frictions literature: see Kennan (2010) and Morales-Jimenez (2017). Comparing the implications of this approach with ours would be of considerable interest.

[11] To be entirely precise, βQ_t is used as the agent's perception of the value in period t of being unemployed and therefore receiving a random wage draw in $t+1$.

[12] The algorithm (13), which is standard in the adaptive learning literature, can be viewed as an implementation of least-squares learning if $\gamma_t = 1/t$. See, for example, Ljung (1977), Marcet and Sargent (1989) and Evans and Honkapohja (2001).

We have presented our algorithm for arbitrary $N \in \mathbb{N}$ for greater generality; however, all of our results hold in particular for $N = 1$, which leads to the useful simplification $\hat{T}(\hat{w}_t, Q_t) = V(\hat{w}_t, Q_t)$, where $\hat{w}_t^1 = \hat{w}_t = \hat{w}_t(1)$. We leverage this simplification by using the calibration $N = 1$ for our later numerical work.

The term $\gamma_t > 0$, known as the gain sequence, is a deterministic sequence that measures the rate at which new information is incorporated into beliefs. Two cases are of particular interest. *Constant-gain* learning sets $\gamma_t = \gamma < 1$, which implies that agents discount past data geometrically at rate $1-\gamma$. This is often used when there is the possibility of structural change, and is discussed in Section 3.4. Under *decreasing-gain* learning $\gamma_t \to 0$ at a rate typically assumed to be consistent with assumption C below. Decreasing gain is often assumed in a stationary environment, and here provides for the possibility of convergence over time to optimal beliefs. The following assumption is made when decreasing gain is employed.

Assumption C: The gain sequence $\gamma_t > 0$ satisfies

$$\sum_{t \geq 0} \gamma_t = \infty \text{ and } \sum_{t \geq 0} \gamma_t^2 < \infty.$$

The conditions identified in Assumption C are entirely standard in the literature: see Bray and Savin (1986), Marcet and Sargent (1989) and Evans and Honkapohja (2001). The first condition guarantees that new information is provided sufficient weight to avoid spurious convergence and the second condition ensures appropriate convergence obtains. A natural example is $\gamma_t = t^{-1}$ in which all observations from $\{1, \dots, t\}$ receive equal weight. For the case at hand this replicates a simple average.

The following theorem is the main theoretical result of our paper.

Theorem 2. ***(Asymptotic rationality)*** *For any Q_0, under Assumptions A, B and C, $Q_t \to Q^*$ almost surely.*

Theorem 2 establishes that in a stationary environment boundedly optimal agents will learn over time to make fully optimal decisions. Section 4 explores the implications of learning when there are structural changes.

A particular limiting case can help highlight the properties of this algorithm further. Consider the algorithm as $N \to \infty$. In this case, we can consider the agent as having full knowledge of the wage offer distribution. In fact, in this case we have $\hat{T}(\hat{w}_t^N, Q_t) \to T(Q_t)$, where T is the T-map defined in the previous section. The evolution of beliefs is then given by

$$Q_{t+1} = Q_t + \gamma_t(T(Q_t) - Q_t).$$

Even though the agent has full knowledge of the offer distribution, she still needs to learn how to behave optimally and she updates beliefs in a deterministic manner over time. In this case, the square summability of the gain sequence is not needed for asymptotic convergence. In fact, the algorithm with $\gamma_t = 1$ is equivalent to iterating the agent's Bellman equation over time. We note that, in this case, convergence obtains for all $\gamma_t = \gamma \in (0, 1]$.

We view the learning rule (13) as providing a particularly appealing model of bounded rationality. Its most striking aspect is its simplicity, i.e. its low demands on agents' cognition and memory: agents only need to store one piece of information each period, namely the current value of Q; and its update requires only a simple average of very simple computations based on observed current wages, i.e. the comparisons of employment and unemployment values identified in equation (6). In particular, it is not necessary for agents to track the history of wages or even perform any sophisticated statistical analysis.

The learning rule addresses several features that make dynamic optimization challenging. For finite N the wage distribution can only be learned asymptotically. If N is large the sample can be viewed as revealing all needed information about the wage distribution; however, computing optimal beliefs Q^* still requires a great deal of sophistication, as noted above. Our learning rule does not require such sophistication; and further, it applies even if agents observe only one wage offer per period.

3.4 Learning with constant gain

Our central analytical result above addresses the case of decreasing gain, which can be viewed as natural if the environment is stationary *and* agents perceive it to be stationary. However, realistic economies are *always* susceptible to structural change, and thus present environments in which our results do not formally apply. We view Theorem 2 as providing credence to the use of otherwise identical algorithms in which agents discount older data in order remain alert to structural change.[13]

Constant gain learning (CGL) algorithms are of particular interest. Analytic results for the small constant-gain limit can be developed using stochastic approximation techniques, though they are local in nature. Informally, for our model, these results imply that for large t and small γ the distribution of beliefs Q is approximately normally distributed around Q^* with variance proportional to γ. Since these formal results are small gain limits, in practice stochastic simulations are used to study systems under CGL.

Adaptive learning in applied models of macroeconomics and finance is often implemented using CGL, in part because it allows the modeler to account for structural change in a natural way, and in part because, in the absence of structural change, it provides for perpetual, *stationary* learning dynamics.[14] The prominent, early work of Sargent (1999) emphasized the role of CGL in monetary policy issues; Bullard and Eusepi (2005) used CGL to understand how the economy reacts to unexpected changes in trend productivity growth; and Milani (2007) estimated DSGE models under the assumption that agents are constant-gain learners. These are just a few examples within a broad literatures in macroeconomics and finance.

[13]This motivation has a venerable history in the literature on stochastic recursive algorithms: see Ljung and Söderström (1983).

[14]In particular, under decreasing gain the learning dynamics vanish asymptotically.

4 Structural Change and Transition Dynamics

We now take our model as populated by many agents; this allows for analysis of interesting aggregates like the unemployment rate. Our application in Section 5 will consider the cross-sectional implications of BR in our model. Here we study the implications for BR following a simple policy change, specifically an unexpected change to unemployment compensation.

We begin with the study of comparative static and dynamic responses analytically under rational expectations and optimal decision-making. We proceed by first studying these dynamics under a fixed belief, Q, and then setting beliefs at the rationally optimal level Q^*. In this manner, we decompose the responses into two terms: the direct effects hold beliefs fixed while the indirect effects come through changes in Q^*. We then show numerically that these two terms allow us to understand the dynamics of the BR agents.

4.1 Preliminaries

We begin by defining the variables of interest. Unemployment and duration, which will be carefully defined below, depend inversely on what we call the "hazard" rate h of leaving unemployment, i.e. the probability per period of an unemployed agent becoming employed. Given beliefs Q and benefits level b the hazard rate is

$$h = h\,(Q, b) = (1 - \alpha)(1 - F(\bar{w}\,(Q, b))).$$

For a given fixed Q, the *perceived* duration δ is defined to be the expected number of periods of consecutive unemployment conditional on being newly unemployed, and a straightforward computation provides that $\delta = h^{-1}$. Finally, we define u to be the corresponding ergodic probability of ending the period unemployed; thus

$$u = u\,(Q, b) = \frac{\alpha}{1 - (1 - \alpha)F(\bar{w}\,(Q, b))} = \frac{\alpha}{h\,(Q, b) + \alpha}. \tag{14}$$

The variables δ and u as just defined capture individual-level behavior; however, they can be connected to aggregate counterparts. If we imagine an economy with a continuum of agents with homogeneous beliefs Q that are held fixed for all t then $u\,(Q, b)$ is the steady-state unemployment rate for the economy. In particular, if $Q = Q^*(b)$ then $u^* = u\,(Q^*(b), b)$ is the long-run employment rate in an economy populated with a continuum of rational agents.

4.2 Comparative statics

We now assume our McCall model is populated by a continuum of rational agents, and consider comparative statics associated with steady-state behavior. The rational counterparts h^*, δ^* (and u^*, as already noted above) are obtained from the above definitions by setting $Q = Q^* = Q^*(b)$. To compute our comparative statics, we continue to adopt Assumption B so that an interior solution exists; it follows from equation (9) that $\frac{\partial \bar{w}}{\partial Q}$ and $\frac{\partial \bar{w}}{\partial b}$ are positive.

In what follows we will compute several derivatives with respect to b. When differentiating any variable other than $Q^* = Q^*(b)$, the symbol "∂" will indicate that beliefs Q are taken as fixed and the symbol "d" will indicate that beliefs Q will vary in accordance with optimality, i.e. $Q^* = Q^*(b)$. Finally, we note direct analysis provides that $\frac{\partial Q^*}{\partial b}$ and $\frac{\partial w^*}{\partial b}$ are positive.

The following Lemma decomposes the comparative statics of the hazard rate into the direct and indirect effects mentioned earlier.

Lemma 2 (***Direct and indirect effects*).** *Under Assumption B,*

$$\frac{dh^*}{db} \equiv \frac{\partial h}{\partial b} + \frac{\partial h}{\partial Q}\frac{\partial Q^*}{\partial b} < 0,$$

with both $\frac{\partial h}{\partial b} < 0$ *and* $\frac{\partial h}{\partial Q}\frac{\partial Q^*}{\partial b} < 0$.

Proofs of the results of this section may be found in Appendix B.

Lemma 2 tells us that the hazard rate of leaving unemployment is decreasing in unemployment benefits. This effect is decomposed into direct effect and indirect effects. $\frac{\partial h}{\partial b}$ captures the direct effect: even if agents do not update their beliefs they will still react to an increase in benefits by raising their reservation wage. Lemma 2 tells us that a rational agent would respond even further by taking into account that higher unemployment benefits also raise the value of Q^*. This is the indirect effect. While the hazard rate for the rational agents exhibits no dynamics, i.e. it jumps from the old steady-state value to the new one, under learning the hazard rate evolves over time as beliefs Q are updated. For this reason, indirect effects are not initially incorporated into the boundedly rational agents' hazard rate.

