Wage Dispersion with Heterogeneous Firm Technologies and Worker Abilities: An Equilibrium Job Search Model for Matched Employer-Employee Data

Fabien Postel-Vinay
INRA-LEA*

Jean-Marc Robin
INRA-LEA and CREST-INSEE†‡

October 2000 — Revised November 2000

Abstract

An equilibrium job search model with on-the-job-search is presented and solved, in which we allow firms to implement optimal wage posting strategies in the sense that they leave no rent to their employees and counter the offers received by their employees from competing firms. Unobserved worker productive heterogeneity is introduced in the form of cross-worker differences in a 'competence' parameter. On the other side of the market, firms also are heterogeneous with respect to their (observable) marginal productivity of labor. The theoretical model can be solved in closed-form and typically delivers a hump-shaped aggregate earnings distribution that reflects both firm- and worker-heterogeneity. The fit to the observed earnings distributions is very good. The model also fits the observed distributions of firm sizes in the populations of workers and firms. Finally, it delivers both between- and within-firm endogenous wage dispersion.

The structural model is estimated using matched employer and employee French panel data. Its fit to the data is good. We then use the results for two applications. The first one is a decomposition of the log-wage means and variances into additive firm and person effects. We find that the share explained by the person effect varies across skill groups, and is generally much smaller than what was found in previous analyses of the same panel. Specifically, this share lies close to 50% for high-skilled white collars, and quickly decreases to 0% as the observed skill level decreases. The second application is a look at the anatomy of the 'matching technology'. We find evidence of nonmonotonic relationships between firm sizes, productivities and recruiting efforts.

Keywords: Labor market frictions, wage dispersion, log wage variance decomposition.
JEL codes: J31, J41, J64

*Address: Laboratoire d'Economie Appliquée de l'INRA, Ecole Normale Supérieure, 48 bld Jourdan, 75014 Paris, France, e-mail: fpostel@java.ens.fr. Internet: www.inra.fr/Internet/Departements/ESR/UR/lea/fpostel.htm.

†Address: Laboratoire d'Economie Appliquée de l'INRA, Ecole Normale Supérieure, 48 bld Jourdan, 75014 Paris, France. E-mail: robin@java.ens.fr.

‡We are grateful for comments received from conference participants at the "search and matching" conference held at the University of Iowa (Aug. 2000), the Econometric Society World Meeting in Seattle (Aug. 2000), and from seminar participants at Boston University, New York University, Brown, Yale, CREST-INSEE, DELTA, Université Catholique de Louvain and University College London. Discussions with Jim Albrecht, Zvi Eckstein, Chris Flinn, Sam Kortum, Francis Kramarz and Dale Mortensen led to significant improvements in the paper. The usual disclaimer naturally applies.

1 Introduction

It is now widely accepted that both worker and employer attributes should be incorporated into properly specified wage functions (e.g. Abowd, Kramarz and Margolis, 1999, Abowd and Kramarz, 1999, for recent examples). The estimation of wage functions would be a straightforward exercise if all relevant attributes were observable to the econometrician. In the more likely situation where unobserved heterogeneity has to be taken into account, at least some structure is needed for the wage function. The existence of non observed worker attributes has motivated and still motivates an important series of econometric contributions to this literature on the determinants of wages (see below; paragraph on *related literature*).

Another acknowledged fact in labor economics is that labor market frictions are a central part of the allocating process of workers into jobs. As such, they also have a crucial bearing on wage determination. On a labor market where worker mobility is hindered by frictions, employees are not generally able to bring all their potential employers into competition to push their wages up to their marginal productivity. As was shown by Burdett and Judd (1983), marginal productivity payments occur only if workers (searching for a job) can simultaneously apply to at least two would-be employers. If job offers do not systematically arrive at least in pairs, then equilibrium wages are equal to neither marginal productivity nor reservation wages but are necessarily dispersed, even among identical workers and firms. Relatedly, Burdett and Mortensen (1998) show that dispersion in equilibrium wages occurs with sequential search if workers can search on the job. Limited or costly job search therefore appears to be a self-reinforcing source of wage dispersion. Costly search gives firms monopsony power which in turn motivates the workers' search activity because there always remains a hope to find a better-paying job.

