DISCUSSION PAPER SERIES

IZA

IZA DP No. 9967

Optimal Unemployment Benefit Policy and the Firm Productivity Distribution

Tomer Blumkin
Leif Danziger
Eran Yashiv

May 2016

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Optimal Unemployment Benefit Policy and the Firm Productivity Distribution

Tomer Blumkin
Ben-Gurion University and IZA

Leif Danziger
Ben-Gurion University and IZA

Eran Yashiv
Tel-Aviv University and IZA

Discussion Paper No. 9967
May 2016

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 9967
May 2016

ABSTRACT

Optimal Unemployment Benefit Policy and the Firm Productivity Distribution

This paper provides a novel justification for a declining time profile of unemployment benefits that does not rely on moral hazard or consumption-smoothing considerations. We consider a simple search environment with homogeneous workers and low- and high-productivity firms. By introducing a declining time profile of benefits, the government can affect the equilibrium wage profile in a manner that enhances the sorting of workers across low- and high-productivity firms. We demonstrate that optimal government policy depends on the dispersion and skewness of the firms' productivity distribution.

JEL Classification: J64, J65

Keywords: unemployment benefit policy, declining unemployment benefits, productivity distribution, skewness, dispersion

Corresponding author:

Leif Danziger
Department of Economics
Ben-Gurion University
84105 Beer-Sheva
Israel
E-mail: danziger@bgu.ac.il

1 Introduction

A prevalent feature of unemployment benefit policy in most OECD countries is a declining benefit schedule. The purpose of the current paper is to provide a normative justification for this pattern. We argue that a declining benefit schedule may serve as a means to internalize matching externalities induced by search frictions in the labor market. Our focus is on endogenous worker allocation across firms with different productivity levels and the associated wage determination. We consider a tractable framework designed to capture the essential feature of a labor-market with search frictions. The economy consists of homogeneous workers and low- and high-productivity firms. Each firm can post a vacancy at a specified wage; there is random assignment of unemployed workers to job vacancies; and each assigned worker decides whether or not to accept the job offer (no bargaining). Once accepting a job offer, a worker does not continue to search (no on-the-job search). There is an exogenous separation rate. The unemployment benefit system is introduced to internalize the induced matching externalities reflected in the steady-state equilibrium.

In the benchmark case with no unemployment benefits, both types of firms will be active and will offer the same wage, which will be accepted by any unemployed worker. As random matching implies an identical matching probability of workers to low- and high-productivity firms, average worker productivity will be relatively low. By introducing unemployment benefits, the government may be able to affect the equilibrium wage profile in a manner that enhances the average worker productivity by shifting workers from low- to high-productivity firms.

In our setting, the government can choose between the benchmark (no-intervention) regime where workers are assigned randomly across firms, and two forms of intervention: a flat regime with high constant unemployment benefits over time, where only high-productivity firms operate so that all workers are allocated to these firms; and a two-tier regime with a decreasing time profile of benefits, where both low- and high-productivity firms operate and relatively more workers are allocated to high-productivity firms. We will henceforth refer to each of the above three configurations according to the degree of "sorting" associated with each configuration, where the term sorting indicates the extent to which workers are shifted from low- to high-productivity firms relative to the benchmark case. We will henceforth refer to the benchmark regime as "no-sorting", a flat regime with high constant unemployment benefits as "full-sorting", and a two-tier regime with a decreasing time profile of benefits as "partial-sorting".

The main result of this paper is that a decreasing time profile of benefits may be preferred to a regime with constant benefits over time. This would be the case when there are moderate differences between firms' productivity levels. The logic is as follows: low-productivity firms offer lower wages than high-productivity firms.[1] Workers at the beginning of their unemployment spell, when benefits are high, turn down offers to work in low-productivity firms. This leads to voluntary unemployment and to more workers who search, shifting employment to high-productivity firms. In this case, employment and average worker productivity fall between what they would be with no unemployment benefits and with high, constant unemployment benefits that would crowd out the low-productivity firms.

We demonstrate that optimal government policy for the time path of unemployment benefits hinges on the properties of the firm productivity distribution. Sorting gains are closely associated with the difference in productivity between low- and high- productivity firms. As may be expected, the larger this difference is, the bigger the gain from the enhanced sorting of workers across firms. That is, the optimal degree of sorting tends to increase with

[1] Mortensen (2003) provides evidence whereby wage dispersion for observationally equivalent workers can be explained by more productive firms paying higher wages. For detailed empirical work, see Davis et al. (1996), Abowd, Kramarz and Margolis (1999), Haltiwanger et al. (1999, 2007) and Bartelsman et al. (2013).

the productivity differences between firms, i.e., first shifting from no-sorting to partial-sorting and then from partial-sorting to full-sorting. Somewhat surprisingly, the proportion of high-productivity firms also plays a key role. Specifically, if the productivity distribution is sufficiently right-skewed, i.e., the proportion of high-productivity firms is small enough, the cost in terms of reduced employment implied by the crowding out of all low-productivity firms from the market is high. Hence, partial sorting dominates full sorting of workers, warranting the implementation of a declining time profile of benefits. In contrast, if the productivity distribution is less right-skewed, i.e., the proportion of high-productivity firms is big enough, full sorting and hence a constant benefits schedule is optimal. The empirical fact that the firm productivity distribution is right-skewed is amply made by the literature, see for example, Feng and Horrace (2012) and Kashara and Lapham (2013) for evidence on the skewness of the firm productivity distribution and Lazear and Shaw (2008) and Heckman and Sattinger (2015) for an overview of the evidence on the skewness of the wage distribution (which is closely related to the productivity distribution).[2] Thus, the current paper makes its key contribution in this empirically-relevant case.

The paper is structured as follows: Section 2 provides a review of the relevant literature and Section 3 describes some empirical observations on actual unemployment insurance systems and their impact on labor market outcomes. Section 4 outlines the set-up. Section 5 presents the benchmark of no unemployment benefits while Section 6 shows two cases of government intervention – a flat unemployment benefit schedule and a declining one. Section 7 outlines the government objective and derives the optimal policy. The proof of the proposition that characterizes the social optimum as a function of the productivity difference between high- and low-productivity firms is in the Appendix. In Section 8 we

[2] Note as well that the much discussed skewness of the distribution of firm size (Luttmer (2007)) is consistent with our model as declining unemployment benefits imply that the expected number of workers per firm increases with firm productivity.

provide a numerical illustration. Section 9 concludes.

2 Literature Review

Our paper relates to two strands of the literature: the search literature and the literature on optimal unemployment compensation policy. The search literature (for a recent survey see Rogerson, Shimer and Wright (2005)), typically identifies three main classes of search models: random matching and bargaining, directed search and wage posting, and random matching and wage posting. The first class (Pissarides (2000)), does not allow for the wage posting behavior of firms and is not geared to explain wage dispersion, which is a salient feature of our analysis. The second class (Moen (1997)), sometimes referred to as "competitive search theory," does share a key feature with the current approach: firms set wages optimally, knowing that the probability of filling a job rises with the wage offer. Additionally, as in Moen (2003), labor market segmentation arises due to the fact that firms cannot condition wage offers on the worker type and workers' productivities differ across matches. But the segmentation in the current paper does not take the form of sub-markets and the operation of market makers; rather, it is due to exogenous productivity dispersion and the effects of unemployment compensation policy. Most importantly, competitive search theory predicts that the resulting allocation of workers across firms would be efficient and would, therefore, obviate the role of the unemployment benefit system in internalizing matching externalities, which is a key feature of our analysis. The current paper combines wage posting with random matching and belongs to the last class of models. This class of models has been widely used in the microeconomic literature, with the prominent example being the seminal work of Burdett and Mortensen (1998)(see also the discussion in Section 6 of Rogerson, Shimer and Wright (2005)). For empirical evidence on the prevalence of this type of wage setting in the US labor market, see Hall and Krueger (2012). Following this strand of literature, the two key elements of the model are wage posting by firms that maximize their discounted expected profit and the determination of reservation wages by risk-neutral workers who maximize their discounted expected income.