The inverse relationship between the hazard rate and both the unemployment rate and duration yields the following Proposition.

Proposition 3 (***Comparative statics*).** *Under Assumption B,* $\frac{du^*}{db} > 0$ *and* $\frac{d\delta^*}{db} > 0$.

4.3 Comparative dynamics under rationality

With rational agents, only the unemployment rate experiences non-trivial transition dynamics; the hazard rate and duration for the newly unemployed simply jump to their new steady state levels. The same would be true for boundedly optimal agents if their beliefs Q were constant over time; however, under learning the evolution over time of beliefs induces transition dynamics in the hazard rate.

To examine unemployment dynamics it is helpful to define the notion of a "quit." We say that an agent employed in time $t-1$ quits in time t, and thereby becomes unemployed, if his wage in time $t-1$ is less that w_t^*. Here the t subscript allows for variations in the optimal reservation wage induced by structural change. We observe that quits can only occur when a structural change between periods $t-1$ and t results in $w_t^* > w_{t-1}^*$. Therefore, to simplify our analysis we will assume that a structure change at time 0 occurs only after a long period of stability so that the economy has reached a long run steady state. We focus on the dynamics of rational agents but, as in the previous section, we decompose changes in unemployment into direct and indirect effects to shed light on the unemployment dynamics with boundedly rational agents.

Let w_{-1} denote the wage of individual drawn randomly in period -1 from the pool of employed individuals. The probability that this individual quits in period 0 is given by

$$q_0 = q\left(w_0^*\right) = \frac{\max\left\{0, F(w_0^*) - F(w_{-1}^*)\right\}}{1 - F(w_{-1}^*)},$$

where we have exploited that the long run distribution of wages will be the distribution of wage offers, $F(\cdot)$, truncated at the reservation wage w_{-1}^*.

Interpreted cross-sectionally, q_0 is the proportion of agents employed in time -1 who quit in time 0. Now let u_t be the proportion of agents who are unemployed in period t . Noting that $1-u$ is the proportion of employed agents, and that $1-h$ is the probability that an unemployed agent remains unemployed, the dynamics of u_t may be written

$$u_t = (1-h_t)\, u_{t-1} + (\alpha + (1-\alpha)q_t)(1-u_{t-1}). \tag{15}$$

We can use equation (15) to assess the impact response of unemployment driven by a change in b, assuming that the economy is initially in steady state. Differentiation of (15) at $t=0$ yields

$$\frac{du_0}{db} = -u^* \frac{dh^*}{db} + (1-\alpha)(1-u^*)\frac{dq}{dw^*} \cdot \frac{dw^*}{db}. \tag{16}$$

It is important to emphasize here that we are differentiating q at the previous steady state reservation wage, and while q is not differentiable at this point it is Gateaux differentiable with

$$dq = \begin{cases} \frac{dF(w_{-1}^*)}{1-F(w_{-1}^*)} dw^* & \text{if } dw^* \geq 0 \\ 0 & \text{if } dw^* < 0 \end{cases}.$$

These considerations lead to the following result:

Proposition 4 (***Comparative dynamics: impact response under RE).***

$$du_0 = \begin{cases} \frac{1}{u^*}\frac{du^*}{db} db & \textit{if } db \geq 0 \\ \frac{\alpha}{u^*}\frac{du^*}{db} db & \textit{if } db < 0. \end{cases}$$

As $\frac{1}{u^*}$ is much larger than one we can conclude that an unexpected increase in unemployment benefits will result in an initial spike in unemployment many times of that of the increase in steady state unemployment. On the other hand α/u^* is necessarily less than one, which implies that a decrease in benefits will result in a fall in unemployment smaller than the fall in steady state.

Just as with the hazard rate, the change in steady-state unemployment can be decomposed into direct and indirect effects:

$$\frac{du^*}{db} \equiv \frac{\partial u^*}{\partial b} + \frac{\partial u^*}{\partial Q}\frac{\partial Q^*}{\partial b}. \tag{17}$$

Further, notice that these effects are both positive. This equation may be used in conjunction with Proposition 4 to decompose the impact response du_0 into direct and indirect effects. With boundedly rational agents, indirect effects are not present on impact, so we can expect the impact response of the BR model to be smaller.

4.4 Comparative dynamics under bounded rationality

We now use numerical methods to study comparative dynamics in our model. While the direct and indirect effects of a benefits change are realized immediately and simultaneously under RE, our boundedly rational agents only learn over time the effect of the change on Q, leading to a delay in the realization of indirect effects.

For the simulations in this Section and the next, we use a baseline calibration close to Hornstein et al (2011). The time unit is months and the discount rate is $\beta = 0.996$, in accordance with an annual interest rate of 5.0%. The monthly separation rate, α, is set to be 3%. We set $b = \$31,200$ to target a replacement rate of 41%. Finally, utility is $CRRA$ with risk aversion parameter $\sigma > 0$, and we set σ to 3.25 to match a job-finding rate of 43%.

The exogenous wage distribution is taken to be lognormal[15] with parameters $\mu, s > 0$, yielding a median wage of e^{μ} and variance $e^{2\mu+s^2}\left(e^{s^2}-1\right)$. We set $\mu = 11.0$, $s = 0.25$. The value of μ corresponds to a median household wage of approximately \$60,000, close to the US value in dollars in 2016. For the choice of s, what is relevant for our model is the distribution of wage draws faced by the individual agent, i.e. not a measure of the population wage distribution.[16] At our baseline value, the interquartile income range is \$50,583 to \$70,871. The lowest decile ends at $w = \$43,460$ and the highest decile begins at $w = \$82,486$. We interpret our calibration as capturing the experience of an individual interacting in a local labor market populated by individuals with similar characteristics.

The simulations in this Section are conducted with a constant gain of $\gamma = 0.1$ and with $N = 1$. Our choice of γ is relatively high, for example compared to calibrated value in Section 5. This is natural given that here we are considering agent behavior in the presence of known structural change, which requires a higher gain in order to more rapidly adapt to the new environment.

Figure 1 presents a simulation of an economy with 100,000 agents who experience an unexpected 10% increase in benefits in period 50. The horizontal (red) dotted lines represent the pre-shock steady-state values values and the horizontal (blue) dashed lines represent post-shock steady-state values.[17]

For fixed beliefs Q, an increase in benefits db results in an increase in the instantaneous return $U'(b)\,db$ to being unemployed, thereby raising the reservation wage. This the direct effect emphasized in the previous Section. The corresponding indirect effect of a rise in

[15] Although lognormal does not impose $w_{\min} > 0$ or $w_{\max} < \infty$, this is numerically indistinguishable from setting $w_{\min}$ small and $w_{\max}$ large.

[16] Our value for s is broadly consistent with the literature. For example, p. 576 of Greene (2012) using a pooled LS estimate of a log wage equation controlling for a number of individual specific characteristics, obtains a residual variance of 0.146, i.e. $s = 0.382$. Krueger et al. (2016), estimate a log-labor earnings process with persistent and transitory shock. They find that the variance of the transitory shocks, which are the shocks more relevant for our model, is 0.0522, i.e. $s = 0.23$. The qualitative features of the simulations are robust to values of s across this range.

[17] All simulations are initialized by providing boundedly rational agents with beliefs in a small neighborhood of the optimal value of Q, and with the percent of agents identified as unemployed corresponding to the rational model's steady-state unemployment rate. To eliminate transient dynamics the model is run for a large number of periods before our simulation begins.

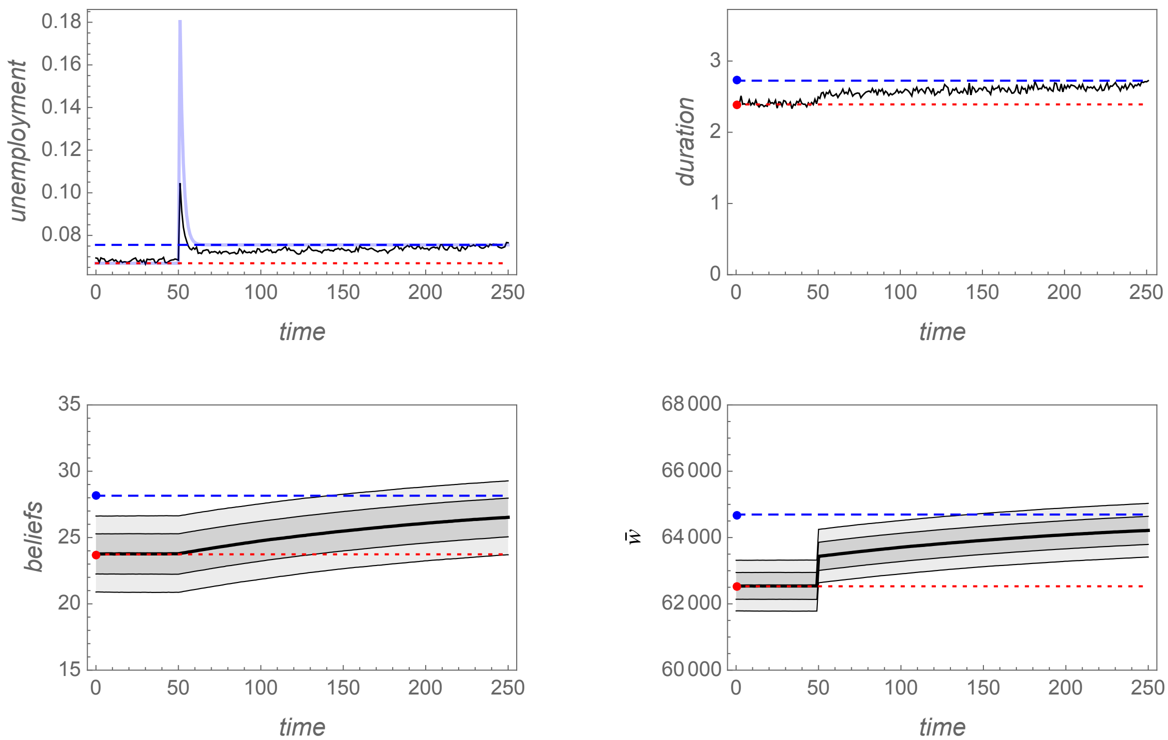

Figure 1: Simulations of the RE and BR models given a 10% increase in benefits in period 50. Dotted and dashed lines indicate pre- and post-shock steady states. The top panels show exaggerated impact responses of unemployment and duration under RE relative to BR. The bottom panels provide the trajectories of mean, quartile, and outer-decile values of agents' beliefs and associated reservation wages.

benefits is that it also raises the optimal present value Q^* of being unemployed. For the rational agents both effects are instantaneous, whereas for the boundedly optimal agents, the initial impact on the reservation wage is only through the increase in the instantaneous return, with the impact from changes in Q developing over time.[18]

The decomposition into direct and indirect effects is evidenced in the bottom-right panel of Figure 1. The dark thicker line represents the cross-sectional mean reservation wage, with the thinner lines identifying quartile and outer-decile values. For the first 50 periods reservation wages are distributed around the pre-shock RE value – the distribution reflects the evolving beliefs of different agents as determined by their idiosyncratic sample draws. At time $t = 50$ there is a sharp increase in the distribution of reservation wages due to the rise in b. Subsequently over time, as evidenced in the bottom-left panel, agents' beliefs converge to a distribution around the new optimal value of Q, and the distribution of corresponding reservation wages evolves to a distribution around the new RE value.

