Equilibrium Wage Dispersion: Monopsony or Sorting?

Giuseppe Moscarini*

Department of Economics & Cowles Foundation for Research in Economics
New Haven, Connecticut, U.S.A.

Perché lavoratori simili ricevono salari diversi? In questo saggio rivisito e confronto due linee di ricerca che hanno compiuto notevoli progressi nella comprensione dei differenziali salariali: (i) *eterogeneità inosservata nel capitale umano ed autoselezione dei lavoratori;* (ii) *potere monopsonistico delle imprese in mercati del lavoro caratterizzati da frizioni di* search. *Ambedue queste ipotesi vedono i differenziali salariali come fenomeno di equilibrio. Nonostante le profonde differenze concettuali e tecniche, esse restano le due principali concorrenti in questa indagine. A differenza di altre ipotesi, esse offrono spiegazioni unificanti dei flussi aggreganti nei mercati del lavoro e della forma della distribuzione dei salari.*

Why are similar workers paid differently? I review and compare two lines of research that have recently witnessed great progress in addressing "unexplained" wage inequality: (i) *worker unobserved heterogeneity in, and sorting by, human capital;* (ii) *firms' monopsony power in labor markets characterized by job search frictions. Both lines share a view of wage differentials as an equilibrium phenomenon. Despite their profound conceptual and technical differences, they remain natural competitors in this investigation. Unlike other hypotheses, they provide natural and unifying explanations for job and worker flows, unemployment duration and incidence, job-to-job quits, and the shape of the wage distribution* [JEL Code: C73; D31; D83; E24; J63; J64].

1. - Introduction

Why are similar workers paid differently? Ever since the required datasets became available about 30 years ago, Mincerian

---

* *<giuseppe.moscarini@yale.edu>.*

human capital regressions have been able to "explain" only a relatively small fraction of the cross-sectional and time-series variance in individual earnings and wages.[1] The existence of sizable and persistent wage differentials among observationally identical workers poses a formidable challenge to the very core of labor economics and of macroeconomics. In labor economics, at stake is the empirical validity of the human capital model. For macroeconomists, wage determination stubbornly remains the key to understand aggregate business fluctuations (Shimer, 2005; Hall, 2005). The mere failure of the law of one price is an intriguing fact that deserves investigation.

In this article I review and compare two lines of research that have addressed «unexplained» wage inequality from two conceptually different viewpoints. These two lines appeal to two well-established ideas, respectively, unobserved heterogeneity in human capital and firms' monopsony power in frictional labor markets. The former approach, inspired by the celebrated Roy (1951) model of sorting in labor markets, and pioneered in a dynamic context by the equally celebrated contributions of Jovanovic (1979, 1984), views a job as an "inspection" and "experience" good. Namely, workers and firms are heterogeneous in ways that not even they, let alone the statistician, can fully describe; the "fit" between their characteristics can be learned only from direct experience, just like in a marriage. Wages are reset based on evolving inside and outside options of workers and firms. The approach based on unilateral wage offers supported by search frictions, since the seminal contribution of Diamond (1971), builds on the

---

[1] A major recent advance in several countries is the availability of longitudinal matched employer-employee datasets. Like panel data, they track workers through their careers and firms as they lose and hire workers; unlike panel data, they also tell us who works for whom. They allow to decompose wage differentials into the contributions of the worker, of the firm they work for, and of the «match» (the residual). The basic message remains (e.g., ABOWD J. et AL. (1999) for France and the US): unobserved worker characteristics, both common to all jobs and match-specific, exist and matter for wages. At the same time, firms differ systematically by productivity, and their heterogeneity is very persistent (FOSTER L. et AL., 2001). In principle, the data can tell us whether this is because some firms attract better workers, or because they have higher TFP, so that reallocation of labor across firms is an important contributor to aggregate productivity growth, or both ("positive assortative matching").

monopsony power that firms possess due to their employees' inability to instantaneously locate alternative employment opportunities.

Recent incarnations of these two lines of research, as exemplified (resp.) by Moscarini (2005) and by the very influential theory of Burdett and Mortensen (1998), converge on the importance of search frictions, but go one step further. They take an *equilibrium* approach, to account not only for observed wage differentials, but also for many additional and potentially related facts. Involuntary unemployment exists and is quantitative important. Workers' separation rates from jobs (Farber, 1994) and propensity to search on the job (Pissarides and Wadsworth, 1994) fall with seniority. Wages rise modestly with seniority (Topel, 1991; Altonji and Williams, 2005). Layoffs predict a sizable and persistent wage loss (Jacobson *et* Al., 1993), and job-to-job quits a significant wage gain (Topel and Ward, 1992). Cross-sectionally, the variance of wages rises with the age of a cohort, the wage distribution is typically unimodal, skewed, with a long Paretian tail, whether one controls or not for observed worker characteristics, unexplained wage differentials are positively correlated with firm size (for a recent review, Mortensen, 2005), with occupational experience (Kambourov and Manovskii, 2002), and correlated with industry membership (Kruger and Summers, 1988).