A major line of research on the optimal design of unemployment compensation policy has focused on issues of moral hazard and consumption smoothing; see Karni (1999) and Tatsiramos and van Ours (2014) for surveys. This literature examines the impact of work disincentives on the design of optimal schemes (the seminal papers are by Baily (1978), Flemming (1978) and Shavell and Weiss (1979)). The main insight provided by the early models was the desirability of a declining schedule, i.e. compensation should decline over the spell of unemployment so as to mitigate the moral hazard effect. The early models have been recently extended in several directions, some of them into general equilibrium frameworks. Hopenhayn and Nicolini (1997), as a notable example, enlarge the set of instruments by allowing for a wage tax after re-employment. This model preserves the sequencing structure of Shavell and Weiss (1979) and attains enhanced consumption smoothing. Other models have incorporated allocation and matching elements. For example, Cremer et al. (1996) show that when individuals are risk averse, they tend to choose unsuitable jobs as they cannot afford waiting for suitable ones in terms of preferences and productivity. Providing a flat unemployment benefit scheme can then enhance welfare by allowing workers to wait for the right offer. This happens at a cost of pushing some workers into permanent unemployment. In order to mitigate the tradeoff between the two, a two-tier declining schedule is shown to be desirable. This paper adds a novel adverse selection argument for a declining schedule, namely reducing the subsidy to the permanently unemployed, who can not be readily distinguished. Acemoglu and Shimer (1999) posit that unemployment compensation generates an increase in output, whereby more productive firms choose to offer higher wages and more workers are assigned to those firms. This model has risk aversion at the heart of the analysis and unemployment compensation has an insurance role. By offering unemployment compensation, apart from the consumption smoothing argument, the policy maker induces risk-averse workers to take on a higher degree of unemployment risk, boosting investment by firms. Their set-up is one with directed search, so externality issues do not arise. Wang and Williamson (2002) evaluate alternative unemployment insurance schemes in a dynamic economy with moral hazard. They consider changes in the size and duration of benefits, and the effects of experience rating, and use a dynamic contracting approach to determine a benchmark optimal allocation. They find that radical changes in the current system increase welfare, but not by much. A move to full experience rating has distributional effects, but aggregate effects are negligible.

The current paper does not belong in the above strand, as it does not consider issues of risk aversion, consumption smoothing, moral hazard, or adverse selection. Rather it focuses on the role unemployment compensation can play in attaining a better match between jobs and workers, deriving optimal policy in the face of productivity dispersion. The seminal contribution in this context has been made by Diamond (1981), who discussed the role of unemployment compensation in enhancing efficiency in the context of a steady state search model. In his model unemployment compensation makes job-taking use more stringent standards, thereby raising the vacancy rate and improving the distribution of job offers.

There are a number of more recent contributions that have dealt with related issues. Marimon and Zilibotti (1999) show that unemployment compensation improves matching between ex-ante heterogenous workers and ex-ante heterogenous firms under random matching. Unemployment compensation serves to reduce worker-job mismatch, as without unemployment compensation, workers would tend to accept unsuitable jobs. Cahuc and Lehmann (2000) ask whether unemployment benefits should decrease with the unemployment spell in a model where both job search intensity and wages are endogenous. The latter are set by collective agreements bargained by insiders. It is shown that a declining time path of unemployment benefits leads to wage increases when the tax rate is given. Such an effect may imply an increase in unemployment and counteracts the response of job search intensity that can be found in standard job search models, with a given wage distribution. Calibration exercises show that it costs twice as much in terms of welfare loss for the long-term unemployed workers to reduce the unemployment rate by 1% when wages are endogenous than in the standard job search model. Fredriksson and Holmlund (2001) use a standard Pissarides (2000) framework to analyze the equilibrium effects of time-varying unemployment compensation. They find that an optimal scheme – under certain conditions – has compensation decline over time. This is so because of an 'entitlement effect', according to which raising the compensation offered to the insured induces additional search effort among the uninsured, bringing them more quickly to employment, which results in future unemployment compensation eligibility. Burkhard (2003) studies the effects of a two-tier unemployment compensation system in a general equilibrium job search model with endogenous distributions of income, wealth, and employment. The model is calibrated to match the German economy. Two key results are that employment is a decreasing function of both unemployment insurance and unemployment assistance, and optimal unemployment compensation follows a declining time path.

The last group of papers typically does not contain the two key ingredients of the current paper together: firm productivity dispersion, which affects both wage offers and matching, coupled with a normative analysis of optimal policy. The current paper is thus able to revisit the case for a declining time profile of unemployment benefits and to provide a novel justification. To summarize our approach: We follow Mortensen (1977), Diamond (1981), Albrecht and Axell (1984) and Marimon and Zilibotti (1999) in viewing unemployment benefits as a search subsidy, and we study the role policy may play in attaining a better match between jobs and workers. The crucial insight is that in a search environment where the assignment of unemployed workers to heterogenous firms is uncoordinated, there is a tradeoff between employment and average worker productivity. As this tradeoff is not adequately reflected in the equilibrium wage profile, unemployment benefits may help to internalize these matching externalities.

Before we conclude this section, two remarks is in order. (1) To the best of our knowledge, this is the first paper that provides a normative framework linking the socially desired time path of unemployment benefits with the moments of the firm productivity distribution. (2) Whereas the mechanism at work seems prima-facie similar – a declining time profile of unemployment benefits induces short-term unemployed workers to reject wage offers accepted by their long-term counterparts – it is important to clarify the key difference between the justification provided by the earlier literature for the declining time pattern and the novel rationale offered in our paper. In earlier contributions (Shavell and Weiss (1979), Cremer et al. (1996) and Hopenhayn and Nicolini (1997), inter-alia) the key role played by unemployment insurance benefits was to provide consumption smoothing to risk-averse workers. Setting a declining time profile served to mitigate the moral hazard entailed by consumption smoothing. In our setting, workers are risk neutral, hence, there is no need for consumption smoothing. The rationale for a declining time profile comes from the sorting gains associated with matching homogenous workers with heterogenous jobs. In this sense, our normative justification is complementary to the one offered by the earlier literature.

3 Empirical Observations on Unemployment Benefits

The current paper focuses on the normative aspect of the design of unemployment insurance benefits. In this section we briefly present some empirical observations on actual unemployment insurance systems and their impact on labor market outcomes.

Unemployment benefits in practice reflect a large number of important policy choices including, amongst others, the time path of benefits, the earnings base, and the eligibility criteria such as means-testing, age and qualifying period. The policy decision on which we focus in this paper is the time path of benefits. A prevalent feature of unemployment benefit policy in most OECD countries is a declining benefit schedule, which can take two forms. The first is to have explicitly declining benefits as, for example, in Italy, the Netherlands, Poland, Spain, Sweden and Switzerland. The second is to have an implicit two-tier system, whereby unemployment benefits are offered for a limited period and are then subsequently replaced by social/income assistance at a lower level. Within both structures, durations vary a lot across countries, from a low of fourteen weeks to a maximum of six years (or even with no limit, as in Australia), but are typically around a quarter to a year.[3]

Regarding the impact on labor market outcomes, the literature finds that more generous benefits have a positive effect on re-employment wages (see Ehrenberg and Oaxaca (1976) for an early documentation) and that when benefits are higher, unemployed workers are more selective and worker productivity is higher (see Marimon and Zilibotti (1999)). Only few empirical studies examine the relationship between the size and the duration of unemployment benefits and the level of employment, with generally ambiguous results (see Hagedron, Manvoskii and Mitman (2015)).

4 The Set-Up

We consider a simple labor market with search and matching frictions. There is a continuum of homogeneous risk-neutral workers with a measure $L > 0$ and a continuum of firms with a measure $M > L$. There is no entry or exit of firms. Firms differ in the technology they

[3] See Tatsiramos and van Ours (2014, Table 2), and for updated statistics, the following website http://www.oecd.org/els/benefitsandwagesstatistics.htm.

possess. A proportion $p \in (0,1)$ of firms have high productivity and worker's output per period is $\overline{x}$, while a proportion $1-p$ have low productivity and worker's output per period is $\underline{x} \in (0, \overline{x})$. High-productivity firms are relatively scarce. In particular, we assume that $p < L/M$, which implies that even in a frictionless labor market all the workers cannot simultaneously work in high-productivity firms.[4] The mean of the firms' productivities is $\mu \equiv p\overline{x} + (1-p)\underline{x}$, and the difference in their productivities is $\theta \equiv \overline{x} - \underline{x}$. Firms maximize their discounted expected profits. Each firm can employ one worker and posts one vacancy at a specified wage and each unemployed worker can apply to one vacancy per period. Due to search frictions, the probability of a match is less than one.[5]

There is no on-the-job search. The assignment of unemployed workers to vacancies occurs at the end of a period. It is random and governed by a constant-returns-to-scale matching function $m(U,V)$, where m denotes the measure of matches, U the measure of unemployed workers, V the measure of vacancies, and $m(U,V) \leq \min(U,V)$.[6] Workers maximize their discounted expected income. A worker who is matched to a vacancy decides whether to accept the firm's offer of a wage which s/he will receive for the duration of employment in the firm. An unmatched worker or a matched worker who rejects a firm's offer remains

[4] Suppose that firms can enter by incurring a fixed entry cost, that an entering firm's productivity is determined by a random draw (productivity is high with a probability p and low with a probability $1-p$), and that entry is determined by a zero discounted expected profit condition. We then conjecture that if the entry cost is sufficiently high, in equilibrium, the total measure of high-productivity firms will be lower than the total measure of the work-force, so that our qualitative results will remain valid.