[18]The delayed onset of the indirect effect reflects in part the simplicity of our learning rule. A more sophisticated agent who has available the history of wage offers could, in principle, learn more quickly.

Turning to unemployment duration, the time series presented in the top-right panel gives, at each point in time, the realized cross-agent average, conditional on being newly unemployed, of the number of periods until the agent is next employed. For rational agents the expected unemployment duration for the newly unemployed jumps to the new steady-state duration level, whereas, because their $\bar{w}$ does not fully adjust immediately, boundedly rational agents are initially more likely than their rational counterparts to take jobs, leading to a more gradual adjustment of the duration.

Finally, we consider the unemployment time series, in the top-left panel, which dramatically illustrates the discrepancy in behavior of the optimal and boundedly optimal agents at the time of the policy change. As noted in the previous Section, an increase in benefits leads to an increase in the rational-agent steady-state unemployment rate. The translucent (blue) path identifies the unemployment rate associated with rational model. This time series exhibits a very large spike at the time of the shock, which reflects the impact response identified in the discussion following Proposition 4. This spike can be explained by the behavior of the associated reservation wage: because optimal agents experience both the direct and indirect effects at the instant of the change in b, their reservation wage rises immediately to the new optimal level, which causes a dramatic rise in unemployment resulting from previously employed agents not accepting their wage offers.

The initial spike for boundedly optimal agents is similarly explained, but is muted by the failure of the indirect effect to materialize immediately. In addition, the time path of unemployment overshoots the new steady state before gradually converging to it from below. This behavior reflects the moderated impact response of $\bar{w}$ relative to the rational case, together with the gradual rise Q, which leads to an associated gradual rise in agents' reservation wages.

A decrease in the benefit rate induces less drama, thus we dispense with the figure: the unemployment rate falls, as one would expect, but there is no overshooting spike in either the rational or boundedly rational case. This is easily understood: when benefit rates rise, employed agents with low wages immediately quit their jobs to capture the increased benefit of being unemployed; however, when benefits fall, all employed agents have increased incentives to retain their jobs and unemployed agents are willing to accept lower wages, but not to the extent that overshooting is implied.

5 Application: frictional wage dispersion

To further illustrate the potential of our approach, we consider a prominent empirical puzzle in the labor-search literature that concerns the size of frictional wage dispersions generated by search models. HKV illustrate this puzzle by constructing a measure of frictional wage dispersion – the mean-min (Mm) ratio – defined as the ratio of the mean of accepted wage offers relative to the minimum of accepted wage offers. HKV document how standard search models fail to generate the Mm ratios observed in the data. In this section we begin by presenting the HKV puzzle in the context of our model populated with rational agents; and we then examine whether the introduction of bounded rationality can serve to mitigate or resolve the puzzle.

HKV derive a simple formula for the Mm ratio using a continuous-time model with risk neutral agents. To derive the analogous formula for our model, we assume rationality and that $\sigma = 0$ (though for simulations we return our calibrated value of σ). HKV assume the level of unemployment benefits are calibrated to be a fraction of the mean observed wage. Letting $\rho \in (0, 1)$ be this proportion, we may write benefits as $\rho E\left(w|w \geq w^*\right)$. Together with the reservation-wage determination equation (8), this benefits structure implies

$$\phi w^* = \rho E\left(w|w \geq w^*\right) + \beta(1-\alpha\phi)Q^*, \tag{18}$$

where we recall that $\phi = (1-\beta(1-\alpha))^{-1}$. Coupling equation (18) with the REE assumption that $Q^* = T(Q^*)$, and letting $\lambda \equiv 1 - F(w^*)$ be the *job-finding rate*, some algebra (provided in Appendix C) yields

$$Mm \equiv \frac{E\left(w|w \geq w^*\right)}{w^*} = \frac{\Delta\lambda + 1}{\Delta\lambda + \rho}, \text{ where } \Delta = \frac{\beta(1-\alpha)}{1-\beta(1-\alpha)}. \tag{19}$$

This is our discrete-time version of the formula derived by HKV.

Using US data, HKV found the empirical range of the Mm ratio to be 1.5 to 2, while using their formula they found $Mm = 1.046$, thus establishing their puzzle. The puzzle persists in our model under rationality: using HKV's calibration, our formula (19) gives $Mm = 1.047$; and, under our calibration with higher risk aversion, it yields $Mm = 1.2$, which is well outside the empirical range HKV find in the data.

Intuitively, the HKV puzzle may be viewed as follows: in the US data, wage dispersion is high and unemployment duration is low. In search models, high wage dispersion induces high unemployment duration: if wage dispersion is high then the potential to eventually get a high wage draw encourages unemployed agents to be selective when considering whether to accept a given wage; this selectively results in longer unemployment spells.

A revision of the formula (19) clarifies this intuition by highlighting an inherent tradeoff between short duration and high dispersion. Let $\delta = ((1-\alpha)\lambda)^{-1}$ be the model-implied unemployment duration. Then

$$\delta = \frac{\Delta}{(1-\alpha)}\left(\frac{Mm - 1}{1 - \rho Mm}\right). \tag{20}$$

We note that, from (19), $Mm > 1 > \rho Mm$ so that $\delta > 0$. A simple computation shows $\delta_{Mm} > 0$, which is the HKV finding that search models predicate a rise in duration resulting from a rise in wage dispersion.

The failure of the standard search model to simultaneously produce high wage dispersion and short unemployment duration is a reflection of the sophisticated, fully-informed behavior of its rational agents. Agents with potential misperceptions about the value of search may make decisions that result in more realistic aggregate outcomes. Our learning framework allows us to evaluate the plausibility of this mechanism.

Our framework thus far has focused attention on perceptions of the wage distribution through the sufficient statistic for decision-making, Q. However it is also natural to assume that agents do not know the other unobserved feature of model, the separation probability α, and instead learn about it through experience. Furthermore, it is plausible that misperceptions about α may serve to mitigate the HKV puzzle: a higher perceived separation rate results in a lower reservation wage by decreasing the value of waiting for a better offer, and a lower reservation wage results in shorter unemployment spells. If the economy is populated with learning agents who have heterogeneous estimates of the separation rate then the economy's mean accepted wage can be similar to that of the rational model while the minimum reservation may be much lower, thus producing a higher Mm ratio.[19]

To examine this matter quantitatively, we modify our model by assuming agents now estimate α as well as Q, and base their decisions on these estimates. If α_t and Q_t are the estimates of α and Q held at the beginning of period t then the agent assesses a wage offer $\hat{w}_t$ via

$$V(\hat{w}_t, Q_t, \alpha_t) = \max\left\{U(b) + \beta Q_t, \phi_t U(\hat{w}_t) + \beta\alpha_t\phi_t Q_t\right\},$$

and sets her reservation wage $\bar{w}_t = \bar{w}(Q_t, \alpha_t)$ to satisfy

$$U(b) + \beta Q_t = \phi_t U(\bar{w}_t) + \beta\alpha_t\phi_t Q_t, \tag{21}$$

where here $\phi_t = (1 - \beta(1 - \alpha_t))^{-1}$. With these assumptions and our benchmark of one observation per month of the random wage draw, $\hat{w}_t$, the updating equation for Q becomes

$$Q_{t+1} = Q_t + \gamma_{t+1}\left(V(\hat{w}_t, Q_t, \alpha_t) - Q_t\right).$$

We observe that the determination of Q_{t+1} and $\bar{w}_t$ depend on the agent's estimate of the separation rate α_t.

We assume that an agent learns about α based on her own job-separation experience, and thus updates her beliefs only if she continues to accept her wage offer (or, if unemployed, accepts a new wage offer), and thus is in a position to observe whether or not she is fired. To operationalize this learning mechanism, recall from equation (1) that s_t is the Bernoulli indicator of job destruction, i.e. $s_t = 1$ if a job separation occurs, and zero otherwise. When updated, the agent's estimate of α evolves according to the recursion

$$\alpha_{t+1} = \alpha_t + \gamma^{\alpha}_{t+1}\left(s_t - \alpha_t\right), \tag{22}$$

otherwise $\alpha_{t+1} = \alpha_t$. We note that Theorem 2 can be modified to allow for learning about the separation rate using the recursion (22). In particular, if agents use decreasing gain then their decision-making is asymptotically optimal.

In line with Section 3.4, we assume agents use constant gain learning algorithms. Algorithms of this type are widely used in the applied literature, and there are several distinct justifications for this.