Therefore, not only should wages depend on firm and worker attributes, but they should also depend on a third random component reflecting the variety of individual career paths.

We acknowledge this conclusion and construct an empirical structural model of individual wage paths using matched employer-employee data similar to the data described by Abowd *et al.* (1999). The theoretical apparatus that we use to this end is a search-theoretic model of the labor market with an original wage setting mechanism that mixes Burdett and Mortensen's (1998) ideas about on-the-job search with Burdett and Judd's (1983) ideas about instant competition of firms for workers through job offer recall. A first version of this model was explored by Postel-Vinay and Robin (1999). The model discussed in the present paper extends this earlier work in two major directions: first, by allowing double—i.e. firm- and worker- —productive heterogeneity, second by exhibiting a more flexible matching procedure than is usually posited by job search models.

Summary of the principal results. We believe this model to be interesting in at least three respects. First, it naturally delivers a structural log-wage equation that we propose as an alternative to that produced by the Roy model. Our equation leads to a decomposition of the log wage variance into three components. The first two are the conventional firm and person effects. The third reflects labor market frictions as we have already explained. The predicted shares of all three effects in the log wage variance can easily be computed after the model has been estimated, and convey new results about the decomposition of log-wage variance. Specifically, we find that the share of total log wage variance explained by a person effect lies around 50% for skilled white collars, and quickly drops down to 0% as the observed skill level decreases (the estimated contribution of the person effect is estimated 0 for the three lower-skilled worker categories, out of seven categories).[1] Perhaps the most striking result for that matter is that for a number of (low-skilled) worker categories, we are able to construct a model that explains the observed variance in earnings without resorting to unobserved exogenous differences in individual abilities.

This contrasts sharply with the results obtained by Abowd *et al.* (1999) and Abowd and

[1] A more careful interpretation goes like this: Either there is indeed no unobserved ability differences for low skilled workers, or some institutional mechanism, like union-employer agreements, forbids the individualization of wages.

Kramarz (1999). Using an error component model, they find an average of more or less 50% of the wage variance due to person effects using the same data as we do. Where did the lost variance go? Our model attributes it to labor market frictions: endogenous worker mobility through the sequential sampling of alternative job offers creates earnings differentials across identical workers working at identical firms. A 'lucky' or 'senior' worker who has gotten one more job offer than his 'unlucky' or 'junior' *alter ego* has also gotten one extra opportunity to bargain for a higher wage. Estimating static error component models when data generating processes are truly dynamic will therefore incorporate all historical differences (in the states of individual wage trajectories at the first observation date) into individual fixed effects. We find that the explanatory power of historical differences is far from negligible (20 to 60% of total wage variance, depending on the skill category).

Secondly, the model potentially improves upon job search models at a number of points. Among these, we can mention the following few: our model generates endogenous within-firm wage mobility (with distinct effects of tenure and competence); it also generates downward (between-firm) wage mobility; it is not incompatible with mobility to firms of smaller size or mean wage; and it is flexible enough to give a perfect fit to the observed distribution of firm sizes, and a very good one to the distribution of earnings. We make very little use of dynamic wage data to estimate the structural model.[2] Only to estimate cross-state transition rates do we use data on spell durations. The simulation of wage mobility therefore provides a true out-of-sample specification testing procedure. Taking account of the fact that the model features no idiosyncratic productivity shocks, one can conclude that the model gives satisfactory predictions of wage variations with and without employer change.

Thirdly, thanks to the flexibility of the matching technology that we first posit, the model has interesting suggestions about the anatomy of the process through which workers and firms

[2]This is another definite advantage of structural estimation. The model has enough structure to allow for its estimation using a very short panel of individual wage trajectories. This opens the way to the estimation of such equilibrium microeconomic models of frictional labor markets when structural parameters are subject to (slow) business cycle changes.

are matched together. Specifically, we come up with a measure of the firms' 'recruiting effort' and find that it is in a decreasing relationship with the firms' productivities: more productive firms devote less effort to hiring, which naturally makes them less efficient in contacting potential new employees. On the other hand, since they generate *ceteris paribus* higher match surpluses, they are more likely to attract the workers that they do contact. Those two opposing forces sum up to a hump-shaped relationship between productivity and firm size.