[5] Instead of assuming that there is a measure pM of high-productivity firms, each being able to employ at most one worker, we could assume, alternatively, that there is less than pM high-productivity firms, each being able to employ more than one worker. Ultimately, what matters for our analysis is that the meaure of high-productivity jobs is given by pM and that firms post vacancies at a specified wage.

Notice that we are invoking a tractable wage determination mechanism which assigns the entire bargaining power to the firm. We conjecture that our qualitative results will remain valid under alternative wage determination mechanisms (such as Nash bargaining).

[6] For empirical evidence on random matching see Petrongolo and Pissarides (2001). The random matching assumption seems an appropriate way of capturing labor market frictions. These underlie the potential gain from government intervention discussed below.

unemployed in the next period. The imputed value of leisure is normalized to zero, with no loss of generality. A successful match terminates with an exogenous probability $s \in (0, 1)$. To close the model, we assume that the firms are owned equally by all the workers who therefore receive an equal share of the profits.

5 The Benchmark Regime: No Unemployment Benefits

We start by characterizing the equilibrium in a benchmark regime without government provision of unemployment benefits.

Following Diamond (1971), as employed workers do not search, a firm's posted wage offer must coincide with the workers' reservation wage. In other words, unemployed workers must be indifferent between accepting and rejecting the wage offer. Unemployed workers will be randomly assigned to vacancies across the two kinds of firms, and any job offer will be accepted. Consequently, the assignment of workers across jobs will be random.

Let V^N and U^N denote, respectively, the measures of vacancies and unemployed workers, and $\beta \in (0, 1)$ the workers' discount factor. The continuation value for an unemployed worker is

$$H = \beta[nJ + (1 - n)H], \tag{1}$$

and the continuation value for an employed worker is

$$J = w + \beta[(1 - s)J + sH], \tag{2}$$

where w is the wage and

$$n = \frac{m(U^N, V^N)}{U^N} \tag{3}$$

is the probability of being matched with a firm.

The continuation value of a filled vacancy in a firm with productivity x is

$$A(x) = x - w + \beta \left[(1 - s) A(x) + sB(x) \right], \tag{4}$$

and the continuation value of an unfilled vacancy in such firm is

$$B(x) = \beta \left[gA(x) + (1 - g) B(x) \right], \tag{5}$$

where

$$g = \frac{m(U^N, V^N)}{V^N}$$

is the probability of filling a vacancy, and the firms' discount factor, β, is identical to that of the workers.

In an equilibrium, as the value of leisure is normalized to zero, it follows that $w = 0$ and that all firms will be active. The flow of successful matches between unemployed workers and firms equals the flow into unemployment of workers due to job separations, i.e.,

$$m(U^N, V^N) = s \left(M - V^N \right). \tag{6}$$

In addition, the measure of filled vacancies is equal to the measure of employed workers, i.e.,

$$M - V^N = L - U^N. \tag{7}$$

6 Government Benefits Policy

We now allow the government to provide unemployment benefits. We assume that government expenditures on unemployment benefits are financed by a lump-sum tax levied on all workers (employed and unemployed) and hence not affecting the choices of workers and firms. This allows us to focus on the inefficiency associated with the matching externalities and the role of the unemployment benefit system in internalizing these.[7]

[7] To ensure that the fiscal system is sustainable, we assume that capital markets are perfect implying that there are no binding liquidity constraints. Without loss of generality we also assume that the government has no revenue needs.

6.1 Flat Unemployment Benefit Regimes

Suppose unemployment benefits, denoted by a, are constant over time. Then, there are two possibilities to consider. If $a < \underline{x}$, all firms will be active, and the assignment of workers and aggregate output in equilibrium will be as in the benchmark case without government intervention. Formally, the equilibrium measures of unemployment and vacancies will be determined by the same flow conditions (6) and (7) as in the benchmark regime without unemployment benefits, and the government's (balanced) budget constraint is given by

$$aU^N = tL,$$

where t denotes a lump-sum tax levied on both employed and unemployed workers. Furthermore, the equilibrium wage will coincide with the workers' reservation wage, which is now a.

In contrast, if $\underline{x} < a < \overline{x}$, only high-productivity firms will be active, unemployed workers will be randomly assigned to vacancies posted by these firms, and all wage offers will be accepted. Hence, the equilibrium measures of unemployment, U^F, and of vacancies, V^F, will be determined by the flow condition

$$m(U^F, V^F) = s\left(pM - V^F\right) \tag{8}$$

together with the condition that the measure of jobs filled is equal to the measure of employed workers, i.e.,

$$pM - V^F = L - U^F. \tag{9}$$

Thus, the equilibrium wage will again be equal to the workers' reservation wage which is a.

The government's (balanced) budget constraint is given by

$$aU^F = tL,$$

where t denotes a lump-sum tax levied on both employed and unemployed workers.

6.2 A Two-Tier Unemployment Benefit Regime

Now suppose newly unemployed workers receive two periods of high unemployment benefits followed by an indefinite period of lower benefits. In such a regime, short-term unemployed workers (one or two periods of unemployment) get unemployment benefit z, whereas long-term unemployed workers get benefit a, where $a < z$. We will derive an equilibrium in which the low-productivity firms offer a low wage, $\underline{w}$, and the high-productivity firms offer a high wage, $\overline{w}$, where $\underline{w} < \overline{w}$. We first characterize the properties of such equilibrium and then show that the equilibrium exists for a proper choice of a and z.

We first consider unemployed workers. These can be divided into those who are (*i*) in their first period of an unemployment spell; (*ii*) in their second period of an unemployment spell; and (*iii*) unemployed for more than two periods during their current unemployment spell.

Let $\underline{V}$ and $\overline{V}$ denote the measures of vacancies posted by firms offering the low and high wage rates, respectively, $V^P \equiv \underline{V} + \overline{V}$ the measure of all vacancies, and U^P the measure of unemployed workers. The probabilities of being matched with firms offering low and high wage rates are then

$$\underline{n} = \frac{m(U^P, V^P)}{U^P} \frac{\underline{V}}{V^P},$$
$$\overline{n} = \frac{m(U^P, V^P)}{U^P} \frac{\overline{V}}{V^P}.$$

The continuation values associated with workers in the first period of their current unemployment spell is

$$H_1 = z + \beta[\overline{n} \max(J_{\overline{w}}, H_2) + \underline{n} \max(J_{\underline{w}}, H_2) + (1 - \overline{n} - \underline{n})H_2], \quad (10)$$

that associated with workers in the second period of their current unemployment spell is

$$H_2 = z + \beta[\overline{n}\max(J_{\overline{w}}, H_{\geq 3}) + \underline{n}\max(J_{\underline{w}}, H_{\geq 3}) + (1 - \overline{n} - \underline{n})H_{\geq 3}], \tag{11}$$

and that associated with workers unemployed for at least three periods in their current unemployment spell is

$$H_{\geq 3} = a + \beta[\overline{n}\max(J_{\overline{w}}, H_{\geq 3}) + \underline{n}\max(J_{\underline{w}}, H_{\geq 3}) + (1 - \overline{n} - \underline{n})H_{\geq 3}], \tag{12}$$

where

$$J_{\underline{w}} = \underline{w} + \beta[(1-s)J_{\underline{w}} + sH_1] \tag{13}$$

$$J_{\overline{w}} = \overline{w} + \beta[(1-s)J_{\overline{w}} + sH_1] \tag{14}$$

denote the continuation values associated with holding low and high wage jobs.[8]

We next consider firms. The continuation value of a filled vacancy in a firm with productivity x offering the wage $w \in \{\underline{w}, \overline{w}\}$ is given by

$$A(x,w) = x - w + \beta\left[(1-s)\,A(x,w) + sB(x,w)\right], \tag{15}$$

and the continuation value of an unfilled vacancy in such firm is

$$B(x,w) = \beta\left\{g(w)A(x,w) + \left[1 - g(w)\right]B(x,w)\right\}, \tag{16}$$

where the probability of filling a high-wage (low-wage) vacancy, respectively, is given by

$$\begin{aligned} g(\overline{w}) &= \overline{g} = \frac{m(U^P, V^P)}{V^P}, \\ g(\underline{w}) &= \underline{g} = \frac{m(U^P, V^P)}{V^P}\left(\frac{U^P - U_1}{U^P}\right) \end{aligned}$$

[8] Note that the subscript of H refers to the length of the current unemployment spell and not to absolute time.

with $U_1 \equiv s(L - U^P)$ denoting the measure of unemployed workers in their first period of unemployment. The term $(U^P - U_1)/U^P$ is less than unity and captures the fact that only a proportion of the matches with low-wage vacancies are successful as first-period unemployed workers turn down low-wage offers. Consequently, offering the higher wage increases the likelihood of filling a vacancy.

