Worker replacement

Guido Menzio[a,*], Espen R. Moen[b†]

[a] University of Pennsylvania and NBER; [b] Norwegian School of Management

Received Date; Received in Revised Form Date; Accepted Date

Abstract

Consider a labor market in which firms want to insure existing employees against income fluctuations and, simultaneously, want to recruit new employees to fill vacant jobs. Firms can commit to a wage policy, i.e. a policy that specifies the wage paid to their employees as a function of tenure, productivity and other observables. However, firms cannot commit to employ workers. In this environment, the optimal wage policy prescribes not only a rigid wage for senior workers, but also a downward rigid wage for new hires. The downward rigidity in the hiring wage magnifies the response of unemployment to negative shocks.

Keywords: Directed Search, Wage Rigidity, Unemployment, Business Cycles

JEL classification: E24, E32, J64

*Corresponding author: gmenzio@sas.upenn.edu

†Menzio: University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104, USA. Moen: Norwegian School of Management, Handelshøyskolen BI, 0442 Oslo, Norway. We are grateful to the editor and to an anonymous referee for their comments. We also received comments from participants at the Society for Economic Dynamics Meeting in Vancouver (July 2006), the NBER Summer Institute in Cambridge (July 2006), and at the macroeconomic seminars at Stanford, Penn, LSE, UCLA and Saint Louis FED. Discussions with Ken Burdett, Dale Mortensen, Guillaume Rocheteau and Randy Wright led to significant improvements in the paper. Menzio gratefully acknowledges the financial support and the hospitality of the Hoover Institution. Moen gratefully acknowledges financial support from the Norwegian Research Council.

1. Introduction

Shimer (2005) documents that, according to the standard search model of the labor market (Pissarides, 1985), productivity shocks account for an implausibly small fraction of the empirical volatility of unemployment and vacancies. This finding has generated a significant research effort devoted to reconciling the predictions of the model with the empirical behavior of the labor market. Hall (2005) and Gertler and Trigari (2009) are among the most prominent examples of this effort. These papers show that, when wage rigidities are introduced into the model, productivity shocks lead to much larger and more plausible fluctuations in unemployment and vacancies. Interestingly, in these papers the source of amplification of productivity shocks is not the rigidity of the wage of workers who are already employed, but the rigidity of the wage of newly hired workers (more precisely, the rigidity of the present value of wages of newly hired workers). Indeed, the fact that the wage offered to a worker in an ongoing employment relationship does not respond to changes in productivity only affects the distribution of rents between the worker and his employer. In contrast, the fact that the wage offered to new hires does not respond to changes in productivity affects the firms' incentives to create vacancies and, in turn, affects unemployment. However, why should the wage of new hires be rigid? This paper answers the question.

We consider a directed search model of the labor market in the spirit of Moen (1997) and Shimer (1996). Firms enter the market and announce their wage policy, i.e. a policy that specifies the wage paid to a worker as a function of his tenure, the idiosyncratic productivity of the firm, and the aggregate state of the economy. Unemployed workers observe the wage policy of new and old firms and choose where to apply for jobs. The application process is subject to matching frictions. Because workers cannot access the capital market, firms have an incentive to insure their employees against income fluctuations. To accomplish this task efficiently,[1] firms would like to offer to each of their workers a constant wage. Because of

[1] That is, to accomplish this task in a way that maximizes profits.

search frictions and turnover, firms also have an incentive to attract new applicants to fill vacant jobs. To accomplish this task efficiently, firms would like to offer to new hires a wage that varies in response to changes in aggregate productivity. Hence, in some states of the world, firms would like to offer a hiring wage lower than the wage paid to existing employees. But what happens if, in one of these states, a firm finds an applicant that is qualified for the job held by one the senior workers? If the firm cannot commit to the employment relationship, it will simply replace the senior worker with the new applicant. This "worker replacement" problem creates a tension between efficient provision of insurance to senior workers and efficient recruitment of new hires.

The first part of the paper characterizes the optimal wage policy of the firm in the presence of the worker replacement problem. We find that, when the firm is hit by a positive productivity shock, the optimal policy prescribes that senior workers should be paid the same wage as in the past (the efficient insurance wage), and that new hires should be paid the wage that maximizes the value of the firm's vacant jobs (the efficient hiring wage). These wages do not induce the firm to replace senior workers with new recruits, because the efficient hiring wage is greater than the efficient insurance wage. When the firm is hit by a negative productivity shock, the optimal policy prescribes that senior workers and new recruits should be paid exactly the same wage. The common wage is smaller than the wage paid to senior workers in the past, and greater than the wage that maximizes the value of the firm's vacant jobs. The common wage guarantees that the firm will not replace senior workers with new recruits, but it distorts the efficient hiring wage up and the efficient insurance wage down. Overall, the worker replacement problem induces a form of downward rigidity in the wage offered to new hires by old firms.

The second part of the paper studies how the worker replacement problem affects the response of the labor market to productivity shocks. Here, the main finding is that the worker replacement problem amplifies the decline in vacancies and the increase in unemployment in response to a negative shock, but has no effect on the response of the labor market to a positive shock. When the economy is hit by a negative shock, the worker replacement problem dampens the decline in the wage offered by old firms to new hires. This downward wage rigidity distorts the flow of applicants towards old firms, amplifies the decline in the number of new firms that enter the labor market, amplifies the decline in vacancies and, ultimately, amplifies the increase in unemployment. In contrast, when the economy is hit by a positive shock, the worker replacement problem does not affect the hiring wage of old firms and, for this reason, it has no effect on vacancies and unemployment. Even though our model is too abstract to carry out a thorough quantitative analysis, a back-of-the-envelope calculation shows that the worker replacement problem can easily account for 20 percent of the increase in unemployment that takes place during a typical recession.

Our theory of downward rigidity in the hiring wage hinges on two elements. First, firms want to offer to each of their employees a constant wage to insure them against income fluctuations. Second, firms do not want to hire new employees for less than the wage of senior employees to avoid the replacement problem. The first element of our theory is the cornerstone of the implicit contract literature (Azariadis, 1975) and is supported by vast empirical evidence. For example, in a survey of US firms, Blinder and Choi (1990) find that the majority of managers agree that wages contain a countercyclical insurance payment. There is also empirical evidence supporting the second, novel element of our theory. For example, in a recent survey of Swedish firms, Agell and Bennmarker (2007) find that 90 percent of managers are not willing to hire new workers for less than the wage of existing workers. This finding is particularly striking considering that the survey was conducted in the midst of a severe recession. Similarly, in a survey of US firms, Bewley (1999) finds that 79 percent of managers refuse to hire undercutters. Moreover, in both surveys, the majority of managers argue (in accordance to our theory) that hiring undercutters would either violate the personnel policy of the firm or some bargaining agreement with the unions.

The theory advanced in this paper is an alternative to recent theories of wage rigidity based on asymmetric information.[2] In Menzio (2005), firms do not want to increase the hiring wage in expansions because they do not want to reveal to existing workers that the gains from trade have increased and, hence, their wages can be renegotiated. In Kennan (2009), workers do not demand higher wages in expansions because they do not want to be turned down by those firms whose productivity did not increase. In Moen and Rosen (2008), firms are reluctant to cut pay in response to negative shocks because of efficiency wage considerations.

2. Model

This section develops a model of directed search à la Moen (1997) and Shimer (1996). In the model, firms have an incentive to insure their employees against income fluctuations because workers are risk averse and cannot access the capital market. Firms also have a desire to attract applicants to fill the jobs that become vacant because of turnover. However, the goals of insuring workers and attracting applicants may be in conflict because the firm cannot commit not to replace senior employees with cheaper new hires.

2.1. Preferences and technologies

There is a continuum of workers with measure one. A worker has preferences given by $\mathbb{E}\sum_{t=0}^{\infty}\beta^t \upsilon(c_t)$, where β is the discount factor, and υ is a strictly increasing and strictly concave utility function. A worker consumes his income each period, so consumption c

[2] Independently from us, Snell and Thomas (2008) develop a related theory of wage rigidity. They consider a frictionless labor market populated by risk neutral firms and risk adverse workers who have no access to the credit market. Under the assumption that a firm cannot pay a different wage to workers with different tenure, they find that the wage policy of the firm prescribes a wage that responds less to aggregate productivity shocks than a spot market wage. Moreover, under the assumption that new firms cannot enter the labor market, they find that there is involuntary unemployment and that involuntary unemployment is countercyclical. Our paper differs from theirs for two main reasons. First, in our paper, the link between the wage of existing workers and the wage of new hires emerges endogenously as part of the optimal solution to the replacement problem. Second, in our paper, the rigidity in the wage of new hires amplifies the response of unemployment because it affects the search strategy of unemployed workers and it crowds out the entry of new firms.

equals the wage w if the agent is working and $b > 0$ if the agent is unemployed.

There is a continuum of entrepreneurs with positive measure. An entrepreneur has preferences given by $\mathbb{E}\sum_{t=0}^{\infty}\beta^t c_t^e$, where c^e is the sum of profits from firms the entrepreneur owns.[3] An entrepreneur can start a firm at the cost $k > 0$. A firm operates a technology that turns labor into output according to the production function $\min\{\overline{n}, n\}y$, where $\overline{n}$ is the measure of jobs at the firm, n is the measure of workers employed by the firm, and y is the firm's productivity. The productivity of a firm in its first period of activity is y_1, where y_1 is a discrete random variable drawn from the probability distribution $\pi_1(y_1|x)$. The productivity of a firm in its second period of activity is y_2, where y_2 is a discrete random variable drawn from the distribution $\pi_2(y_2|x)$. To simplify the analysis, we assume that firms exit the market after two periods. The distribution of firm's productivities depends on the aggregate state of the economy x, which follows a Markov chain with transition probabilities $\pi_x(x'|x)$.

A period is divided into four stages: separation, entry, search and production. At the separation stage, an employed worker becomes unemployed for exogenous reasons with probability δ. Moreover, at this stage, an employed worker has the option to leave his job and become unemployed, and a firm has the option to dismiss any of its workers.

At the entry stage, an entrepreneur chooses how many new firms to create. After realizing the productivity y_1, a new firm announces the wage policy $\sigma = (w_1, w_{2i})$. The first element of σ is the wage paid to a worker who is employed in the first period of the firm's life. The second element is a function $w_{2i}(x', y_2', \mu_2')$ which returns the wage paid to a worker who is employed in the second period of the firm's life, conditional on the worker's tenure i, the aggregate state of the economy x', the firm's productivity y_2', and a sunspot μ_2'. The sunspot is a random variable drawn from the uniform distribution with unit support.[4]

[3] As discussed in Rudanko (2009), the assumption that entrepreneurs and workers have a different preference for risk can be motivated in a number of ways. First, if becoming an entrepreneur or a worker is a choice, it is natural to expect less risk averse agents to become entrepreneurs and more risk averse agents to become workers. Second, if entrepreneurs are wealthier than workers, they might behave as if they were less risk averse. Finally, it might be the case that entrepreneurs have better access to asset markets and, hence, can insure away risk.