When taken together, these facts impose formidable restrictions. Contemporary macroeconomics almost universally embraces the so called Flow Approach, a view of the labor market as a tremendously dynamic place, where transitions between employment states and wage determination constantly interact. This is in stark contrast to the competitive view where, absent adjustment costs and frictions, stocks of demand and supply confront each other in a sequence of essentially unrelated spot markets. Albeit economic theory is the art of finding the innocuously simplifying assumptions to focus on a subset of facts, in this case we should be wary of counterfactual predictions for the other facts, either wage dispersion, dynamics, or/and worker turnover. The basic philosophy of the Flow Approach is that these phenomena are intimately inter-related and cannot be explained in isolation,

or in the context of a single employment relationship surrounded by a market vacuum.

The two classes of explanations for wage differentials that I focus on, job matching and pure search frictions, hardly exhaust the spectrum of possibilities. Effciency wages, insurance, dual labor markets, and many other hypotheses are still alive and well. But, they have never been developed in a full-fledged general equilibrium context. They are typically partial equilibrium and/or static stories, that have little to say about the observed joint dynamics of quantities and prices in aggregate labor markets. For example, they have a hard time explaining why separation rates fall with tenure and seniority, why the variance of wages in a cohort rises with age, why the wage distribution has a thick right tail, and so on.

In the job-matching model, workers keep searching on and off the job until they find a match that looks good enough. Wages must reflect expected match quality (productivity) to induce the worker to search. Because jobs take time to locate, involuntary unemployment exists and is quantitative important. Workers' separation rates from jobs and their propensity to search on the job fall with seniority, because long-tenured workers are likely to be those who are well matched and no longer need to shop for jobs. For the same reasons, wages rise with seniority, when a match dissolves exogenously (a layoff, say) all the matching human capital accumulated through trial-and-error goes lost, so a layoff predicts a sizable and persistent wage loss, and job-to-job quits only occur if the worker gains. As learning and sorting take place, some workers are luckier than others and their wages grow faster, so the wages of a given cohort «fan out» as time goes by. The whole sorting process shifts workers to good matches and creates the thick right tail in the wage distribution.

Different industries may be subject to different noise which makes the sorting process more or less efficient, thus creating equilibrium wage differentials. If the most important match for a worker is with her career and not with the specific firm, then occupational tenure matters most. The firm size-wage premium is difficult to explain in this context.

In the simplest monopsony-frictional model, identical firms commit to wage offers and trade off the labor cost of each worker against firm size through the effects of wages on hiring and retention. The key is that workers can search on the job, so firms will not extract all rents with their unilateral wage offers, because that would increase turnover. In equilibrium, identical firms must offer different wages, that make them indifferent: if they all offered the same wage, a single firm could gain lots of workers by paying just a little bit more than the others. Again, because jobs take time to locate, involuntary unemployment exists and is quantitative important. Over time, workers climb the wage ladder through job-to-job quits, so long-tenured workers must be those who have already found a high wage and are unlikely to separate and search on the job. When a match dissolves exogenously the worker has to restart from scratch his ladder-climbing. Quits, again, occur only to firms that pay better. As ladder-climbing is a random process, some workers are luckier than others and their wages grow faster, so the wages of a given cohort «fan out» as time goes by. The firm size-wage premium is the natural outcome of this equilibrium. Inter-industry wage differentials may exist because in some industries, say, all firms prefer to pay high wages and have a large labor force; thus, wage premia go hand-in-hand with inter-industry size differentials. Similarly for different occupations: occupational tenure may be good for wages just because correlated with working for a high-wage firm, although this is a quite contrived rationale. Explaining the shape of the wage distribution has proven elusive in this context.

In the following sections, I lay out a barebone model of equilibrium wage dispersion generated by search frictions, along the lines of Burdett and Mortensen (1998), and I summarize a job matching model of equilibrium wage dispersion. The goal is to present them back-to-back in a common notation and in their simplest forms, to bring out the similarities and differences. Finally, I evaluate their respective pros and cons. I conclude with a summary of what we have learned and with some suggestions for future research.

2. - Monopsony Power in Search Equilibrium

This section draws from Burdett and Mortensen (1998), a simple continuous time search model of the labor market without recall but with sequential offers to employed workers.