As employed workers do not search, a firm's posted wage offer must coincide with one of the reservation wage rates. That is, workers unemployed for more than one period must be indifferent between accepting and rejecting the low wage offer, i.e., $J_{\underline{w}} = H_{\geq 3}$, whereas workers unemployed in the first period must be indifferent between accepting and rejecting the high wage offer, i.e., $J_{\overline{w}} = H_2$.

Using equations (10)-(14) we obtain that in equilibrium the wage rates are implicitly given by

$$z - a = \frac{\overline{w} - \underline{w}}{1 - \beta(1 - s)}, \tag{17}$$

$$\overline{w} = (z - a)[1 - \beta + \beta\overline{n} - \beta^2 s(1 - \overline{n})] + a. \tag{18}$$

In equilibrium the following flow conditions must also hold

$$\overline{g}\overline{V} = s\left(pM - \overline{V}\right), \tag{19}$$

$$\underline{g}\underline{V} = s[(1 - p)M - \underline{V}]. \tag{20}$$

Condition (19) states that the flow of successful matches between unemployed workers and high-productivity firms (the left-hand side) equals the flow into unemployment of workers due to separations from high-productivity firms (the right-hand side). Similarly, Condition (20) states that the flow of successful matches between unemployed workers and low-productivity firms equals the flow into unemployment of workers due to separations from low-productivity firms.[9]

In addition, in equilibrium the measure of filled vacancies is equal to the measure of employed workers,

$$M - V^P = L - U^P. \tag{21}$$

As low-productivity firms must find it optimal to post a low wage offer, i.e., the discounted expected profits associated with paying the low wage weakly exceeds that associated with paying the high wage, we have that

$$B\left(\underline{x}, \underline{w}\right) \geq B(\underline{x}, \overline{w}). \tag{22}$$

Similarly as high-productivity firms must find it optimal to post a high wage offer,

$$B(\overline{x}, \overline{w}) \geq B\left(\overline{x}, \underline{w}\right). \tag{23}$$

By properly choosing the policy parameters a and z, we can ensure that there exists a two-tier equilibrium. Formally, manipulating conditions (15) and (16) yields

$$B(x, w) = K(w)(x - w),$$

where

$$\begin{aligned} K(\overline{w}) &\equiv \frac{\beta \overline{g}}{(1-\beta)\left[1-\beta(1-s)+\beta\overline{g}\right]}, \\ K(\underline{w}) &\equiv \frac{\beta \underline{g}}{(1-\beta)\left[1-\beta(1-s)+\beta\underline{g}\right]}. \end{aligned}$$

Now, let $\underline{w} = \underline{x} - \epsilon$, and $\overline{w} = \underline{x} + \epsilon$, where $\epsilon > 0$. Inequality (22) is satisfied as $\overline{w} > \underline{x}$ so that low-productivity firms will never choose to offer the high wage. Inequality (23) holds if

[9] Fully differentiating equations (19) and (20) with respect to s we can obtain an expression for $\partial U^P/\partial s$. For technical reasons we assume that $\lim_{s\to 0}(\partial U^P/\partial s) < \infty$. This property holds for a large class of matching functions.

and only if

$$\epsilon \leq \frac{[K(\overline{w}) - K(\underline{w})](\overline{x} - \underline{x})}{K(\overline{w}) + K(\underline{w})}.$$

As $\overline{g} > \underline{g}$, it follows that $K(\overline{w}) - K(\underline{w}) > 0$. Hence, there exist values of ϵ that satisfy both (22) and (23). After substituting for $\underline{w}$ and $\overline{w}$, the policy parameters a and z are implicitly given by equations (17) and (18).

As a consequence, there exists a two-wage equilibrium with two-tier unemployment benefits. In this equilibrium, conditions (19), (20), and (21) determine the measures of unemployment, U^P, and of vacancies posted by firms offering high and low wage rates, $\overline{V}$ and $\underline{V}$.

The lump-sum tax, t, is set so at to satisfy the government's budget constraint which is now given by

$$a\left[U^P - U_1(2-\overline{n})\right] + zU_1(2-\overline{n}) = tL.$$

7 The Government Objective and Optimal Policy

A worker's utility coincides with his net income, which is the sum of his labor income, the net benefits received from the government, and his share of distributed profits. We assume that the government seeks to maximize the sum of the workers' utilities, which is equivalent to maximizing aggregate output.[10]

Due to matching frictions, the allocation of workers obtained under the benchmark setting does not generally achieve this aim. The equilibrium wage offered by both types of firms will be the same and therefore will not reflect firms' productivities. Firms are unable to signal their productivities via wage posting with the result that sorting externalities emerge. There

[10] Maximizing aggregate output is a common assumption in the search literature (e.g., Albrecht and Axell (1984)). Given the linearity of utility in income, an allocation is second-best efficient (given the matching friction) if and only if it maximizes the sum of utilities.

is no sorting of workers across low- and high-productivity firms with the random matching of workers, although some sorting may be desirable when productivities are sufficiently dispersed. As we show below, unemployment benefits may serve to internalize these sorting externalities.

In the benchmark case both types of firms are active and all unemployment is involuntary. There is no sorting of workers across the two types of firms, so aggregate output is given by

$$\begin{aligned} W^N &= (L - U^N)[p\overline{x} + (1-p)\underline{x}] \\ &= (L - U^N)\mu. \end{aligned} \tag{24}$$

In light of the characterization of possible equilibria delineated above, there are two alternative configurations of unemployment benefits that need to be considered. One possibility is a flat regime whereby the benefit level is set high enough so that only the high-productivity firms are active, all unemployment is involuntary and any wage offer is accepted.[11] Workers are fully sorted across low- and high-productivity firms, and as a consequence aggregate output is given by

$$W^F = (L - U^F)\overline{x} \tag{25}$$

$$= (L - U^F)[\mu + (1-p)\theta].$$

A second possibility is a two-tier regime of unemployment benefits that supports a two-wage equilibrium in which both types of firms are active and voluntary unemployment emerges with low wage offers rejected by the short-term unemployed. Workers are partially sorted over low- and high-productivity firms, and the associated aggregate output is

[11] Given that the government objective is to maximize aggregate output, and hence sets aside redistributive concerns, a flat regime whereby the benefit level is set so low such that all firms are active, coincides with the benchmark regime without unemployment benefits.

given by

$$W^P = (L - U^P)[q\overline{x} + (1-q)\underline{x}] \tag{26}$$
$$= (L - U^P)[\mu + (q-p)\theta],$$

where $q \equiv \left(pM - \overline{V}\right) / \left(M - V^P\right)$ is the fraction of employed workers that work in high-productivity firms. Conditions (19) and (20) imply that $p < q < 1$, where p and unity are the fractions of workers assigned to high productivity firms under no sorting and full sorting, respectively. Thus, the equilibrium associated with a two-tier regime (declining benefits) features partial sorting of workers across firms.

In attempting to maximize aggregate output, there is a tradeoff between employment and average worker productivity. There are three possible results, depending on the value θ of the difference in the firms' productivities:

Under no sorting of workers, all firms are active and there is no voluntary unemployment; therefore, employment is the highest possible. However, the quality of matches is relatively poor due to the random nature of the matching process, which implies the same matching probability in low- and high-productivity firms. For a given mean μ of the firms' productivities, when the difference in the firms' productivities θ is small, much is to be gained from increased employment and little to be lost from a reduction in average worker productivity due to there being no sorting of workers. In this case, W^N exceeds both W^F and W^P so that no sorting maximizes aggregate output. Thus, no intervention is called for.

In contrast, under full sorting of workers, low-productivity firms are inactive. Therefore, employment is low while average worker productivity is the highest possible, as all matches involve high-productivity firms. When the difference in the firms' productivities θ is large, the sorting consideration prevails. In this case, W^F exceeds both W^N and W^P so that full sorting of workers maximizes aggregate output.

When the difference in the firms' productivities θ is in an intermediate range, the gain from high employment with no sorting of workers is not large enough to justify not increasing average worker productivity by enhanced sorting, and the gain from full sorting of workers is not large enough to justify crowding out the low-productivity firms with its associated reduction in employment. Partial sorting of workers then constitutes a fitting compromise between no and full sorting. It implies that W^P exceeds W^N and W^F so that partial sorting maximizes aggregate output. Low-productivity firms remain active, but have a lower probability of filling a vacancy than their high-productivity counterparts. This would result in less employment than under no sorting but more than under full sorting of workers.

The following proposition characterizes the social optimum as a function of the productivity difference, θ. We focus on the empirically-relevant case of low separation rates.[12]

Proposition *For given p and μ and a low s, there exist cutoff levels $\underline{\theta}$ and $\overline{\theta}$ of θ, $0 < \underline{\theta} < \overline{\theta} < \mu/p$, such that:*

(i) The no-sorting configuration is uniquely optimal for $\theta \in (0, \underline{\theta})$;

(ii) The partial-sorting configuration is uniquely optimal for $\theta \in (\underline{\theta}, \overline{\theta})$;

(iii) The full-sorting configuration is uniquely optimal for $\theta \in (\overline{\theta}, \mu/p)$.