1. **Structural change and model misspecification.** As is well-known and easy to see, constant gain algorithms place more weight on current data while geometrically

[19]The possibility that wage dispersion arises from heterogenous values of search was first explored by Albrecht and Axell (1984), in the context of rational agents with heterogenous benefits levels.

discounting older data. This feature induces several statistical advantages; in particular, as emphasized by Williams (2019), CGL algorithms are "…known to work well in non-stationary environments, and are good predictors even when the underlying model is misspecified."

2. **Perception of potential structural change.** Constant gain learning provides agents a mechanism by which they can protect against possible future structural change of a type not yet experienced. This behavior can be viewed as form of robust decision-making resulting from the perception of possible structural change, real or imagined. Evans, Honkapohja and Williams (2010) write, "…if the correct specification of the model is not known, then one may want to choose an estimator that is robust in that it performs well across a range of alternatives," and they proceed to demonstrate that CGL algorithms can provide just such a robust optimal estimator.[20]

3. **Availability bias.** This term from the behavioral economics literature refers to the tendency of agents to weight more heavily experiences that readily come to mind: if they easily remember it, it must be important. For a discussion, see example, Kahneman (2011). The manifestation within the context of our model is for agents to overweight their most recent employment experience.

We anchor our calibration strategy to these justifications.

To calibrate the gain for the α-recursion, we start by examining the first of these justifications using data on monthly separation rates (employment to non-employment) for the US from 1994:01 – 2019:08.[21] For privacy reasons, the data are unavailable from 1995:05 – 1995:08. We exploit this by using the data from 1994:01 – 1995:04 to initialize our estimation procedure, which is then applied to the data from 1995:09 – 2019:08. These data allow us to use simulations to compute the *mean-square error (MSE) gain curve*. To this end, we let $\hat{\alpha}_t$ be the period-t empirical separation rate (obtained directly from the data), and fix a finite number of agents I. We construct independent, boolean-valued random variables $\{s_{it}\}$, where $i \in \{1, \ldots, I\}$ and t varies across the sample period, and where $s_{it} = 1$ with probability $\hat{\alpha}_t$. We interpret s_{it} as the separation outcome for agent i in period t, used to update her estimate α_{it}.

Next, fix $i \in \{1, \ldots, I\}$ and $\gamma_\alpha \in [0, 1]$, and for $t \geq$ 1995:09, set

$$\alpha_{it+1} = \alpha_{it} + \gamma_\alpha \left(s_{it} - \alpha_{it} \right). \tag{23}$$

To initialize (23) we must specify α_{it_0} for $t_0 =$ 1995:08. This done by setting it equal to the mean of a realization of s_{it} for $t =$ 1994:01 – 1995:04. The MSE for agent i associated with gain γ_α is $MSE_i(\gamma_\alpha) = (288)^{-1} \sum_t \left(s_{it} - \alpha_{it} \right)^2$. The *MSE gain curve* is the average MSE across agents as a function of the (common) gain:

$$MSE(\gamma_\alpha) = \frac{1}{I} \sum_i MSE_i(\gamma_\alpha), \tag{24}$$

[20] See also, for example, Sargent (1999), Hommes and Zhu (2014) and Cho and Kasa (2004).

[21] Data taken from https://www.federalreserve.gov/econresdata/researchdata/feds200434.html

and the optimal gain, γ_α^*, is the value of $\gamma_\alpha \in [0,1]$ that minimizes MSE as given by (24). Figure 2 plots this curve with $I = 10^5$. The convex shape reflects the well-known tradeoff between tracking and filtering, emphasized by, e.g. Ljung and Söderström (1983): a higher gain enables rapid responses to structural change, while a lower gain averages out random disturbances.

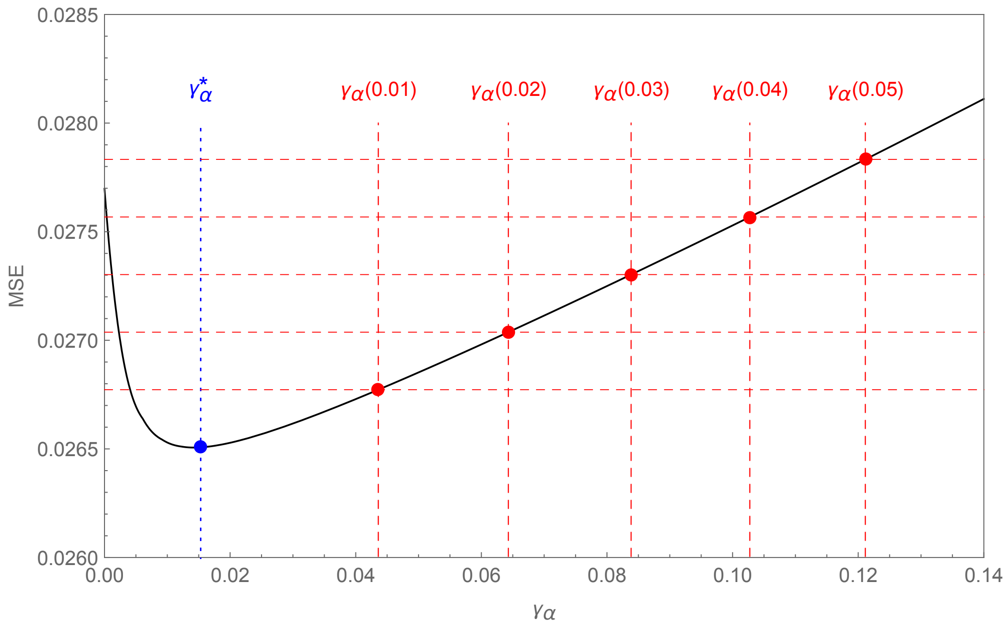

Figure 2: MSE gain curve. The vertical (red) dashed lines, labeled $\gamma_\alpha(p)$, identify the upper bound on the gains given tolerance level p.

The (blue) vertical dotted line identifies the estimated optimal gain $\gamma_\alpha^* \approx 0.0165$, which we take as reflecting considerations identified in item one: structural change and model misspecification. We note that the MSE gain curve indicates there is a broad range of gains compatible with the data. For example, Marcet and Nicolini (2003) argue that it is reasonable for agents to choose any gain that yields an MSE within a "small" increment, or *tolerance*, here measured as a percent increase, above the minimum achievable MSE. The horizontal (red) dashed lines in Figure 2 mark the MSE upper bounds corresponding to tolerances $p \in \{0.01, 0.02, 0.03, 0.04, 0.05\}$; their intersections with the MSE gain curve, identified by the vertical (red) dashed lines and labeled $\gamma_\alpha(p)$, give the maximum gains consistent with these tolerances.

Based on this figure, a tolerance level of 2%, which we view as reasonable, justifies gains as high as approximately 0.064 as being consistent with the data. However, as already emphasized, this procedure is only designed to address the presence of structural change and model misspecification. The additional justifications for the use of constant gain – items 2 and 3 above – suggest that we should select a calibrated value that is at the upper end of the gains identified as reasonable using this narrow statistical criterion.

Turning now to the gain for Q, we have no data to construct a corresponding estimate, though it would be natural to assume agents use the same gain for each algorithm. We will show below that estimates of the Mm ratio under bounded rationality are not sensitive to the calibration of the Q-gain. With these considerations in mind we choose 0.05 to be the benchmark calibrated value for both gains.

To evaluate the Mm ratio in the boundedly rational model we first simulate the stationary joint distribution of beliefs, wages and unemployment, and then compute the ratio of the mean wage to the minimum wage within the ergodic distribution. Apart from the gains, we continue to use the same parameter values as in the Section 4.[22] This results in a benchmark value for the Mm ratio of 1.53, which is significantly higher that the value of 1.2 observed in the RE model and is consistent with the empirical 1.5 - 2.0 range identified by HKV.