[4] Because of the sunspot, a change in the second period wages $w_{2i}(x', y_2', \mu_2')$ has a marginal effect on the

If a worker was unemployed at the beginning of the period, he has the opportunity to apply for a job at the search stage. For simplicity, we assume that a worker cannot apply for a job if he is employed or if he became unemployed during the separation stage. The application process is directed, in the sense that a worker applies for a particular job at a particular firm.[5] If the worker applies for a job of type W, his application is successful with probability $p(\theta(W))$, where $\theta(W)$ is the ratio of applicants to jobs of type W and p is a strictly decreasing function such that $p(0) = 1$ and $p(\infty) = 0$. Following Acemoglu and Shimer (1999), we refer to $\theta(W)$ as the queue length for the job. Symmetrically, the firm finds a successful applicant for a job of type W with probability $q(\theta(W))$, where q is a strictly increasing function such that $q(\theta) = p(\theta)\theta$, $q(0) = 0$, $q(\infty) = 1$. Moreover, the elasticity of q wrt θ, $\epsilon_q(\theta)$, is such that $q(\theta)\epsilon_q(\theta)/[1 - \epsilon_q(\theta)]$ is a decreasing function of θ.[6] When a firm finds a successful applicant for a job, it has the option to hire him and dismiss any other worker who might have been holding the job. The type of a job is the value that it offers to a successful applicant.

At the production stage, production and consumption take place. Then nature draws the aggregate state of the economy x', the idiosyncratic productivity of the firm y_2', and the sunspot μ_2' for next period. The draws of the productivity and the sunspot are independent across firms.

The model incorporates some new elements in a model of directed search. First, the model assumes that entrepreneurs are risk neutral and that workers are risk averse and cannot access the capital market. The assumption is common in the literature on implicit contracts and

---

lifetime utility W_1 provided to a worker hired in the first period of the firm's life. This property allows to establish the necessary conditions (K) and (R), which are used to characterize the optimal wage policy. As an alternative to the sunspot, one could assume that the productivity of the firm is a continuous random variable.

[5] The qualitative results of our paper are robust to an alternative specification of the search process in which workers apply to firms rather than to specific jobs.

[6] The assumptions on $p(\theta)$ and $q(\theta)$ are satisfied by many standard matching processes. For example, they are satisfied by the urn-ball matching process, $p(\theta) = \theta(1 - \exp(-1/\theta))$ and $q(\theta) = 1 - \exp(-1/\theta)$, and by the telephone-line matching process, $p(\theta) = \theta/(1 + \theta)$ and $q(\theta) = 1/(1 + \theta)$.

it implies that firms have an incentive to insure their workers against income fluctuations (Azariadis, 1985). Second, the model assumes that firms have limited commitment, in the sense that firms can commit to a wage policy but cannot commit to employ any of their workers. The assumption is also present in some of the literature on implicit contracts (see e.g. MacLeod and Malcomson 1989 or Thomas and Worrall 1988). It is usually motivated by the fact that a third party (say a court of law or a labor union) can easily verify how much a worker is paid, but not whether a worker who has left the firm has done so voluntarily, because of poor performance, because his job ceased to exist, or, in our case, because the firm replaced him with a new hire. Lack of verifiability also motivates the assumption that the wage paid by the firm to a worker cannot depend on whether the firm has found a replacement for him, or on whether the worker has been hired as a replacement of somebody else.[7]

Rudanko (2009) also introduces elements from the implicit contract literature in a model of directed search. The main difference between her model and ours is the size of a firm.[8] In her model, a firm has only one job. If the job is vacant, the only goal of the firm is to recruit. If the job is filled, its only goal is to insure the worker. In our model, a firm has many jobs. Therefore, the firm needs to insure existing workers and recruit new workers at the same time. Because of limited commitment, there may be a tension between these two goals. The resolution of this tension is the key to our theory of rigidity for the wage of new hires.

[7]Note that we do not allow for wage renegotiation when the firm finds a replacement for the worker. It is easy to justify this modeling choice. Assume that the firm privately observes whether a replacement for the worker is available. Under this assumption, the firm has the incentive to announce that it has found a replacement whether this is actually true or not. Hence, the worker ignores the announcement of the firm and renegotiation does not take place. An analogous argument is used in many models of search on the job to justify the assumption that wage renegotiation does not take place when an employed worker finds an alternative employer (see e.g. Burdett and Mortensen 1998).

[8]Another difference between Rudanko's model and our model is the life span of a firm. In her model, a firm has a constant probability of exiting the labor market in each period. In our model a firm exits the labor market deterministically after two periods.

2.2. Competitive search equilibrium

Let $Z(x)$ be the lifetime utility of an unemployed worker at the search stage, given that the worker has the opportunity to apply for a job and the state of the economy is x. We refer to Z as the value of search. Let $U(x) \leq Z(x)$ be the lifetime utility of an unemployed worker at the production stage, given that the state of the economy is x. We refer to U as the value of unemployment. Given U and Z, the queue length for a job of type W is

$$\theta(W,x) = \begin{cases} \theta : p(\theta)W + (1-p(\theta))U(x) = Z(x), \text{ if } W > Z(x), \\ 0, \text{ if } W \leq Z(x). \end{cases} \tag{1}$$

If the value of the job to a successful applicant, W, is greater than the value of search, Z, the queue length is driven up to the point where workers are indifferent between applying for the job and searching somewhere else. If W is smaller than Z, the queue length is zero. Notice that, if a worker applies for a filled job at a firm with a wage policy that specifies $w_{21} \geq w_{22}$, the worker will not be hired even if his application is successful. Hence, the value of this job to a successful applicant is equal to the value of unemployment U, and the queue length is zero. In contrast, if a worker applies for a filled job at a firm with a wage policy that specifies $w_{21} < w_{22}$, the worker will be hired if his application is successful. Hence, the queue length of this job may be positive.

Given the value of search, Z, and the queue length for different types of jobs, θ, one can compute the value of unemployment and employment to the worker, as well as the value of a firm to the entrepreneur. The value of unemployment to the worker is

$$U(x) = \upsilon(b) + \beta \mathbb{E} Z(x'). \tag{2}$$

In the current period, the worker consumes b units of output. At the search stage of next period, the worker has the option of sending an application and his continuation utility is Z.

The value to the worker from being employed by an old firm at the wage w is

$$W_2(w,x) = \upsilon(w) + \beta \mathbb{E} Z(x'). \tag{3}$$

In the current period, the worker consumes w units of output. At the end of the current period, the firm exits the labor market and the worker moves into unemployment. At the search stage of next period, the worker has the option of sending a job application and his continuation utility is Z.

The value to the worker from being employed by a new firm with wage policy σ is

$$W_1(\sigma,x) = \upsilon(w_1) + \beta \mathbb{E} \left\{ U(x') + (1-d)(1-\rho\tilde{q}(w_{21},x')) \left[W_2(w_{22},x') - U(x') \right] \right\} \tag{4}$$

where

$$d = \{\delta \text{ if } w_{22} \in [b, y_2'],\ 1 \text{ else}\},\ \rho = \{0 \text{ if } w_{22} \leq w_{21},\ 1 \text{ else}\},$$

$$\tilde{q}(w_{21},x') = q(\theta(W_2(w_{21},x')),x'),$$

and the dependence of w_{21}, w_{22}, d and ρ on (x', y_2', μ_2') is omitted for brevity. In the current period, the worker consumes w_1 units of output. At the separation stage of next period, the worker moves into unemployment and gets the continuation utility U with probability d, where d equals δ if $b \leq w_{22} \leq y_2'$ and d equals 1 otherwise. Intuitively, the worker moves into unemployment for exogenous reasons with probability δ, he voluntarily moves into unemployment if $w_{22} < b$, and the firm dismisses him if $w_{22} > y_2$. At the search stage, the worker is replaced by a new hire and gets the continuation utility U with probability $(1-d)\rho\tilde{q}(w_{21},x')$, where ρ equals 1 if $w_{22} > w_{21}$ and 0 otherwise. Intuitively, if $w_{22} > w_{21}$, the firm has an incentive to replace the worker and will succeed in doing so with probability $\tilde{q}(w_{21},x')$. At the production stage of next period, the worker is still employed with probability $(1-d)(1-\rho\tilde{q}(w_{21},x'))$. In this case, the worker's continuation utility is $W_2(w_{22})$.

The value to the entrepreneur of a new firm with wage policy σ is

$$\begin{aligned} F(\sigma,x,y_1) &= n_1(y_1-w_1)+\beta\mathbb{E}\left\{n_1(1-d)(1-\rho\tilde{q}(w_{21},x'))(y_2'-w_{22})\right\} \\ &\quad +\beta\mathbb{E}\left\{\left[\overline{n}-n_1(1-d)(1-\rho)\right]\tilde{q}(w_{21},x')(y_2'-w_{21})\right\} \end{aligned} \tag{5}$$

where

$$n_1=\overline{n}q(\theta(W_1(\sigma,x),x)).$$

At the search stage, the firm attracts $\theta(W_1(\sigma))$ applicants for each one of its $\overline{n}$ vacant jobs. Hence, at the production stage, the firm employs $n_1=\overline{n}q(\theta(W_1(\sigma)))$ workers. At the search stage of next period, the firm has $\overline{n}-n_1(1-d)$ vacant jobs and $n_1(1-d)$ filled jobs. The firm receives $\theta(W_2(w_{21}))$ applications for each of its vacant jobs. In addition, the firm receives $\theta(W_2(w_{21}))$ applications for each of its filled jobs if $w_{21}<w_{22}$. Hence, at the production stage, the firm employs $n_1(1-d)$ senior workers and $(\overline{n}-n_1(1-d))\tilde{q}(w_{21})$ new hires if $w_{21}\geq w_{22}$, and $n_1(1-d)(1-\tilde{q}(w_{21}))$ senior workers and $\overline{n}\tilde{q}(w_{21})$ new hires if $w_{21}<w_{22}$.

We define an equilibrium along the lines of Moen (1997), Acemoglu and Shimer (1999), and Rudanko (2009).

Definition 1 *A competitive search equilibrium consists of a search value $Z(x)$, a queue length $\theta(W,x)$ and a wage policy $\sigma(x,y_1)$ with the following properties:*
(i) Optimal application: For all (W,x), $\theta(W,x)$ satisfies (1).
(ii) Profit maximization: For all $(x,y_1,\tilde{\sigma})$, $F(\sigma(x,y_1),x,y_1)\geq F(\tilde{\sigma},x,y_1)$.
(iii) Zero profit: For all x, $\mathbb{E}_{y_1}\left[F\left(\sigma(x,y_1),x,y_1\right)\right]-k=0$.

Condition (i) ensures that the queue length is consistent with the optimal application strategy of the workers. Condition (ii) ensures that a new firm chooses its wage policy to maximize profits. Condition (iii) ensures that free entry drives the maximized profits of a new firm down to zero. Formally, condition (i) pins down the queue length as a function of Z, condition (ii) pins down the wage policy as a function of Z, and condition (iii) pins down the equilibrium value of Z.

Note that the competitive search equilibrium is block recursive. That is, the system of equations that pins down the equilibrium value of search, the queue length, and the wage policy of new firms is independent of the distribution of workers across different employment states and the distribution of old firms across different wage policies. As explained in Menzio and Shi (2009a, b), the block recursive nature of the equilibrium is a property common to many models of directed search.

