Online Appendix for:
A Labor Market Sorting Model of Scarring and Hysteresis[*]

Edoardo Maria Acabbi[1], Andrea Alati[2], and Luca Mazzone[3]

[1]*Universidad Carlos III de Madrid*
[2]*Bank of England*
[3]*International Monetary Fund*

March 2023

Contents

[*]Acabbi: eacabbi@emp.uc3m.es
Alati: andrea.alati@bankofengland.co.uk
Mazzone: lmazzone@imf.org

1 Introduction

The following Online Appendix expands and complements the discussion of the model properties in **Section 4** of the paper. In addition, it provides more details on the empirical results discussed in the validation of the paper.

2 Discussion

In this section we briefly discuss the properties of the equilibrium of the model economy developed in the previous sections. All propositions and corresponding proofs are reported in **Appendix 3** and **4**.

2.1 Workers optimal behavior

In the following proposition we summarize the main results regarding the behavior of the workers and their objective functions.

Proposition 2.1. *Given the worker search problem, the following properties hold:*

(i) The returns to search, $p(\theta(h,\tau,\iota,v;\Omega))\big[v-V\big]$*, are strictly concave with respect to promised utility,* v*.*

(ii) The optimal search strategy

$$v^*(h,\tau,\iota,V;\Omega) \in \arg\max_{v} \big\{p(\theta(h,\tau,\iota,v;\Omega))\big[v-V\big]\big\}$$

is unique and weakly increasing in V*.*

(iii) For all promised utilities, the search gain $R(h,\tau,\iota,V;\Omega)$ *is positive, weakly decreasing in* V*.*

(iv) The survival probability of the match, given the optimal choice of the worker, is increasing in the value of current and future promised utilities, so $\widetilde{p}_t(h,\tau,\iota,v;\Omega)$ *is increasing in* v *and* V*.*

Proof. See **Appendix Section 3.1**. □

The first statement implies that the marginal returns of searching towards better firms are decreasing. The intuition is that as workers search for work at firms granting better values, their job-finding probability decreases as better employment prospects are also subject to higher competition.

As a consequence of the strict concavity established in the first statement, workers' optimal search strategy is unique. The search strategy is also (weakly) increasing in the value of lifetime utilities granted by the current contract, which is the outside option for the worker.

The third statement follows from the fact that marginal returns to search are decreasing and the set of feasible utility promises is compact. The intuition is that employees at firms with higher utility promises have a relatively fewer chances of improving their position. Given a high outside option, the utility gain from moving is relatively lower, whereas the probability of matching with any firm does not depend on the *current* utility promise per se, but on the future promise offered by the vacancy.

The fourth statement finally follows from considering the implication of the previous ones. Given that the optimal search strategy is increasing in V workers' probability of leaving the firm at any time ends up depending negatively on V. This guarantees a longer expected duration of the match at higher current promised utility V, thus retention probabilities that are increasing in promised utilities v.

As human capital accumulation is tightly linked to the quality of the employer, workers that are able to start their working careers in good times have a greater chance of finding themselves on an higher path of human capital growth. As worker careers are limited and human capital accumulation follows a slow-moving process, business cycle effects on human capital quality fade only slowly and the quality of initial matches, both in terms of lifetime utility and firm quality, bears a long-standing effect on workers' careers.

2.2 Characteristics of the optimal contract

The optimization in the contracting problem balances a trade-off between insurance provision and profit maximization for firms. The contract implicitly takes into account workers' search incentives and their inability to commit to stay. The following proposition characterizes workers' incentives along the business cycle from the firms' standpoint.

Proposition 2.2. *The Pareto frontier $J(h, \tau, \iota, W, y; a, \mu)$ is increasing in the aggregate productivity shock a, while retention probabilities, $\widetilde{p}(h, \tau, \iota, W; a, \mu)$ decrease in aggregate productivity.*

Proof. See **Appendix Section 4.1**. □

The intuition behind this proposition relies on the observation that higher productivity realization are associated not only with better outcomes on impact but also to better future prospects, given that the productivity process is an increasing Markov chain.

A key property of the model is that it allows to characterize the workers' optimal behaviour along the business cycle. The following proposition summarizes how the search strategy changes depending on the aggregate productivity realization.

Corollary 2.1. *The optimal search strategy of the workers is increasing in aggregate productivity.*

Proof. The claim follows directly from **Proposition 2.2**. □

Firm value $J(\cdot)$ is increasing in a, while retention is decreasing. As firm make more profits, they can marginally increase offered utilities to workers to maximize retention. This in turn will imply that in better times workers will search for better firms, given their greater outside options. The way in which offered wages w and thus values W are tied to profits J is described in **Proposition 2.3**.

Proposition 2.2 and Corollary 2.1 have an important implication regarding firms' vacancy posting and workers' search decisions. The fact that at the posting stage profits J are increasing in aggregate productivity implies that more entry will take place in good times, and ceteris paribus more entrepreneurs will open up vacancies across the whole firms' distribution.[1] The resulting higher tightness impacts workers' optimal search behaviour as the job finding probability increases in all submarkets. As a consequence, workers respond optimally to the productivity increase searching in submarkets that guarantee higher lifetime utility promises.

Firms utility promises depend on the structure of the optimal contract. The contract provides insurance to workers through wage paths that are downward rigid, and at the same time allows firms to profit as wages only partially adjust to productivity realizations.

The following propositions provide a clear picture of the growth path prescribed by the optimal contract for a continuing firm. First, let us define the productivity threshold that determines whether a worker-firm match does not survive.

Corollary 2.2. *There exists a productivity threshold $a^*(h, \tau, \iota, W, y)$ below which firms will not continue the operate.*

The intuition of why this has to be the case is linked to the fact that the Pareto frontier is strictly increasing in a and decreasing in the level of promised utilities to the worker. Hence once the aggregate state realizes a firm is able to perfectly predict whether next period it will exit the market or stay in (given the timing, the decision is based on expected profits, and is thus *not* state-contingent to next period's productivity). The

[1] In our model a better firm is a more productive firm. We do not specifically model the determinant of quality heterogeneity but we take the existence of profound differences in firm quality as a fact (Arellano-Bover, 2020, **?**).

choice is taken *before* new realizations of productivity, so it is possible that a firm makes negative profits for at most one period.

Proposition 2.3. *For each state in which the firm is willing to continue the contract, the optimal contract delivers a wage path that follows firms profits according to the wage Euler equation:*

$$\frac{\partial \widetilde{p}(\Theta)}{\partial W_i'} \frac{J'(\Theta)}{\widetilde{p}(\Theta)} = \frac{1}{u'(w_i)} - \frac{1}{u'(w)} \tag{1}$$

with $\Theta \equiv (\phi(h, y), \tau + 1, \iota, W_i'; \Omega')$ *being the definition of the relevant state and* w_i *is the wage paid in the future state.*

Proof. See **Appendix Section 4.3**. □

The optimal contract links the wage growth to the realization of firms profits. The right hand side of Equation 1 shows that, in providing insurance to the worker, the firm links wage growth to profits and to the incentive to maximize retention, incorporated in $\frac{\partial \log \widetilde{p}}{\partial W}$, the semi-elasticity of the retention probability to the utility offer. As the production stage takes place *after* exit choices are taken by the incumbent firms, the wage growth related to the continuation value of the contract is bound to be (weakly) positive, hence workers enjoy a non-decreasing wage profile under the optimal contract.[2]

A feature that the optimal contract derived in our model shares with the literature on long-term contracts with lack of commitment on the worker side is thus the backloading of wages.[3] Workers in our model make search decisions that affect the survival probability of the match. They do not however appropriate the full future value of the current match while making these search decisions (unless the firm makes zero profits). This makes it optimal for the firm to front-load profits and back-load wages. The reason is that the firm provides insurance and income smoothing to the worker, but given its risk neutrality it prefers to front-load its profits while providing an increasing compensation path to maximize retention. The contract thus optimally balances the consumption smoothing motives (i.e. the insurance provision of the contract) with the commitment problem of the worker.

Special case with log-utility. The wage Euler equation discussed in **Proposition 2.3** can be simplified to a more intuitive interpretation in the log-utility case. In case of log-utility, in fact, $u'(w_{i,\Omega}) = \frac{1}{w_{i,\Omega}}$. Multiplying and dividing by wage levels and rearranging,

[2]As the exit decision takes place by considering *expected* profits next period, a firm operating at low but positive expected profits might end up, at most for a period, to have a negative continuation value. This would imply that wage growth *can* be negative before a firm's closure, which is actually a common finding in empirical studies (firstly observed in Ashenfelter (1978)).