A unit measure of identical risk-neutral workers and a measure $m > 1$ of identical firms maximize payoffs, discounted at rate $r > 0$ by workers and undiscounted by firms. Output is produced by firms and workers with a Leontief technology, in 1:1 matches of same productivity μ. Each firm can hire as many workers as desired and as allowed by the frictional search process. An unemployed worker receives a payoff b and contacts a firm at Poisson rate λ, at no cost. The contact rate λ is taken as exogenous, but it can be easily endogenized through a standard matching function, given unemployment and the number of firms hiring. An employed worker is paid a wage w and contacts another potential employer at rate $\zeta\lambda$, at no cost. Matches separate exogenously at rate δ, and the worker becomes unemployed; they separate endogenously when a worker quits to another job. Firms make unilateral and unconditional (on the employment status of the applicant) wage offers and, by assumption, can commit to them.

When making and committing to ("posting") a wage offer, a firm weighs the higher wage bill against the increased ability to hire and retain workers, therefore the larger size and output it can attain. Burdett and Mortensen (1998) study the steady-state sequential Nash equilibrium of this wage-posting dynamic game, which results in an endogenous distribution of wage offers and of wages paid to workers.

If $\zeta = 0$, i.e. a worker must go through unemployment to find another job, Diamond (1971) established the well-known paradox that the unique equilibrium features wage offers all equal to the reservation wage $w = b$. I.e., the extreme monopsony outcome occurs for any arbitrarily small friction $\lambda < \infty$, in contrast to the competitive allocation that would give all rents to the workers in the frictionless case ($\lambda \to \infty$). If search has any cost for workers, the market shuts down, because workers can never recover their investment in search. The root of this result is the combination of

monopsony power and the inability of the firms to commit to wage offers before receiving applications. To break the Diamond paradox, Burdett and Mortensen assumed $\zeta > 0$, on-the-job search. For simplicity, here I set $\zeta = 1$, so job search is equally productive from unemployment or employment. A firm may either post a wage or wait until applications actually arrive to make its wage offers. After the offer is made and accepted, the firm cannot change it.

Let u denote the unemployment rate, F denote the c.d.f. of offered wages, and G the c.d.f. of paid wages. The densities of F and G, if they exist, are (resp.) f and g. It is intuitive, so I omit the proof, that the lower bound of either wage distribution must be b, the worker's reservation value. Only if $\zeta > 1$, i.e. if employment made search for other wage offers more effective, would an unemployed worker accept a wage below b, so the support of F would extend below that level. By the same token, b must be in the support of the wage distribution G, because always acceptable to unemployed workers.

A key observation is that F cannot have atoms. If there was an atom of firms offering the same wage w, then any firm could post a wage $w + \varepsilon$ for some small ε, pay its workers just a little more, but win the competition for all workers employed at wage w who search on the job, a discrete mass. So firm size would jump and the unit labor cost would not, raising profits. Ergo, F must be continuous.

When a firm posts a wage w, it attains a steady state labor force of size $L(w)$ and makes average profits $(\mu - w)L(w)$. It is easy to see that:

$$L(w) = \lim_{\varepsilon \downarrow 0} \frac{[G(w) - G(w - \varepsilon)](1 - u)}{[F(w) - F(w - \varepsilon)]m} = \frac{g(w)}{f(w)} \frac{1 - u}{m}$$

where the first fraction has in the numerator the number of worker being paid wages between $w - \varepsilon$ and w, and the denominator is the number of firms offering those wages, so the ratio is the number of workers per firm at those wages.

The two distributions are linked in steady state by an equality of flows. Let u denote the unemployment rate. Then the num-

ber of unemployed workers who find a job at a wage less than w equals the number of workers employed at those wages who either flow into unemployment or upgrade to a higher wage.

$$\lambda F(w)\, u = [\delta + \lambda\,(1 - F(w))]G(w)\,(1 - u)$$

This guarantees that the number of workers who are employed and paid less than w, namely $G(w)$, does not change over time. At the highest possible wage $\bar{w} \le \mu$, we have clearly $F(\bar{w}) = G(\bar{w}) = 1$, so I can solve for the stationary unemployment rate:

$$u = \frac{\delta}{\delta + \lambda}$$

Substituting and rearranging:

$$G(w) = \frac{\delta F(w)}{\delta + \lambda - \lambda F(w)}$$

Taking derivatives on both sides, the densities exist and

$$g(w) = G'(w) = \frac{\delta f(w)}{\delta + \lambda - \lambda F(w)} + \frac{\lambda f(w)\delta F(w)}{[\delta + \lambda - \lambda F(w)]^2}$$

and finally the labor force size of a firm paying a wage w equals:

$$L(w) = \frac{g(w)}{f(w)} = \frac{\delta(\delta + \lambda)}{[\delta + \lambda - \lambda F(w)]^2}$$

All firms, including those posting the lowest wage b, must make in equilibrium the same profits: for all wages in the support of the wage offer distribution F (all wages actually offered to job applicants):

$$\pi(w) \equiv \frac{g(w)}{f(w)}(\mu - w) = \frac{\delta(\delta + \lambda)(\mu - w)}{[\delta + \lambda - \lambda F(w)]^2} = \frac{\delta}{\delta + \lambda}(\mu - b)$$

while, for all other wages, $\pi \le (\mu - b)\,\delta/(\delta + \lambda)$. We obtain a unique steady state wage offer distribution, which then determines the supremum $\bar{w}$ of the support,