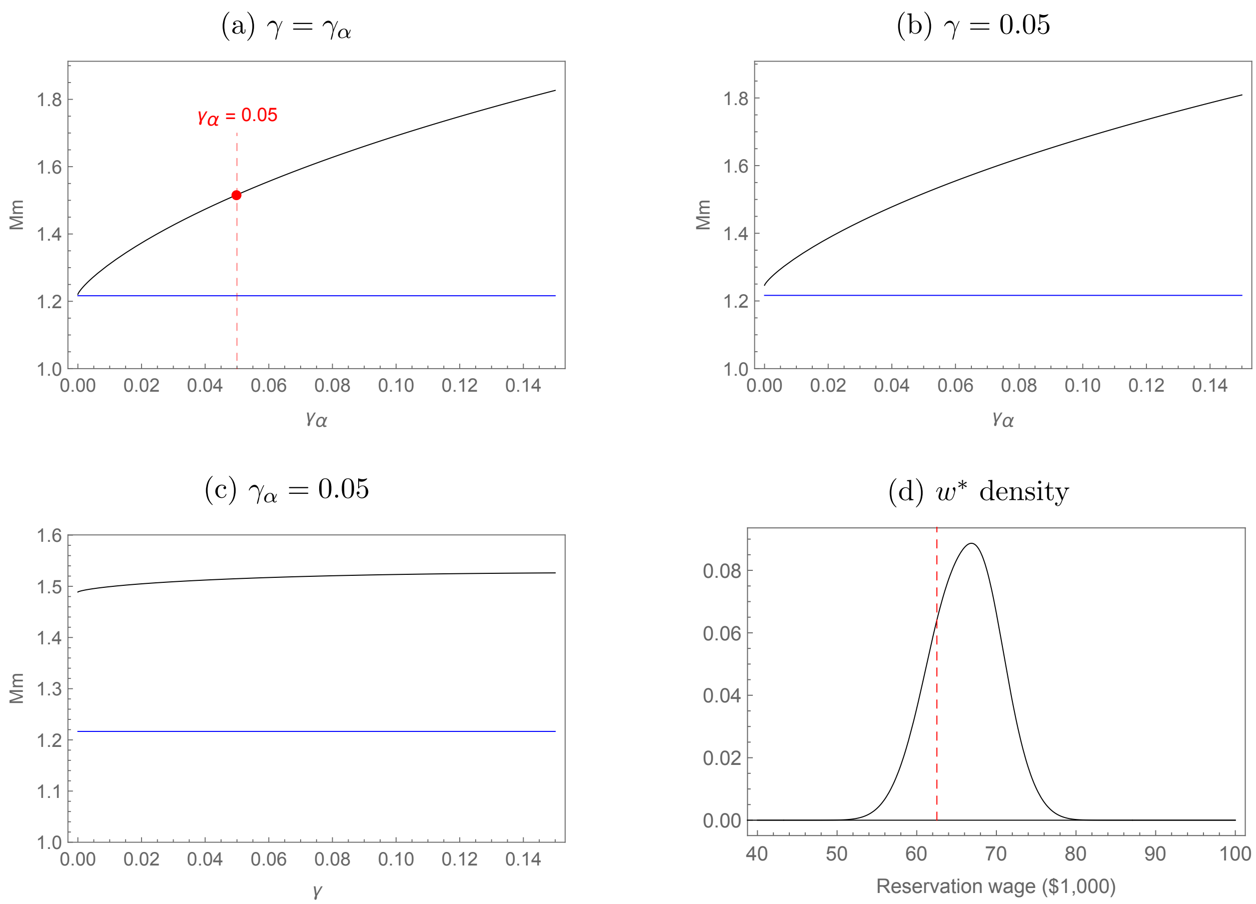

Figure 3: Panel (a) presents the Mm ratio as a function of the common gain $\gamma = \gamma_\alpha$. Panels (b) and (c) allow γ_α and γ, respectively, to vary, holding fixed the companion gain at the benchmark value 0.05. Panel (d) plots the density of the ergodic distribution of reservation wages for $\gamma = \gamma_\alpha = 0.05$.

Figure 3 presents further details of our numerical results, which are computed using

[22]In particular, the separation rate is taken as constant at $\alpha = 0.03$. One could add actual structural change, but for the HKV puzzle the relevant object is the cross-sectional variation in agents' estimates.

the associated ergodic distributions. Panel (a) presents the Mm ratio as a function of the common gain $\gamma = \gamma_\alpha$, with the blue horizontal line recording the Mm ratio under RE. From this panel we see, for example, that a gain of 0.1 yields an Mm ratio of approximately 1.7. Based solely on justification 1 above, this gain is consistent with a tolerance of less that 4%. We remark that extending the range of common gains to 0.90 results in a graph of the same shape: Mm increases monotonically at a decreasing rate, with values approaching 3.0.

Panels (b) and (c) provide information on the relative importance of γ_α and γ. Panel (c), which takes $\gamma_\alpha = 0.05$ and allows γ to vary, shows that our results are robust to a wide range of Q-gains, as anticipated above.[23] Panel (b), which takes $\gamma = 0.05$ and allows γ_α to vary, indicates the potential of learning about α to mitigate or resolve the HKV puzzle. Finally, panel (d) presents the ergodic distribution of reservation wages under the benchmark gains. Intuitively, equation (21) implies that the reservation wage $\bar{w}_t$ depends on α_t both directly and also indirectly through the dependence of $\bar{w}_t$ on Q_t; thus the combined learning dynamics can be expected to induce variation in the reservation wage across agents.

The variation in the reservation wage across agents seen in panel (d) of Figure 3 is in sharp contrast to the distribution of reservation wages in the rational expectations equilibrium, which is a mass point represented by the red line. The learning behavior of the agents induces a non-trivial distribution of beliefs and reservations wages around their rational exceptions values. A direct consequence of this is a larger Mm ratio relative to RE.

6 Conclusions

We consider boundedly optimal behavior in a well known partial-equilibrium model of job search. Boundedly optimal decision-making depends on a univariate sufficient statistic that summarizes the perceived value to the job-seeker of receiving a random wage draw. Following the adaptive learning literature, agents update their perceived values over time based on their current perceptions and observed wage draws. We show that, in a stationary environment and under natural assumptions, this learning algorithm is globally stable: given any initial perception, our boundedly optimal agents learn over time to make optimal decisions.

Our approach shows the potential of bounded rationality to obtain plausible dynamics in response to structural change as a result of the gradual adjustment of beliefs. This is illustrated by our comparative dynamic analysis of a change unemployment benefits.

This implementation of bounded optimality also has the potential to mitigate empirical puzzles in the search literature. When extended to allow for learning about the separation rate, our model results in empirically plausible measures of frictional wage dispersion. While the results of this paper were obtained in a partial equilibrium setting, embedding this learning behavior in a general equilibrium environment would be feasible, and is an important direction for future research.

[23]Extending the range of γ to 0.9 extends the pattern evidenced in panel (c): increasing the Q-gain increases the Mm ratio modestly, with higher gains leading to Mm ratios above 1.6. We also find that if α is assumed to be known to agents, then the Mm ratio is monotonically increasing in γ. However it does not increase enough to resolve the HKV puzzle.

Appendix A: Proofs of results in Section 3

Proof of Lemma 1. First, observe that the agent rejects the wage offer w if and only if

$$\phi U(w) \leq U(b) + \beta(1 - \alpha\phi)Q. \tag{25}$$

The argument is completed by addressing the following three cases:

1. If $Q > Q_{\max}$ then condition (25) always holds; thus $\bar{w}(Q) = w_{\max}$, the agent rejects any offer and receives $U(b) + \beta Q$.

2. If $Q < Q_{\min}$ then condition (25) never holds; thus $\bar{w}(Q) = w_{\min}$, the agent accepts any offer w and receives $\phi U(w) + \beta\alpha\phi Q$.

3. Finally, if $Q_{\min} \leq Q \leq Q_{\max}$ then

$$\phi U(w_{\min}) \leq U(b) + \beta(1 - \alpha\phi)Q \leq \phi U(w_{\max}). \tag{26}$$

   Since $U'(w) > 0$ it follows that for each $Q \in [Q_{\min}, Q_{\max}]$ there is a unique $\bar{w}(Q) \in [w_{\min}, w_{\max}]$ such that

$$\phi U(\bar{w}(Q)) = U(b) + \beta(1 - \alpha\phi)Q,$$

   and further that, in this case, condition (25) holds if and only if $w \leq \bar{w}(Q)$.

   It remains to show that, so defined, $\bar{w}$ is differentiable on $(Q_{\min}, Q_{\max})$. Since u is C^2, by the implicit function theorem, it follows that for each $Q \in (Q_{\min}, Q_{\max})$ there is an open set $U(Q) \subset (Q_{\min}, Q_{\max})$ and a differentiable function $g_Q : U(Q) \to [w_{\min}, w_{\max}]$ such that for all $Q' \in U(Q)$,

$$\phi U(g_Q(Q')) = U(b) + \beta(1 - \alpha\phi)Q',$$

   and, further, by uniqueness of $\bar{w}(Q')$, we may conclude that $\bar{w} = g_Q$ on $U(Q)$. Since the $U(Q)$ cover $(Q_{\min}, Q_{\max})$ the proof is complete. ■

To establish Theorem 1 we need the following technical result:

Lemma A.1. *If $f : \mathbb{R} \to \mathbb{R}$ is continuous, if f is differentiable except at perhaps a finite number of points, and if the derivative of f, when it exists, is positive except at perhaps a finite number of points, then f is strictly increasing.*

Proof: In the context of this proof, we say that x_0 is *anomalous* if either $f'(x_0)$ does not exist or $f'(x_0) \leq 0$. We begin by assuming f has only one anomalous point x_0. Because the derivative is positive for $x \neq x_0$, it suffices to show that if $x < x_0$ then $f(x) < f(x_0)$ and if $x > x_0$ then $f(x) > f(x_0)$. Suppose $x < x_0$. By the mean value theorem applied to $[x, x_0]$, which requires that f be continuous on $[x, x_0]$ and differentiable on (x, x_0), there exists $x^* \in (x, x_0)$ such that

$$\begin{aligned} \frac{f(x_0) - f(x)}{x_0 - x} &= f'(x^*), \text{ or} \\ f(x_0) - f(x) &= f'(x^*)\,(x_0 - x) > 0. \end{aligned}$$

An analogous argument holds if $x > x_0$. Finally, this argument is easily generalized to account for a finite number of anomalous points. ■

The following Lemma, which is referenced in the main text, establishes important properties of the T-map, including an upper bound on its derivative.

Lemma A.2. *The map given by (10) is continuous on* $\mathbb{R}$, *differentiable everywhere except possibly* $Q_{\min}$ *and* $Q_{\max}$, *and* $0 < DT \leq \beta < 1$ *whenever it exists.*

Proof. Using Lemma 1, direct computation yields the following formulation of the T-map:

$$T(Q) = \begin{cases} \alpha\beta\phi Q + \phi \int_{w_{\min}}^{w_{\max}} U(w)dF(w) & \text{if} \quad Q < Q_{\min} \\ \begin{array}{l} (U(b) + \beta Q)\, F(\bar{w}(Q)) + (1 - F(\bar{w}(Q)))\, \beta\alpha\phi Q \\ +\phi \int_{\bar{w}(Q)}^{w_{\max}} U(w)dF(w) \end{array} & \text{if} \quad Q_{\min} \leq Q \leq Q_{\max} \\ U(b) + \beta Q & \text{if} \quad Q > Q_{\max} \end{cases}.$$

Clearly $DT(Q) > 0$. It further follows from Lemma 1 that the map T is continuous on $\mathbb{R}$ and differentiable everywhere except possibly $Q_{\min}$ and $Q_{\max}$. Next we compute an upper bound on DT. If $Q < Q_{min}$ then $DT(Q) = \beta\alpha\phi < \beta$, where the inequality follows from $\alpha\phi \in (0, 1)$. If $Q > Q_{\max}$ then $DT(Q) = \beta$. Finally, if $Q_{\min} < Q < Q_{\max}$ we may compute

$$\begin{aligned} DT(Q) &= (U(b) + \beta Q)\, dF(\bar{w}) \frac{\partial \bar{w}}{\partial Q} + \beta F(\bar{w}) - (\phi U(\bar{w}) + \beta\alpha\phi Q) dF(\bar{w}) \frac{\partial \bar{w}}{\partial Q} \\ &\quad + (1 - F(\bar{w}))\, \beta\alpha\phi \\ &= \beta \left(F(\bar{w}(Q)) + (1 - F(\bar{w}(Q)))\, \alpha\phi \right) < \beta, \end{aligned}$$

where the second equality exploits the definition of $\bar{w}$. ■

Proof of Theorem 1. We begin the proof by establishing that the T-map has a unique fixed point. Let

$$\hat{Q} \leq \min\left\{ \frac{\phi U(w_{\min})}{1 - \alpha\beta\phi}, Q_{\min} \right\}.$$