3. Microeconomics of worker replacement

This section characterizes the wage policy of the firm in the presence of the worker replacement problem. From a technical point of view, this task is not trivial because the objective function of the firm (5) is discontinuous with respect to the wage offered to new hires in the second period. In fact, if the hiring wage is greater than the wage of senior workers, the firm does not attract any applications for its filled jobs. However, if the hiring wage is even a penny less than the wage of senior workers, the firm attracts a non-negligible number of applications to each of its filled jobs. Because of this discontinuity, one cannot use standard optimality conditions to characterize the optimal wage policy.

3.1. Necessary condition for optimality

Let $\sigma = (w_1, w_{2i})$ be the optimal wage policy of the firm, and let n_1 be the number of workers employed by the firm in its first period of activity. Then, (w_{21}, w_{22}) must maximize the weighted sum of the firm's profits and the n_1 senior workers' utility in the firm's second period of activity, where the weight on the firm's profits is 1 and the weight on the senior workers' utility is $1/\upsilon'(w_1)$. The intuition behind this necessary condition for optimality is simple and can be given here. A formal derivation of this condition is given in the proof of lemma 2 in the appendix. Intuitively, the choice of (w_{21}, w_{22}) has not only a direct effect on the profits of the firm in the second period, but also an indirect effect on the profits of the firm in the first period, and the magnitude of this indirect effect is precisely the utility of senior workers weighted by $1/\upsilon'(w_1)$. In fact, a choice of (w_{21}, w_{22}) that increases the utility of senior workers in the second period by one unit allows the firm to lower the wage w_1 by $1/\upsilon'(w_1)$ dollars without affecting the lifetime utility W_1 delivered to workers hired in the first period and, hence, without affecting the number of workers n_1 employed by the firm in the first period. For this reason, the optimal wage policy maximizes the weighted sum of the firm's profits and the senior workers' utility in the second period.

Now, restrict attention to the set of wages that do not induce the firm to replace senior workers with new hires, i.e. $w_{22} \leq w_{21}$, and do not induce the firm and the senior workers to voluntarily break up at the separation stage, i.e. $b \leq w_{22} \leq y_2$. Over this set of wages, the maximized sum of the firm's profits and the senior workers' utility is[9]

$$
\begin{aligned}
V_k = \max_{(w_{21}, w_{22})} \quad & n_1 \left[\delta \upsilon(b) + (1-\delta)\upsilon(w_{22})\right] / \upsilon'(w_1) \\
& + n_1(1-\delta)(y_2 - w_{22}) + \left[\overline{n} - n_1(1-\delta)\right] \tilde{q}(w_{21})(y_2 - w_{21}), \\
\text{s.t.} \quad & b \leq w_{22} \leq y_2, \quad w_{22} \leq w_{21}.
\end{aligned}
\tag{K}
$$

The first term on the rhs of (K) is the utility of senior workers weighted by $1/\upsilon'(w_1)$. The utility of each of the n_1 senior workers is $\upsilon(b)$ with probability δ and $\upsilon(w_{22})$ with probability $1-\delta$. The second term and third terms on the rhs of (K) are the profits of the firm. The firm employs $n_1(1-\delta)$ senior workers and earns a profit of $y_2 - w_{22}$ on each of them. Also, the firm employs $\left[\overline{n} - n_1(1-\delta)\right] \tilde{q}(w_{21})$ new hires and earns a profit of $y_2 - w_{21}$ on each of them. Denote as (w_{21}^k, w_{22}^k) the solution to (K).

Next, restrict attention to the set of wages that induce the firm to replace senior workers with new hires, i.e. $w_{22} > w_{21}$, but do not induce the firm and the senior workers to voluntarily break up at the separation stage, i.e. $b \leq w_{22} \leq y_2$. Over this set of wages, the

[9]For the sake of brevity, the dependence of V_k, V_r and $\hat{V}$ on w_1, n_1 and x is omitted.

maximized sum of the firm's profits and the senior workers' utility is

$$\begin{aligned} V_r = \max_{(w_{21},w_{22})} \quad & n_1 \left\{ \upsilon(b) + (1-\delta)(1-\tilde{q}(w_{21})) \left[\upsilon(w_{22}) - \upsilon(b) \right] \right\} / \upsilon'(w_1) \\ & + n_1(1-\delta)(1-\tilde{q}(w_{21}))(y_2 - w_{22}) + \overline{n}\tilde{q}(w_{21})(y_2 - w_{21}), \\ \text{s.t.} \quad & b \leq w_{22} \leq y_2, \quad w_{22} > w_{21}. \end{aligned} \tag{R}$$

The first term on the rhs of (R) is the utility of senior workers weighted by $1/\upsilon'(w_1)$. The utility of each of the n_1 senior workers is $\upsilon(b)$ with probability $\delta + (1-\delta)\tilde{q}(w_{21})$, and $\upsilon(w_{22})$ with complementary probability. The second and third terms on the rhs of (R) are the profits of the firm. The firm employs $n_1(1-\delta)(1-\tilde{q}(w_{21}))$ senior workers and earns a profit of $y_2 - w_{22}$ on each of them. Also, the firm employs $\overline{n}\tilde{q}(w_{21})$ new hires and earns a profit of $y_2 - w_{21}$ on each of them. Denote as (w_{21}^r, w_{22}^r) the solution to (R).

Since the optimal wage schedule maximizes the sum of the profits of the firm and the utility of the senior workers over all (w_{21}, w_{22}), the following result obtains.

Lemma 2 *Let $\sigma^* = (w_1^*, w_{2i}^*)$ be the optimal wage policy for a new firm. Then, σ^* satisfies the following necessary conditions for optimality: (i) If $b \leq w_{22}^* \leq y_2$ and $w_{22}^* \leq w_{21}^*$, then $(w_{21}^*, w_{22}^*) = (w_{21}^k, w_{22}^k)$ and $V_k \geq V_r$; (ii) If $b \leq w_{22}^* \leq y_2$ and $w_{22}^* > w_{21}^*$, then $(w_{21}^*, w_{22}^*) = (w_{21}^r, w_{22}^r)$ and $V_r \geq V_k$; (iii) If and only if $y_2 \geq b$, $b \leq w_{22}^* \leq y_2$.*

The proof of this and other lemmas can be found in an online supplementary appendix.

3.2. Wage policy with and without replacement

Lemma 2 suggests a simple procedure for characterizing the optimal wage policy of the firm. First, characterize the solution to (K). That is, characterize the wages that maximize the sum of the firm's profits and the senior workers' utility subject to the no-replacement constraint, $w_{22} \leq w_{21}$, and the individual rationality constraint, $b \leq w_{22} \leq y_2$. Second, characterize the solution to (R). That is, characterize the wages that maximize the sum of the firm's profits and the senior workers' utility subject to the replacement constraint, $w_{22} > w_{21}$, and the individual rationality constraint. Finally, compare the two maximized sums to identify the prescriptions of the optimal wage policy. The following pages carry out this procedure under the maintained assumption that w_1 and y_2 are greater than z, where z is the consumption equivalent of the flow value of search, i.e. $v(z) = Z(x) - \beta \mathbb{E} Z(x')$.

3.2.1. Wage policy under commitment

As a preliminary step, it is useful to characterize the prescriptions of the optimal wage policy if the firm could commit not to replace senior workers with new hires. In this case, one can show that the optimal wage policy solves the problem

$$
\begin{aligned}
\hat{V} = \max_{(w_{21}, w_{22})} \quad & n_1 \left[\delta v(b) + (1-\delta) v(w_{22})\right] / v'(w_1) \\
& + n_1(1-\delta)(y_2 - w_{22}) + \left[\overline{n} - n_1(1-\delta)\right] \tilde{q}(w_{21})(y_2 - w_{21}), \\
\text{s.t.} \quad & b \le w_{22} \le y_2,
\end{aligned}
\tag{C}
$$

where the objective function is the weighted sum of the firm's profits and the senior workers' utility in the second period given that senior workers are never replaced by new hires.

The optimal hiring wage $\hat{w}_{21}$ satisfies the first order condition

$$
\tilde{q}'(\hat{w}_{21})(y_2 - \hat{w}_{21}) = \tilde{q}(\hat{w}_{21}). \tag{6}
$$

Since the firm can commit not to replace senior workers, the hiring wage $\hat{w}_{21}$ affects the profits of the firm, but does not affect the utility of senior workers. In particular, an increase in $\hat{w}_{21}$ has a positive effect on the profits of the firm because each vacant job is filled with higher probability. This effect is measured by the lhs of (6). On the other hand, an increase in $\hat{w}_{21}$ has a negative effect on the profits of the firm because it reduces the profit margin on each new hire. This effect is measured by the rhs of (6). The optimal hiring wage equates these two effects. The lhs of (6) is positive for $\hat{w}_{21} < y_2$ and negative otherwise. The rhs of (6) is zero for $\hat{w}_{21} \leq z$ and positive otherwise. Hence, the optimal hiring wage lies between the flow value of search z and the productivity of the firm y_2. Moreover, $\tilde{q}'(\hat{w}_{21})$ is a decreasing function of $\hat{w}_{21}$. Hence, the optimal hiring wage is increasing in y_2.

The optimal wage for senior workers $\hat{w}_{22}$ satisfies the first order condition

$$\upsilon'(\hat{w}_{22})/\upsilon'(w_1) \geq 1, \tag{7}$$

and $\hat{w}_{22} \leq y_2$ with complementary slackness. On the one hand, an increase in $\hat{w}_{22}$ raises the weighted utility of each senior worker by $\upsilon'(\hat{w}_{22})/\upsilon'(w_1)$. On the other hand, an increase in $\hat{w}_{22}$ reduces the profit that the firm earns on each senior worker by 1 dollar. The optimal wage for senior workers equates these two effects as long as it satisfies the constraint $\hat{w}_{22} \leq y_2$. Hence, $\hat{w}_{22}$ is equal to the past wage of senior workers if the firm's productivity is greater than w_1, and is equal to y_2 otherwise. From the properties of $\hat{w}_{21}$ and $\hat{w}_{22}$, it follows that there exists a $\hat{y}$ such that the optimal wage for new hires, $\hat{w}_{21}$, is greater than the optimal wage for senior workers, $\hat{w}_{22}$, if the firm's productivity is greater than $\hat{y}$, and such that $\hat{w}_{21}$ is smaller than $\hat{w}_{22}$ if the firm's productivity is less than $\hat{y}$. The properties of $\hat{w}_{21}$ and $\hat{w}_{22}$ are summarized in the following proposition and illustrated in figure 1.

Proposition 3 *Let $(\hat{w}_{21}, \hat{w}_{22})$ denote the solution to (C). (i) For $y_2 > z$, the wage for new hires $\hat{w}_{21}$ lies in the interval (z, y_2), and the wage for senior workers $\hat{w}_{22}$ equals* $\min\{w_1, y_2\}$*. (ii) There exists a $\hat{y}$ such that $\hat{w}_{21} < (>) \hat{w}_{22}$ if $y_2 < (>) \hat{y}$.*

3.2.2. Wage policy without replacement

Now, we return to the case in which the firm cannot commit not to replace senior workers with new hires. The following proposition characterizes the solution to (K), i.e. the second period wages that maximize the weighted sum of the firm's profits and the senior workers' utility subject to the no replacement constraint, $w_{22} \leq w_{21}$, and the individual rationality constraint, $b \leq w_{22} \leq y_2$.