[3]See for instance, Thomas and Worral (1988), Tsuyuhara (2016) and Balke and Lamadon (2022).

we can express the elasticity of retention probability to offered utility as

$$\varepsilon_{\widetilde{p},W_y} = \underbrace{\frac{(w_i - w)}{w}}_{\text{Wage growth}} \underbrace{\frac{w}{J(\Theta)}}_{\substack{\text{Ratio of wage}\\ \text{to match value}}} . \tag{2}$$

with $\varepsilon_{\widetilde{p},W} \equiv \frac{\partial \widetilde{p}(\Theta)}{\partial W_i \widetilde{p}(\Theta)}$.

The interpretation of this result is of interest to analyses that relate labor market dynamism to wage dynamics, like Engbom (2020). This is because $\varepsilon_{\widetilde{p},W_y}$, being a function of structural parameters of the matching technology, γ, search frictions λ_e, and measures of labor market tightness θ, provides us with a good proxy of labor market fluidity. The right hand side of (2), is composed entirely of observable quantities, as the ratio of wages to match value is a function of factor shares in value added. The quantity can then be used to compare the dynamism of different local, regional or national labor markets.
The next proposition, instead, confirms our initial conjecture that in equilibrium firm qualities and utility promises are related to a one-to-one mapping.

Proposition 2.4. *The mapping defined by the function $f_v : \mathcal{Y} \rightarrow \mathcal{V}$ is an injective function for each worker characteristic (h, τ).*

Proof. See **Appendix Section 4.2**. □

The proof is based on the fact that the Pareto frontier J is concave, the vacancy filling probability q is weakly positive and vacancy costs are both weakly positive and increasing in y. As shown in **Appendix 4** these features are enough to guarantee that only one kind of firm y, given workers' characteristics h, τ, can optimally offer a given lifetime utility promise W, and that this mapping is monotonically increasing. We thus obtain a unique monotonic solution in which higher quality firms offer higher lifetime utility promises to workers.

Finally, we provide the alternative recursive formulation for the contracting problem described in the paper. The saddle-point functional equation that can be alternatively used to define the recursive contract in **Equation** (5) is expressed in the following proposition.

Proposition 2.5. *The solution to the contracting problem in* ***Equation*** (5) *is the same as the solution to the following saddle-point functional equation:*

$$\begin{aligned}\mathcal{P}_t(h_t, \tau_t, y_t, a_t, \gamma_t) = &\underset{\gamma_t}{inf}\, \underset{w_t}{sup}(f(a_t, y_t, h_t) - w_t) + \mu_t^1 W_{y,t} - \gamma_t^1 (W_{y,t} - u(w_t)) + \\ &\beta \mathbb{E}_t(\lambda U_{t+1} + (1-\lambda)\lambda_e p_{t+1} v_{t+1}^*) + \beta \mathbb{E}_t \widetilde{p}_{t+1} \mathcal{P}_{t+1}(h_{t+1}, \tau_{t+1}, y_{t+1}, a_{t+1}, \gamma_{t+1})\end{aligned}$$

with $\mu_t = \gamma_{t_1}$ for some starting γ_0.

Proof. See **Appendix 5** for the details of the derivation of the SPFE following Marcet and Marimon (2019). □

3 Properties of worker optimal behavior

For compactness of notation, we omit the dependence on education level, which is a fixed characteristic, and the idiosyncratic human capital shock, which is additive, from the proofs in Appendices. The logic of the proofs follows without loss of generality.

The following propositions characterize the properties of workers' optimal search strategies that solve the search problem in (9), restated here for convenience:

$$R(h, \tau, V; \Omega) = \sup_{v} \Big[p(\theta(h, \tau, v; \Omega)) \big[v - V \big] \Big]. \tag{3}$$

Lemma 3.1. *The composite function $p(\theta(h, \tau, v; \Omega))$ is strictly decreasing and strictly concave in v.*

Proof. For this proof we follow closely Menzio and Shi (2010), Lemma 4.1 (ii). From the properties of the matching function we know that $p(\theta)$ is increasing and concave in θ, while $q(\theta)$ is decreasing and convex. Consider that the equilibrium definition of $\theta(\cdot)$ is

$$\theta(h, \tau, v; \Omega) = q^{-1} \left(\frac{c(y)}{J(h, \tau, v, y; \Omega)} \right),$$

and that the first order condition for the wage and the envelope condition on V of the optimal contract problem in (5) implies

$$\frac{\partial J(h, \tau, v, y; \Omega)}{\partial v} = -\frac{1}{u'(w)}.$$

so that as $u'(\cdot) > 0$, $J(\cdot)$ is decreasing in v.

From the equilibrium definition of $\theta(\cdot)$ and noting that $q^{-1}(\cdot)$ is also decreasing due to the properties of the matching function we have that

$$\frac{\partial \theta(h, \tau, v; \Omega)}{\partial v} = \frac{\partial q^{-1}(\xi)}{\partial \xi} \bigg|_{\xi = \frac{c(y)}{J(h,\tau,v,y;\Omega)}} \cdot \left(- \frac{\partial J(h, \tau, v, y; \Omega)}{\partial v} \right) \cdot \frac{c(y)}{(J(h, \tau, v, y; \Omega))^2} < 0,$$

which, in turn, implies that

$$\frac{\partial p(\theta(h, \tau, v; \Omega))}{\partial v} = \frac{\partial p(\theta)}{\partial \theta} \bigg|_{\theta = \theta(h,\tau,v;\Omega)} \cdot \frac{\partial \theta(h, \tau, v; \Omega)}{\partial v} < 0.$$

Suppressing dependence on the states (h, τ, y, Ω) for readability, to prove that $p(\theta(v))$

is concave, consider that $J(v)$ is concave[4] and a generic function $\frac{c}{v}$ is strictly convex in v. This implies that with $\alpha \in [0,1]$ and $v_1, v_2 \in \mathcal{V}$, $v_1 \neq v_2$:

$$\frac{c}{J(\alpha v_1 + (1-\alpha)v_2)} \leq \frac{c}{\alpha J(v_1) + (1-\alpha)J(v_2)} < \alpha \frac{c}{J(v_1)} + (1-\alpha)\frac{c}{J(v_2)}.$$

As $p(q^{-1}(\cdot))$ is strictly decreasing the inequality implies that

$$\begin{aligned} p\left(q^{-1}\left(\tfrac{c}{J(\alpha v_1+(1-\alpha)v_2)}\right)\right) &\geq p\left(q^{-1}\left(\tfrac{c}{\alpha J(v_1)+(1-\alpha)J(v_2)}\right)\right) \\ &> \alpha p\left(q^{-1}\left(\tfrac{c}{J(v_1)}\right)\right) + (1-\alpha)p\left(q^{-1}\left(\tfrac{c}{J(v_2)}\right)\right), \end{aligned}$$

and as $\theta(v) = q^{-1}(\frac{c}{J(v)})$:

$$p(\theta(\alpha v_1 + (1-\alpha)v_2)) > \alpha p(\theta(v_1)) + (1-\alpha)p(\theta(v_2))$$

so that $p(\theta(v))$ is strictly concave in v. □

3.1 Proof of Proposition 2.1

Proof. The proofs follow closely Shi (2009), Lemma 3.1 and Menzio and Shi (2010), Corollary 4.4. More formally, for each triplet (h, τ, Ω) given at each search stage, we can re-define the search objective function as $K(v, V) = p(\theta(v))(v - V)$ and $v^*(V) \in \arg\max_v K(v, V)$ as the function that maximises the search returns (i.e. the optimal search strategy of the worker) and prove the following

(i) To show that $K(v, V)$ is strictly concave in v consider two values for v, $v_1, v_2 \in \mathcal{V}$ such that $v_2 > v_1$ and define $v_\alpha = \alpha v_1 + (1-\alpha)v_2$ for $\alpha \in [0,1]$.

Then by definition:

$$\begin{aligned} K(v_\alpha, V) &= p(\theta(v_\alpha))(v_\alpha - V) \\ &\geq [\alpha p(\theta(v_1)) + (1-\alpha)p(\theta(v_2))][\alpha(v_1 - V) + (1-\alpha)(v_2 - V)] \\ &= \alpha K(v_1, V) + (1-\alpha)K(v_2, V) + \alpha(1-\alpha)[(p(\theta(v_1)) - p(\theta(v_2))](v_2 - v_1) \\ &> \alpha K(v_1, V) + (1-\alpha)K(v_2, V) \end{aligned}$$

where the first inequality follows from the concavity of $p(\theta(\cdot))$ (this is true if $J(\cdot)$ concave with respect to V) and the second inequality stems from the fact that $p(\theta(\cdot))$ is strictly decreasing hence $\alpha(1-\alpha)[(p(\theta(v_1)) - p(\theta(v_2))](v_2 - v_1) > 0$.