Related literature. The literature on wage determination, wage dispersion and the estimation of wage functions is simply huge, for a representative overview of which this introductory discussion is certainly not the place. The papers mentioned below are the ones we think of as most directly related to ours.

The theoretical model that we use is an equilibrium job search model with on-the-job search. As we already mentioned, its innovations mainly concern the specification of the wage setting mechanism[3] and that of the matching technology. It thus borrows from and adds to the literature on job search (see Mortensen and Pissarides, 1998 and the references therein).

The idea that wage equations should include both firm and worker heterogeneity components may be viewed as the central message of a long history of theoretical research on hedonic wages and assignment models (see Sattinger, 1993, for a survey). In this respect, our paper is related to that strand of literature.

On the empirical front, a recent series of papers by Abowd, Kramarz and Margolis (1999), Abowd and Kramarz (1999), and Abowd, Finer and Kramarz (1999) estimate error component models on French and US longitudinal data. As we argued in the previous paragraph, this approach is essentially descriptive and lacks the structure needed to interpret the true nature of the individual fixed effect. Our main contribution consists in constructing a tractable equilibrium model of the labor market that provides this structure.

[3] An early version of which was set up in Postel-Vinay and Robin (1999). A somewhat similar idea was independently developed by Dey and Flinn (2000).

Structural estimations of equilibrium models of the labor market with double productive heterogeneity are rather uncommon. In this field, the contributions most commonly referred to use the Roy (1951) model of self selection and earnings inequality where heterogeneous workers sort themselves across various sectors requiring sector-specific tasks (e.g. Heckman and Sedlacek, 1985, Heckman and Scheinkman, 1987, and Heckman and Honoré, 1990). The original Roy model is Walrasian and thus abstracts from labor market frictions. As a result of perfect labor mobility between sectors, it does not deliver any *observed* worker mobility in equilibrium because all workers instantaneously go to their elected sector and stay there forever. This is clearly at odds with empirical evidence and again pleads in favor of models like ours that take account of the existing obstacles to labor mobility. Search frictions are incorporated into the Roy model of self selection in a recent paper by Moscarini (2000), who mainly focuses on the differences in search strategies across workers. Although very promising, this approach still involves too much analytical complexity to be empirically implementable. Our model has the drawback of treating the search strategies as essentially exogenous, its advantage being its ability to deliver quantitatively realistic earnings distributions and wage dynamics which can successfully be confronted to the data.[4]

Finally, a number of equilibrium job search models with heterogeneous workers and/or firms have been estimated. This literature was initiated by Eckstein's and Wolpin's (1990) celebrated estimation of the Albrecht and Axell (1984) model. Recent additions include the estimation of the Burdett and Mortensen (1998) model by Van den Berg and Ridder (1998) and Bontemps, Robin and Van den Berg (1999, 2000) (again, see Mortensen and Pissarides, 1998, for a survey). As far as we are aware, though, our paper is the first one in this literature to use matched firm and worker data, as is advocated by Mortensen (1999). This type of data indeed allow to clearly identify firm and worker effects. Another distinguishing feature of our contribution is

[4]The model presented in Moscarini (2000) only has two types of jobs and therefore cannot deliver quantitatively realistic wage dynamics. Increasing the number of job types, although in principle not particularly difficult, has an enormous cost in terms of tractability.

its departure from the assumption of wage posting. In particular, as we mentioned in previous paragraph, the alternative wage setting mechanism that we propose is able to generate simulated individual wage dynamics that offer a satisfactory fit to those observed in the data.

Outline. The rest of the paper is divided into two Parts: Part 1 details the theoretical model, Part 2 describes the data, details the structural estimation procedure and reports and discusses the estimation results. A final Section concludes on the successes and failures of our model, and points to some ideas to improve on the latter. Some proofs are gathered in a final Appendix.

2 Theory

In this first section we describe a labor market in which search frictions matter and imply a non-Walrasian wage formation process.

2.1 Behavioral and technological assumptions

Workers. We consider the market for a homogeneous profession (manual workers, administration employees, managers, ...), in a steady state and in which a measure M of atomistic workers face a continuum of competitive firms, with a mass normalized to 1, that produce one unique multi-purpose good. Workers face a constant birth/death rate μ,[5] and firms live forever.