We claim that $T(\hat{Q}) > \hat{Q}$. Indeed,

$$T(\hat{Q}) = \alpha\beta\phi\hat{Q} + \phi \int_{w_{\min}}^{w_{\max}} U(w)dF(w) > \alpha\beta\phi\hat{Q} + \phi U(w_{\min}) \geq \hat{Q}.$$

Next, let $h : \mathbb{R} \to \mathbb{R}$ be defined as

$$h(Q) = T(\hat{Q}) + \beta(Q - \hat{Q}).$$

We claim $Q \geq \hat{Q}$ implies $h(Q) \geq T(Q)$. Indeed let $H(Q) = h(Q) - T(Q)$. Then H is continuous and $H'(Q) > 0$ except perhaps at $Q_{\min}$ and $Q_{\max}$. Thus by Lemma A.1, H is strictly increasing. The claim follows from the fact that $H(\hat{Q}) = 0$.

Finally let $\check{Q} \equiv (1 - \beta)^{-1} \left(T(\hat{Q}) - \beta\hat{Q} \right)$. Then

$$Q \geq \check{Q} \Rightarrow h(Q) < Q \Rightarrow T(Q) < Q.$$

Thus we have $T(\hat{Q}) > \hat{Q}$ and $T(\check{Q}) < \check{Q}$. Since T is continuous, the existence of a fixed point Q^* is guaranteed by the intermediate value theorem. Finally, let $S(Q) = Q - T(Q)$. Then S is continuous and $S'(Q) > 0$ except perhaps at $Q_{\min}$ and $Q_{\max}$. Thus by Lemma A.1, S is strictly increasing, from which it follows that the fixed point of T is unique.

Now we turn to connecting Q^* to the Bellman functional equation (1), which we repeat here for convenience:

$$\begin{aligned} V(w) &= \max_{a\in\{0,1\}} U\left(c(a,w)\right) + \beta E\left(V(w')|a,w\right) \\ & \quad w' = g(w,a,\hat{w},s). \end{aligned}$$

The binary nature of the choice variable makes this problem accessible. Specifically,

$$\begin{aligned} E\left(V(w')|0,w\right) &= \int V(\hat{w})dF(\hat{w}) \\ E\left(V(w')|1,w\right) &= (1-\alpha)V(w) + \alpha \int V(\hat{w})dF(\hat{w}). \end{aligned}$$

It follows that

$$a = 0 \implies V(w) = U(b) + \beta \int V(\hat{w})dF(\hat{w}) \tag{27}$$

$$a = 1 \implies V(w) = U(w) + \beta(1-\alpha)V(w) + \alpha\beta \int V(\hat{w})dF(\hat{w}), \text{ or}$$

$$a = 1 \implies V(w) = \phi U(w) + \phi\alpha\beta \int V(\hat{w})dF(\hat{w}), \tag{28}$$

where $\phi = (1-\beta(1-\alpha))^{-1}$. We conclude that the Bellman functional equation may be rewritten as

$$V(w) = \max\left\{U(b) + \beta \int V(\hat{w})dF(\hat{w}), \phi U(w) + \phi\alpha\beta \int V(\hat{w})dF(\hat{w})\right\}. \tag{29}$$

Now define $\tilde{Q} = \int V(\hat{w})dF(\hat{w})$, which may be interpreted as the value of having a random draw from the exogenous wage distribution. Then equation (29) becomes

$$V(w) = \max\left\{U(b) + \beta\tilde{Q}, \phi U(w) + \phi\alpha\beta\tilde{Q}\right\}, \tag{30}$$

from which it follows that

$$\tilde{Q} = \int V(w)dF(w) = \int \left(\max\left\{U(b) + \beta\tilde{Q}, \phi U(w) + \phi\alpha\beta\tilde{Q}\right\}\right) dF(w). \tag{31}$$

Using Lemma 1 we may write

$$\begin{aligned} & \int \left(\max\left\{U(b) + \beta\tilde{Q}, \phi U(w) + \phi\alpha\beta\tilde{Q}\right\}\right) dF(w) \\ = \ & (U(b) + \beta\tilde{Q})F\left(\bar{w}\left(\tilde{Q}\right)\right) + \phi \int_{\bar{w}(\tilde{Q})}^{w_{\max}} U(w)dF(w) + \phi\alpha\beta\tilde{Q}\left(1 - F\left(\bar{w}\left(\tilde{Q}\right)\right)\right). \end{aligned}$$

We conclude that equation (31) can be written

$$\tilde{Q} = (U(b) + \beta\tilde{Q})F\left(\bar{w}\left(\tilde{Q}\right)\right) + \phi\int_{\bar{w}(\tilde{Q})}^{w_{\max}} U(w)dF(w) + \phi\alpha\beta\tilde{Q}\left(1 - F\left(\bar{w}\left(\tilde{Q}\right)\right)\right) = T(\tilde{Q}),$$

where the last equality follows from the definition of T. Since the T-map has a unique fixed point Q^*, we conclude that $\tilde{Q} = Q^*$. By equation (30) $\tilde{Q}$, and hence Q^*, uniquely identifies V, the solution to the Bellman system. It follows from equation (6) that $V(w) = V(w, Q^*)$. Finally, Proposition 1 implies $w^* = \bar{w}(Q^*)$. ■

To prove Theorem 2, we require the following technical Lemma:

Lemma A.3. *Suppose that γ_n is a sequence of positive numbers satisfying $\sum_n \gamma_n^2 < \infty$. The following are equivalent:*

a. $\sum_n \gamma_n = \infty$.

b. There exists $\lambda > 0$ such that $\prod_n (1 - \lambda\gamma_n) = 0$.

c. $\prod_n (1 - \lambda\gamma_n) = 0$ for all $\lambda > 0$.

Proof. Denote by $\{\gamma_n^N\}$ the N-tail of $\{\gamma_n\}$, that is, $\gamma_n^N = \gamma_{N+n}$. It will be helpful to observe that since $\gamma_n \to 0$, given $\varepsilon > 0$ there is an $N > 0$ so that $\gamma_n^N < \varepsilon$ for all $n > 0$.
($a \Rightarrow c$). Let $\lambda > 0$ and choose $N_2(\lambda) > 0$ so that $\lambda\gamma_n^{N_2} < 1$ for all $n > 0$. By the concavity of the logarithm, we have that

$$\log\left(1 - \lambda\gamma_n^{N_2}\right) < -\lambda\gamma_n^{N_2}.$$

Now define

$$P_M^{N_2}(\lambda) = \prod_{n=1}^{M}(1 - \lambda\gamma_n^{N_2}),$$

and observe that

$$\log P_M^{N_2}(\lambda) \quad < \quad -\lambda\sum_{n=1}^{M}\gamma_n^{N_2}.$$

Since by assumption $\sum_{n=1}^{\infty}\gamma_n^{N_2} = \infty$, it follows that $\log P_M^{N_2}(\lambda) \to -\infty$, or $P_M^{N_2}(\lambda) \to 0$ as $M \to \infty$. Finally, notice that

$$\prod_{n}^{\infty}(1 - \lambda\gamma_n) = \prod_{n=1}^{N_2-1}(1 - \lambda\gamma_n)\lim_{M\to\infty} P_M^{N_2}(\lambda) = 0,$$

establishing item c.
($b \Rightarrow a$). Suppose $\lambda > 0$ is so that $\prod_n (1 - \lambda\gamma_n) = 0$. Choose $N_1 > 0$ so that $\lambda\gamma_n^{N_1} < 1$ for all $n > 0$. Let $\hat{\gamma} = \sup_n \gamma_n^{N_1} < \lambda^{-1}$, and write

$$\log\left(1 - \lambda\gamma_n^{N_1}\right) = -\lambda\gamma_n^{N_1} + \left(\lambda\gamma_n^{N_1}\right)^2 F(\lambda\gamma_n^{N_1}),$$

where F is a continuous function on $[0, \hat{\gamma}]$. Define

$$P_M^{N_1}(\lambda) = \prod_{n=1}^{M}(1 - \lambda\gamma_n^{N_1}),$$

and observe that

$$\log P_M^{N_1}(\lambda) = -\lambda \sum_{n=1}^{M} \gamma_n^{N_1} + \sum_{n=1}^{M} \left(\lambda\gamma_n^{N_1}\right)^2 F(\lambda\gamma_n^{N_1}).$$

Let

$$\hat{F} = \sup_{\gamma \in [0,\hat{\gamma}]} |F(\lambda\gamma)| < \infty.$$

It follows that

$$\sum_{n=1}^{\infty} \left(\lambda\gamma_n^{N_1}\right)^2 |F(\lambda\gamma_n^{N_1})| \leq \hat{F}\lambda^2 \sum_{n=1}^{\infty} \left(\gamma_n^{N_1}\right)^2 < \infty,$$

and thus there exists $\delta \in \mathbb{R}$ so that

$$\sum_{n=1}^{M} \left(\lambda\gamma_n^{N_1}\right)^2 F(\lambda\gamma_n^{N_1}) \to \delta \text{ as } M \to \infty.$$