(ii) **Weakly Increasing.** Consider a worker employed in a job that gives lifetime

[4] $J(\cdot)$ concave give the two-point lottery in the structure of the contract. See Menzio and Shi (2010) Lemma F.1.

utility V. Given that $v \in [\underline{v}, \overline{v}]$, and that submarkets are going to open depending on realizations of the aggregate productivity, a, there is only one region in the set of promised utilities where the search gain is positive. This set is $[V, v(a)]$ with $v(a)$ being the highest possible offer that a firm makes in the submarket for the worker (h, τ). Any submarket that promises higher than $v(a)$ is going to have zero tightness. Therefore, the optimal search strategy for $V \geq v(a)$ is $v^*(V) = V$, as $K(V, v(a)) = K(V, V) = K(\overline{v}, V) = 0$ (the search gain is null given the current lifetime utility V). For $V \in [V, v(a)]$, instead, as $K(v, V)$ is bounded and continuous, the solution $v^*(V)$ has to be interior and therefore respect the following first order condition

$$V = v^*(V) + \frac{p(\theta(v^*(V)))}{p'(\theta(v^*(V)) \cdot \theta'(v^*(V))}. \tag{4}$$

Now consider two arbitrary values V_1 and V_2, $V_1 < V_2 < \overline{v}$ and their associated solutions $W_i = v^*(V_i)$ for $i = 1, 2$. Then, V_1 and V_2 have to generate two different values for the right-hand side of (4). Hence, $v^*(V_1) \cap v^*(V_2) = \emptyset$ when $V_1 \neq V_2$.

This also implies that the search gain evaluated at the optimal search strategy is higher than the gain at any other arbitrary strategy so that $K(W_i, V_i) > K(W_j, V_i)$ for $i \neq j$. This implies that

$$\begin{aligned} 0 >&[K(W_2, V_1) - K(W_1, V_1)] + [K(W_1, V_2) - K(W_2, V_2)] \\ &= (p(\theta(W_2)) - p(\theta(W_1)))(V_2 - V_1), \end{aligned}$$

thus, $p(\theta(W_2)) < p(\theta(W_1))$. As $p(\theta(\cdot))$ is strictly decreasing (see Lemma 3.1), then $v^*(V_1) < v^*(V_2)$.

Lipschitz continuous. Consider generic v_2, v_1 such that $v_i = v^*(V_i)$. Given that $p(\cdot)$ is a concave function, one can write $p(v_1) + p'(v_1)(v_2 - v_1) > p(v_2)$. Using equation(4), we obtain $V_2 - V_1$:

$$\begin{aligned} V_2 - V_1 &= v_2 - v_1 + \frac{p(v_2)}{p'(v_2)} - \frac{p(v_1)}{p'(v_1)} \\ &> 2(v_2 - v_1) + p(v_2)\frac{p(v_1') - p(v_2')}{p'(v_2)p'(v_1)} > 2(v_2 - v_1) \end{aligned}$$

This condition proves that v^* is Lipschitz.

Unique. Uniqueness follows directly from strict concavity shown in (i).

(iii) The Bellman equation for the search problem is:

$$R(h, \tau, V; \Omega) = \sup_v \Big[p(\theta(h, \tau, v; \Omega))\big[v - V\big]\Big]$$

hence a simple envelope argument shows that

$$\frac{\partial R(h,\tau,V;\Omega)}{\partial V} = -p(\theta(h,\tau,v;\Omega)) \leq 0,$$

as the job finding probability is weakly positive for all utility promises.

As $p(\theta(\cdot)) \geq 0$, $v^*(\cdot) \in [\underline{v}, \overline{v}]$ then $R(\cdot) \geq 0$.

(iv) Given the optimal search strategy, $v^*(h,\tau,V;\Omega)$, we can define the survival probability of the match as in (12):

$$\widetilde{p}(h,\tau,V;\Omega) \equiv (1-\lambda)(1-\lambda_e p(\theta(h,\tau,v^*;\Omega))).$$

Then, given (h,τ,Ω)

$$\frac{\partial \widetilde{p}(V)}{\partial V} = -\beta(1-\lambda)\lambda_e \left.\frac{\partial p(\theta)}{\partial \theta}\right|_{\theta=\theta(v^*)} \left.\frac{\partial \theta(v)}{\partial v}\right|_{v=v^*(V)} \frac{\partial v^*(V)}{\partial V} > 0,$$

because $p(\cdot)$ and $v^*(\cdot)$ are both increasing functions while $\theta(\cdot)$ is a decreasing function in promised utilities.

□

4 Properties of the optimal contract

Lemma 4.1. *The Pareto frontier $J(h,\tau,W_y,y;\Omega)$ is concave in W_y.*

Proof. This is a direct consequence of using a two-point lottery for $\{w_i, W'_{iy}\}$ as shown by Menzio and Shi (2010), Lemma F.1. □

Lemma 4.2. *The Pareto frontier $J(h,\tau,W_y,y;\Omega)$ is increasing in y.*

Proof. The intuition for this proof follows the fact that a higher y firm, once the match exists, can always deliver a certain promise V and have resources left over. Within a dynamic contract, future retention is already optimized as the match is formed. This means that the promise V can be delivered by the greater capacity on the part of producing with respect to a close y firm. In presence of human capital accumulation, the worker is compensated through greater option values in the future, which again means that, even with lower retention, the firm cashes in more profits while decreasing wages (and respecting the V promise).

One can get to the same conclusion by starting from time T, noticing that the function J is trivially increasing in y in the last period, and the stepping back. At $T-1$, given V, any higher y function can make greater profits with the same delivery of value V, given the contract's optimal promise, which is a fortiori true with human capital accumulation (the option value is greater, so the firm can decrease w as a response). A more formal argument goes as follows: start from the optimal policies, as per Eq. (5) of a firm that has $y = \bar{y}$ and assume that one could exogenously increase its installed capital to $\bar{y} + \varepsilon$. We want to know whether, keeping policies constant, this would increase the flow of profits while keeping the worker indifferent. If this is true, then *a fortiori* it will be true that the firm value function J will be increasing in y. With a slight abuse of notation and for conciseness, we refer to the future J in Eq. (5) as J'.

This amounts to calculating:[5]

$$\left.\frac{d\bar{J}}{dy}\right|_{W_y,\{\pi_i,w_i,W'_{iy}\}} = \frac{\partial f(\cdot)}{\partial y} + \beta\mathbb{E}_\Omega\left[\left.\frac{\partial \tilde{p}(\cdot)}{\partial y}\right|_{W_y,\{\pi_i,w_i,W'_{iy}\}} \bar{J}'\right]$$

This first order condition presents the trade-off discussed in words above, namely that an increase in y will be instantaneously beneficial to production, but might also potentially have a longer term adverse effect on profits through decreased retention. Finding the sign of the derivative on the LHS hinges on understanding the sign of the derivative of the second element on the RHS, since $\frac{\partial f(\cdot)}{\partial y} > 0$ given the properties of the production function $f(\cdot)$. The change in y would affect search objective v_y through the variation in h due to the human capital accumulation dynamics, even taking the current firm policies as given.

Notice that:

$$\frac{\partial \tilde{p}(\cdot)}{\partial y} = -\lambda\lambda_E \frac{\partial p(\cdot)}{\partial \theta}\frac{\partial \theta(\cdot)}{\partial y}$$

Remember that $\frac{\partial p(\cdot)}{\partial \theta} > 0$ due to the properties of the job-finding probability function $p(\cdot)$. Assume first the case in which $\frac{\partial \theta(\cdot)}{\partial y} \leq 0$. In this case $\frac{\partial \tilde{p}(\cdot)}{\partial y} > 0$, which in turn implies, as argued, that J has to be increasing in y.