Workers can either be employed or unemployed, and the unemployment rate of a given category of labor is denoted by u. Newborn workers begin their working life as unemployed. The pool of unemployed workers is steadily fueled by layoffs that occur at the exogenous rate δ, and by the constant flow μM of newborn workers.

Workers are homogeneous with respect to the set of observable characteristics defining their profession—or equivalently the particular market on which they operate—, but may differ in their personal 'abilities'. A given worker's ability is measured by the amount ε of efficiency units of labor she/he supplies per unit time. The workers' ability parameters ε are exogenously

[5] The birth-death process adds very little to the theory and could be discarded for that matter. Yet we shall see that it provides a simple way of modelling the attrition observed in the data.

distributed among the total population of workers according to the cdf H over $[\varepsilon_{\min}, \varepsilon_{\max}]$, both positive numbers. Newborn workers are assumed to draw their value of ε randomly from the distribution H. We only consider continuous ability distributions and further denote the corresponding density by h.[6]

A type-ε unemployed worker has an income flow of εb, with b a positive constant[7], which he has to forgo from the moment he finds a job. We can give b at least two interpretations. The more traditional one is to think of it as the sum of possible sources of income and search expenses during unemployment. These incomes may consist e.g. of unemployment benefits and the worker's valuation of 'home production'. (Although unemployment benefits *per se* may not be systematically related to skills, the adopted simple specification makes sense under the 'home production' interpretation. One can also argue that UI payments are typically fractions of earlier wages which, as we shall see, are positively related to personal ability.) An alternative interpretation is to think of b as some measure of the unemployed workers' 'bargaining power'. As will become clear below, being unemployed is equivalent to working at a 'virtual' firm of labor productivity equal to b that would operate in a frictionless competitive labor market, therefore paying each employee their marginal productivity, εb. The more productive this virtual firm is, the more rent the workers can extract from their future matches with 'actual' employers.

Firms and matches. Firms differ in the technologies that they operate. We make the simplifying assumption of constant returns to labor.[8] More specifically, we assume that firms differ by an exogenous technology parameter p with cdf Γ across firms over a bounded support $[p_{\min}, p_{\max}]$. This distribution is assumed continuous with density γ. The marginal productivity of the match (ε, p) of a worker with ability ε and a firm with technology p is $p\varepsilon$. A type-p firm's total per period output is consequently equal to p times the sum of its employees' abilities.

[6] We fully characterize unobserved individual heterogeneity by a scalar index. This strong restriction greatly simplifies both the theory and the estimation.

[7] The—admittedly restrictive—assumption that a worker's productivities 'at home' and at work are both proportional to ε greatly simplifies the upcoming analysis.

[8] Exploration of the more general, yet formally equivalent case of perfect additive substitutability of workers (within all professional categories but maybe not across professions) is left to later work.

Matching. Contrary to what is assumed in competitive models of heterogeneous markets (of which the Roy, 1951 model is an example) we do not assume that workers can freely choose which type of firm to apply to or firms which type of workers to contact. Rather, we assume that firms and workers are brought together pairwise through a (possibly two-sided) search process, that search takes time, is sequential, and is random.

Specifically, unemployed workers sample job offers sequentially at a Poisson rate λ_0. As in the original paper by Burdett and Mortensen (1998), employees may also search for a better job while employed. The arrival rate of offers to on-the-job searchers is λ_1. The type (mpl) p of the firm from which a given offer originates is assumed to be randomly selected in $[p_{\min}, p_{\max}]$ according to a *sampling distribution* with cdf F (and $\overline{F} \equiv 1 - F$) and density f. Unlike Burdett and Mortensen,[9] though, we assume no *a priori* connection between f and the density of firm types γ. (Were p uniformly distributed across firms, then it might still be that different firm would have different probabilities of been contacted by workers.) Also, the sampling distribution is the same for all workers irrespective of their ability and employment status.