By assumption, $P_M^{N_1}(\lambda) \to 0$ and thus $\log P_M^{N_1}(\lambda) \to -\infty$ as $M \to \infty$. Thus

$$\begin{aligned} -\infty &= \lim_{M\to\infty} \log P_M^{N_1}(\lambda) = \lim_{M\to\infty} \left(-\lambda \sum_{n=1}^{M} \gamma_n^{N_1} + \sum_{n=1}^{M} \left(\lambda\gamma_n^{N_1}\right)^2 F(\lambda\gamma_n^{N_1}) \right) \\ &= -\lim_{M\to\infty} \lambda \sum_{n=1}^{M} \gamma_n^{N_1} + \lim_{M\to\infty} \sum_{n=1}^{M} \left(\lambda\gamma_n^{N_1}\right)^2 F(\lambda\gamma_n^{N_1}) \\ &= -\lambda \lim_{M\to\infty} \sum_{n=1}^{M} \gamma_n^{N_1} + \delta. \end{aligned}$$

It follows that

$$\infty = \lim_{M\to\infty} \sum_{n=1}^{M} \gamma_n^{N_1} < \sum_{n=1}^{\infty} \gamma_n,$$

thus establishing item a.
That $(c \Rightarrow b)$ is trivial and the proof is complete. ■

Proof of Theorem 2. Define

$$\overline{Q} = \max\left\{ \frac{\phi U(w_{\max})}{1 - \beta\alpha\phi}, \frac{U(b)}{1-\beta} \right\} \text{ and } \underline{Q} = \max\left\{ \frac{\phi U(w_{\min})}{1 - \beta\alpha\phi}, \frac{U(b)}{1-\beta} \right\},$$

where we note that by Assumption B $\underline{Q} < \overline{Q}$. It is clear from equation (7) of Lemma 1 that $\hat{T}(\overline{Q}, \hat{w}_t^N) < \overline{Q}$ and $\hat{T}(\underline{Q}, \hat{w}_t^N) > \underline{Q}$ for all samples $\hat{w}_t^N$. It follows that for any initial Q

the sequence is eventually in $[\underline{Q}, \overline{Q}]$. Thus, without loss of generality, we can assume that $Q_0 \in [\underline{Q}, \overline{Q}]$ and therefore that $Q_t \in [\underline{Q}, \overline{Q}]$ for all $t \geq 1$.

From equation (12) we have that

$$Q_{t+1} - Q^* = Q_t - Q^* + \gamma_{t+1} \left(\hat{T}(Q_t, \hat{w}_t^N) - Q_t \right).$$

Denote by $E_t(\cdot)$ the expectations operator conditional on all information available before the time t wage sample is drawn. Observe that

$$\begin{aligned} E_t \left(\hat{T}(Q_t, \hat{w}_t^N) \right) &= N^{-1} \sum_{k=1}^{N} E_t \max \left\{ \begin{array}{c} \phi U \left(\hat{w}_t(k) \right) + \beta \alpha \phi Q_t \\ U(b) + \beta Q_t \end{array} \right\} \\ &= N^{-1} \sum_{k=1}^{N} E_t V(\hat{w}_t(k), Q_t) = N^{-1} \sum_{k=1}^{N} T(Q_t) = T(Q_t). \end{aligned}$$

The second equality follows from (6) and the third equality follows from (10) and the random sample assumption. Using this observation we may compute

$$E_t[(Q_{t+1} - Q^*)^2] = (Q_t - Q^*)^2 + 2\gamma_{t+1}(Q_t - Q^*)(T(Q_t) - Q_t) + \gamma_{t+1}^2 E_t \left[\left(\hat{T}(Q_t, \hat{w}_t^N) - Q_t \right)^2 \right].$$

As $[\underline{Q}, \overline{Q}]$ is compact and $\hat{T}$ is continuous in Q there exists $M > 0$ such that

$$E_t \left[\left(\hat{T}(Q_t, w_{t+1}) - Q_t \right)^2 \right] \leq M$$

for all $Q_t \in [\underline{Q}, \overline{Q}]$.

Note that if $f : [a, b] \to \mathbb{R}$ is continuous and is differentiable everywhere except at a finite number of points $a < x_1 < \cdots < x_n < b$, and, where defined, if $f'(x) < \beta$ then for all $a < x < y < b$ we have that

$$\frac{f(y) - f(x)}{y - x} \leq \beta.$$

To see this, suppose, for example, that $a < x < x_1 < y < x_2$. Then

$$\begin{aligned} \frac{f(y) - f(x)}{y - x} &= \frac{f(y) - f(x_1) + f(x_1) - f(x)}{y - x} \\ &\leq \frac{\beta(y - x_1) + \beta(x_1 - x)}{y - x} = \beta. \end{aligned}$$

The general result is then easily verified.

Applying this observation to T, and using the facts that $T'(Q) \leq \beta$ for all Q except possibly at $Q_{\max}$ and $Q_{\min}$, and that $T(Q^*) = Q^*$, it follows that

$$\frac{T(Q) - Q}{Q - Q^*} \leq \beta - 1$$

for all Q. Define $\lambda = -2(\beta - 1) > 0$. Then

$$\begin{aligned} E_t[(Q_{t+1} - Q^*)^2] &\leq (Q_t - Q^*)^2 + 2\gamma_{t+1}(Q_t - Q^*)(T(Q_t) - Q_t) + \gamma_{t+1}^2 M \\ &\leq \left(1 + 2\gamma_{t+1}\frac{T(Q_t) - Q_t}{Q_t - Q^*}\right)(Q_t - Q^*)^2 + \gamma_{t+1}^2 M \\ &\leq (1 - \lambda\gamma_{t+1})(Q_t - Q^*)^2 + \gamma_{t+1}^2 M. \end{aligned} \tag{32}$$

Following the proof strategy of Bray and Savin (1986), define

$$c_t = (Q_t - Q^*)^2 + \left(\sum_{k=t}^{\infty} \gamma_{t+1}^2\right) M.$$

From Equation (32) we know that c_t is a sub-martingale since

$$\begin{aligned} E_t c_{t+1} = E_t[(Q_{t+1} - Q^*)^2] + \left(\sum_{k=t+1}^{\infty} \gamma_{t+1}^2\right) M \\ \leq (1 - \lambda\gamma_{t+1})(Q_t - Q^*)^2 + \gamma_{t+1}^2 M + \left(\sum_{k=t+1}^{\infty} \gamma_{t+1}^2\right) M \\ \leq (Q_t - Q^*)^2 + \left(\sum_{k=t}^{\infty} \gamma_{t+1}^2\right) M = c_t. \end{aligned}$$

As c_t is bounded from below by 0, we apply the Martingale Convergence Theorem to conclude that c_t converges to some random variable $\tilde{c}$ almost surely. This immediately implies that $(Q_t - Q^*)^2$ converges to some random variable $\tilde{D}$ almost surely. It remains to be shown that $\tilde{D} = 0$ almost everywhere, and thus $Q_t \to Q^*$ almost surely.

Suppose not, then $E(\tilde{D}) > 0$. Convergence almost surely then implies that there exists $L > 0$ and $t^* > 0$ such that $E(Q_t - Q^*)^2 \geq L$ for all $t \geq t^*$. Taking expectations of Equation (32) we have that

$$E[(Q_{t+1} - Q^*)^2] \leq (1 - \lambda\gamma_{t+1})E[(Q_t - Q^*)^2] + \gamma_{t+1}^2 M.$$

Since $\gamma_t \to 0$, we can choose any $N > t^*$ such that $\gamma_{t+1} \leq \frac{L\lambda}{2M}$ for all $t \geq N$. It follows that

$$E[(Q_{t+1} - Q^*)^2] \leq \left(1 - \frac{\lambda}{2}\gamma_{t+1}\right) E[(Q_t - Q^*)^2]$$

for all $t \geq N$. We therefore conclude that

$$E[(Q_t - Q^*)^2] \leq E[(Q_N - Q^*)^2] \prod_{k=N}^{t-1} \left(1 - \frac{\lambda}{2}\gamma_{k+1}\right)$$

for all $t \geq N$. By Lemma A.3, Assumption C implies that $\prod_{k=N}^{\infty}\left(1-\frac{\lambda}{2}\gamma_{k+1}\right)=0$ and thus

$$E(\tilde{D})=\lim_{t\to\infty}E[(Q_t-Q^*)^2]=0,$$

which is a contradiction. Therefore, we conclude that $Q_t \to Q^*$ almost surely. ■

Appendix B: Proofs of results in Section 4

Computation of $\delta(Q,b)$. Let

$$\psi=\psi(Q,b)\equiv F\left(\bar{w}(Q,b)\right)+\alpha\left(1-F\left(\bar{w}(Q,b)\right)\right),$$

which is the probability of being unemployed at the end of the current period conditional on being unemployed at end of the previous period. Then

$$\begin{aligned}\delta(Q,b) &= 1\cdot(1-\psi)+2\cdot\psi\cdot(1-\psi)+3\cdot\psi^2\cdot(1-\psi)+\dots\\ &= (1-\psi)\sum_{n\geq 0}(n+1)\psi^n=\frac{1}{(1-\alpha)\left(1-F\left(\bar{w}(Q,b)u\right)\right)}.\ ■\end{aligned}$$

Proof of Lemma 2. We begin by showing that $Q_b^*>0$. Implicit differentiation yields $Q_b^*=(1-DT(Q^*))^{-1}T_b(Q^*)>0$. As shown in the proof of Lemma A.2, $DT(Q)\in(0,1)$. Also, since Q^* is in the interior, the T-map is given locally by

$$T(Q)=\left(U(b)+\beta Q\right)F(\bar{w})+\beta\alpha\phi Q(1-F(\bar{w}))+\phi\int_{\bar{w}(Q)}^{w_{\max}}U(w)dF(w). \tag{33}$$