Now consider the second case, namely $\frac{\partial \theta(\cdot)}{\partial y} > 0$. By the free entry condition, we obtain:

$$\frac{\partial \theta(\cdot)}{\partial y} = \frac{\partial q^{-1}(c(y)/J(y))}{\partial y} = \frac{1}{q'\left(q^{-1}\left(c(y)/J(y)\right)\right)}\frac{c'(y)J(y) - \frac{\partial J(y)}{\partial y}c(y)}{J(y)^2}$$

The first term in the result is negative, given the properties of function $q(\cdot)$. Given our assumption on $\frac{\partial \theta(\cdot)}{\partial y}$ the second term in the result has to be negative as well. This

[5] Without loss of generality, we assume that J' is constant with respect to y. Alternatively, one may start proving the result for contracts offered to workers just one period before retirement (for which $\frac{\partial J'}{\partial y}$ is trivially positive), and generalize the result with backward induction.

requires: $c'(y)J(y) - \frac{\partial J(y)}{\partial y}c(y) < 0$, or $\frac{\partial J(y)}{\partial y} > \frac{c'(y)J(y)}{c(y)} > 0$. □

Corollary 4.1. *The retention probability $\widetilde{p}(h, \tau, V_{y,\Omega}; \Omega)$ is decreasing in h.*

Proof. Similarly to what shown above:

$$\frac{\partial \tilde{p}(\cdot)}{\partial h} = -\lambda\lambda_E \frac{\partial p(\cdot)}{\partial \theta}\frac{\partial \theta(\cdot)}{\partial h}$$

The sign of $\frac{\partial \theta(\cdot)}{\partial h}$ can be again obtained by looking at the free entry condition. This time:

$$\frac{\partial \theta(\cdot)}{\partial h} = \frac{1}{q'\left(q^{-1}\left(c(y)/J(\cdot)\right)\right)} \frac{-\frac{\partial J(h)}{\partial h}c(y)}{J(\cdot)^2} > 0$$

Both terms are negative - it is straightforward to show that $\frac{\partial J(h)}{\partial h} > 0$ using a similar logic as in **Lemma 4.2**. Intuitively: if $\frac{\partial J(h)}{\partial h} < 0$ then retention increases in h. But so do flow revenues, which would contradict the initial assumption. Bringing these elements together we get $\frac{\partial \theta(\cdot)}{\partial h} < 0$ which proves the overall result. □

4.1 Proof of Proposition 2.2

Proof. We proceed by backward induction.[6] Following the logic of the proof of **Lemma 4.2**, the proposition is trivially true for workers T periods old. Given that the firm increases its production while keeping the worker at least indifferent, J is at least weakly increasing in a. However, the firm can also feasibly increase the worker's wage by ε, with $\varepsilon < \frac{\partial f(\cdot)}{\partial a}$. J is thus strictly increasing in y.

Consider now a worker who is $T-1$ periods old. A firm matched to a worker in submarket $\{h, T-1, y, W_{y,\Omega}\}$ will face the following Pareto frontier

$$\begin{aligned} J(h, T-1, y, W_y; a, \mu) = \sup_{w, W'_y} \Big(& f(h, y; a) - w \\ & + \mathbb{E}_\Omega \left[\widetilde{p}(h', T, W'_y; a', \mu')(f(h', y; a') - w')\right] \Big) \end{aligned}$$

Analogously to the proof in **Lemma 4.2**, assume that aggregate productivity increases from $\bar{a}$ to $\bar{a} + \varepsilon$. Assume that the firm keeps its policies constant once again. We aim at proving that, even in such a case, firm value increases while keeping the worker at least

[6] For compactness of notation, we omit without loss of generality the two-point lottery in the equations in the proof.

indifferent. If this is the case, it is *a fortiori* true that J increases in a after reoptimizing firms' policies.

We are now interested in the sign of:[7]

$$\left.\frac{\partial \bar{J}}{\partial a}\right|_{W_y,\pi,w,\{W_y'\}} = \frac{\partial f(\cdot)}{\partial a} + \beta \mathbb{E}_\Omega \left[\left.\frac{\partial \tilde{p}(\cdot)}{\partial a}\right|_{W_y,\pi_i,w,\{W_y'\}} \bar{J}' \right]$$

Now notice that, in equilibrium,

$$\frac{\partial \widetilde{p}(\theta)}{\partial a} \propto -\frac{\partial p(\theta)}{\partial a} = \underbrace{\frac{\partial p(\theta)}{\partial \theta}}_{>0} \cdot \underbrace{\frac{\partial \theta}{\partial J(\cdot)}}_{>0} \cdot \frac{\partial J(\cdot)}{\partial a}$$

where the sign of the second derivative on the right hand side comes from the free entry condition and the properties of vacancy filling probability function $q(\cdot)$. Given this, it has to be that $\frac{\partial p(\theta)}{\partial a}$ and $\frac{\partial J(\cdot)}{\partial a}$ have the same sign in equilibrium. Now, if both are strictly positive, both statements of our proposition are immediately true. Let's now assume they are both negative or zero. If this is the case, then $\frac{\partial \tilde{p}(\cdot)}{\partial a} \geq 0$. But this implies $\frac{\partial \bar{J}}{\partial a} > 0$, which is a contradiction. □

Corollary 4.2. *There exists a productivity threshold $a^*(h, \tau, W_y, y)$ below which firms will not continue the operate.*

Proof. The proof follows immediately from **Proposition 2.2** and the timing of the shock. Given the timing of the shock, exit is fully determined by the current productivity shock and incumbent firms know in advance whether they are willing to produce in the next period.

Therefore, as the Pareto frontier is strictly increasing in a, firms are willing to continue the contract if $\mathbb{E}_\Omega[J(h', \tau+1, W'_{y,}, y; a', \mu')|h, \tau, W_y, y, a, \mu] \geq 0$, so that the threshold that determines exit is

$$a^*(h, \tau, W_y, y) : \mathbb{E}_\Omega[J(h', \tau+1, W'_y, y; a', \mu')|h, \tau, W_y, y, a, \mu] = 0.$$

□

Corollary 4.3. *The productivity threshold $a^*(h, \tau, y, W_y)$ below which firm y in match with worker (h, τ) and promised utility W_y exits the market in the aggregate state Ω is increasing in y.*

Proof. Consider two firms characterized by y_1, y_2 with $y_1 < y_2$. Consider the threshold for firm y_1, $a_1^* = a^*(h, \tau, , W_{y_1}, y_1)$. Firm y_1 makes 0 profits if state a_1^* materializes next

[7]We assume that $J' = f(h', y; a') - w'$ is constant with respect to a. It is possible to prove, by backward induction, that this assumption is without loos of generality for the sake of the proof.

period. Consider firm y_2 trying to mimic the current contract offered by y_1 to (h, τ). We know that J is increasing in y from **Lemma 4.2**, which implies that the firm is making a profit at a_1^*. This completes the proof. □

Lemma 4.3. *The Pareto frontier $J(h, \tau, W_y, y; \Omega)$ is strictly concave in y.*

Proof. **No human capital accumulation.** The proof without human capital accumulation is straightforward. Assuming there is no dependency of h' on y, one can write:

$$\frac{dJ}{dy} = f_y + \widetilde{p}\frac{dJ'}{dy}$$
$$\frac{d^2J}{dy^2} = f_{yy} + \widetilde{p}\frac{d^2J'}{dy^2}$$

where we take J' to represent next period's firm value $J(\cdot)$, and the dependence of all controls on y is ignored by virtue of the envelope condition. One can readily observe by induction that, given the concavity of $f(\cdot)$ in y, $\frac{d^2J}{dy^2} < 0$.

Human capital accumulation. With human capital accumulation the concavity of the function J on y depends on further assumptions on the concavity of the human capital accumulation function $g(\cdot)$. As these assumptions involve equilibrium objects, we state them here and verify ex-post numerically that they are always respected in our setting.

$$\frac{dJ}{dy} = f_y + g_y\frac{\partial \widetilde{p}J'}{\partial h} + \widetilde{p}\frac{dJ'}{dy}$$
$$\frac{d^2J}{dy^2} = f_{yy} + g_{yy}\frac{\partial \widetilde{p}J'}{\partial h'} + g_y^2\frac{\partial^2 \widetilde{p}J'}{\partial h'^2} + 2g_y\left(\frac{\partial \widetilde{p}}{\partial h'}f_y + \widetilde{p}f_{hy}\right) + \widetilde{p}\frac{d^2J'}{dy^2}$$

This condition implies that the concavity of $J(\cdot)$ necessarily imposes limits in equilibrium on the properties of $g(\cdot)$, the deterministic component of human capital accumulation. We find this condition to be respected for reasonable parameterization and choices of functional form.

□

4.2 Proof of Proposition 2.4

Proof. Note: throughout the proof we drop the dependence of the functions to the state (h, τ, Ω) to ease readability.

If the function f_v is an injective function then it defines a one-to-one mapping between $\mathcal{Y}$ and $\mathcal{V}$ so that for $(y_1, y_2) \in \mathcal{Y}$, and $f_v(y_1) = W_1$ and $f_v(y_2) = W_2$, $(W_1, W_2) \in \mathcal{V}$,

$f_v(y_1) = f_v(y_2) \Rightarrow y_1 = y_2$.[8] We proceed by contradiction. To begin, assume that $f_v(y_1) = f_v(y_2)$ and $y_1 \neq y_2$.