Assuming that all workers have the same sampling distribution independently of their ability and employment status may seem strong. A possible rationale is that it would go against anti-discrimination regulations for a firm to post an offer specifying a range of acceptable values of worker types, except for those which have been agreed upon by the collective agreements defining the marketed profession. Our assumption typically rules out the existence of help-wanted ads reading "Economist wanted; three-digit-IQed applicants only". However, it does not imply that employers do not discriminate between workers since we shall assume that employers condition their wage offers on worker characteristics. Firms are therefore unable to select workers *ex ante*, but they can do so *ex post.*

One may thus think of the search process as follows: workers go to job agencies and take the job offers posted by the highest-p firms, because higher p's generate higher surpluses (see

[9] Who assume that they are equal. See below in this paragraph and paragraph 3.4.7 for more on this point.

below). The probability for a given worker to contact a firm of a given p thus only depends on the (steady-state) number of ads posted by these firms. The sampling weights $f(p)/\gamma(p)$ can be interpreted as the average flows of ads (or hiring effort) posted by firms of productivity p per unit time. This hiring effort is likely to be a decision variable of the firms. Yet, we only consider here the partial equilibrium conditional on a given distribution of these ratios across firms. We leave these ratios unrestricted and provide no theory to endogenize them. We just refer to two recent papers by Mortensen (1998, 1999) who brings together the search and matching strands of the microeconomic and macroeconomic literature on labor in a way that provides foundations for the individual employer/employee match formation process we assume here.

2.2 Wage setting

Objectives and strategies. Workers discount the future at an exogenous and constant rate $\rho > 0$ and seek to maximize the expected discounted sum of future utility flows. The instantaneous utility flow enjoyed from a flow of income x is $U(x)$.[10]

Firms seek to minimize wage costs. We make the following important three assumptions on wage strategies:

1. Firms can vary their wage offers according to the characteristics of the particular worker they meet;
2. Firms can counter the offers received by their employees from competing firms;
3. Wage contracts are long-term contracts that can be renegotiated by mutual agreement only.

The first two assumptions are a departure from the standard Burdett and Mortensen (1998) model. Their implications are explored by Postel-Vinay and Robin (1999) in a model where

[10] We rule out intertemporal transfers and savings and assume incomplete insurance markets. Risk averse individuals who want to smooth consumption over time should therefore want to save and borrow. But this is a source of additional complexity that we cannot yet afford.

workers are all equally productive, but differ in their opportunity cost of employment. They naturally arise from the assumption of perfect information about the individual characteristics of matching counterparts. This is a disputable assumption. Yet, recruitment interviews definitely reveal some information about worker ability. Moreover, even in countries like France, where strict regulations constrain the firms' layoff policies, the labor legislation generally allows for trial periods during which firms are free to let go their hirees at no (or minimal) cost. We therefore claim that perfect information is a valid alternative to the blindness of interacting agents in the Burdett-Mortensen model.

Second, even if information is perfect, their might exist limits to the extent to which firms can vary the wage they offer to workers. These limits could be legal restrictions like a minimum wage decided by the government or negotiated by trade unions. We leave to further work the analysis of the effects of such restrictions on the wage setting mechanism. These restrictions could also be self-imposed (one can e.g. think of a firm willing to avoid moral hazard problems with the rest of its other employees). Although we recognize the importance of these effects, analyzing them within the context of a general dynamic equilibrium model is a formidable task that we shall not undertake here. Moreover, we shall see in the empirical part of this paper that there is no evidence of any restriction to wage dispersion in the data since it will reveal necessary to allow for very long right tails of the distribution of firm productivities to explain the observed huge dispersion at the right end of the individual wage distribution.

Third, when an employee receives an outside offer of a wage greater than her current wage but lower than her marginal productivity, there is no reason why her current employer should let her leave the firm although this kind of passive behavior is clearly sub-optimal. Allowing for counter-offers thus provides the equilibrium search model with a greatly extended amount of flexibility.

Finally, assumption 3 is more standard and only ensures that a firm cannot unilaterally cancel a promotion obtained by one of its employees after having received an outside job offer, once the worker has eventually turned down that offer. It follows that wage cuts within the firm are not permitted. Note that firms will never fire any workers because nothing can change in the firm's environment which would render a wage contract unprofitable to the firm if it previously was.

Wage contracts. We now exploit the preceding series of assumptions to derive the precise values of the wage resulting from the various forms of employer-employee contacts.