Furthermore,

(iv) Both the no- and the partial-sorting configurations are optimal for $\theta = \underline{\theta}$;

(v) Both the partial- and the full-sorting configurations are optimal for $\theta = \overline{\theta}$.

Proof See Appendix.

The optimal sorting of workers depends on the dispersion of productivities as measured by

[12] U.S. data on monthly separation rates indicate an average of about 1.5%, which is consistent with the example worked out in the next section (see, for example, Yashiv (2007, Table 1b) and Elsby et al. (2013, Figure 2)).

θ. Given the proportion p of high-productivity firms, the optimal degree of sorting increases in θ. That is, if θ is low enough, no sorting is called for. Increasing θ moderately shifts the economy into a region where partial sorting is desirable. As θ becomes high enough, full sorting becomes optimal.

In particular, there exist combinations of p and θ for which social welfare is maximized under partial sorting of workers. As we have shown that a regime with declining unemployment benefits is needed in order to obtain such partial-sorting equilibrium, the implication is that an unemployment benefit policy with a decreasing time profile is optimal in this case. Thus, we obtain the result that declining benefits may be optimal without having to invoke the standard argument in the literature that declining benefits serve as a means to mitigate the tradeoff between consumption smoothing and moral hazard.[13]

8 A Numerical Illustration

To further explore the properties of the social optimum, we consider the following numerical example. We set $M = 100$, $L = 70$, $s = 0.01$, and $\mu = 10$, and let the matching function take the form $m(U, V) = 0.1 * U^{0.5}V^{0.5}$ in the relevant range.[14] The figure depicts the optimal sorting configuration for various combinations of the proportion p of high-productivity firms and the productivity difference θ between high- and low- productivity firms. As the definitions of μ and θ imply that $\underline{x} = \mu - p\theta$, the fact that $\underline{x} > 0$ implies that $\theta < \mu/p$. The feasible combinations of p and θ therefore lie below the dashed curve $\theta = 10/p$.

Figure

[13] When no sorting is optimal, there is no need for unemployment benefits. In contrast, when full sorting is optimal, a flat regime with sufficiently high unemployment benefits is warranted.

[14] The constant-returns-to-scale Cobb-Douglas matching function has wide empirical support (see Petrongolo and Pissarides (2001)). Yashiv (2000) and Borowczyk-Martins at al. (2013) provide detailed estimates and discussion. Of course, the matching function can only take the form $0.1 * U^{0.5}V^{0.5}$ in the range where $0.1 * U^{0.5}V^{0.5} \leq \min(U, V)$. Since $V = U + 30$, the relevant range is $30/99 \leq U$.

Inspection of the figure reveals the following:

(i) There exists a non-empty range of parameter values for which partial sorting of workers maximizes social welfare.

(ii) The welfare dominance of the partial sorting regime occurs over an intermediate range of productivity dispersion.

(iii) Given a θ for which at least some sorting is called for, there exists a threshold level of p below which partial sorting is desirable and above which full sorting is the optimal choice. The reason is that for a sufficiently right-skewed productivity distribution (that is, for a sufficiently small p), shifting from partial to full sorting by crowding out all low-productivity firms is too costly in terms of the associated reduction in employment. In contrast, with a less right-skewed productivity distribution, the sorting-employment tradeoff tilts in the other direction and calls for implementation of the full-sorting regime.

(iv) The curve separating the regions where partial and full sorting constitute optimal policy consists of the combinations of p and θ for which the government is indifferent between implementing the partial and the full sorting regimes. The curve is downward sloping, reflecting the fundamental tension between dispersion and skewness of the productivity distribution: the larger is the dispersion of productivities, the stronger is the case for enhancing the sorting by shifting from a partial- to a full-sorting regime; in contrast, a more right-skewed productivity distribution works in the direction of shifting policy from full to partial sorting.

9 Concluding Remarks

This paper has shown that the commonly observed policy of declining unemployment benefits may be an efficient way of internalizing externalities that are generated by the sorting of workers across firms in a labor market with matching frictions. Accordingly, a two-tier declining unemployment benefit system may be desirable even in the absence of consumption-smoothing and moral-hazard considerations. For a wide range of parameter values, declining unemployment benefits will be preferred to flat unemployment benefits as the former maximize aggregate output by striking an optimal balance between employment and average worker productivity.

We focus on the role of unemployment benefits as a means to internalize matching externalities. Assuming non-directed search (random matching), wage posting, and no on-the-job search render the model tractable and the presence of composition externalities more manifest. Nonetheless, we conjecture that our key qualitative results would remain valid in a setting that partially relaxes these assumptions. In such a more general framework there would be wage dispersion even in the benchmark case without unemployment benefits, with high-productivity firms offering higher wages and being more likely to fill their vacancies than low-productivity firms. Such wage dispersion might reduce the attainable gain from the introduction of unemployment benefits but is likely to maintain the tension between average worker productivity and employment.

The analysis highlights the close link between the properties of the productivity distribution and optimal policy choice. We emphasize the key role played by the asymmetric nature of technological dispersion, namely the extent to which the productivity distribution is skewed to the right, on the sorting-employment tradeoff faced by the government. In particular, when the productivity distribution is sufficiently skewed to the right, choosing full rather than partial sorting is too costly in terms of reduced employment, with the result that a declining unemployment benefit policy is to be preferred.

Appendix

Proof of the Proposition

Fix p and μ, and consider a small value of s. Equations (24)-(26) show that the social welfare associated with the no-, partial- and full-sorting configurations are different and linear in θ. The linearity implies the single-crossing property that for each two of the three configurations there exists at most one θ for which the two configurations yield the same welfare. Hence, the set of θ's for which each configuration maximizes social welfare is given by a distinct (possibly empty) interval.

The Proposition will be proved in four steps. In step 1 we characterize the limiting values of dU^P/ds and $\overline{V}$ as $s \to 0$. In step 2 we establish that there exists a $\underline{\theta} > 0$ such that no sorting maximizes welfare if and only if $\theta \in (0, \underline{\theta}]$. In step 3 we use the limiting values determined in step 1 to establish that there exists a $\overline{\theta} < \mu/p$ such that full sorting maximizes welfare if and only if $\theta \in [\overline{\theta}, \mu/p)$. Finally, in step 4 we use the limiting value of dU^P/ds determined in step 1 to establish that $\underline{\theta} < \overline{\theta}$ so that partial sorting maximizes welfare if and only if $\theta \in [\underline{\theta}, \overline{\theta}]$.

Step 1: Characterization of the Limiting Values of dU^P/ds and $\overline{V}$ as $s \to 0$

Equations (19) and (20) imply that

$$\frac{(pM - \overline{V})\underline{V}}{[(1-p)M - \underline{V}]\,\overline{V}} = \frac{U^P}{U^P - s\,(L - U^P)}. \tag{27}$$

Since $\lim_{s\to 0} U^P = 0$ and hence $\lim_{s\to 0}\left[U^P - s\left(L - U^P\right)\right] = 0$, by applying l'Hospital's rule the limit of the right-hand side of (27) as $s \to 0$ is

$$\begin{aligned} & \lim_{s\to 0}\left[\frac{U^P}{U^P - s\left(L - U^P\right)}\right] \\ = \quad & \frac{\lim_{s\to 0}\left(dU^P/ds\right)}{\lim_{s\to 0}\left[(1+s)(dU^P/ds) + U^P - L\right]} \\ = \quad & \frac{\lim_{s\to 0}\left(dU^P/ds\right)}{\lim_{s\to 0}\left(dU^P/ds\right) - L}. \end{aligned} \tag{28}$$

To facilitate the exposition, we will establish step 1 by use of the following five claims.

Claim 1: $\lim_{s\to 0}\left(dU^P/ds\right) \geq L$

Let $t(s) \equiv U^P - s\left(L - U^P\right)$. By construction, as $\lim_{s\to 0} U^P = 0$, we have that $\lim_{s\to 0} t(s) = 0$ and $\lim_{s\to 0} t'(s) = \lim_{s\to 0}(dU^P/ds) - L$. For $s > 0$, since some workers are unemployed and some unemployed workers have been unemployed for more than one period, we have that $U^P > 0$ and $U^P - s\left(L - U^P\right) > 0$. Thus, $t(s) > 0$ for all $s > 0$. Using that $\lim_{s\to 0} t(s) = 0$, it follows that $\lim_{s\to 0} t'(s) = \lim_{s\to 0}(dU^P/ds) - L \geq 0$, i.e., $\lim_{s\to 0}\left(dU^P/ds\right) \geq L$.