As the optimal contract is a concave function in firm quality, we know that the tangents at each point are above the graph of the function. Thus, we can define the tangents at the two points y_1, y_2 as

$$T_1(y) \equiv J(y_1) + \left.\frac{\partial J(y)}{\partial y}\right|_{y=y_1} (y - y_1) \quad \text{and} \quad T_2(y) \equiv J(y_2) + \left.\frac{\partial J(y)}{\partial y}\right|_{y=y_2} (y - y_2).$$

Without loss of generality, consider the case in which $y_2 > y_1$. Knowing that $T_i(y) \geq J(y)$ for $i = 1, 2$ due to the concavity of $J(\cdot)$, we can define the following inequalities:

$$T_1(y_2) - J(y_2) \geq 0 \quad \text{and} \quad T_2(y_1) - J(y_1) \geq 0.$$

Using the definitions for the tangents at y_1 and y_2 they imply that

$$\frac{J(y_2) - J(y_1)}{y_2 - y_1} \leq \left.\frac{\partial J(y)}{\partial y}\right|_{y=y_1} \quad \text{and} \quad \frac{J(y_2) - J(y_1)}{y_2 - y_1} \geq \left.\frac{\partial J(y)}{\partial y}\right|_{y=y_2},$$

hence combining the inequalities we get that

$$\left.\frac{\partial J(y)}{\partial y}\right|_{y=y_2} \leq \frac{J(y_2) - J(y_1)}{y_2 - y_1} \leq \left.\frac{\partial J(y)}{\partial y}\right|_{y=y_1}. \tag{5}$$

However, the free-entry condition in vacancy posting implies that in the submarket (h, τ, W) both firms must be respecting $c(y_i) = q(\theta)\beta J(y_i)$ for $i = 1, 2$. As $c(y_i)$ is a linear function of firm quality $\frac{\partial c(y_i)}{\partial y_i} = c$ for $i = 1, 2$ and therefore from the free-entry condition:

$$c = q(\theta)\beta \left.\frac{\partial J(y)}{\partial y}\right|_{y=y_i}$$

which is a contradiction of the slopes of the two tangents being decreasing as shown in Equation (5). Note that if $c(y)$ is convex and twice differentiable, then the derivatives of $c(y)$ are increasing in y while the derivatives of $J(\cdot)$ are decreasing leading again to a contradiction. The proof for the case in which $y_1 > y_2$ follows the same arguments and leads to a similar contradiction on the implied slopes of the optimal contract and those implied by the free entry condition. □

Lemma 4.4. *Given a state (y, h, τ, W) the optimal contract implies that*

$$-\frac{\partial J_t(h, \tau, y, W; \Omega)}{\partial W} = \frac{1}{u'(w)}$$

[8] As the contrapositive of Definition 2.2 in Rudin (1976), that defines a one-to-one mapping for $(x_1, x_2) \in A$ as $x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2)$.

so that promised utilities and wages move in the same direction.

Proof. The proof follows directly from the envelope theorem and the concavity of the utility function $u(\cdot)$, as discussed in the proof of Proposition 2.3. □

Corollary 4.4. *The Pareto frontier* $J(h, \tau, y, W; \Omega)$ *is decreasing in promised utilities* W.

Proof. The envelope condition in **Lemma 4.4** and note that $u'(\cdot) \geq 0$. □

Proposition 4.1. *Then utility promises are unique and increasing in* y, $\frac{\partial W}{\partial y} > 0$.

Proof. Uniqueness follow directly from **Lemma 4.3** and **Proposition 2.4**, given the existence of an injective mapping $f_v(\cdot)$ and the strict concavity of the objective function in Equation (2). One can prove that W is increasing in y by analyzing the first order condition in y of the problem.

Conditional on χ the entrepreneur has to choose the optimal value y to offer to workers.

For the rest of the proof we consider as given the dependence of the functions on (χ) and consider directly the function $q(\theta(W))$ as $q(W)$. The optimization involves a trade-off which respects the following first order condition:

$$c_y - qJ_y = 0 \tag{6}$$

We also know that the second order condition is strictly negative:

$$-c_{yy} + qJ_{yy} < 0 \tag{7}$$

By the implicit function theorem, the derivative of quation 6 with respect to W is:

$$(-c_{yy} + qJ_{yy} < 0)\frac{\partial y}{\partial W} + q_W J_y + qJ_{Wy} = 0 \tag{8}$$

The first term in parenthesis is negative, as shown above. The second term is positive, given **Lemma 4.2** and the fact that q_W is positive (**Lemma 3.1**). The third term is 0 because of the envelope condition for the firm problem being independent of y (**Lemma 4.4**). This means that, in order for the equality to be respected, $\frac{\partial y}{\partial W} > 0$, which given $f_v(\cdot)$ also implies $\frac{\partial W}{\partial y} > 0$.

□

4.3 Proof of Proposition 2.3

Proof. Consider the firm problem in Equation (5), restated here for convenience

$$J(h,\tau,W,y;\Omega) = \sup_{\{\pi_i,w_i,W_i\}} \sum_{i=1,2} \pi_i \Big(f(y,h;\Omega) - w_i$$
$$+\mathbb{E}_\Omega \left[\widetilde{p}(h',\tau+1,W_i;\Omega')J(h',\tau+1,W_i,y;\Omega')\right] \Big)$$

$$s.t.\ [\lambda]\ W = \sum_{i=1,2} \pi_i \left(u(w_i) + \mathbb{E}_\Omega \widetilde{r}(h',\tau+1,W_i;\Omega')\right),$$
$$\sum_{i=1,2} \pi_i = 1, \quad h' = \phi(h,y).$$

For $i=1,2$, the first order conditions with respect to the wage and the promised utilities are:

$$[w_i]:\ \lambda = \frac{1}{u'(w_i)} \tag{9}$$
$$[W_i]:\ \frac{\partial \widetilde{p}(\cdot)}{\partial W_i} J(\cdot) + \widetilde{p}(\cdot)\frac{\partial J(\cdot)}{\partial W_i} + \lambda \frac{\partial \widetilde{r}(\cdot)}{\partial W_i} = 0. \tag{10}$$

Note that by definition,

$$\widetilde{r}(h,\tau,V;\Omega) \equiv \lambda U(h,\tau;\Omega) + (1-\lambda)\Big[W + \lambda_e \max\{0, R(h,\tau,V;\Omega)\}\Big]$$

therefore we can use the envelope theorem as in Benveniste and Scheinkman (1979), Theorem 1 and the definition in Equation (12) to derive an expression for the derivative of the employment value in $t+1$ as the period ahead of the following:

$$\frac{\partial \widetilde{r}(h,\tau,W;\Omega)}{\partial W} = \widetilde{p}(h,\tau,W;\Omega).$$

Similarly, using the envelope condition on the firm problem and the first order condition for the wage, we can establish that

$$\frac{\partial J(h,\tau,y,W;\Omega))}{\partial W} = -\lambda \ \therefore\ \frac{\partial J(h,\tau,W,y;\Omega))}{\partial W} = -\frac{1}{u'(w)}. \tag{11}$$

Moving these two expressions one period ahead, substituting them in (10), focusing on $\widetilde{p(\cdot)} > 0$ and $\pi_i > 0$ and rearranging we have that:

$$\frac{\partial \widetilde{p}(\Theta)}{\partial W_i}\frac{J(\Theta)}{\widetilde{p}(\Theta)} = \frac{1}{u'(w_i')} - \frac{1}{u'(w)},$$

with $\Theta \equiv (\phi(h,y),\tau+1,W;\Omega')$ and where w' is the wage next period in state Ω'. □

5 Derivation of recursive contract SPFE

Solving the optimal contract and the overall model given the recursive structure obtained by following the promised utility method of Spear and Srivastava (1987) is computationally infeasible. This is due to the fact that the optimal contract requires to define a valid recursive domain and codomain of promised values that respects all the future forward looking constraints. Known solution methods for these kinds of models (Abreu, Pearce and Stacchetti, 1990), although robust, easily become computationally unmanageable as the number of states of the model increases. We thus follow Marcet and Marimon (2019) in deriving a recursive expression for the optimal contract in which the Lagrange multiplier for the promise keeping constraint **Equation 11** is added as a co-state of the model, and allows us to circumvent the problem of searching for valid promised values domains altogether.

The reason why the recursive contracts method in Marcet and Marimon (2019) simplifies our problem is simple. As shown in **Equation 11**, wage growth and levels in any next period and at every node are determined by the state-contingent multiplier on tomorrow's promise keeping constraints. This considerably reduces the complexity of the problem, as by definition Lagrange multipliers are defined over $\mathbb{R}^+$.