10

$$\overline{w} = \mu - \left(\frac{\delta}{\delta + \lambda}\right)^2 (\mu - b)$$

and the wage distributions F and G:

$$F(w) = \frac{\delta + \lambda}{\lambda}\left(1 - \sqrt{\frac{\mu - w}{\mu - b}}\right) \ \text{ and } \ G(w) = \frac{\delta}{\lambda}\left(\sqrt{\frac{\mu - b}{\mu - w}} - 1\right)$$

The wage density:

$$g(w) = \frac{\delta\sqrt{\mu - b}}{2\lambda(\lambda - w)^{\frac{3}{2}}}$$

is increasing: in order to convince some firms to pay higher wages to all workers, the number of competitors who are crowded out by the higher wage must rise fast enough.

This is the unique equilibrium of the wage-posting game. Identical firms must offer different wages, to create sufficient wage dispersion and sufficient incentives to support it. With monopsony power and on-the-job search, turnover is so central to a firm's profits that a unique wage paid by identical firms to identical workers is not sustainable. Pure wage dispersion obtains solely as a result of equilibrium interaction. The uniqueness of the equilibrium makes this prediction particularly compelling. While the idea of "effciency wages" as a resolution of the firm's turnover problem has been around for a long time, its discussions have always been vague, ambiguous, and unproductive, because typically framed in partial equilibrium and rarely backed by rigorous formal analysis. In short, until Burdett and Mortensen's article, this line of thinking was, essentially, just an interesting hypothesis with no empirical content. The model brings out the surprising implication that wage dispersion is a necessary outcome, and exposes both the solid and the flawed parts of past debates.

To recap. Workers match randomly with firms. Each firm offers the same wage w to all applicants, and commits to any accepted wage. So the firm follows a "wage policy". An unemployed worker accepts an offer w and keeps searching for better

offers from $F(w')/[1-F(w)]$. So a worker's wage drifts up, the probability of a quit to another firm (of a better offer) drifts down, actual wages for a cohort of workers "fan out". Firms offer different wages to sustain an equilibrium where no firm has an incentive to deviate to steal workers from its competitors. When a match is hit by a separation shock at rate δ, the worker has to restart from scratch: he faces a wage distribution $F(w')$, which is worse than $F(w')/[1-F(w)]$, so he matches faster but is paid less on average than employed workers.

3. - Sorting and Job-Matching in Search Equilibrium

This section draws from Moscarini (2005). Consider the same setup, but now assume that the average productivity or "quality" of each match, μ, is specific and *ex ante* uncertain: upon matching, nature draws μ, independent of past events, from the lottery $p_0 = \Pr(\mu = \mu_H) = 1 - \Pr(\mu = \mu_L) \in (0, 1)$, where μ_L denotes a "bad" match and $\mu_H(> \mu_L)$ a "good" match. Worker and firm do not know μ. This can be thought of as the result of combining hard-to-describe and privately known firm and worker characteristics. The cumulative output of a match of duration t is a normal random variable with mean μ and known variance σ^2:

$$X_t \sim \mathbb{N} \ (\mu t, \ \sigma^2 t)$$

Gaussian white noise keeps μ hidden and creates an inference problem. Over time, parties observe output realizations and update in a Bayesian fashion their belief from the prior p_0 to the posterior $p_t \equiv \Pr(\mu = \mu_H \mid X_t)$. Firms are in excess supply (m is unbounded above), they also discount payoffs at rate r like workers, and $b \in [\mu_L, (1 - p_0)\mu_L + p_0\mu_H]$, so that a match is always accepted but sometimes later discarded, if suffciently unproductive. In practice, the firm and the worker perform a sequential probability ratio test of simple hypotheses on the viability of the match. Critically, now firm and worker cannot commit to a wage contract/offer, but they split match rents according to a generalized Nash bargaining rule,

assigning a geometric weight β to the worker's surplus. When a worker receives an outside offer, an ascending auction between the two firms (employer and poacher) ensues, followed by renewed Nash bargaining between the worker and the winner of the auction. The other assumptions about job finding rates $\lambda \in (0, \infty)$, $\zeta = 1$ and job destruction rate $\delta > 0$ are unchanged. Again, I study a steady state equilibrium of this economy.