Direct computation yields

$$\begin{aligned}T_b(Q^*) &= F(w^*)U'(b)+(U(b)+\beta Q^*-\beta\alpha\phi Q^*)dF(w^*)w_b^*-\phi U(w^*)dF(w^*)w_b^*\\ &= F(w^*)U'(b)+\left[U(b)+\beta Q^*-(\phi U(w^*)+\beta\alpha\phi Q^*)\right]dF(w^*)w_b^*=F(w^*)U'(b)>0,\end{aligned}$$

where the term in square brackets equals zero by (8). It follows that $Q_b^*>0$

To conclude, we need only establish that $\frac{\partial h}{\partial b}<0$ and $\frac{\partial h}{\partial Q}<0$. Since $h=(1-\alpha)(1-F)$, we may compute

$$\frac{\partial h}{\partial b}=-(1-\alpha)dF(w^*)\frac{\partial}{\partial b}\bar{w}\left(Q^*,b\right)<0 \text{ and } \frac{\partial h}{\partial Q}=-(1-\alpha)dF(w^*)\frac{\partial}{\partial Q}\bar{w}\left(Q^*,b\right)<0,$$

where the inequalities follow from equation (9). ■

Proof of Proposition 4. First observe that

$$du_0=-u^*dh+(1-\alpha)(1-u^*)q_w dw. \tag{34}$$

Next, notice that

$$u^* = \frac{\alpha}{h+\alpha} \implies -u^* dh = \alpha \frac{du^*}{u^*}. \tag{35}$$

If $db < 0$ then $dq = q_w dw = 0$. It follows from equations (34)-(35) that $du_0 = \alpha \frac{du^*}{u^*}$. Turning now to the case $db \geq 0$, and using the definition of h to get $dh = -(1-\alpha)dF$ and that $dq = (1-F)^{-1} dF$, we have that $dh = -(1-\alpha)(1-F)dq = -hdq$. Plugging into equation (34) we find

$$\begin{aligned} du_0 &= \alpha \frac{du^*}{u^*} - \frac{(1-\alpha)(1-u^*)}{h} dh = \alpha \frac{du^*}{u^*} - \left(\frac{1-\alpha}{\alpha}\right) u^* dh \\ &= \alpha \frac{du^*}{u^*} + \left(\frac{1-\alpha}{\alpha}\right) \alpha \frac{du^*}{u^*} = \frac{du^*}{u^*} \end{aligned}$$

as desired. ■

Appendix C: Derivation of Mm ratio in Section 5

Let $\bar{w}(Q)$ be the reservation wage consistent with beliefs Q. Then, with $\sigma = 0$, equation (8) yields

$$\phi \bar{w}(Q) = \rho E\left(w | w \geq \bar{w}(Q)\right) + \beta(1-\alpha\phi)Q. \tag{36}$$

The T-map becomes

$$T(Q) = (\rho \bar{w} + \beta Q) F\left(\bar{w}(Q)\right) + \alpha\beta\phi Q(1 - F\left(\bar{w}(Q)\right)) + \phi(1 - F\left(\bar{w}(Q)\right)) E\left(w | w \geq \bar{w}(Q)\right).$$

Using $1 - \alpha\beta\phi = (1-\beta)\phi$ and $1 - \alpha\phi = (1-\beta)(1-\alpha)\phi$, we find that $T(Q^*) = Q^*$ implies $Q^* = \psi E\left(w | w \geq w^*\right)$, where

$$\psi = \frac{1}{\phi(1-\beta)} \left(\frac{\rho F(w^*) + \phi(1 - F(w^*))}{1 - \beta(1-\alpha)F(w^*)}\right) = \frac{1}{\phi(1-\beta)} \left(\frac{\rho(1-\lambda) + \phi\lambda}{1 - \beta(1-\alpha)(1-\lambda)}\right).$$

Here $\lambda = 1 - F(w^*)$ is the job-finding rate. Using $Q^* = \psi E\left(w | w \geq w^*\right)$ and $\bar{w}(Q) = w^*$ in equation (36), we get

$$\begin{aligned} Mm &= \frac{\phi}{\rho + \beta(1-\beta)(1-\alpha)\phi\psi} = \frac{\phi}{\rho + \beta(1-\alpha)\left(\frac{\rho(1-\lambda)+\phi\lambda}{1-\beta(1-\alpha)(1-\lambda)}\right)} \\ &= \frac{\phi\left(1 - \beta(1-\alpha)(1-\lambda)\right)}{\left[(1 - \beta(1-\alpha)(1-\lambda)) + \beta(1-\alpha)(1-\lambda)\right]\rho + \beta(1-\alpha)\phi\lambda} = \frac{\left(\frac{\beta(1-\alpha)}{1-\beta(1-\alpha)}\right)\lambda + 1}{\left(\frac{\beta(1-\alpha)}{1-\beta(1-\alpha)}\right)\lambda + \rho} \end{aligned}$$

as desired.

References

[1] Adam, K., Marcet, A. and Beutel, J. (2017), Stock price booms and expected capital gains, American Economic Review 107, 2352-2408.

[2] Albrecht J.W., and Axell, B. (1984), An equilibrium model of search unemployment, Journal of Political Economy 92, 824-840.

[3] Burdett, K. and Vishwanath, T. (1988), Declining reservation wages and learning, Review of Economic Studies 55:4, 655-666.

[4] Bray, M.M. and Savin, N.E. (1986), Rational expectations equilibria, learning, and model specification, Econometrica 54, 1129-1160.

[5] Bullard, J. and Eusepi, S. (2005), Did the great inflation occur despite policy maker commitment to a Taylor rule?, Review of Economic Dynamics, 8, 324 - 359.

[6] Cho, I. and Kasa, K. (2004), Learning, large deviations, and recurrent currency crises, International Economic Review, 45:1, 141-173.

[7] Cogley and Sargent (2008), Anticipated utility and rational expectations as approximations of Bayesian decision making, International Economic Review, 49, 185-221.

[8] Damdinsuren, E. and Zaharievia, A. (2018), Expectation formation and learning in the labour Market with on-the-job search and Nash bargaining, Bielefeld Working Papers in Economics and Management, 10.

[9] Evans, G.W. and Honkapohja, S. (2001), Learning and Expectations in Macroeconomics. Princeton University Press, Princeton.

[10] Evans, G.W., Honkapohja, S. and Williams, N. (2010), Generalized stochastic gradient learning, International Economic Review, 51:1, 237 - 262.

[11] Evans, G.W. and McGough, B. (2018a), Learning to optimize, mimeo., University of Oregon.

[12] Evans, G.W. and McGough, B. (2018b), Agent-level learning in general equilibrium: the shadow-price approach.

[13] Greene, W.H. (2012), Econometric Analysis, 7th ed. Prentice Hall, Boston MA.

[14] Hommes, C. and Zhu, M. (2014), Behavioral learning equilibria, Journal of Economic Theory, 150:C, 778-814.

[15] Hornstein, A., Krusell, P. and Violante G.L. (2011), Frictional wage dispersion in search models: a quantitative assessment, American Economic Review 101, 873-898.

[16] Kahneman, D. (2011), Thinking fast and slow, Farrar, Straus and Giroux, New York.

[17] Kennan, J., (2010), Private information, wage bargaining and employment fluctuations, Review of Economic Studies 77, 633-664.

[18] Kreps, D. (1998), Anticipated utility and dynamic choice, in Frontiers of Economic Research: the Nancy L. Schwartz Memorial Lectures, eds. Jacobs, D.P., Kalai, E. and Kamien, M. Cambridge University Press, Cambridge UK.

[19] Krueger, D., Mitman, K. and Perri, F. (2016), Macroeconomics and household heterogeneity, Ch 11 in Handbook of Macroeconomics, Volume 2A, 843-921, eds. Taylor, J.B. and Uhlig, H., Elsevier.

[20] Ljung, L. (1977), Analysis of recursive stochastic algorithms, IEEE Transactions on Automatic Control, 22:551 - 575.

[21] Ljung, L. and Söderström, T. (1983), Theory and Practice of Recursive Identification, MIT Press, Cambridge MA.

[22] Ljungqvist, L. and Sargent, T.J. (2012), Recursive Macroeconomic Theory, 3rd ed., MIT Press, Cambridge,MA.

[23] Lucas, R. and Prescott, E., (1974), Equilibrium search and unemployment, Journal of Economic Theory 7, 188-209.

[24] Marcet, A. and Nicolini, J.P. (2003), Recurrent hyperinflations and learning, American Economic Review, 93, 1476-1498.

[25] Marcet, A. and Sargent, T.J. (1989), Convergence of least squares learning mechanisms in self-referential linear stochastic models, Journal of Economic Theory 48, 337-368.

[26] Marimon, R., McGrattan E. and Sargent, T.J. (1990), Money as a medium of exchange with artificially intelligent agents, Journal of Economic Dynamics and Control 14, 329-373.

[27] McCall, J.J. (1970), Economics of information and job search, Quarterly Journal of Economics 84, 113-126.

[28] Milani, F. (2007), Expectations, learning and macroeconomic persistence, Journal of Monetary Economics 54, 2065-2082.

[29] Morales-Jimenez, C., (2017), The cyclical behavior of unemployment and wages under information frictions, Finance and Economics Discussion Series 2017-047. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2017.047.

[30] Preston, B. (2005), Learning about monetary policy rules when long-horizon expectations matter, International Journal of Central Banking 1, 81-126.

[31] Pissarides, C. (1985), Short-run equilibrium dynamics of unemployment vacancies, and real wages, American Economic Review 75, 676-690.

[32] Rogerson, R., Shimer, R., and Wright, R. (2005), Search-theoretic models of the labor market: a survey, Journal of Economic Literature 43, 959-988.

[33] Rotemburg, J. J. (2017), Group learning, wage dispersion and non-stationary offers, Economica, 84:335, 365-392.

[34] Sargent, T. (1999), The conquest of American inflation, Princeton Univ. Press, Princeton NJ.

[35] Sutton, R. S. and Barto, A. G. (2011), Reinforcement learning: An introduction, MIT Press, Cambridge, MA.

[36] Watkins, C. (1989), Learning from Delayed Rewards. PhD Thesis, University of Cambridge, Cambridge, UK.

[37] Watkins C. and Dayan, P. (1992), Technical Note: Q-learning," Machine Learning, 8, 279-292.

[38] Williams, N. (2019), Escape dynamics in learning models, Review of Economic Studies, 86: 882-912