We follow Marcet and Marimon (2019) (hereby MM) and their terminology to define how a recursive saddle point functional equation (SPFE) can be obtained from the sequential formulation of the problem. For the present exposition of the constructive method to obtain the SPFE, for simplicity and without loos of generality, we ignore the randomization of the contract over the lotteries and the limited liability constraint. The latter choice, in particular, does not create any problem in terms of thinking about of developing the sequential problem over time: our choice of timing of exit decision is such as that exiting firms know form the start of their period whether the productivity level is below the critical one $a^*_{h,\tau,\iota,y}$ for the match (h, τ, ι, W, y), and thus whether they will exit or not. The lack of uncertainty and optimization over the next periods makes the problem of these firms, at some low states, equivalent to the problem of a firm with a lower maximum length (which is T, the retirement age, in general). At an exiting state t the firm knows *with certainty* that any $J_j = 0$ for $j > t$, match with a worker of age T.

Consider the problem[9]

[9] Without loss of generality, we ignore the level of education ι in the proof, as it is a fixed worker characteristic.

$$J_t(h_t, \tau_t, y_t, W_t, a_t) = \sup_{w_t, \{W_{s^{t+1}}\}} \Big(f(a_t, y_t, h_t) - w_t$$
$$+\mathbb{E}_{s^t} \left[\widetilde{p}(h_{t+1}, \tau_{t+1}, W_{s^{t+1}}, a_{s^{t+1}})(J_{t+1}(h_{t+1}, \tau_{t+1}, y_t + 1, W_{s^{t+1}}, a_{s^{t+1}}) \right] \Big) \tag{12}$$

$$s.t.\ W_t = u(w_t) + \beta \mathbb{E}_{s^t} \Big(\lambda U_t(h_{t+1}, \tau_{t+1}, a_{t+1}) +$$
$$(1-\lambda)(\lambda_e p_{t+1}(h_{t+1}, \tau_{t+1}, W_{s^{t+1}}, a_{s^{t+1}}) v^*(h_{t+1}, \tau_{t+1}, W_{s^{t+1}}, a_{s^{t+1}})$$
$$+ (1 - \lambda_e p_{t+1}(h_{t+1}, \tau_{t+1}, W_{s^{t+1}}, a_{s^{t+1}})) W_{s^{t+1}}) \Big) \tag{13}$$

We define as endogenous states $\mathbf{x}_t = [h_t, \tau_t, y_t, W_t]$, controls $\mathbf{c}_t = [w_t, W_{s^{t+1}}]\ \forall t, s^{t+1}$, whereas the only exogenous state is a_t. The endogenous states follow the law of motion

$$\mathbf{x}_{t+1} = \begin{bmatrix} h_{t+1} \\ \tau_{t+1} \\ y_{t+1} \\ W_{s^{t+1}} \end{bmatrix} = l(\mathbf{x}_t, \mathbf{c}_t, a_{s^{t+1}}) = \begin{bmatrix} \phi(h_t, y_t) \\ \tau_t + 1 \\ y_t \\ W_{s^{t+1}} \end{bmatrix} \tag{14}$$

In the subsequent notation, where appropriate, we omit listing all states on which elements in the equation, and subsume their dependence under just listing the time t. J can be rewritten, by developing forward the recursion until time T, at which the match surely dissolves, as

$$J_t(\{h_t, \tau_t, y_t, W_t, a_t\}_{t=t_0}^{T-t_0}) = \mathbb{E}_{t_0} \sum_{t=t_0}^{T} \beta^{t-t_0} \prod_{i=0}^{t-t_0} \widetilde{p}_{t_o+i} \Big(f(a_t, y_t, h_t) - w_t \Big) \tag{15}$$

where $\widetilde{p}_{t_0} = 1$. Notice that the forward-looking constraint in **Equation 13** is state contingent and an instance of it applies at *every* node of any possible history $s^t\ \forall t$ given the prevailing W_y promised at that node. The equilibrium is an instance of subgame perfect Nash equilibrium in which an agent chooses its strategies while anticipating the best response of the following agent, as common in dynamic games with a leader-follower component introduced by Von Stackelberg (1934). The structure of the problem and the solution also shares some commonality with Ramsey optimal policy problems in which a policy maker (in this case the firm) optimizes the utility of all agents according to some weights and taking into account their optimal behavior.[10]

[10] In the terminology of MM, we treat constraints coming from **Equation 13** as a set of one period ahead forward looking constraint, which makes the analysis of our case akin to their case where one have $j = 1$ forward looking constraints, and $N_1 = 0$. The difference with their problems, however, is that our problem features finite time, and thus each one period ahead forward looking constraint technically applies to a *different* function j_t (indexed by t).

We can redefine the problem:

$$V_{t_0}(\mathbf{x}_t, a_t) = \underset{\{w_{s^t}, W_{s^t}\}}{sup} \mathbb{E}_{t_0} \sum_{t=t_0}^{T} \beta^{t-t_0} \prod_{i=0}^{t-t_0} \widetilde{p}_{t_o+i} \Big(f(a_t, y_t, h_t) - w_t \Big) \tag{16}$$

$$s.t.\ [j=0]: \sum_{t=t_0}^{T} \beta^{t-t_0} \prod_{i=0}^{t-t_0} \widetilde{p}_{t_o+i} \Big(f(a_t, y_t, h_t) - w_t \Big) - R \geq 0 \tag{17}$$

$$[j=1, s^t]:\ W_{s^t} - u(w_{s^t}) - \beta \mathbb{E}_{s^t} \Big(\lambda U_{s^{t+1}} + (1-\lambda)(\lambda_e p_{s^{t+1}} v^*_{s^{t+1}}) + \widetilde{p}_{s^{t+1}} W_{s^{t+1}}) \Big) \geq 0 \tag{18}$$

where the constraint 24 is a slack participation constraint for a sufficiently small R, so that the principal (the firm) is willing to enter the contract in the first place.

In the terminology of MM we can label

$$h_0^0(\mathbf{x}_t, \mathbf{c}_t, a_t) = f(a_t, y_t, h_t) - w_t \tag{19}$$

$$h_1^0(\mathbf{x}_t, \mathbf{c}_t, a_t) = f(a_t, y_t, h_t) - w_t - R \tag{20}$$

$$h_0^1(\mathbf{x}_t, \mathbf{c}_t, a_t) = W_t \tag{21}$$

$$h_1^1(\mathbf{x}_t, \mathbf{c}_t, a_t) = W_t - u(w_t) + \beta \mathbb{E}_t(\lambda U_{t+1} + (1-\lambda)\lambda_e p_{t+1} v^*_{y,t+1}) \tag{22}$$

and define the Pareto problem ($\mathbf{PP}_\mu$)

$$\mathbf{PP}_\mu : V_{\mu,t_0}(\mathbf{x}_t, a_t) = \underset{\{w_{s^t}, W_{s^t}\}}{sup} \mathbb{E}_{t_0} \sum_{t=t_0}^{T} \beta^{t-t_0} \prod_{i=0}^{t-t_0} \widetilde{p}_{t_0+i} \mu^0 \Big(f(a_t, y_t, h_t) - w_t \Big) + \mu^1 W_{t_0} \tag{23}$$

$$s.t.\ [j=0;\ \gamma^0]: \sum_{t=t_0}^{T} \beta^{t-t_0} \prod_{i=0}^{t-t_0} \widetilde{p}_{t_o+i} \Big(f(a_t, y_t, h_t) - w_t \Big) - R \geq 0 \tag{24}$$

$$[j=1, s^t;\ \gamma^1_{s^t}]:\ W_{s^t} - u(w_{s^t}) - \beta \mathbb{E}_{s^t} \Big(\lambda U_{s^{t+1}} + (1-\lambda)(\lambda_e p_{s^{t+1}} v^*_{s^{t+1}}) + \widetilde{p}_{s^{t+1}} W_{s^{t+1}}) \Big) \geq 0 \tag{25}$$

Still following the notation from Marcet and Marimon (2019), we can define the Saddle Point Problem ($\mathbf{SPP}_\mu$) as:

$$
\begin{aligned}
\mathbf{SPP}_\mu : SV_{\mu,t_0}(\mathbf{x}_{t_0}, a_{t_0}) = & \underset{\{\gamma\in\mathbb{R}^l_+\}}{inf} \underset{\{w_{st}, W_{s t_0}\}}{sup} \mu^0 \Big(f(a_{t_0}, y_{t_0}, h_{t_0}) - w_{t_0} \Big) + \mu^1 W_{t_0} + \\
& + \beta \mathbb{E}_t \left(\phi(\mu,\gamma) \sum_{i=0}^{T-t_0} \left[\beta^{t_0+i} \prod_{i=0}^{T-t_0-1} \widetilde{p}_{t_0+1+i} \left(f(a_{t_0+i}, y_{t_0+i}, h_{t_0+i}) - w_{t_0+i} \right) + W_{t_0+i} \right] \right) + \\
& + \gamma^1 \left(u(w_{t_0} + \beta \mathbb{E}_{t_0} \left(\lambda U_{t_0+1} + (1-\lambda)\lambda_e p_{t_0+1} v^*_{y,t_0+1} \right) \right) + \\
& + \gamma^0 \left(f(a_{t_0}, y_{t_0}, h_{t_0}) - w_{t_0} - R \right) \qquad (26)
\end{aligned}
$$