A sufficient statistic for output history, which determines the future prospects of a match, thus also the natural state variable of the bargaining game, is the posterior belief p_t that the match was a success ($\mu = \mu_H$). Because match quality is specific, we can interpret the accumulated knowledge p_t to be "firm-specific human capital". Conditional on the output process X, the posterior probability of a good match evolves from any prior $p_0 \in (0, 1)$ as a martingale diffusion solving:

$$(3.1) \qquad\qquad dp_t = p_t(1 - p_t)sd\overline{Z}_t$$

where

$$s \equiv \frac{\mu_H - \mu_L}{\sigma}$$

is the signal/noise ratio of output, and

$$\overline{Z}_t \equiv \frac{1}{\sigma}\left[X_t - \mu_L t - (\mu_H - \mu_L)\int_0^t p_s ds\right] \sim \mathbb{N}(0, t)$$

is the innovation process, the normalized difference between realized and unconditionally expected flow output.[2] Intuitively, beliefs move faster the more uncertain match quality (the term $p(1-p)$ peaks at $p = 1/2$), and the more informative production, as measured by the signal/noise ratio s. When flow output dX_t per unit time is above current expectations $[p_t\mu_H + (1-p_t)\mu_L]dt$, beliefs rise and the worker-firm pair becomes more optimistic that their match "works".

Let $W(p)$ denote the discounted total payoffs that a worker receives in the equilibrium of the bargaining-and-search game, when

---

[2] Formally, $\overline{Z}$ is Wiener with respect to the filtration induced by output observations.

employed in a match that is successful with current posterior chance p and when paid an equilibrium wage $w(p)$. Similarly, let U denote the worker's value of unemployment, independent of p because of the match-specific nature of match quality μ, $J(p)$ the rents of the firms, V the value to the firm of holding an open vacancy. I seek an equilibrium where W and J are strictly increasing in p.

For the worker, these values solve the Bellman equations:

$$rU = b + \lambda[W(p_0) - U]$$

$$rW(p) = w(p) + \Sigma(p)W''(p) - \delta[W(p) - U] + \lambda \max \langle W(p_0) - W(p), 0 \rangle$$

where:

$$\Sigma(p) \equiv \frac{1}{2} s^2 p^2 (1 - p)^2$$

is half the ex ante variance of the change in posterior beliefs. Roughly speaking, this term measures the "speed of Bayesian learning" about match quality: if posterior beliefs are not expected to change in the next instant, the variance is zero and nothing is learned.

To understand the Bellman equation for U, notice that the opportunity cost of unemployment, rU, equals its flow benefit b plus the capital gain $W(p_0) - U$ from a new match, which has prior belief p_0 of being successful, accruing at rate λ. Similarly, the opportunity cost $rW(p)$ of working in a job that is successful with posterior chance p equals the flow wage $w(p)$, plus a diffusion-learning term $\Sigma(p)W''(p)$, minus the capital loss following exogenous separation at rate δ. The learning speed $\Sigma(p)$ is converted into consumption payoffs by the convexity of the Bellman value $W''(p)$, because information (here in the form of output) spreads posterior beliefs and empowers more informed decisions by the worker. The worker optimally quits to unemployment at every belief $\underline{p}_W \in [0, 1]$ such that $W(\underline{p}_W) = U$ (value matching) and $W'(\underline{p}_W) = 0$ (*smooth pasting*). Similarly, the worker stops searching on the job when beliefs that the match works exceed the belief p_0 of a new match. This behavior imposes analogous optimality conditions at p_0.

The problem of the firm is similar. The free entry condition $V = 0$ is used to close the general equilibrium. The value to the employer $J(p)$ of an active match that is successful with posterior chance p solves the Bellman equation:

$$rJ(p) = \bar{\mu}(p) - w(p) + \Sigma(p)J''(p) - J(p)\left\{\delta + \mathbb{I}(W(p) < W(p_0))\right\}$$

The opportunity cost of production $rJ(p)$ equals expected flow output:

$$\bar{\mu}(p) \equiv p\mu_H + (1 - p)\mu_L$$

minus the wage $w(p)$, plus the return from learning the quality of the match $\Sigma(p)J''(p)$, minus the expected capital loss due to exogenous separation $(\delta J(p))$ or to a worker's quit to another firm (which happens when the worker gains from a quit: $W(p) < W(p_0)$; here $\mathbb{I}$ is an indicator function). The firm optimally fires the worker at every $\underline{p}_J \in [0, 1]$ such that $J(\underline{p}_J) = 0$ and $J'(\underline{p}_J) = 0$.

The generalized Nash bargaining solution selects a wage on the Pareto frontier:

$$w(p) \in \arg\max_w \; [W(p) - U]^\beta \, [J(p)]^{1-\beta}$$

for some worker bargaining power $\beta \in (0, 1)$ exogenously given. After some algebra, we obtain a simple and intuitive expression:

$$w(p) = (1 - \beta)\, b + \beta\left[\bar{\mu}(p) + \lambda J(p_0)(1 - \psi\mathbb{I}[J(p) < J(p_0)])\right]$$

The worker receives a wage that weighs with the bargaining share his flow outside option, the opportunity cost of time b, and his inside option, flow expected output $\bar{\mu}(p)$, plus the continuation value of unemployed job search $\beta\lambda J(p_0) = (1 - \beta)[W(p_0) - U]$, reduced by the prospect of a quit that reduces match surplus but raises the worker's outside option.