Proposition 4 *Let (w_{21}^k, w_{22}^k) denote the solution to (K). (i) For $y_2 > \hat{y}$, the wage for new hires w_{21}^k equals $\hat{w}_{21}$, and the wage for senior workers w_{22}^k equals $\hat{w}_{22}$. (ii) There exists a $y_k \in (z, \hat{y})$ such that, for $y_k < y_2 < \hat{y}$, w_{21}^k and w_{22}^k are both equal to $w_2^k \in (\hat{w}_{21}, \hat{w}_{22})$. (iii) For $z \leq y_2 \leq y_k$, w_{21}^k and w_{22}^k are both equal to $w_2^k = y_2$.*

The results in proposition 4 are intuitive. When the productivity of the firm is greater than $\hat{y}$, the no-replacement constraint does not bind at $(\hat{w}_{21}, \hat{w}_{22})$. In this case, the optimal hiring wage w_{21}^k equals $\hat{w}_{21}$, the wage that the firm would choose if it could commit not to replace workers. Similarly, the optimal wage for senior workers w_{22}^k equals $\hat{w}_{22}$, the wage that the firm would choose under commitment. When the productivity of the firm is between y_k and $\hat{y}$, the no-replacement constraint is violated at $(\hat{w}_{21}, \hat{w}_{22})$. In this case, the wage for new hires and the wage for senior workers are both set equal to w_2^k. This firm-wide wage satisfies the first order condition

$$n_1(1-\delta)\left[\upsilon'(w_2^k)/\upsilon'(w_1) - 1\right] + \left[\overline{n} - n_1(1-\delta)\right]\left[\tilde{q}'(w_2^k)(y_2 - w_2^k) - \tilde{q}(w_2)\right] = 0. \tag{8}$$

The first term on the rhs of (8) is the effect of an increase in w_2^k on the utility of senior workers and on the profits that the firm earns on them. The second term is the effect of an increase in w_2^k on the profits that the firm earns on new recruits. The optimal firm-wide wage equates the sum of these two effects to zero. The first term on the rhs of (8) is positive for $w_2^k < w_1$, and negative otherwise. The second term is negative for $w_2^k > \hat{w}_{21}$. Since the constraint $w_2^k \leq y_2$ does not bind for $y_2 \in [y_k, \hat{y}]$, it follows that w_2^k is greater than $\hat{w}_{21}$ and smaller than $\hat{w}_{22} = \min\{w_1, y_2\}$. That is, the optimal firm-wide wage is greater than the wage for new hires and smaller than the wage for senior workers that the firm would choose under commitment. Finally, when the productivity of the firm falls below y_k, the constraint

$w_2 \leq y_2$ begins to bind. In this case, the optimal firm-wide wage equals the productivity of the firm. The properties of (w_{21}^k, w_{22}^k) are illustrated in figure 2.

3.2.3. Wage policy with replacement

The following proposition characterizes the solution to (R), i.e. the second period wages that maximize the weighted sum of the firm's profits and the senior workers' utility subject to the replacement constraint, $w_{21} < w_{22}$, and the individual rationality constraint.

Proposition 5 *Let (w_{21}^r, w_{22}^r) denote the solution to (R). For $z \leq y_2 \leq \hat{y}$, the wage for new hires w_{21}^r lies in the interval $[z, \hat{w}_{21})$, and the wage for senior workers w_{22}^r equals $\hat{w}_{22}$.*

The optimal hiring wage w_{21}^r satisfies the first order condition

$$\overline{n}\left[\tilde{q}'(w_{21}^r)(y_2 - w_{21}^r) - \tilde{q}(w_{21}^r)\right] = n(1-\delta)\tilde{q}'(w_{21}^r)\left\{y_2 - w_{22}^r + \left[\upsilon(w_{22}^r) - \upsilon(b)\right]/\upsilon'(w_1)\right\}. \quad (9)$$

On the one hand, an increase in w_{21}^r affects the profits that the firm earns on new recruits. This effect is measured by the lhs of (9). On the other hand, an increase in w_{21}^r increases the probability of replacement for senior workers and, hence, it reduces their utility as well as the profits that the firm earns from them. This effect is measured by the rhs of (9). The optimal hiring wage equates these two effects. The term on the lhs is positive for $w_{21}^r < \hat{w}_{21}$ and negative otherwise. The term on the rhs is positive. Hence, w_{21}^r is less than $\hat{w}_{21}$. That is, the optimal hiring wage is lower than the wage that the firm would offer under commitment. In contrast, the optimal wage for senior workers w_{22}^r equals $\hat{w}_{22}$, the wage that the firm would choose under commitment. Intuitively, even though the possibility of replacement affects the utility of senior workers and the profits that the firm makes off of them, it does not affect the optimal resolution of the trade-off. The properties of (w_{21}^r, w_{22}^r) are illustrated in figure 3.

3.3. The optimal wage policy

Theorem 6 *Let $\sigma^* = (w_1^*, w_{2i}^*)$ be the optimal wage policy. (i) For $y_2 \geq \hat{y}$, the wage for new hires w_{21}^* equals w_{21}^k and the wage for senior workers w_{22}^* equals w_{22}^k, where $w_{2i}^k = \hat{w}_{2i}$. (ii) There exists a $\overline{y} \in [z, \hat{y})$ such that, for $\overline{y} \leq y_2 < \hat{y}$, w_{21}^* and w_{22}^* are both equal w_2^k, where $w_2^k \in (\hat{w}_{21}, \hat{w}_{22})$.*

The theorem characterizes the properties of the optimal wage policy of the firm in the presence of the replacement problem. First, consider the case in which the productivity of the firm is greater than $\hat{y}$. In this case, the wage policy without replacement prescribes the wages (w_{21}^k, w_{22}^k) where $(w_{21}^k, w_{22}^k) = (\hat{w}_{21}, \hat{w}_{22})$. Under this policy, senior workers have the same probability of employment and the same wage as under commitment. Moreover, under this policy, the firm recruits the same number of new hires for its vacant jobs and pays them the same wage as under commitment. Since this policy distorts neither the provision of insurance to senior workers nor the recruitment of new hires, $V_k = \hat{V}$. In turn, since V_k, $V_r \leq \hat{V}$, this implies that the policy without replacement is optimal.

Next consider the case in which the productivity of the firm is less than $\hat{y}$. In this case, the policy without replacement prescribes the wage w_2^k for both new hires and senior workers, where $\hat{w}_{21} < w_2^k < \hat{w}_{22}$. Under this policy, senior workers have the same probability of employment but a lower wage than they would under commitment. Moreover, under this policy, the firm hires more workers for its vacant jobs and pays them more than in the commitment case. In contrast, the wage policy with replacement prescribes the wage w_{21}^r for new hires and w_{22}^r for senior workers, where $w_{21}^r < \hat{w}_{21}$ and $w_{22}^r = \hat{w}_{22}$. Under this policy, senior workers have the same wage as under commitment, but a lower probability of employment. Moreover, the firm hires fewer workers for its vacant jobs and pays them less than it would under commitment. Overall, when the productivity of the firm is less than $\hat{y}$, both policies introduce some distortions in the provision of insurance to senior workers and in the recruitment of new hires, i.e. V_k, $V_r < \hat{V}$ In principle, either one of these two policies may be optimal.

However, if the productivity of the firm is not too far below $\hat{y}$, the optimal wage policy is always the one without replacement. The intuition for this result is straightforward. For $y_2 \to \hat{y}^-$, the distortions introduced by the policy without replacement disappear because $w_2^k \to \hat{w}_{21}$, $\hat{w}_{22}$. In contrast, for $y_2 \to \hat{y}^-$, the distortions introduced by the policy with replacement do not vanish. If w_{21}^r converges to a value greater than z, the firm attracts some applicants to its filled jobs and the probability of employment for senior workers is too low. If w_{21}^r converges to z, the firm does not attract any applicants and the number of filled vacancies is too low. Therefore, there exists an interval $[\overline{y}, \hat{y}]$ for y_2 where the wage policy without replacement is the optimal one. Whether $\overline{y}$ is equal to z (and the optimal policy is always the one without replacement) or it is greater than z (and the optimal policy is the one with replacement for sufficiently low y_2) depends on the fundamentals of the model. For the parameter values presented at the end of the next section, we find that $\overline{y} = z$. Figure 4 illustrates an optimal wage policy in which $\overline{y} > z$.

The key implication of theorem 6 is that the contractual solution to the worker replacement problem induces a form of downward rigidity in the wage of new hires. The firm would like to offer to its senior workers a wage that remains constant over time, and to its new hires a wage that varies in response to changes in productivity. When the productivity of the firm falls below a critical threshold, the wage that the firm would like to offer to its new hires is less than the wage offered to senior workers in the past. Since at these wages the firm would have an incentive to replace senior workers with new hires, the firm chooses instead to offer the same wage to everybody (as long as y_2 is not too low). This firm-wide wage successfully prevents replacement, but it distorts the wage of senior workers down and the wage of new hires up. Hence, the optimal solution to the worker replacement problem dampens the response to negative productivity shocks of the wage offered by old firms to new hires.

4. Macroeconomics of worker replacement

This section studies how the worker replacement problem affects the response of unemployment and vacancies to aggregate productivity shocks. To this aim, the section compares the predictions of our model with those of a model in which the replacement problem does not arise because firms can commit not to substitute senior workers with new recruits.

4.1. Environment

For this section, we restrict attention to a particular stochastic process for the aggregate state of the economy x and for the idiosyncratic productivity of the firm y. The aggregate state of the economy can take one of two values $x \in \{x_\ell, x_h\}$, where x_ℓ is a state in which firms have low productivity (a recession), and x_h is a state in which firms have high productivity (an expansion). In particular, when $x = x_j$, $j = \ell, h$, a new firm has productivity y_1^j and an old firm has productivity y_2^j, where $b < y_1^\ell < y_1^h$ and $b < y_2^\ell < y_2^h$. The aggregate state of the economy is very persistent, in the sense that $\pi_x(x_j|x_j) \to 1$ for $j = \ell, h$. This specification of the stochastic process for x and y affords a clear-cut comparison between the predictions of our model and those of a model in which firms can commit not to replace workers.

We also restrict attention to particular parameter values. Let $(\hat{Z}, \hat{\theta}, \hat{\sigma})$ be the equilibrium of a version of the model in which firms can commit not to replace workers. We choose the productivity of the firm in the first and second period of activity so that the wage policy $\hat{\sigma}(x_j) = (\hat{w}_1^j, \hat{w}_{2i}^j(x'))$ prescribes a hiring wage that remains constant throughout the life of the firm (conditional on a given state of the economy). That is, we choose (y_1^j, y_2^j) so that $\hat{w}_1^j = \hat{w}_{21}^j(x_j)$ for $j = \ell, h$. Moreover, we choose the cost of entry k so that number of applicants attracted by a vacant job remains constant during the life of the firm. That is, we choose k so that $\hat{\theta}(W_1(\hat{\sigma}(x_\ell), x_\ell)) = \hat{\theta}(W_2(\hat{w}_{21}^\ell(x_\ell), x_\ell))$.