The problem can be restated as a saddle-point problem over a Lagrangian equation

$$
\begin{aligned}
\underset{\gamma_t}{inf} \underset{\{w_{st}, W_{st}\}}{sup} & \mu^0 \Big(f(a_{t_0}, y_{t_0}, h_{t_0}) - w_{t_0} \Big) + \mu^1 W_{t_0} + \\
& \gamma^0 \Big(\big(f(a_{t_0}, y_{t_0}, h_{t_0}) - w_{t_0} \big) - R \Big) + \\
& \gamma^1_{t_0} \Big(- W_{t_0} + u(w_{t_0}) + \beta \mathbb{E}_{t_0} (\lambda U_{t_0+1} + (1-\lambda)(\lambda_e p_{t_0+1} v^*_{t_0+1} + \widetilde{p}_{t_0+1} W_{t_0+1}) \Big) + \\
& \beta \mathbb{E}_{t_0} \Bigg[(\mu_0 + \gamma_0) \sum_{t=t_0+1}^{T} \beta^{t-t_0-1} \prod_{i=0}^{T-t_0-1} \widetilde{p}_{t_0+1+i} \Big(f(a_t, y_t, h_t) - w_t \Big) + \\
& \sum_{t=t_0+1}^{T} \mathbb{E}_t \beta^{t-t_0-1} \prod_{i=0}^{t-t_0-1} \widetilde{p}_{t_0+1+i} \gamma^1_t \Big(- W_t + u(w_t) + \\
& \beta(\lambda U_{t+1} + (1-\lambda)(\lambda_e p_{t+1} v^*_{t+1} + \widetilde{p}_{t+1} W_{t+1}) \Big) \Bigg] \qquad (27)
\end{aligned}
$$

which, thanks to some algebra and the law of iterated expectations becomes

$$
\begin{aligned}
\underset{\gamma_t}{inf} \underset{\{w_{st}, W_{st}\}}{sup} & - \gamma^0 R + \mathbb{E}_{t_0} \sum_{t=t_0}^{T} \beta^{t-t_0} \prod_{i=0}^{t-t_0} \widetilde{p}_{t_0+i} \Bigg[\Big(\mu^0_t + \gamma^0_t \Big) \Big(f(a_t, y_t, h_t) - w_t \Big) + \mu^1_t W_t - \\
& \gamma^1_t \Big(W_t - u(w_t) - \beta(\lambda U_{t+1} - (1-\lambda)\lambda_e p_{t+1} v^*_{t+1} \Big) \Bigg] \qquad (28)
\end{aligned}
$$

where $\mu^0_t = \mu^0 = 1$, $\gamma^0_t = \gamma_0 = 0$, $\mu^1_t = \gamma^1_{t-1}$ for some starting $\gamma^1_{t_0-1}$.

The problem can now be written in recursive form. Define

$$
\mathcal{P}_t(h_t, \tau_t, y_t, a_t, \gamma_t) = \underset{W_t}{sup}\, J_t(h_t, \tau_t, y_t, W_t, a_t) + \mu^1_t W_t \qquad (29)
$$

Given **Equation 28** the SPFE of the problem can be written as

$$\mathcal{P}_t(h_t, \tau_t, y_t, a_t, \gamma_t) = \underset{\gamma_t}{inf}\,\underset{w_t}{sup}(f(a_t, y_t, h_t) - w_t) + \mu_t^1 W_t - \gamma_t(W_t - u(w_t)) + \\ \beta \mathbb{E}_t(\lambda U_{t+1} + (1-\lambda)\lambda_e p_{t+1} v_{t+1}^*) + \beta \mathbb{E}_t \widetilde{p}_{t+1} \mathcal{P}_{t+1}(h_{t+1}, \tau_{t+1}, y_{t+1}, a_{t+1}, \gamma_{t+1}) \tag{30}$$

One can easily verify that the solution of this equation is the same we found in the maximization of **Equation 5** in the main text. Take the first order conditions and compute the envelope condition:

$$[FOC\ w_t]:\ -1 + \gamma_t u'(w_t) = 0 \tag{31}$$

$$[ENV\ W_t]:\ \frac{\partial \mathcal{P}_t}{\partial W_t} = \mu_t^1 - \gamma_t \tag{32}$$

$$[FOC\ W_{t+1}]:\ -\widetilde{p}_{t+1} W_{t+1} \gamma_t + \frac{\partial \widetilde{p}_{t+1}}{\partial W_{t+1}} \mathcal{P}_{t+1} + \widetilde{p}_{t+1} \frac{\partial \mathcal{P}_{t+1}}{\partial W_{t+1}} = 0 \tag{33}$$

where **Equation33** is obtained by adding and subtracting from **Equation 30** $\beta \gamma_t \widetilde{p}_{t+1} W_{t+1}$. The reader should also keep in mind that the condition in **Equation 33** is actually state contingent and applied to *all* future states next period, with a different set of co-states $\gamma_{s^{t+1}}$ for each realization of a_{t+1}.

Some rearranging of the **Equation 33** leads to the following result

$$\frac{\partial \log \widetilde{p}_{t+1}}{\partial W_{t+1}} \Big(\mathcal{P}_{t+1} - \gamma_t W_{t+1} \Big) = \gamma_{t+1} - \mu_{t+1}^1 \tag{34}$$

which, given the law of motion of the co-states and the definition in **Equation 29** can be re-written as:

$$\frac{\partial \log \widetilde{p}_{t+1}}{\partial W_{t+1}} J_{t+1} = \frac{1}{u'(w_{t+1})} - \frac{1}{u'(w_t)} \tag{35}$$

which is exactly **Equation 14**, namely the Euler equation that governs the behavior of wage setting and disciplines the provision of insurance within the contract.

6 Data construction and sample selection

We rely on two main data sources provided by INPS through the VisitINPS Program: i) data on employment relationships in Italy from 1996 to 2018 (*Uniemens* dataset) and ii) balance sheet data for incorporated Italian firms from 1996 to 2018 (*Cerved* dataset).

Starting from data on virtually the universe of Italian private employment relationships at the contract level, we calculated for each worker their yearly gross real earnings and for the workers with multiple contracts we selected the information associated to the highest paid contract in the year and we find the associated annualized income using information on the number of actual weeks worked. For workers that experience a job flow or are promoted after 2009 we obtain the education level from the *Comunicazioni Obbligatorie* dataset, provided to INPS by the Ministry of Labor, and identify graduates and non-graduates. We label as graduates in the data both workers getting undergraduate education or above, or workers getting specialized diplomas after high-school.

We restrict our focus on workers employed under either full-time or part-time working schemes between 16 and 65 years old. Moreover, in order to compute AKM fixed effects and maintain the estimation computationally feasible, we analyze the connected set of firms and workers across firms with more than 15 employees. As in Bonhomme, Lamadon and Manresa (2019) we first cluster firms by means of a weighted K-means clustering based on deciles of their wage distribution. The weight for the clustering is the yearly number of employees.

From *Cerved*, we have access to firm balance sheet data. From this dataset we obtain information of value added, overall labor costs and calculate quantiles of value added per employee. All monetary values (wages, value added, total compensation) are trimmed at the 1% level and deflated by the consumer price index for Italy.

7 Model solution and estimation

Model solution. For each education level, we solve the model by backward induction on a regular grid of human capital, wage levels, firm quality and aggregate shocks.[11] In particular, starting from the last period of workers lives we compute the terminal wage in each state, then solve the search problems for both the employed and the unemployed, which allows us to compute the market tightness in each submarket as well as the value of the contract and of unemployment. With these objects in hand we proceed backwards until workers labor market entry.

In the model simulations, we populate each cohort of agents with 42 individuals, 30 non-graduates and 12 graduates. Graduates enter the labor market in a staggered manner every quarter for the first three years of their life. We then run the model for 600 periods (180 burn in) and construct a panel with 419 periods and $7,560$ individuals per period.