The wage is affine and increasing in the posterior belief. Also, the separation cutoff $\underline{p}$ is agreed upon: any inefficient separation can be avoided by an appropriate wage raise/cut. However, when

a worker quits to another better match, the current employer loses and cannot do anything about it.

Replacing the wage expression back in the firm's Bellman equation allows to solve for the firm's value J, the separation cut-off $\underline{p}$, and the wage function. The value to a firm J and to a worker W is convex in the belief p that the match is successful.

To recap. A firm and a worker match randomly. They know that their match is successful with chance p_0. They start producing and update beliefs about their match quality based on output performance. The match-specific rents that develop are split in a bargaining game. Wages reflect not directly output, but rather its permanent component, as summarized by beliefs or "promise" of the match: the better the match looks, the higher the wage (and the firm's expected future profits). When moderately pessimistic, the firm and the worker keep producing but the worker accepts outside offers, which restart new matches that are more promising than what the current one turned out to be. When sufficiently pessimistic, the firm and the worker separate to unemployment.

A fundamental property of Bayesian learning is that beliefs are martingales: they (thus here wages) are not expected to rise or fall on average. This is the same reason why asset prices are unforecastable in efficient markets. But only good matches survive.

Thus, low output realizations and beliefs are discarded and not reflected in actual market wages. Therefore, in a continuing job, the wage rises on average (although not for sure) with tenure, while the propensity to quit to unemployment or to another job declines with seniority. This fundamental prediction is the result of the interaction of efficient learning and sorting/selection.

The wage offer distribution F in this model is very simple: it is just an atom at $w(p_0)$. Raising the offer to attract workers is impossible, because now the firm can no longer commit to a wage offer, and can only bargain *ex post* with negotiation power $1 - \beta$. Over time, the worker makes transitions in and out of employment, and sees his wage rise or fall with output and the resulting inference about match quality. The ergodic distribution of beliefs has a density ϕ that solves the Kolmogorov forward equation:

$$0 = \frac{d^2}{dp^2}\left[\Sigma(p)\phi(p)\right] - (\delta + \lambda\mathbb{I}\{\underline{p} \le p < p_0\})\phi(p)$$

subject to appropriate boundary conditions, ensuring the constancy of unemployment and employment in steady state. In a large economy, this density is also the cross-sectional density of workers by beliefs about match quality. This differential equation can be solved analytically for ϕ. Inverting the equilibrium map from beliefs to wages, we can solve for the steady state wage density:

$$(3.2) \quad g(w) = \begin{cases} c_{0f}(w - \omega_0)^{-\frac{1}{2} - \sqrt{\frac{1}{4} + \frac{2(\delta+\lambda)}{s^2}}} (1 - w + \omega_0)^{-\frac{3}{2} + \sqrt{\frac{1}{4} + \frac{2(\delta+\lambda)}{s^2}}} \\ \left[\left(\frac{1 - \underline{p}}{\underline{p}}\right)^{\sqrt{\frac{1}{4} + \frac{2(\delta+\lambda)}{s^2}}} - 1\right] & \text{for } p \le p_0 \\[2em] c_{1f}(w - \omega_1)^{-\frac{3}{2} - \sqrt{\frac{1}{4} + \frac{2\delta}{s^2}}} (1 - w + \omega_1)^{\sqrt{\frac{1}{4} + \frac{2\delta}{s^2}} - \frac{3}{2}} & \text{for } p > p_0 \end{cases}$$

where c_{if} and ω_i, $i = 0, 1$, are known coeffcients. The support of the wage distribution is $[w(\underline{p}), 1 + \omega_1]$.

This expression has two important implications. First, the theoretical equilibrium wage distribution $g(w)$ may potentially replicate the typical shape of an empirical wage distribution, including its well-known Paretian right tail. Quits to other jobs and to unemployment weed out disproportionately bad matches, censor the left tail, and skew the distribution.

The distribution has in fact a globally declining right tail, which gives it overall an empirically accurate shape, if $\delta \ge s^2$, that is, if matches are destroyed exogenously before selection can place too many workers in their ideal matches. Moscarini (2003) presents a detailed quantitative evaluation of a discrete time version of this model at a monthly frequency, and shows that the restriction $\delta \simeq s^2 = 0.011$ is required to match aggregate empirical evidence on US labor market transitions.