The parametric restrictions above can be interpreted as normalizations. If a firm had an infinite horizon and a constant productivity, the optimal wage policy under commitment would prescribe a constant hiring wage and would attract a constant number of applicants to each vacancy. However, for the sake of tractability, our model assumes that firms have a finite horizon. Hence, if a firm in our model had a constant productivity, the optimal wage policy under commitment would prescribe a time-varying hiring wage and would attract a time-varying number of applicants. The parametric restrictions above make sure that the behavior of a firm in our model reproduces the behavior of a firm with an infinite horizon.

4.2. Competitive search equilibrium

Lemma 7 *Let* $(\hat{Z}, \hat{\theta}, \hat{\sigma})$ *be the competitive search equilibrium of the model in which firms can commit not to replace workers. Then* (Z, θ, σ) *is a competitive search equilibrium of our model, where: (i)* $Z = \hat{Z}$ *and* $Z(x_\ell) < Z(x_h)$*; (ii)* $\theta = \hat{\theta}$ *and* $\theta(W, x_\ell) > \theta(W, x_h)$*; (iii)* $\sigma(x_\ell) = \hat{\sigma}(x_\ell)$ *and* $\sigma(x_h) = (w_1^h, w_{2i}^h(x'))$*, where* $w_1^h = \hat{w}_1^h$ *and* $w_{2i}^h(x_h) = \hat{w}_{2i}^h(x_h)$ *for* $i = 1, 2$*. Moreover, as long as* $|y_1^h - y_1^\ell|$ *is not too large,* $w_{2i}^h(x_\ell) = w_2$ *for* $i = 1, 2$*, where* $\hat{w}_{21}^h(x_\ell) < w_2 < \hat{w}_{22}^h(x_\ell)$*.*

Lemma 7 characterizes the properties of the equilibrium of our model.[10] The wage policy of a firm that enters the labor market during a recession is the same that the firm would choose if it could commit not to replace workers. This result is intuitive. The wage policy under commitment prescribes the wage $\hat{w}_1^\ell$ in the first period. In the second period, the policy prescribes that senior workers should be paid the same wage as in the first period, and that new hires should be paid the wage $\hat{w}_{21}^\ell(x')$ that maximizes the value of the firm's vacant jobs. Since $\hat{w}_1^\ell = \hat{w}_{21}^\ell(x_\ell)$ and $\hat{w}_{21}^\ell(x_\ell) < \hat{w}_{21}^\ell(x_h)$, the optimal wage policy under commitment never tempts the firm to replace senior workers with new hires. Hence, it is also the optimal policy under limited commitment.

The wage policy of a firm that enters the labor market during an expansion is $\sigma(x_h)$. In the first period, the policy prescribes the wage $\hat{w}_1^h$. If in the second period the economy is

[10] The proof of Lemma 7 is avaliable upon request.

still in an expansion, the policy prescribes the wages $\hat{w}_{21}^{h}(x_h)$ and $\hat{w}_{22}^{h}(x_h)$, the same wages that the firm would choose under commitment. These wages are optimal because they do not induce the firm to replace senior workers with new hires. If the economy turns into a recession, the wages $\hat{w}_{21}^{h}(x_\ell)$ and $\hat{w}_{22}^{h}(x_\ell)$ would induce the firm to replace workers because $\hat{w}_{21}^{h}(x_\ell) < \hat{w}_{22}^{h}(x_\ell)$. For this reason, the wage policy prescribes the wage w_2 for both new hires and senior workers, where $\hat{w}_{21}^{h}(x_\ell) < w_2 < \hat{w}_{22}^{h}(x_\ell)$.

Notice that, when a firm enters the labor market, its expected profits are the same as under commitment. Intuitively, if the firm enters the labor market during a recession, its expected profits are the same as under commitment because the wage policy is the same. If the firm enters the labor market during an expansion, the wage policy is the same as under commitment as long as the economy remains in an expansion. Since the probability of this event is 1, the profits of the firm are the same as under commitment. In turn, from the free-entry condition, it follows that, if the expected profits of a new firm are the same as under commitment, so are the worker's value of search Z and the expected queue length θ.

Overall, the only difference between the equilibrium of our model and the equilibrium of the model with commitment are the wages that old firms pay to new hires and senior workers when the economy moves from an expansion to a recession.[11] Then, it is clear that the worker replacement problem will not affect the response of the labor market to a positive shock to aggregate productivity, but it will have an effect on the response of unemployment, vacancies and other labor market variables to a negative shock. In the following pages we shall characterize these effects.

[11] If the transition probability of x from one state to the other were greater than zero, the worker replacement problem would also have an indirect effect on the value of search Z and on the queue length θ. This paper does not explore how these indirect effects impact the response of unemployment and vacancies to productivity shocks.

4.3. Unemployment and vacancies

Suppose that the economy moves from an expansion to a recession. In the economy there are f_2 old firms, where f_2 is determined by the history of realizations of the aggregate productivity shock. Every old firm has the wage policy $\sigma(x_h)$ which prescribes that both new hires and senior workers should be paid the wage w_2 in the current period. Every old firm has $n_1(1-\delta)$ filled jobs and $\overline{n} - n_1(1-\delta)$ vacant jobs. At the search stage, each one of these vacant jobs attracts θ_2 applicants and is filled with probability $q(\theta_2)$, where $\theta_2 = \tilde{\theta}(w_2, x_\ell)$. Hence, at the production stage, every old firm employs $n_1(1-\delta)$ senior workers and $[\overline{n} - n_1(1-\delta)]q(\theta_2)$ new hires.

Entrepreneurs create f_1 new firms, where f_1 is determined endogenously. Every new firm chooses the wage policy $\sigma(x_\ell)$ which prescribes that workers should be paid the wage $\hat{w}_1^\ell$ in the current period. Every new firm has $\overline{n}$ vacant jobs. At the search stage, each one of these vacant jobs attracts θ_1 applicants and is filled with probability $q(\theta_1)$, where $\theta_1 = \tilde{\theta}(\hat{w}_{21}^\ell(x_\ell), x_\ell)$ because of our normalizations. Hence, at the production stage, every new firm employs $\overline{n}q(\theta_1)$ workers.

The number of new firms clears the market for applicants. That is, the number of new firms is such that the number of job applications received by new and old firms is equal to the number of workers who apply for a job, a. Since every new firm receives $\overline{n}\theta_1$ applications and every old firm receives $[\overline{n} - n_1(1-\delta)]\theta_2$ applications, the number of new firms is

$$f_1 = \{a - f_2[\overline{n} - n_1(1-\delta)]\theta_2\}/\,\overline{n}\theta_1. \tag{10}$$

The vacancy rate, the job-finding rate and the unemployment rate are

$$
\begin{aligned}
v &= f_1\overline{n} + f_2[\overline{n} - n_1(1-\delta)], \\
h &= \left(\tfrac{f_1\overline{n}\theta_1}{a}\right) p(\theta_1) + \left(1 - \tfrac{f_1\overline{n}\theta_1}{a}\right) p(\theta_2), \\
u &= 1 - f_1\overline{n}q(\theta_1) - f_2[\overline{n} - n_1(1-\delta)]q(\theta_2) - f_2 n_1(1-\delta).
\end{aligned} \tag{11}
$$

The vacancy rate v is the sum of vacant jobs at new and old firms. The job-finding rate h is the weighted average of the job-finding rate for workers who apply for jobs at new and old firms. The unemployment rate u is the difference between the number of workers in the economy and the number of workers employed at new and old firms.

4.4. Worker replacement as an amplification mechanism

To understand how the worker replacement problem affects the response of the labor market to a negative productivity shock, we compare the predictions of our model with those of a model in which firms can commit not to replace workers. As shown in lemma 7, when the economy is hit by a negative productivity shock, the wage offered by old firms to new hires falls less in our model than it would in a model with commitment. Specifically, this wage falls to w_2 rather than to $\hat{w}_{21}^{\ell}(x_\ell)$, where $w_2 > \hat{w}_{21}^{\ell}(x_\ell)$. Because of this downward wage rigidity, the number of workers who apply for a vacant job at an old firm increases more in our model than it would in a model with commitment. Specifically, this number increases to θ_2 rather than $\tilde{\theta}(\hat{w}_{21}^{\ell}(x_\ell), x_\ell)$, which is equal to θ_1 because of our normalizations.

From the above observations and equations (10) and (11), it follows that, when the economy is hit by a negative productivity shock, the vacancy rate in our model falls more than it would in a model with commitment. Specifically, the vacancy rate falls by an additional

$$
\Delta_v = f_2\left[\overline{n} - n_1(1-\delta)\right] \cdot \left[\left(\theta_2 - \theta_1\right)/\theta_1\right] > 0. \tag{12}
$$

The first term on the rhs of (12) is the number of vacant jobs at old firms. The second term is the relative difference between the number of applicants attracted by each of these jobs in our model and in a model with commitment. This expression is intuitive. For every additional worker who applies to old firms, there are $1/\overline{n}\theta_1$ fewer firms that find it profitable to enter the labor market and, hence, there are $1/\theta_1$ fewer vacancies.

From equations (10) and (11), it also follows that, in response to a negative productivity shock, the job-finding rate in our model falls more than it would in a model with commitment. Specifically, the job-finding rate falls by an additional

$$\begin{aligned} \Delta_h \quad &= f_2[\overline{n} - n_1(1-\delta)]\,[q(\theta_1)\theta_2/\theta_1 - q(\theta_2)]/\,a \\ &= \{f_2[\overline{n} - n_1(1-\delta)]\theta_2/a\} \cdot [p(\theta_1) - p(\theta_2)] > 0, \end{aligned} \tag{13}$$

where the second line makes use of $q(\theta) = p(\theta)\theta$. This expression is intuitive. The first term on the rhs of (13) is the fraction of applicants who look for vacant jobs at old firms. The second term is the difference between the probability that an applicant finds one of these jobs in our model and in a model with commitment. Finally, equations (10) and (11) imply that unemployment in our model increases more than it would under commitment, where the additional increase is

$$\Delta_u = f_2\theta_2\,[\overline{n} - n_1(1-\delta)]\,[p(\theta_1) - p(\theta_2)] > 0. \tag{14}$$

Overall, the worker replacement problem amplifies the response of the vacancy, job-finding and unemployment rates to a negative shock to aggregate productivity. In contrast, it is straightforward to see that the replacement problem has no effect on how the labor market responds to a positive shock. Intuitively, the worker replacement problem leads to a form of downward rigidity in the wage of new hires at old firms. When the economy is hit by a negative productivity shock, this downward wage rigidity amplifies the increase in the number of applicants attracted by old firms, it amplifies the decline in the number of firms that enter the market and, ultimately, the decline in vacancies and the increase in unemployment. When the economy is hit by a positive productivity shock, the wage of new hires at old firms is unaffected and so are labor market outcomes. These findings are summarized in the following theorem.