The lifetime utility of an unemployed worker with competence ε (a worker of type ε, for short) is denoted by $V_0(\varepsilon)$, and that of the same worker when employed at a firm of type p and paid a wage w is $V(\varepsilon, w, p)$. A type-p firm is able to employ a type-ε unemployed worker if the match is productive enough to at least compensate the worker for his forgone unemployment income, *i.e.* $\varepsilon p \geq \varepsilon b$. Therefore, the infimum of Γ's support, $p_{\min}$, has to be no less than b, for a firm less productive than b would never attract any worker. Whenever that condition is met, any type-p firm will want to hire any type-ε unemployed worker upon 'meeting' him on the search market. To this end, the type-p firm optimally offers to the type-ε unemployed worker the wage $\phi_0(\varepsilon, p)$ that exactly compensates this worker for his opportunity cost of employment, which is defined by

$$V(\varepsilon, \phi_0(\varepsilon, p), p) = V_0(\varepsilon). \tag{1}$$

Because a given employed worker's future employment prospects depend on both the mpl of the firm he works at and his personal ability, the minimum wage at which a type-ε unemployed worker is willing to work at a given type-p firm depends on p and ε, as shown by equation (1).

When a given type-p firm's employee receives an outside offer from a type-p' firm both firms enter in a Bertrand competition won by the most competitive firm. Let $\phi(\varepsilon, p, p')$ denote the optimal wage that the challenging firm $p' > p$ has to propose to a worker (of type ε) employed at a firm with mpl p, and that the worker is willing to accept. Since it is willing to extract a positive marginal profit out of every match, the best the firm of type p can do for its employee is to set his wage exactly equal to εp. The highest level of utility the worker can attain by staying at the type-p firm is therefore $V(\varepsilon, \varepsilon p, p)$. Accordingly, he accepts to move to a potentially better match with a firm of type p' if the latter offers at least the wage $\phi(\varepsilon, p, p')$ defined by

$$V\left(\varepsilon, \phi\left(\varepsilon, p, p'\right), p'\right) = V(\varepsilon, \varepsilon p, p). \tag{2}$$

Any less generous offer on the part of the type-p' firm is successfully countered by the type-p firm. If p' is less than p, then $\phi(\varepsilon, p, p') \geq \varepsilon p'$, in which case the type-p' firm will never raise its offer up to this level. Rather, the worker will stay at his current firm, and be promoted to the wage $\phi(\varepsilon, p', p)$ that makes him indifferent between staying and working at the type-p' firm.

The precise value of $\phi(\cdot)$ is derived in Appendix A as:

$$U\left(\phi\left(\varepsilon, p, p'\right)\right) = U(\varepsilon p) - \frac{\lambda_1}{\rho + \delta + \mu} \cdot \int_p^{p'} \overline{F}(x)\, \varepsilon U'(\varepsilon x) dx. \tag{3}$$

Note that in distinction to standard search theory, we get an explicit definition of $\phi(\varepsilon, p, p')$ from (3) instead of an implicit definition as the solution to an integral equation. This will greatly simplify the numerical computations in the empirical analysis.

Expression (3) has some rather intuitive features. First, the wage paid by firm p' is less than the maximal wage firm p can afford to pay, i.e. the marginal productivity of the match εp. The difference between the two, measured by the integral term in (3) represents the *option value* of turning down the type p firm to work at the type p' firm. This option value increases with the productivity difference $\varepsilon(p' - p)$. Workers indeed accept lower wages to work at more productive firms because $p\varepsilon$ being an upper bound on any wage offer resulting from the competition between the incumbent employer p and any challenger p', workers agree to trade a lower wage now for increased chances or higher wages tomorrow. It is thus more difficult to draw a worker out of a more productive firm, and equivalently workers are more easily willing to work at more productive firms.

It is then straightforward to understand why the option value positively depends on the frequency of outside offers (λ_1) and the likelihood of high-p draws.[11] It negatively depends on the overall job termination rate $\delta + \mu$, which tends to reduce the probability that an outside wage offer arrives before the match breaks up. Finally, the amount of intertemporal transfer negatively depends on the discount rate and the coefficient of relative risk aversion. More myopic and more risk-averse individuals[12] are indeed less keen on accepting such risky transfers.