Claim 2: $\lim_{s\to 0}\left(dU^P/ds\right) = L$ *implies that* $\lim_{s\to 0} \overline{V} = 0$

Suppose that $\lim_{s\to 0}\left(dU^P/ds\right) = L$ and $\lim_{s\to 0} \overline{V} > 0$. By virtue of (27), $\lim_{s\to 0}\left(dU^P/ds\right) = L$ implies that (28), the right-hand side of (27), diverges to infinity as $s \to 0$. As the measure of unfilled vacancies in low-productivity firms cannot exceed the measure of such firms, i.e., $\underline{V} \leq (1-p)M$, and the measure of unfilled vacancies in high-productivity firms is nonnegative, i.e., $\overline{V} \geq 0$, it follows that the numerator of the left-hand side of (27) is bounded from above. Hence, in order for the left-hand side of (27) to diverge to infinity as $s \to 0$, the denominator must converge to zero as $s \to 0$; that is, $\lim_{s\to 0}\left\{\left[(1-p)M - \underline{V}\right]\overline{V}\right\} = 0$. Since by presumption $\lim_{s\to 0} \overline{V} > 0$, it follows that $\lim_{s\to 0} \underline{V} = (1-p)M$.

As $\lim_{s\to 0} U^P = 0$, it follows from (21) that $\lim_{s\to 0} V^P = M - L + \lim_{s\to 0} U^P = M - L$. Hence, $\lim_{s\to 0} \overline{V} = \lim_{s\to 0} V^P - \lim_{s\to 0} \underline{V} = M - L - (1-p)M = pM - L < 0$ which contradicts that $\lim_{s\to 0} \overline{V} > 0$. Therefore, $\lim_{s\to 0}\left(dU^P/ds\right) = L$ implies $\lim_{s\to 0} \overline{V} = 0$.

Claim 3: $\lim_{s\to 0}\left(dU^P/ds\right) > L$ *implies that* $\lim_{s\to 0}\overline{V} > 0$

Suppose that $\lim_{s\to 0}\left(dU^P/ds\right) > L$ and $\lim_{s\to 0}\overline{V} = 0$. Since $\lim_{s\to 0}\left(dU^P/ds\right) > L$, (28), and hence the limit of the right-hand side of (27), is finite. Since by presumption $\lim_{s\to 0}\overline{V} = 0$, it follows from (21) that $\lim_{s\to 0} V^P = M - L + \lim_{s\to 0} U^P = M - L$. Hence, $\lim_{s\to 0}\underline{V} = \lim_{s\to 0} V^P - \lim_{s\to 0}\overline{V} = M - L$. By substituting for $\lim_{s\to 0}\underline{V}$ and $\lim_{s\to 0}\overline{V}$, we see that the left-hand side of (27) diverges to infinity as $s \to 0$. We thus obtain a contradiction. Therefore, $\lim_{s\to 0}\left(dU^P/ds\right) > L$ implies $\lim_{s\to 0}\overline{V} > 0$.

Claim 4: $\lim_{s\to 0}\left(dU^P/ds\right) > L$

Suppose that $\lim_{s\to 0}\left(dU^P/ds\right) = L$. To establish Claim 4, by virtue of Claim 1 it suffices to demonstrate that this presumption leads to a contradiction. Consider condition (20) which states that the flow of successful matches between unemployed workers and low-productivity firms equals the flow into unemployment of workers due to separations from low-productivity firms. Differentiating (20) with respect to s yields

$$\begin{aligned}
&\left(\left\{\frac{V^P\left[m_U(U^P,V^P)+m_V(U^P,V^P)\right]-m(U^P,V^P)}{V^{P2}}\right\}\frac{\partial U^P}{\partial s}\underline{V}+\frac{m(U^P,V^P)}{V^P}\frac{\partial \underline{V}}{\partial s}\right)\\
&\left(1+s-\frac{sL}{U^P}\right)+\frac{m(U^P,V^P)}{V^P}\underline{V}\left(1-\frac{L}{U^P}+\frac{sL}{U^{P2}}\frac{\partial U^P}{\partial s}\right)\\
=\ &(1-p)M-\underline{V}-s\frac{\partial \underline{V}}{\partial s},
\end{aligned} \tag{29}$$

where we have used that $M - V^P = L - U^P \Rightarrow \partial V^P/\partial s = \partial U^P/\partial s$ and subscripts denote partial derivatives. Eq. (29) can be written as

$$\begin{aligned}
& \left[m_U(U^P,V^P)+m_V(U^P,V^P)\right]\frac{\underline{V}}{V^P}\frac{\partial U^P}{\partial s}\left(1+s-\frac{sL}{U^P}\right) \\
& -\frac{m(U^P,V^P)}{V^P}\frac{\underline{V}}{V^P}\frac{\partial U^P}{\partial s}\left(1+s-\frac{sL}{U^P}\right) \\
& +\frac{m(U^P,V^P)}{V^P}\frac{\partial \underline{V}}{\partial s}\left(1+s-\frac{sL}{U^P}\right) \\
& +\frac{m(U^P,V^P)}{U^P}\frac{\underline{V}}{V^P}\left(U^P-L+\frac{sL}{U^P}\frac{\partial U^P}{\partial s}\right) \\
= \; & (1-p)M-\underline{V}-s\frac{\partial \underline{V}}{\partial s}. \qquad (30)
\end{aligned}$$

Taking the limit of (30) as $s \to 0$ we have

$$\begin{aligned}
& \lim_{s\to 0}\left[m_U(U^P,V^P)+m_V(U^P,V^P)\right]\lim_{s\to 0}\left(\frac{\underline{V}}{V^P}\right)\lim_{s\to 0}\left(\frac{\partial U^P}{\partial s}\right)\lim_{s\to 0}\left(1+s-\frac{sL}{U^P}\right) \\
& -\lim_{s\to 0}\left[\frac{m(U^P,V^P)}{V^P}\right]\lim_{s\to 0}\left(\frac{\underline{V}}{V^P}\right)\lim_{s\to 0}\left(\frac{\partial U^P}{\partial s}\right)\lim_{s\to 0}\left(1+s-\frac{sL}{U^P}\right) \\
& +\lim_{s\to 0}\left[\frac{m(U^P,V^P)}{V^P}\right]\lim_{s\to 0}\left(\frac{\partial \underline{V}}{\partial s}\right)\lim_{s\to 0}\left(1+s-\frac{sL}{U^P}\right) \\
& +\lim_{s\to 0}\left[\frac{m(U^P,V^P)}{U^P}\right]\lim_{s\to 0}\left(\frac{\underline{V}}{V^P}\right)\left[\lim_{s\to 0}U^P-L+\lim_{s\to 0}\left(\frac{sL}{U^P}\right)\lim_{s\to 0}\left(\frac{\partial U^P}{\partial s}\right)\right] \\
= \; & (1-p)M-\lim_{s\to 0}\underline{V}-\lim_{s\to 0}\left(s\frac{\partial \underline{V}}{\partial s}\right). \qquad (31)
\end{aligned}$$

We will make use of the following properties:

- $\lim_{s\to 0} U^P = 0$ and $m(U^P,V^P) \le U^P$, which imply that $\lim_{s\to 0} m(U^P,V^P) = 0$;
- $\lim_{s\to 0}\left(\partial U^P/\partial s\right) = L$ (by the presumption we aim to contradict), which implies by virtue of Claim 2 that $\lim_{s\to 0}\overline{V} = 0$; hence, $\lim_{s\to 0}\underline{V} = \lim_{s\to 0} V^P$;
- $\lim_{s\to 0}\overline{V} = 0$ which implies that $\lim_{s\to 0}\left(\partial\overline{V}/\partial s\right) \ge 0$;[15]

[15] If it were the case that $\lim_{s\to 0}\left(\partial\overline{V}/\partial s\right) < 0$, then $\overline{V} < 0$ for sufficiently small $s > 0$, which contradicts that $\overline{V} \ge 0$.

- $M - V^P = L - U^P$ which together with $\lim_{s\to 0} U^P = 0$ imply that $\lim_{s\to 0} V^P = M - L$;

- $\lim_{s\to 0} \left(\partial V^P/\partial s\right) = L$ and $\lim_{s\to 0} \left(\partial \overline{V}/\partial s\right) \geq 0$ which imply that $\lim_{s\to 0} (\partial \underline{V}/\partial s) < \infty$;

- The constant-returns-to-scale of the matching function which implies that $\lim_{s\to 0} \left[m_U(U^P, V^P) + m_V(U^P, V^P)\right] \leq 1$;[16]

- $U^P \geq m(U^P, V^P)$ which implies that $\lim_{s\to 0} \left[m(U^P, V^P)/U^P\right] \leq 1$;

- $\lim_{s\to 0} U^P = 0$ which by l'Hospital's rule implies that $\lim_{s\to 0}(sL/U^P)$ $= L \lim_{s\to 0}(\partial s/\partial s)/ \lim_{s\to 0}(\partial U^P/\partial s) = L/L = 1$.