Theorem 8 *The worker replacement problem amplifies the response of v, h and u to a negative shock to the aggregate productivity x. However, the worker replacement problem does not affect the response of v, h or u to a positive shock to aggregate productivity.*

The model cannot be used for a thorough quantitative analysis of business cycles because of the assumption that firms operate for two periods only. Nevertheless, the model can still be used to get a sense of the importance of the worker replacement problem as an amplification mechanism. To this aim, note that the expected utility of a worker who applies for a job at an old firm is the same in our model and in a model with commitment, i.e.

$$\begin{aligned} & p(\theta_2)\left[\upsilon(w_2)-\upsilon(b)\right]+U(x_\ell)=Z(x_\ell) \\ & =\hat{Z}(x_\ell)=p(\theta_1)\left[\upsilon(\hat{w}_{21}^{\ell}(x_\ell))-\upsilon(b)\right]+U(x_\ell). \end{aligned}$$

The above equation implies that $p(\theta_2)\left[\upsilon(w_2)-\upsilon(b)\right]$ is equal to $p(\theta_1)\left[\upsilon(\hat{w}_{21}^{\ell}(x_\ell))-\upsilon(b)\right]$. Using this fact, we can rewrite Δ_u as

$$\begin{aligned} \Delta_u \quad & = f_2\left[\overline{n}-n_1(1-\delta)\right]\theta_2\cdot p(\theta_1)\cdot\frac{\upsilon(w_2)-\upsilon(\hat{w}_{21}^{\ell}(x_\ell))}{\upsilon(w_2)-\upsilon(b)} \\ & \approx f_2\left[\overline{n}-n_1(1-\delta)\right]\theta_2\cdot p(\theta_1)\cdot\frac{w_2-\hat{w}_{21}^{\ell}(x_\ell)}{w_2-b}. \end{aligned} \tag{15}$$

The first term on the rhs of (15) is the measure of workers who apply to old firms. The second term is the job-finding probability for a worker who applies to a new firm. The third term measures the extent of the wage rigidity caused by the replacement problem as the ratio between $w_2-\hat{w}_{21}^{\ell}(x_\ell)$ and w_2-b. Now, suppose that the measure of workers who apply to old firms is 30% per period, that the job-finding probability is 90%, that $w_2 - \hat{w}^{\ell}_{21}(x_\ell)$ is 0.5% of w_2 and $w_2 - b$ is 30% of w_2.[12]. Given these numbers, the worker replacement problem amplifies the response of unemployment to a negative productivity shock by approximately 1/2 of a percentage point.[13] Considering that in a typical recession the unemployment rate increases by 3 percentage points and that presumably only a fraction of this increase is attributable to the decline in labor productivity, the worker replacement problem appears to be a quantitatively important mechanism through which negative productivity shocks are transmitted into the labor market.

5. Conclusions

The paper advances a theory of downward rigidity for the wage on new hires in the context of a labor market with search frictions. Firms want to insure existing workers against income fluctuations, because workers are risk averse and have no access to the capital market. To perform this task efficiently, firms would like to keep the wage of each worker constant over time. Firms also want to attract applicants to fill the jobs that become vacant because of turnover. To perform this task efficiently, firms would like to offer a hiring wage that varies over time with the state of the economy. Generally, firms cannot perform both tasks efficiently because they cannot commit not to replace senior workers with new hires.

The first part of the paper studied the effect of the worker replacement problem on the optimal design of the wage policy of the firm. We found that, because of the worker replacement problem, the wage policy prescribes a hiring wage that does not vary as much as the efficient hiring wage in response to negative productivity shocks. The downward rigidity

[12] It is easy to construct examples that deliver similar numbers. For instance, suppose that the preferences of the worker are given by $\beta = .95$, $b = .46$, and $v(c) = c^{1-\gamma}/(1-\gamma)$ with $\gamma = 2$. Suppose that the technology of the firm is given by $\delta = .1$, $y^{\ell}_2 = 1$ and $y^h_2 = 1.05$. Suppose that the matching technology is given by $p(\theta) = M\theta^{\eta-1}$, $q(\theta) = M\theta^{\eta}$, where $M = 1/10$ and $\eta = 1/2$. Then, given our normalizations for y^{ℓ}_1, y^h_1 and k, the measure of workers applying to old firms is 39%, the job-finding probability at new firms is 90%, the difference $w_2 - \hat{w}^{\ell}_{21}(x_\ell)$ is .5% of w_2, and the difference $w_2 - b$ is 28% of w_2.

[13] By comparison, in the standard search model of Pissarides (1985), a 5 percent decline in productivity increases unemployment by 0.1 percentage points.

in the hiring wage allows the firm to keep the wage profile of existing workers smooth without having the temptation to replace them with new recruits. This finding may help explain why, in the real world, managers are reluctant to hire new workers for less than the wage of existing workers even during recessions.

The second part of the paper studies how the worker replacement problem affects the response of unemployment and vacancies to aggregate productivity shocks. We found that the worker replacement dampens the decline in the wage of new hires at old firms in response to a negative shock. In turn, this wage rigidity amplifies the increase in the number of unemployed workers who look for jobs at old firms, it amplifies the decline in the entry of new firms in the labor market and, ultimately, it amplifies the increase in the unemployment rate in response to a negative shock. In contrast, the worker replacement problem does not affect the response of the labor market to positive aggregate productivity shocks. A back-of-the-envelope calculation suggests that the worker replacement problem may help explain why small aggregate productivity shocks are associated with large unemployment fluctuations.

References

Acemoglu, D., Shimer, R., 1999. Efficient unemployment insurance. Journal of Political Economy 107, 893–928.

Agell, J., Bennmarker, H., 2007. Wage incentives and wage rigidity: A representative view from within. Labour Economics 14, 347–69.

Azariadis, C., 1975. Implicit contracts and underemployment equilibria. Journal of Political Economy 83, 1183–202.

Bewley, T., 1999. Why wages don't fall during a recession?. Harvard University Press, Cambridge, MA.

Blinder, A., Choi, D. 1990. A shred of evidence on theories of wage stickiness. Quarterly Journal of Economics 105, 1003–15.

Burdett, K., Mortensen, D., 1998. Wage differentials, employer size, and unemployment. International Economic Review 39, 257–73.

Gertler, M., Trigari, A., 2009. Unemployment fluctuations with staggered nash bargaining. Journal of Political Economy 117, 38–86.

Hall, R., 2005. Employment fluctuations with equilibrium wage stickiness. American Economic Review 95, 50–65.

Kennan, J., 2009. Private information, wage bargaining and employment fluctuations. Review of Economic Studies (forthcoming).

MacLeod, B., Malcomson, J., 1989. Implicit contracts, incentive compatibility, and involuntary unemployment. Econometrica 57, 447–80.

Menzio, G., 2005. High-frequency wage rigidity. Unpublished Manuscript, University of Pennsylvania.

Menzio, G., Shi, S., 2009a. Efficient search on the job and the business cycle. NBER Working Paper.

Menzio, G., Shi, S., 2009b. Block recursive equilibria for stochastic models of search on the job. Journal of Economic Theory (forthcoming).

Moen, E., 1997. Competitive search equilibrium. Journal of Political Economy 105, 694–723.

Moen, E., Rosen, Å., 2008. Incentives in competitive search equilibrium. Unpublished Manuscript, Norwegian School of Management.

Pissarides, C. 1985. Short-run equilibrium dynamics of unemployment, vacancies, and real wages. American Economic Review 75, 676–690.

Rudanko, L., 2009. Labor market dynamics under long-term wage contracting. Journal of Monetary Economics 56, 170–83.

Shimer, R., 1996. Contracts in a frictional labor market. Unpublished Manuscript, M.I.T.

Shimer, R., 2005. The cyclical behavior of unemployment and vacancies. American Economic Review 95, 25–49.

Snell, A., Thomas, J., 2008. Labor contracts, equal treatment, and wage-unemployment dynamics. Unpublished Manuscript, University of Edinburgh.

Thomas, J., Worrall, T., 1988. Self-enforcing wage contracts. Review of Economic Studies 55, 541–53.

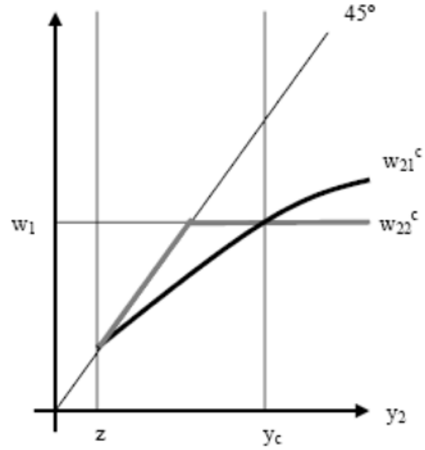

Figure 1: Wage policy under commitment

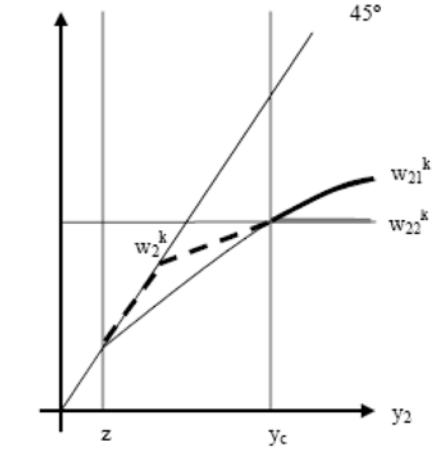

Figure 2: Wage policy without replacement

Notes: z is the flow value of search; y_2 is the productivity of the firm in the second period; w_1 is the first period wage; (w_{21}^{c}, w_{22}^{c}) are the second period wages under commitment; (w_{21}^{k}, w_{21}^{k}) are the second period wages subject to the no replacement constraint, which are denotes as w_2^{k} when equal.

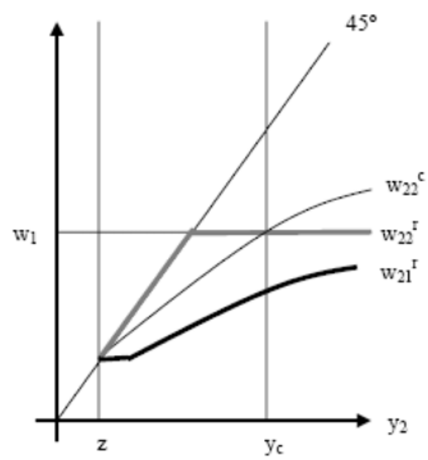

Figure 3: Wage policy with replacement

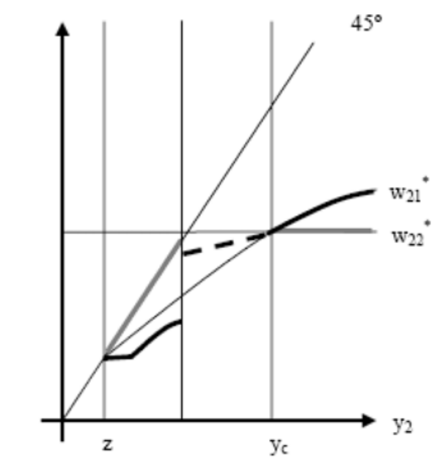

Figure 4: Optimal wage policy

Notes: z is the flow value of search; y_2 is the productivity of the firm in the second period; w_1 is the first period wage; (w_{21}^{c}, w_{22}^{c}) are the second period wages under commitment; (w_{21}^{r}, w_{21}^{r}) are the second period wages subject to the replacement constraint, (w_{21}^{*}, w_{21}^{*}) are the optimal second period wages.