The second result is that the right tail of the wage distribution

$g(w)$ decays faster the larger the ratio δ/s^2 between the exogenous match dissolution rate δ and the (squared) informativeness of output s^2. Intuitively, when jobs are at high risk of exogenous destruction (δ is large), or when the output process is very noisy and uninformative, so beliefs move slowly (σ is large and the signal/noise ratio s is low), the learning-selection process has no time to produce its effects. A "noisy" economy is "sclerotic": high idiosyncratic output uncertainty unrelated to firm and worker characteristics (high δ and σ) clouds the intrinsic inequality in productivities (μ) and prevents it from being reflected by equilibrium prices. Wages remain concentrated around their starting value $w(p_0)$; income inequality tends to be dampened, rather than enhanced, by high idiosyncratic output risk.

4. - Discussion

The two views of labor markets analyzed in this article share many common elements, such as frictional matching, on the job search and involuntary unemployment. But they are conceptually very different. The main difference is wage-setting. In the monopsony approach, only firms can determine wages, but they can credibly promise workers to maintain announced wages. In the job-matching approach, the firm has no such commitment power, so the worker must be endowed with some exogenous bargaining power to receive any rents.

As a consequence, the sources of wage dynamics, turnover and wage dispersion are also quite different. In the monopsony approach, wage dispersion in a market of identical agents is a necessary outcome of strategic interaction. On the job search is necessary to temper the firm's monopsony power, and to make hiring and retention depend on the wage offer. Workers slowly and randomly, but monotonically, climb the ladder from small, low-wage firms to large, high-paying employers. Any exogenous displacements resets the clock and restarts the climbing process. Burdett and Coles (2003) let firms offer (and commit to) wage-tenure profiles, as opposed to constant wages, and repeat the

exercise. Now wages rise deterministically but constantly with seniority within a job: backloading wages without changing their present discounted value to the worker reduces turnover and helps the firm.

In the job matching approach, on the job search simply reinforces a sorting/selection process that would take place anyway through transitions via unemployment. Since unemployment is painful, job-to-job quits are an integral part of the story, but only quantitatively. Workers and firms slowly learn whether their heterogeneous characteristics combine well or not. *Ex post* heterogeneity in productivity is the engine of wage dynamics, wage dispersion, and turnover. Equilibrium selection keeps alive only good matches, that pay high wages. Exogenous displacements and workers' quits to other jobs destroy all job-specific human capital.

The two models explain the same facts in very different ways. The job matching approach features more interesting and realistic dynamics "on-the-job", and similar dynamics across jobs; but firm size is indeterminate, due to the match-specific nature of productivity and to the assumption of constant returns to scale. The monopsony model has a richer description of a firm's problem, where turnover and firm size take centerstage. The wage plays more of an allocative role, rather than just imperfectly reflecting productivity. But the within-firm wage dynamics are more stylized. In short, dynamics and wages are driven by selection across heterogeneous firms in the job matching approach, by competition among identical firms in the monopsony approach.

The efficiency implications are also quite different. In the monopsony approach, job-to-job quits have a pure rent-shifting purpose, and do not change output. The model can be enriched to allow for permanent differences in firms' productivites μ, but then it resembles more a sorting model. In the latter approach, quits and separations are always welfare and output-improving, because they originate from the need to eliminate unsuccessful matches and to try something better.

Both models appear quite successful at matching this wealth of evidence. Which are their main empirical weaknesses? The monopsony approach has struggled to produce unimodal wage

distributions with a declining and thick right tail. Although wage dispersion is its central implication, it is of the wrong kind (in equation *(2.1)*, *g* is increasing). The introduction of firm heterogeneity (Postel-Vinay and Robin, 2002*a*), *ex post* competition for employed workers (Postel-Vinay and Robin, 2002*b*), and continuous on-the-job search effort (Christensen *et* Al., 2005) goes a long way, but at the cost of introducing free and fundamentally unobservable parameters. The job matching approach, conversely, can replicate even the Pareto shape of the distribution, under the fairly natural assumption of Gaussian noise in production (see equation *(3.2)*). But, it cannot naturally explain the firm size-wage premium, which is a very robust property of all market economies and an unavoidable outcome of the monopsony model.

It is plausible that reality contains a bit of both hypotheses, as well as of other factors affecting wage dispersion, dynamics and worker turnover. The question is really which model best approximates reality. This is a quantitative question, that can be settled only by direct structural estimation of these models, using matched employer-employee data. Work in this area has just begun.

5. - Conclusions

In this article, I have reviewed in parallel and evaluated two leading views of equilibrium wage dispersion, of wage dynamics and worker turnover: the job-matching theory of sorting and the monopsony theory of sequential search with outside offers. Until recently, the two traditions have evolved virtually ignoring each other. I have tried to demonstrate that they address the same phenomena, they explain them quite differently, but they also share a common core of assumptions. The relative strengths and weaknesses of the two approaches make them valid competitors, while other aspects that are likely to be important in labor markets (such as asymmetric information, incentives, insurance) so far lack an equally consistent and tractable equilibrium analysis, and therefore cannot be judged by the same standards.