From these properties we obtain that

$$\begin{aligned}
\lim_{s\to 0} \left[m_U(U^P, V^P) + m_V(U^P, V^P)\right] \lim_{s\to 0} \left(\frac{\underline{V}}{V^P}\right) \lim_{s\to 0} \left(\frac{\partial U^P}{\partial s}\right) \lim_{s\to 0} \left(1 + s - \frac{sL}{U^P}\right) &= 0; \\
\lim_{s\to 0} \left[\frac{m(U^P, V^P)}{V^P}\right] \lim_{s\to 0} \left(\frac{\underline{V}}{V^P}\right) \lim_{s\to 0} \left(\frac{\partial U^P}{\partial s}\right) \lim_{s\to 0} \left(1 + s - \frac{sL}{U^P}\right) &= 0; \\
\lim_{s\to 0} \left[\frac{m(U^P, V^P)}{V^P}\right] \lim_{s\to 0} \left(\frac{\partial \underline{V}}{\partial s}\right) \lim_{s\to 0} \left(1 + s - \frac{sL}{U^P}\right) &= 0; \\
\lim_{s\to 0} \left[\frac{m(U^P, V^P)}{U^P}\right] \lim_{s\to 0} \left(\frac{\underline{V}}{V^P}\right) \left[\lim_{s\to 0} U^P - L + \lim_{s\to 0} \left(\frac{sL}{U^P}\right) \lim_{s\to 0} \left(\frac{\partial U^P}{\partial s}\right)\right] &= 0; \\
(1-p)M - \lim_{s\to 0} \underline{V} - \lim_{s\to 0} \left(s\frac{\partial \underline{V}}{\partial s}\right) &= L - pM.
\end{aligned}$$

Substituting into (31) and noting that $L - pM > 0$, we obtain the desired contradiction. We conclude that $\lim_{s\to 0} \left(dU^P/ds\right) > L$.

---

[16] The constant-returns-to-scale implies that

$$m_U(U^P, V^P)U^P + m_V(U^P, V^P)V^P = m(U^P, V^P).$$

As $M - V^P = L - U^P$ and $M > L$ it follows that $V^P > U^P \geq m(U^P, V^P)$. Therefore, $m_U(U^P, V^P) + m_V(U^P, V^P) < 1$ and hence $\lim_{s\to 0} \left[m_U(U^P, V^P) + m_V(U^P, V^P)\right] \leq 1$.

Claim 5: $\lim_{s\to 0} \overline{V} > 0$

Follows from Claims 3 and 4.

Step 2: No Sorting Maximizes Welfare for Small Values of θ

Since the no-sorting configuration maximizes employment, for $\theta \to 0$ it also maximizes social welfare. Thus, by continuity, this configuration maximizes social welfare for sufficiently small values of θ. As the set of θ's for which the no-sorting configuration maximizes social welfare is given by a non-empty interval, there exists a cutoff level $\underline{\theta} > 0$ defined by $\underline{\theta} \equiv \sup_\theta \left\{\theta \mid W^N > \max\left(W^P, W^F\right)\right\}$ such that the no-sorting configuration is uniquely optimal for $\theta \in (0, \underline{\theta})$.

Step 3: Full Sorting Maximizes Welfare for Large Values of θ

Taking the limit as θ approaches its upper bound μ/p (i.e., as $\underline{x}$ approaches zero), and using the fact that $\lim_{s\to 0} U^N = \lim_{s\to 0} U^P = 0$ and $\lim_{s\to 0} U^F = L - pM$,[17] we obtain that $\lim_{s\to 0, \theta\to\mu/p} W^N = L\mu$, $\lim_{s\to 0,\theta\to\mu/p} W^F = M\mu$, and $\lim_{s\to 0,\theta\to\mu/p} W^P = (M - \lim_{s\to 0} \overline{V}/p)\mu$. As $M > L$ and it was shown in Claim 5 of step 1 that $\lim_{s\to 0} \overline{V} > 0$, it follows that for $\theta \to \mu/p$ the full-sorting configuration maximizes social welfare. Hence, by continuity, this configuration maximizes social welfare for sufficiently high values of θ. As the full-sorting configuration maximizes social welfare in a non-empty interval of θ's, there exists a cutoff level $\overline{\theta} < \mu/p$ defined by $\overline{\theta} \equiv \inf_\theta \left\{\theta \mid W^F > \max\left(W^N, W^P\right)\right\}$ such that the full-sorting configuration is uniquely optimal for $\theta \in (\overline{\theta}, \mu/p)$.

Step 4: Partial Sorting Maximizes Welfare for Intermediate Values of θ

In step 1 and 2 it was shown that there exist two cutoff levels for θ, $\underline{\theta}$ and $\overline{\theta}$, satisfying $\theta < \underline{\theta} \leq \overline{\theta} < \mu/p$, such that the no-sorting configuration uniquely maximizes welfare if

[17] When $s \to 0$, there is no frictional unemployment so that employment equals the smaller of the measure of workers and the measure of active firms. Thus, employment is L in the no- and partial-sorting configurations, and is pM in the full-sorting configuration.

$\theta \in (0, \underline{\theta})$, and the full-sorting configuration uniquely maximizes welfare if $\theta \in (\overline{\theta}, \mu/p)$. To complete the proof it remains to be shown that $\underline{\theta} < \overline{\theta}$, so that there exist a non-empty set $(\underline{\theta}, \overline{\theta})$ of θ's for which only the partial-sorting configuration maximizes social welfare, while both the no- and partial-sorting configurations maximize welfare if $\theta = \underline{\theta}$, and both the partial- and full-sorting configurations maximize welfare if $\theta = \overline{\theta}$.

Since $\lim_{s\to 0} W^N = L\mu$ and $\lim_{s\to 0} W^F = pM\left[\mu + (1-p)\theta\right]$, it follows that $\lim\limits_{s\to 0} W^N > \lim\limits_{s\to 0} W^F$ if and only if

$$\frac{(L - pM)\mu}{(1-p)pM} > \theta. \tag{32}$$

Now, consider a θ which satisfies this inequality. Note that (26) implies that $\lim_{s\to 0} W^P = L\left[\mu + (\lim_{s\to 0} q - p)\theta\right]$, where q is the fraction of employed workers that work in high-productivity firms. Since $\lim_{s\to 0} W^N = L\mu$, in order to show that partial sorting maximizes social welfare, it therefore suffices to show that $\lim_{s\to 0} q > p$. Since it was shown in Claim 4 of step 1 that $\lim_{s\to 0}(dU^P/ds) > L$ and by assumption $\lim_{s\to 0}(dU^P/ds) < \infty$, it follows that (28) exceeds unity. Therefore, by (27),

$$\begin{aligned} & \lim_{s\to 0} \frac{(pM - \overline{V})\underline{V}}{\left[(1-p)M - \underline{V}\right]\overline{\overline{V}}} && > 1 \\ \Leftrightarrow \quad & p \lim_{s\to 0} \underline{V} && > (1-p) \lim_{s\to 0} \overline{V}, \end{aligned}$$

and hence

$$\begin{aligned} \lim_{s\to 0} q &= \frac{pM - \lim_{s\to 0} \overline{V}}{M - \lim_{s\to 0} \underline{V} - \lim_{s\to 0} \overline{\overline{V}}} \\ &> p. \end{aligned}$$

Thus, for a value of θ satisfying (32), the partial-sorting configuration strictly dominates the no-sorting and hence also the full-sorting configuration. Consequently, $\underline{\theta} < \overline{\theta}$ so that $W^P > \max\left(W^N, W^F\right)$ for $\theta \in (\underline{\theta}, \overline{\theta})$, while $W^N = W^P$ for $\theta = \underline{\theta}$ and $W^F = W^P$ for $\theta = \overline{\theta}$.

□

References

Abowd, J.M., F. Kramarz, and D.N. Margolis,1999. High Wage Workers and High Wage Firms. **Econometrica** 67, 251-233.

Acemoglu, D. and R. Shimer, 1999. Efficient Unemployment Insurance. **Journal of Political Economy** 107, 5, 893-928.

Albrecht, J. and B. Axell, 1984. An Equilibrium Model of Search Unemployment. **Journal of Political Economy** 92, 824-840.

Baily, M., 1978. Some Aspects of Optimal Unemployment Insurance. **Journal of Public Economics**, 10, 379-402.

Bartelsman, E., J. C. Haltiwanger and S.Scarpetta, 2013. Cross-Country Differences in Productivity: the Role of Allocation and Selection. **American Economic Review** 103,1, 305-334.