Finally, we also show in Appendix A that the wage offered by a firm of type p to a type-ε unemployed worker is $\phi_0(\varepsilon, p) = \phi(\varepsilon, b, p)$. Unemployed workers of all types are thus prepared to work for a wage ϕ_0 that is *less* than the opportunity cost of employment εb for the same intertemporal arbitrage motive. Moreover, the reservation wage does not depend on the arrival rate of offers λ_0. In conventional search theory, reservations wages do depend on λ_0, because the wage offers are not necessarily equal to the reservation wage. A longer search duration may thus increase the value of the eventually accepted job. Here, this does not happen: Firms always pay the reservation wage to workers. Therefore, there is no gain to expect from rejecting an offer and waiting for the following one.

Wage mobility and job mobility. The following mobility patterns then naturally emerge from these wage setting mechanisms: There exists a threshold $q(\varepsilon, w, p)$ defined by the equality

$$\phi(\varepsilon, q(\varepsilon, w, p), p) = w$$

such that:

1. An type-ε employee of a type-p firm at a wage w receiving an offer from a firm of type $p' \leq q(\varepsilon, w, p)$ does not gain anything from this contact because the challenging firm is not productive enough to grant the worker a positive wage raise.

[11] For two sampling distributions F_1 and F_2, if F_1 first-order stochastically dominates F_2 then the wedge $p\varepsilon - \phi(\varepsilon, p, p')$ is greater for F_1 than for F_2.

[12] These two concepts, time discounting and risk aversion, play a somewhat similar role in this model. Less risk averse individuals will have similar trajectories than more risk averse workers if they are at the same time more myopic.

2. If $p \geq p' > q(\varepsilon, w, p)$ then the current employer can match any offer of the challenging firm and the worker profits from the Bertrand competition between p and p' by getting a wage raise in firm p equivalent, in present value, to being paid his marginal productivity $p'\varepsilon$ in the type-p' firm. (Note that it is a dominant strategy for the weaker firm p' to challenge firm p. Indeed it loses nothing if p counters and wins the worker if not.)

3. If $p < p'$ then firm p is no match to p' and lets its employee leave to firm p' at a wage that is equivalent to being paid at his previous marginal productivity $p\varepsilon$. If p' is large enough the worker may even accept a wage that is lower than his previous wage w.

The wage setting mechanism that we assume in this paper thus delivers nice earnings profiles. First, individual tenure profiles of within-firm earnings are non-decreasing and concave. A longer tenure increases the probability of raising a good outside offer. On the other hand, workers with long tenures, who on average have received more offers and therefore get higher wages, are less likely to receive an attractive offer that would result in a promotion. Second, the model can generate firm-to-firm mobilities with wage cuts when the tenure profile in the new firm is expected to be increasing over a very long time span. Note that this wage mobility occurs although there is no human capital accumulation (abilities ε do not change over time). The model therefore offers an alternative to the usual explanation of tenure effects.

2.3 Steady-state equilibrium

Let $\ell(\varepsilon, p)$ be the density of type ε employees at type-p firms and let $\ell(p) = \int_{\varepsilon_{\min}}^{\varepsilon_{\max}} \ell(\varepsilon, p) d\varepsilon$ be the density of employees working at type-p firms. We also denote with a capital letter $L(\varepsilon, p)$ and $L(p)$ the corresponding cumulated distribution functions. Let $G(w|\varepsilon, p)$ be the cdf of the (not absolutely continuous, as we shall see) conditional distribution of wages within the pool of workers of ability ε within type-p firms. We now proceed to derive the steady-state instantaneous flow equations for those various worker stocks.

Unemployment. The equality between the flows into- and out of unemployment implies that

$$\delta(1-u)+\mu=(\lambda_0+\mu)u \Longleftrightarrow (\delta+\mu)(1-u)=\lambda_0 u.$$