The question that opens this article has been narrowed down considerably. Are similar workers paid differently because they are inherently different, in a way that only a slow and frictional process can reveal even to them? Or because, for purely strategic reasons also supported by search frictions, firms adopt «wage policies», possibly accepting a lower profit margin to hire and retain a larger labor force? Answering this question is a priority in labor and macroeconomics. We need more theoretical analysis, to flesh out additional "robust" and testable implications of these two different views, and above all we need much empirical work to test the numerous predictions that we already know. The central role that general equilibrium effects and the wage distribution play in the recent incarnations of these theories suggests that reduced-form regressions must be abandoned, in favor of structural estimation building on the equilibrium restrictions.

BIBLIOGRAPHY

ABOWD J. - KRAMARZ F. - MARGOLIS D., «High Wage Workers and High Wage Firms», *Econometrica*, vol. 67, 1999, pp. 251-333.

ALTONJI J. - WILLIAMS N., «Do Wages Rise with Job Seniority? A Reassessment», Forthcoming in the *Industrial and Labor Relations Review*, 2005.

BURDETT K. - COLES M., «Equilibrium Wage-Tenure Contracts», *Econometrica*, vol. 71, 2003, pp. 1377-1404.

BURDETT K. - MORTENSEN D., «Wage Differentials, Employer Size, and Unemployment», *International Economic Review*, vol. 39, 1998, pp. 257-273.

CHRISTENSEN B.J. - LENTZ R. - MORTENSEN D. - NEUMANN G. - WERVATZ A., «On the Job Search and the Wage Distribution», *Journal of Labor Economics*, January, 2005.

DIAMOND P., «A Model of Price Adjustment», *Journal of Economic Theory*, vol. 3, 1971, pp. 156-168.

FARBER H., «The Analysis of Interfirm Worker Mobility», *Journal of Labor Economics*, vol. 12, 1994, pp. 554-593.

FOSTER L. - HALTIWANGER J. - KRIZAN C.J., «Aggregate Productivity Growth: Lessons from Microeconomic Evidence», in HULTEN C. - DEAN E. - HARPER M. (eds.), *New Developments in Productivity Analysis*, pp. 303-63. NBER Studies in Income and Wealth, vol. 63, Chicago and London, University of Chicago Press, 2001.

HALL R.J., «Employment Fluctuations with Equilibrium Wage Stickiness», *American Economic Review*, March, 2005.

JACOBSON L. - LALONDE R. - SULLIVAN D., «Earnings Losses of Displaced Workers», *American Economic Review*, vol. 83, no. 4, 1993, pp. 685-709.

JOVANOVIC B., «Job Matching and the Theory of Turnover», *Journal of Political Economy*, vol. 87, 1979, pp. 972-990.

— - —, «Matching, Turnover and Unemployment», *Journal of Political Economy*, vol, 92, 1984, pp. 108-122.

KAMBOUROV G. - MANOVSKII I., «Occupational Specificity of Human Capital», mimeo University of Pennsylvania, Department of Economics, 2004.

KRUEGER A. - SUMMERS L., «Efficiency Wages and the Inter-industry Wage Structure», *Econometrica*, vol. 56, no. 2, 1988, pp. 259-93.

MORTENSEN D., *Wage dispersion: Why are similar workers paid differently?,* Zeuthen Lecture Book Series, Cambridge and London, MIT Press, 2003, pp. XII-143, 2005.

MOSCARINI G., «Skill and Luck in the Theory of Turnover,» mimeo Yale University, Department of Economics, 2003.

— - —, «Job Matching and the Wage Distribution», *Econometrica*, vol. 72, no. 2, 2005, pp. 481-516.

PISSARIDES C. - WADSWORTH J., «On-the-job Search: Some Empirical Evidence from Britain», *European Economic Review*, vol. 38, 1994, pp. 385-401.

POSTEL-VINAY F. - ROBIN J.-M., «Equilibrium Wage Dispersion with Worker and Employer Heterogeneity», *Econometrica*, vol. 70, 2002*a*, pp. 2295-2350.

— - —, «The Distribution of Earnings in an Equilibrium Search Model with State - Dependent Offers and Counteroffers», *International Economic Review*, vol. 43, 2002*b*, pp. 989-1016.

ROY A., «Some Thoughts on the Distribution of Earnings», *Oxford Economic Papers* (New Series), vol. 3, 1951, pp. 135-146.

SHIMER R., «The Cyclical Behavior of Equilibrium Unemployment and Vacancies», *American Economic Review*, March, 2005.

TOPEL R., «Specific Capital, Mobility and Wages: Wages Rise with Job Seniority», *Journal of Political Economy*, vol. 99, 1991, pp. 145-176.

TOPEL R. - WARD M., «Job Mobility and the Careers of Young Men», *Quarterly Journal of Economics*, vol. 107, no. 2, 1992, pp. 439-79.