Burdett, K. and D.T. Mortensen, 1998. Wage Differentials, Employer Size and Unemployment, **International Economic Review** 39, 2, 257-273.

Burkhard H., 2003. Employment and Welfare Effects of a Two-Tier Unemployment Compensation System. **International Tax and Public Finance** 10, 2, 147-68.

Cahuc, P., and E. Lehmann, 2000. Should Unemployment Benefits Decrease with the Unemployment Spell? **Journal of Public Economics** 77,1, 135-153.

Cremer, H., M. Marchand, and P.Pestieau, 1996. The Optimal Level of Unemployment Insurance Benefits in a Model of Employment Mismatch. **Labour Economics** 2, 407-420.

Davis, S.J, J.C. Haltiwanger, and S. Schuh, 1996. **Job Creation and Destruction.** MIT Press, Cambridge.

Diamond, P., 1971. A Model of Price Adjustment. **Journal of Economic Theory** 3, 156-168.

Diamond, P., 1981. Mobility Costs, Frictional Unemployment and Efficiency. **Journal of Political Economy** 89, 798-812.

Ehrenberg, R. and R.L. Oaxaca, 1976. Unemployment Insurance, Duration of Unemployment, and Subsequent Wage Gain. **American Economic Review** 66, 754–766.

Elsby, M.W.L., B. Hobijn, and A. Sahin, 2013. On the Importance of the Participation Margin for Market Fluctuations. **San Francisco Fed WP 2013-05.**

Feng, Q. and W.C. Horrace, 2012. Alternative Technical Efficiency Measures: Skew, Bias and Scale. **Journal of Applied Econometrics** 27, 253-268.

Flemming, J., 1978. Aspects of Optimal Unemployment Insurance: Search, Leisure, Savings and Capital Market Imperfections. **Journal of Public Economics**, 10, 403-425.

Fredriksson, Peter and Bertil Holmlund, 2001. Optimal Unemployment Insurance in Search Equilibrium, **Journal of Labor Economics**,19, 2, 370-399.

Hagedorn, M., I. Manovskii and K. Mitman, 2015. The Impact of Unemployment Benefit Extensions on Employment: The 2014 Employment Miracle? NBER Working Paper no. 20884.

Hall, R.E. and A.B. Krueger, 2012. Evidence on the Incidence of Wage Posting, Wage Bargaining, and On-the-Job Search, **American Economic Journal: Macroeconomics** 4,4,56-67.

Haltiwanger, J.C., J.I. Lane, and J.R. Spletzer, J.R., 1999. Productivity Differences Across Employers: The Role of Employer Size, Age, and Human Capital. **American Economic Review Papers and Proceeding**s 99, 94–98.

Haltiwanger, J.C., J.I. Lane, and J.R. Spletzer, 2007. Wages, Productivity, and the Dynamic Interaction of Businesses and Workers. **Labour Economics** 14, 575-602.

Heckman, J.J., and M. Sattinger, 2015. Introduction to the Distribution of Earnings and of Individual Output, by A.D. Roy. **The Economic Journal**, 125, 378-402.

Hopenhayn, H., and J. Nicolini, 1997. Optimal Unemployment Insurance. **Journal of Political Economy** 105, 2, 412-438.

Karni, E., 1999. Optimal Unemployment Insurance – A Survey. **Southern Economic Journal,** 66, 442-465

Kasahara, H., and B. Lapham, 2013. Productivity and the Decision to Import and Export: Theory and Evidence. **Journal of International Economics,** 89, 297-316.

Lazear E.P. and K.L. Shaw (eds)., 2008. **The Structure of Wages: An International Comparison.** University of Chicago Press, Chicago and London.

Luttmer E.G.J., 2007. Selection, Growth, and the Size Distribution of Firms. **Quarterly Journal of Economics** 122, 1103-1144.

Marimon, R., Zilibotti, F., 1999. Unemployment vs. Mismatch of Talents: Reconsidering Unemployment Benefits. **The Economic Journal** 109, 266-291.

Moen, E., 1997. Competitive Search Equilibrium. **Journal of Political Economy**, 105, 385-411.

Moen, E., 2003. Do Good Workers Hurt Bad Workers or is it the Other Way Round? **International Economic Review,** 44,779-799.

Mortensen, D., 1977. Unemployment Insurance and Job Search Decisions. **Industrial and Labor Relations Review** 30, 505-517.

Mortensen, D.T., 2003. **Wage Dispersion**. MIT Press, Cambridge.

Neal, D. and S. Rosen, 2000. Theories of the Distribution of Earnings. In A.B.Atkinson and F. Bourguignon (eds). **Handbook of Income Distribution**, Vol. 1, North Holland.

OECD, 2007. **Benefits and Wage**s. OECD, Paris.

Petrongolo B., and C.A. Pissarides, 2001. Looking into the Black Box: A Survey of the Matching Function. **Journal of Economic Literature** 39, 390-431.

Pissarides, C.A., 2000. **Equilibrium Unemployment Theory**. 2nd edition, MIT Press, Cambridge.

Rogerson, R., R. Shimer and R. Wright, 2005. Search Theoretic Models of the Labor Market: A Survey. **Journal of Economic Literature,** 43,4,959-988.

Shavell, S. and L. Weiss, 1979. The Optimal Payment of Unemployment Insurance Over Time. **Journal of Political Economy**, 87, 1347-1362.

Tatsiramos, K., and J.C. van Ours, 2014. Labor Market Effects of Unemployment Insurance Design. **Journal of Economic Surveys** 28, 2, 284-311**.**

Wang, C. and S.D. Williamson, 2002. Moral hazard, optimal unemployment insurance, and experience rating, **Journal of Monetary Economics** 49(7), 1337-1371.

Yashiv, E., 2000. The Determinants of Equilibrium Unemployment. **American Economic Review** 90, 1297-1322.

Yashiv, E., 2007. U.S. Labor Market Dynamics Revisited. **Scandinavian Journal of Economics** 109, 779-806.

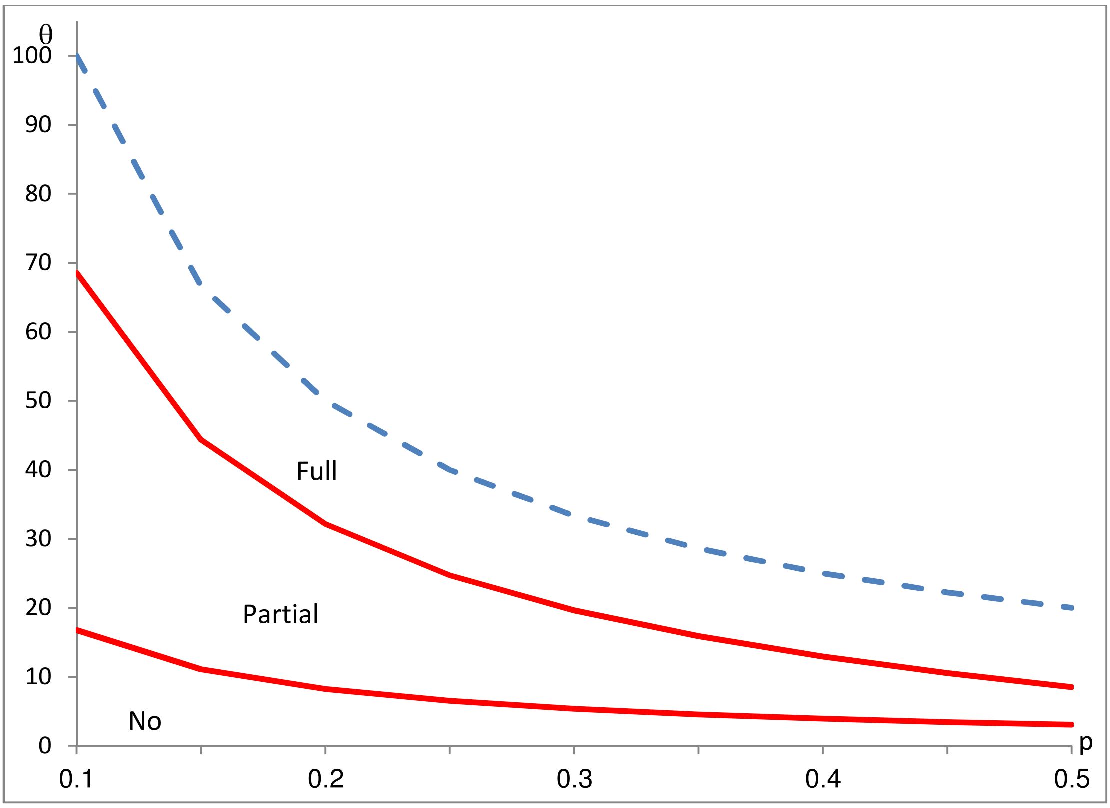

Figure 1: The Optimal Sorting