[11] We use a grid of $20 \times 15 \times 25 \times 7$ for each education level and we consider 180 quarters of workers life.

Estimation. To pin down the 18 internal parameters in the model, $\boldsymbol{\theta}$, we use Simulated Method of Moments to minimize the absolute error between model simulated moments, $m(\boldsymbol{\theta})$, and their empirical counterparts, $\mathbf{d}$. More formally, we pick the vector of parameters

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \left\{ (|m(\boldsymbol{\theta}) - \mathbf{d}|)' W (|m(\boldsymbol{\theta} - \mathbf{d}|) \right\}.$$

In $\mathbf{d}$ we stack the profiles of E-E and E-U transitions by age (9 age groups); wage growth by education level and experience (9 experience groups); the level of unemployment; the correlations of labor market flows with the cyclical component of output; the average inactivity rate; average sorting and the ratios of the second and third quintile of value added per employee to the first. Jointly, these moments deliver 44 restrictions for 18 parameters. We weight each moment equally, meaning that each profile, as a whole, has the same weight of the other moments and, within each profile, we give more weight to age groups that have a higher demographic share in the Italian population.[12] We collect this weighting scheme in the matrix W.

In order to globally minimize the distance between model-generated and data-generated moments, we adopt two complementary strategies. The first is to minimize the distance between model-generated and data-generated moments using a global solver on the largest feasible domain; for this, we rely on the particle swarm algorithm contained in the library developed by (Blank and Deb, 2020). The second is to solve and simulate the model over a sparse grid of the parameter space.[13] In both cases, we use the combination of parameters that delivers the smallest error as the starting point of a local minimization routine (for this, we rely on the Nelder-Mead algorithm).

8 Additional Figures and Tables

In order to estimate the effect of entering the labor market in a recession we use an age-cohort-period model in which we break the collinearity among the three set of fixed effects by proxying the cohort fixed effects with the cyclical component of real GDP (Hamilton filtered). In particular, we estimate the following yearly model:

$$\log(w)_{t,c,e} = \Phi_t + \Phi_e + \beta_e \widetilde{Y}_c \times \Phi_e + u_{t,c,e}, \tag{36}$$

[12] In other words, the EE rate for 18-22 years old has 1/9 times the share of 18-22 years old in the Italian population (0.09) weight than the unemployment rate.

[13] The advantage of selecting the parameters with this procedure is that the grid is built so that the function domain is optimally covered with the least amount of possible points compared to other forms of approximation (e.g. equi-spaced or random grids). The sparse grid is built using the"Tasmanian" libraries (Stoyanov, 2015).

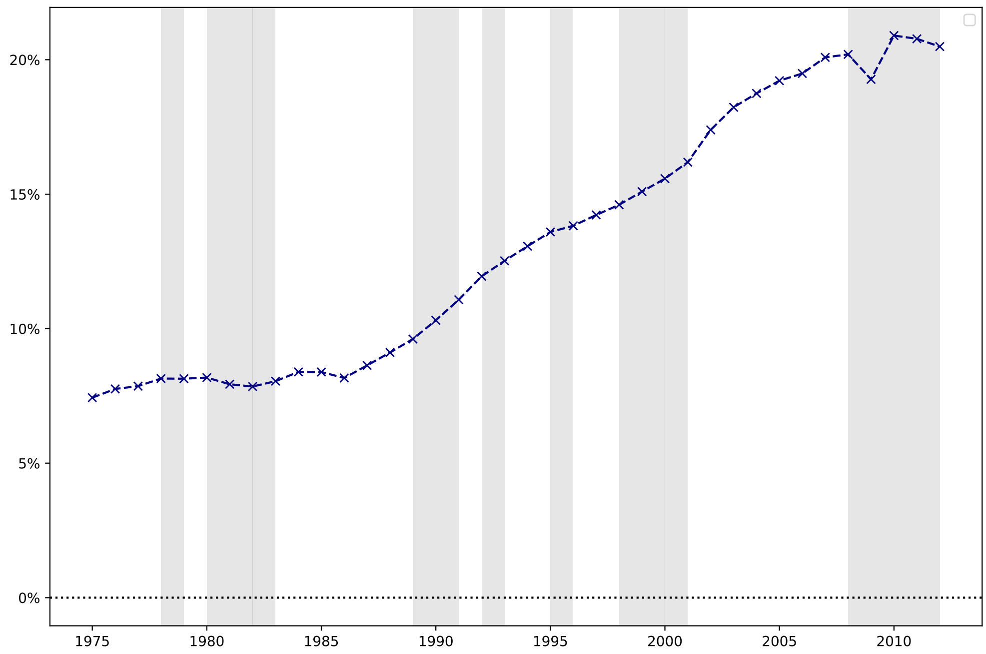
where Φ_t, Φ_e, are dummies for calendar years and labor market experience, and $\widetilde{Y}_c$ is the cyclical realization of real GDP for cohort c at time of their labor market entry. The set of coefficients β_e, therefore, estimate the effect of aggregate conditions on real wages at each year of labor market experience.

Table 1. Effects of initial aggregate conditions along the experience profile and experience growth profile

Dep.Variable: Log-Wage	Experience × Cycle	Experience
Experience Dummy		
0	1.968	
	(0.452)	
1	1.463	0.150
	(0.238)	(0.011)
2	1.473	0.246
	(0.283)	(0.014)
3	1.239	0.304
	(0.343)	(0.014)
4	1.250	0.342
	(0.349)	(0.015)
5	1.239	0.375
	(0.349)	(0.015)
6	1.301	0.399
	(0.287)	(0.015)
Age FE	✓	✓
Year FE	✓	✓
Sex FE	✓	✓
LLM FE	✓	✓
R^2	0.89	0.89
N	254,000,000	254,000,000

Note: The table reports regression coefficients for the empirical estimates in the data used to construct the profiles in Figure 7.

Figure 1. Tertiary School Enrollment and the Business Cycle



Note: The figure plots the ratio of students enrolled in tertiary education to population aged 16-29 in every year. Source: Italian National Institute of Statistics (ISTAT). Shaded areas indicate the OECD based Recession Indicators for Italy.

References

Abreu, Dilip, David Pearce, and Ennio Stacchetti. 1990. "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring." *Econometrica*, 58(5): 1041–1063. 19

Arellano-Bover, Jaime. 2020. "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size." Working Paper. 5

Ashenfelter, Orley. 1978. "Estimating the effect of training programs on earnings." *The Review of Economics and Statistics*, 47–57. 6

Balke, Neele, and Thibaut Lamadon. 2022. "Productivity Shocks, Long-Term Contracts and Earnings Dynamics." *American Economic Review*, 112(7): 2139–2177. 6

Benveniste, L.M., and J.A. Scheinkman. 1979. "On the Differentiability of the Value Function in Dynamic Models of Economics." *Econometrica*, 47(3): 727–732. 18

Blank, Julian, and Kalyanmoy Deb. 2020. "Pymoo: Multi-objective optimization in python." *IEEE Access*, 8: 89497–89509. 25

Bonhomme, Stephane, Thibaut Lamadon, and Elena Manresa. 2019. "A Distributional Framework for Matched Employer Employee Data." *Econometrica*, , (87): 1–71. 24

Engbom, Niklas. 2020. "Labor market fluidity and human capital accumulation." *New York University mimeo*. 7

Marcet, Albert, and Ramon Marimon. 2019. "Recursive Contracts." *Econometrica*, 87(5): 1589–1631. 8, 19, 21

Menzio, Guido, and Shouyong Shi. 2010. "Block recursive equilibria for stochastic models of search on the job." *Journal of Economic Theory*. 8, 9, 11

Rudin, Walter. 1976. *Principles of Mathematical Analysis.* McGraw-Hill Book Company. 16

Shi, Shouyong. 2009. "Directed search for equilibrium wage-tenure contracts." *Econometrica*, 77(2): 561–584. 9

Spear, Stephen E., and Sanjay Srivastava. 1987. "On Repeated Moral Hazard with Discounting." *The Review of Economic Studies*, 54(4): 599. 19

Stoyanov, M. 2015. "User Manual: TASMANIAN Sparse Grids." Oak Ridge National Laboratory ORNL/TM-2015/596, One Bethel Valley Road, Oak Ridge, TN. 25

Thomas, Jonathan, and Tim Worral. 1988. "Self-enforcing wage contracts ." *The Review of Economic Studies*, 55(4): 541–553. 6

Tsuyuhara, Kunio. 2016. "Dynamic Contracts With Worker Mobility Via Directed on-the-Job Search." *International Economic Review*, 57(4): 1405–1424. 6

Von Stackelberg, Heinrich. 1934. *Marktform und gleichgewicht.* J. springer. 20