Appendix

A Proof of Lemma 2

(i) Let $\sigma^* = (w_1^*, w_{2i}^*)$ be the optimal wage policy for a new firm, given that the aggregate state of the economy is x, and the productivity of the firm is y_1. Let $\overline{w}_{21}^*$ denote $w_{21}^*(s)$ for $s \in S^*$, where S^* is the set $\{(x', y_2', \mu_2') : x' = \overline{x},\ y_2' = \overline{y}_2,\ \mu_2' \in [\underline{\mu}_2, \overline{\mu}_2]\}$. Similarly, let $\overline{w}_{22}^*$ denote $w_{22}^*(s)$ for $s \in S^*$. Suppose that $\overline{w}_{22}^* \leq \overline{w}_{21}^*$ and $\overline{w}_{22}^* \in [b, y_2]$.

Next, let $\sigma = (w_1, w_{2i})$ denote an alternative wage policy such that: (i) for all $s \notin S^*$, $w_{2i}(s)$ is equal to $w_{2i}^*(s)$; (ii) for all $s \in S^*$, $w_{2i}(s)$ is equal to $\overline{w}_{2i}$; (iii) $W_1(\sigma, x) = W_1(\sigma^*, x)$. In words, the alternative wage policy σ is such that: (i) in all states $s \notin S^*$, the wages for new hires and senior workers are the same as under the policy σ^*; (ii) in all states $s \in S^*$, the wages for new hires and senior workers may be different than under the policy σ^*; (iii) in its first period of activity, the firm offers the same lifetime utility as under the policy σ^*. The last property implies that, in its first period of activity, the firm employs the same number of workers as under the policy σ^*.

First, consider the case in which $\overline{w}_{22} \leq \overline{w}_{21}$ and $\overline{w}_{22} \in [b, y_2]$. In this case, the difference between the profits of the firm under the wage policy σ^* and the alternative policy σ

$$\begin{aligned} & F(\sigma^*, x, y_1) - F(\sigma, x, y_1) \\ = \;& n_1(w_1 - w_1^*) + \beta \Pr(S^*) n_1 (1-\delta)(\overline{w}_{22} - \overline{w}_{22}^*) \\ + \;& \beta \Pr(S^*)[\overline{n} - n_1(1-\delta)][\tilde{q}(\overline{w}_{21}^*)(y_2 - \overline{w}_{21}^*) - \tilde{q}(\overline{w}_{21})(y_2 - \overline{w}_{21})] \geq 0, \end{aligned} \tag{1}$$

where

$$w_1 = v^{-1}\left\{v(w_1^*) + \beta \Pr(S^*)(1-\delta)\left[v(\overline{w}_{22}^*) - v(\overline{w}_{22})\right]\right\}. \tag{2}$$

After substituting (2) into (1), we obtain

$$\begin{aligned}
& lim_{\overline{\mu}_2 \rightarrow \underline{\mu}_2} \left[\frac{F(\sigma^*, x, y_1) - F(\sigma, x, y_1)}{\beta \Pr(S^*)} \right] \\
= \; & n_1(1-\delta) \left\{ \left[\upsilon(\overline{w}^*_{22}) - \upsilon(\overline{w}_{22}) \right] / \upsilon'(w^*_1) + \overline{w}_{22} - \overline{w}^*_{22} \right\} \\
+ \; & [\overline{n} - n_1(1-\delta)][\tilde{q}(\overline{w}^*_{21})(y_2 - \overline{w}^*_{21}) - \tilde{q}(\overline{w}_{21})(y_2 - \overline{w}_{21})] \geq 0.
\end{aligned} \tag{3}$$

Denote as $\upsilon_k(w_{21}, w_{22})$ the objective function in the maximization problem (K). Then, it is straightforward to verify that the above limit is equal to the difference between $\upsilon_k(\overline{w}^*_{21}, \overline{w}^*_{22})$ and $\upsilon_k(\overline{w}_{21}, \overline{w}_{22})$.

Next, consider the case in which $\overline{w}_{21} < \overline{w}_{22}$ and $\overline{w}_{22} \in [b, y_2]$. In this case, the difference between the profits of the firm under the policy σ^* and the alternative policy σ is

$$\begin{aligned}
& F(\sigma^*, x, y_1) - F(\sigma, x, y_1) \\
= \; & n_1(w_1 - w^*_1) + \beta \Pr(S^*) n_1 (1-\delta)[\tilde{q}(\overline{w}_{21})\overline{w}_{21} + (1 - \tilde{q}(\overline{w}_{21}))\overline{w}_{22} - \overline{w}^*_{22}] \\
+ \; & \beta \Pr(S^*)[\overline{n} - n_1(1-\delta)][\tilde{q}(\overline{w}^*_{21})(y_2 - \overline{w}^*_{21}) - \tilde{q}(\overline{w}_{21})(y_2 - \overline{w}_{21})] \geq 0,
\end{aligned} \tag{4}$$

where

$$w_1 = \upsilon^{-1} \left\{ \upsilon(w^*_1) + \beta \Pr(S^*)(1-\delta) \left[\upsilon(\overline{w}^*_{22}) - \tilde{q}(\overline{w}_{21})\upsilon(b) - (1 - \tilde{q}(\overline{w}_{21}))\upsilon(\overline{w}_{22}) \right] \right\}. \tag{5}$$

After substituting (5) into (4), we obtain

$$\begin{aligned}
& lim_{\overline{\mu}_2 \rightarrow \underline{\mu}_2} \left[\frac{F(\sigma^*, x, y_1) - F(\sigma, x, y_1)}{\beta \Pr(S^*)} \right] \\
= \; & n_1(1-\delta) \left[\upsilon(\overline{w}^*_{22}) - \tilde{q}(\overline{w}_{21})\upsilon(b) - (1 - \tilde{q}(\overline{w}_{21}))\upsilon(\overline{w}_{22}) \right] / \upsilon'(w^*_1) \\
+ \; & n_1(1-\delta) \left[\tilde{q}(\overline{w}_{21})\overline{w}_{21} + (1 - \tilde{q}(\overline{w}_{21}))\overline{w}_{22} - \overline{w}^*_{22} \right] \\
+ \; & [\overline{n} - n_1(1-\delta)][\tilde{q}(\overline{w}^*_{21})(y_2 - \overline{w}^*_{21}) - \tilde{q}(\overline{w}_{21})(y_2 - \overline{w}_{21})] \geq 0.
\end{aligned} \tag{6}$$

Denote as $\upsilon_r(w_{21}, w_{22})$ the objective function in the maximization problem (R). Then, it is straightforward to verify that the above limit is equal to the difference between $\upsilon_k(\overline{w}^*_{21}, \overline{w}^*_{22})$ and $\upsilon_r(\overline{w}_{21}, \overline{w}_{22})$..

Finally, note that inequality (3) holds for all wages $(\overline{w}_{21}, \overline{w}_{22})$ such that $\overline{w}_{22} \leq \overline{w}_{21}$ and $b \leq \overline{w}_{22} \leq y_2$. Moreover, note that inequality (6) holds for all wages $(\overline{w}_{21}, \overline{w}_{22})$ such that

$\overline{w}_{21} < \overline{w}_{22}$ and $b \le \overline{w}_{22} \le y_2$. Hence, (3) and (6) imply that $v_k(\overline{w}^*_{21}, \overline{w}^*_{22}) = V_k \ge V_r$ and $(\overline{w}^*_{21}, \overline{w}^*_{22}) = (w^k_{21}, w^k_{22})$.

(ii) We omit the proof of part (ii) because it is similar to the proof of part (i). We omit the proof of part (iii) because it is straightforward. All details are available upon request.

B Proof of Proposition 3

The maximization problem (C) can be written as

$$\begin{aligned}\hat{V} &= [\overline{n} - n_1(1-\delta)] \max_{w_{21}} \{\tilde{q}(w_{21})(y_2 - w_{21})\} + n_1 \upsilon(b)/\upsilon'(w_1) \\ &+ n_1(1-\delta) \max_{b \le w_{22} \le y_2} \{[\upsilon(w_{22}) - \upsilon(b)]/\upsilon'(w_1) + y_2 - w_{22}\}.\end{aligned} \tag{7}$$

(i) Consider the first maximization problem in (7). The objective function of this problem is continuous in w_{21}. It is equal to zero for $w_{21} = z$, strictly positive for $w_{21} \in (z, y_2)$, and negative for $w_{21} \ge y_2$. Hence, the solution to this problem, $\hat{w}_{21}$, exists and belongs to the interval (z, y_2). Now, consider the second maximization problem in (7). It is straightforward to verify that the solution of this problem, w_{22}, is equal to $\min\{w_1, y_2\}$.

(ii) Return to the first maximization problem in (7). For all $w_{21} \in (z, y_2)$, the derivative of the objective function with respect to w_{21} is

$$\begin{aligned}&\frac{d[\tilde{q}(w_{21})(y_2 - w_{21})]}{dw_{21}} \\ =\ & \tilde{q}'(w_{21})(y_2 - w_{21}) - \tilde{q}(w_{21}) \\ =\ & q'(\theta(W_2(w_{21})))\theta'(W_2(w_{21}))\upsilon'(w_{21})(y_2 - w_{21}) - q(\theta(W_2(w_{21}))), \\ =\ & q(\theta(W_2(w_{21})))\tfrac{\epsilon_q(\theta(W_2(w_{21})))}{1-\epsilon_q(\theta(W_2(w_{21})))}\tfrac{\upsilon'(w_{21})}{\upsilon(w_{21})-\upsilon(b)}(y_2 - w_{21}) - q(\theta(W_2(w_{21}))),\end{aligned} \tag{8}$$

where the third line makes use of the definitions of $\tilde{q}(w)$ and $W_2(w)$, and the fourth line makes use of the equilibrium condition for $\theta(W)$. The first term on the rhs of (8) is strictly decreasing in w_{21}, because $q(\theta)\epsilon_q(\theta)/(1-\epsilon_q(\theta))$ is a strictly decreasing function of θ and $\theta(W_2(w_{21}))$ is strictly increasing in w_{21}. The second term is also strictly decreasing in w_{21}, because $q(\theta)$ is a strictly increasing function of θ. Hence, the objective function is strictly concave in w_{21}. This implies that $\hat{w}_{21}$ is the unique solution to

$$\frac{\epsilon_q(\theta(W_2(w_{21})))}{1-\epsilon_q(\theta(W_2(w_{21})))}\frac{\upsilon'(w_{21})}{\upsilon(w_{21})-\upsilon(b)}(y_2-w_{21})-1=0. \tag{9}$$

The lhs of (9) is strictly increasing in y_2 and strictly decreasing in w_{21}. Hence, $\hat{w}_{21}$ is strictly increasing in y_2. For any given w_{21}, the lhs of (9) is strictly positive when y_2 is sufficiently large. Hence, $\lim_{y_2\to\infty}\hat{w}_{21}=\infty$. From these properties of $\hat{w}_{21}$ and from the fact that $\hat{w}_{21}\in(z,y_2)$ and $\hat{w}_{22}=\min\{w_1,y_2\}$, it follows that there exists a unique $\hat{y}$ such that $\hat{w}_{21}>(<)\hat{w}_{22}$ if $y_2>(<)\hat{y}$.

C Proof of Proposition 4

The maximization problem (K) can be written as

$$\begin{aligned} V_k = \max_{w_{21},w_{22}} \quad & [\overline{n}-n_1(1-\delta)]\,\tilde{q}(w_{21})(y_2-w_{21})+n_1\upsilon(b)/\upsilon'(w_1) \\ & n_1(1-\delta)\left\{[\upsilon(w_{22})-\upsilon(b)]/u'(w_1)+y_2-w_{22}\right\}, \\ \text{s.t.} \quad & b\le w_{22}\le w_{22}\le y_2. \end{aligned} \tag{10}$$

First, notice that the objective function in (10) is continuous in (w_{21},w_{22}) and the the feasible set is non-empty and compact. Hence, a solution to the maximization problem (10) exists. Second, notice that the objective function is strictly concave in (w_{21},w_{22}) and the feasible set is convex. Hence, the solution to the maximization problem (10) is unique. Moreover, the solution (w_{21}^k,w_{22}^k) and the multiplier λ^k satisfy the following necessary and sufficient conditions for optimality with respect to (w_{21},w_{22},λ):

(i) the wage w_{21} is such that

$$[\overline{n}-n_1(1-\delta)]\left[\tilde{q}'(w_{21})(y_2-w_{21})-\tilde{q}(w_{21})\right]+\lambda\ge 0$$

and $w_{21} \leq y_2$, with complementary slackness;

(ii) the wage w_{22} is such that

$$n_1(1-\delta)\left[\upsilon'(w_{22})/\upsilon'(w_1) - 1\right] - \lambda \leq 0$$

and $b \leq w_{22}$, with complementary slackness;

(iii) the multiplier λ is such that $\lambda \geq 0$ and $w_{21} - w_{22} \geq 0$, with complementary slackness.

For all $y_2 \geq \hat{y}$, it is immediate to verify that the triple $(\hat{w}_{21}, \hat{w}_{22}, 0)$ satisfies the necessary and sufficient conditions for optimality (i)–(iii). For all $y_2 < \hat{y}$, it is immediate to verify that the optimality conditions (i)–(iii) are satisfied by the triple $(w_2^k, w_2^k, \lambda^k)$, where λ^k is equal to $n_1(1-\delta)[\upsilon'(v_2)/\upsilon'(w_1) - 1]$ and w_2^k is the solution for w_2 to the following inequalities

$$n_1(1-\delta)\left[\frac{\upsilon'(w_2)}{\upsilon'(w_1)} - 1\right] + \left[\overline{n} - n_1(1-\delta)\right]\left[\tilde{q}'(w_2)(y_2 - w_2) - \tilde{q}(w_2)\right] \geq 0 \tag{11}$$

and $w_2 \leq y_2$, with complementary slackness.

To characterize the properties of w_2^k, it is useful to denote as y_k the solution for y_2 to the following equation

$$n_1(1-\delta)\left[\frac{u'(y_2)}{u'(w_1)} - 1\right] - \left[\overline{n} - n_1(1-\delta)\right]\tilde{q}(y_2) = 0.$$

Clearly, y_k is strictly greater than z and strictly smaller than $\hat{y}$. First, consider the case where $y_2 \leq y_k$. In this case, the lhs of (11) is strictly decreasing in w_2, and it is positive at $w_2 = y_2$. Hence, w_2^k equals y_2. Second, consider the case where $y_2 \in (y_k, \hat{y})$. In this case, the first term on the lhs of (11) is strictly decreasing in w_2, and it is equal to zero at $w_2 = w_1$. The second term is strictly decreasing in w_2, and it is equal to zero at $w_2 = \hat{w}_{21}$, where $\hat{w}_{21} < w_1$. Moreover, the sum of the first and the second terms is strictly negative at

$w_2 = y_2$. Hence, w_2^k is strictly greater than $\hat{w}_{21}$ and strictly smaller than $\hat{w}_{22} = \min\{w_1, y_2\}$.

D Proof of Proposition 5

Without the replacement constraint $w_{12} < w_{22}$, the maximization problem (R) can be written as

$$\overline{V}_r = \max_{w_{21}} \left\{ \begin{array}{l} \overline{n}\tilde{q}(w_{21})(y_2 - w_{21}) + n_1 \upsilon(b)/\upsilon'(w_1) + \\ +n_1(1-\delta)(1-\tilde{q}(w_{21})) \max\limits_{b \le w_{22} \le y_2} \{[\upsilon(w_{22}) - \upsilon(b)]/\upsilon'(w_1) + y_2 - w_{22}\} \end{array} \right\} \quad (12)$$

Consider the inner maximization problem in (12). As established in Proposition 3, the solution to this problem, $\overline{w}_{22}^r$, is equal to $\min\{w_1, y_2\}$. Next consider the outer maximization problem in (12). The objective function is continuous and differentiable in w_{21}. Also, the derivative of the objective function with respect to w_{21} is equal to zero for all $w_{21} \le z$ and is strictly negative for all $w_{21} \ge \hat{w}_{21}$. Hence, the solution to this problem, $\overline{w}_{21}^r$, exists and belongs to the interval $[z, \hat{w}_{21})$. Finally, note that $\overline{w}_{21}^r < \overline{w}_{22}^r$ because $\hat{w}_{21} < \min\{w_1, y_2\}$ for all $y_2 < \hat{y}$. Hence, the solution to the maximization problem (12) is also the solution to (R).

E Proof of Theorem 6

The maximization problem (C) can be written as

$$\begin{aligned} \hat{V}(y_2) = \max_{w_{21}, w_{22}} \quad & [\overline{n} - n_1(1-\delta)]\,[\tilde{q}(w_{21})(y_2 - w_{21})] + n_1\upsilon(b)/\upsilon'(w_1) \\ & + n_1(1-\delta)\,\{[\upsilon(w_{22}) - \upsilon(b)]/\upsilon'(w_1) + y_2 - w_{22}\} \\ \text{s.t.} \quad & b \le w_{21}, w_{22} \le y_2, \end{aligned} \quad (13)$$

where we made the dependence of $\hat{V}$ on y_2 explicit. The objective function in (13) is continuous in w_{21}, w_{22} and y_2. The feasible set in (13) is non-empty, bounded and continuous in y_2. Hence, by the Theorem of the Maximum, the value function $\hat{V}(y_2)$ is continuous in y_2.

The maximization problem (K) can be written as

$$
\begin{aligned}
V_k(y_2) = \max_{w_{21}, w_{22}} \quad & [\overline{n} - n_1(1-\delta)]\,[\tilde{q}(w_{21})(y_2 - w_{21})] + n_1 \upsilon(b)/\upsilon'(w_1) \\
& + n_1(1-\delta)\,\{[\upsilon(w_{22}) - \upsilon(b)]\,/\upsilon'(w_1) + y_2 - w_{22}\} \\
\text{s.t.} \quad & b \leq w_{21}, w_{22} \leq y_2,
\end{aligned} \tag{14}
$$

The objective function in (14) is continuous in w_{21}, w_{22} and y_2. The feasible set in (14) is non-empty, bounded and continuous in y_2. Hence, by the Theorem of the Maximum, the value function $V_k(y_2)$ is continuous in y_2. Now, notice that the objective function in (14) is the same as in (13) and the feasible set in (14) is a subset of the feasible set in (13). Hence, $V_k(y_2) \leq \hat{V}(y_2)$. Moreover, note that (w_{21}^k, w_{22}^k) equals $(\hat{w}_{21}, \hat{w}_{22})$ for all $y_2 \geq \hat{y}$. Hence, $V_k(y_2) = \hat{V}(y_2)$ for all $y_2 \geq \hat{y}$.

The value function $V_r(y_2)$ is smaller than $\overline{V}_r(y_2)$, where

$$
\begin{aligned}
\overline{V}_r(y_2) = \max_{w_{21}} \quad & \overline{n}\tilde{q}(w_{21})(y_2 - w_{21}) + n_1 \upsilon(b)/\upsilon'(w_1) + \\
& + n_1(1-\delta)(1-\tilde{q}(w_{21}))\,\{[\upsilon(\hat{w}_{22}) - \upsilon(b)]\,/\upsilon'(w_1) + y_2 - \hat{w}_{22}\} \\
\text{s.t.} \quad & z \leq w_{21} \leq y_2.
\end{aligned} \tag{15}
$$

The objective function in (15) is continuous in w_{21} and y_2. The feasible set in (15) is non-empty, bounded and continuous in y_2. Hence, by the Theorem of the Maximum, the value function $\overline{V}_r(y_2)$ is continuous in y_2.

The difference between $\hat{V}(y_2)$ and $\overline{V}_r(y_2)$ is

$$
\begin{aligned}
& \hat{V}(y_2) - \overline{V}_r(y_2) \\
= \quad & \min_{z \leq w_{21} \leq y_2} n_1(1-\delta)\tilde{q}(w_{21})\,\{[\upsilon(\hat{w}_{22}) - \upsilon(b)]\,/\upsilon'(w_1) + w_{21} - \hat{w}_{22}\} \\
& + [\overline{n} - n_1(1-\delta)][\tilde{q}(\hat{w}_{21})(y_2 - \hat{w}_{21}) - \tilde{q}(w_{21})(y_2 - w_{21})].
\end{aligned} \tag{16}
$$

From the concavity of υ, it follows that $\upsilon(\hat{w}_{22}) - \upsilon(b)$ is strictly greater than $\upsilon'(\hat{w}_{22})\,(\hat{w}_{22} - b)$. From the inequalities $\hat{w}_{22} \leq w_1$ and $b \leq z \leq w_{21}$, it follows that $\upsilon'(\hat{w}_{22})\,(\hat{w}_{22} - b)$ is strictly greater than $\upsilon'(w_1)(\hat{w}_{22} - w_{21})$. Hence, the first term on the rhs of (16) is strictly positive for all $w_{21} > z$, and it is equal to zero for $w_{21} = z$. Moreover, the second term on the rhs of

(16) is strictly positive for all $w_{21} \neq \hat{w}_{21}$, and it equals zero for $w_{21} = \hat{w}_{21}$. Since $\hat{w}_{21} > z$, it follows that the rhs of (16) is strictly positive. Hence, $\hat{V}(y_2) > \overline{V}_r(y_2)$.

In the above paragraphs, we have established that $V_k(y_2) = \hat{V}(y_2)$ and $\overline{V}_r(y_2) < \hat{V}(y_2)$ for all $y_2 \geq \hat{y}$. Since the value functions $V_k(y_2)$, $\hat{V}(y_2)$ and $\overline{V}_r(y_2)$ are continuous with respect to y_2, there exists an $\overline{y} \in [z, \hat{y})$ such that $V_k(y_2) > \overline{V}_r(y_2)$ for all $y_2 \in (\overline{y}, \hat{y})$. Since $\overline{V}_r(y_2) \geq V_r(y_2)$, the previous observations imply that the optimal wage policy prescribes the wages (w_{21}^k, w_{22}^k) for all $y_2 > \overline{y}$.