Hence the unemployment rate:

$$u=\frac{\delta+\mu}{\delta+\mu+\lambda_0}. \tag{4}$$

Firm sizes. We then show how to derive the steady-state distribution of workers across firm types. In an infinitesimal fragment of time, a fraction $\delta+\mu+\lambda_1\overline{F}(p)$ of the workforce $(1-u)M\cdot\ell(p)$ of type p firms leaves these firms. This fraction consists of those workers who are either fired—which occurs at rate δ—or retire—which occurs at rate μ—or get an offer from a more attractive firm—which occurs at rate $\lambda_1\overline{F}(p)$. On the inflow side, the measure of workers entering this employment stock is $[\lambda_0 uM+\lambda_1(1-u)M\cdot L(p)]\cdot f(p)$, where $\lambda_0 uM=(\delta+\mu)(1-u)M$ is the fraction of unemployed workers receiving a job offer and $\lambda_1(1-u)M\cdot L(p)$ is the cumulated workforce of all firms with productivity less than p receiving an offer. The steady-state equality of the flows into and out of the stock of workers employed by firms with mpl p therefore writes as:

$$\left[\delta+\mu+\lambda_1\overline{F}(p)\right]\cdot\ell(p)=[\delta+\mu+\lambda_1 L(p)]\cdot f(p). \tag{5}$$

A result that will prove crucial in the empirical applications can then be derived from the above linear differential equation. It indeed provides the relation between the distribution of firm heterogeneity across employees L and the sampling distribution F. The following relationship therefore holds in equilibrium:

$$1+\kappa_1\overline{F}(p)=\frac{1+\kappa_1}{1+\kappa_1 L(p)}, \tag{6}$$

with $\kappa_1=\frac{\lambda_1}{\delta+\mu}$, and it follows from differentiation of equation (6) that

$$\ell(p)=\frac{(1+\kappa_1)}{\left[1+\kappa_1\overline{F}(p)\right]^2}\cdot f(p). \tag{7}$$

Equation (6) expresses the equilibrium distribution of workers across firms (as characterized by their technological parameters p), L and ℓ, as a function of κ_1, which characterizes the relative chances for an employee of getting an outside offer (λ_1), being laid off (δ), or quitting the market (μ), and the sampling distribution F. It is a particularly useful prediction of the model which will allow us to back out the sampling distribution F from its empirical counterpart L (p being observed and κ_1 known).

Within-firm wage and worker ability distributions. The $G(w|\varepsilon,p)\,\ell(\varepsilon,p)\,(1-u)M$ workers of type ε, employed at firms of type p, and paid less than $w \in [\phi_0(\varepsilon,p), \varepsilon p]$ leave this category either because they are laid off (rate δ), or because they retire (rate μ), or finally because they receive an offer from a firm with mpl $p \geq q(\varepsilon,w,p)$ which grants them a wage increase or induces them to leave their current firm (rate $\lambda_1 \overline{F}[q(\varepsilon,w,p)]$). On the inflow side, workers entering the category (ability ε, wage $\leq w$, mpl p) come from two distinct sources. Either they are hired away from a firm less productive than $q(\varepsilon,w,p)$, or they come from unemployment.

The steady-state equality between flows into and out of stocks $G(w|\varepsilon,p)\,\ell(\varepsilon,p)$ thus takes the form:

$$\begin{aligned}
&\left\{\delta+\mu+\lambda_1\overline{F}[q(\varepsilon,w,p)]\right\} G(w|\varepsilon,p)\,\ell(\varepsilon,p)\,(1-u)M \\
&= \left\{\lambda_0 uM\, h(\varepsilon) + \lambda_1(1-u)M \int_{p_{\min}}^{q(\varepsilon,w,p)} \ell(\varepsilon,x)dx\right\} f(p) \\
&= \left\{(\delta+\mu)h(\varepsilon) + \lambda_1 \int_{p_{\min}}^{q(\varepsilon,w,p)} \ell(\varepsilon,x)dx\right\}(1-u)Mf(p), \qquad (8)
\end{aligned}$$

since $\lambda_0 u = (\delta+\mu)(1-u)$.

Solving this equation for $G(w|\varepsilon,p)$ and $\ell(\varepsilon,p)$ is not as difficult as it might seem at first sight. First, the maximal wage that firm p can pay to a worker ε is $w = p\varepsilon$, in which case $q(\varepsilon,w,p) = p$, and equation (8) thus becomes:

$$\left[\delta+\mu+\lambda_1\overline{F}(p)\right]\ell(\varepsilon,p) = \left[(\delta+\mu)\,h(\varepsilon) + \lambda_1 \int_{p_{\min}}^{p} \ell(\varepsilon,x)dx\right] f(p). \qquad (9)$$

Term-by-term identification with equation (5) immediately shows that

$$\ell(\varepsilon, p) = h(\varepsilon)\ell(p). \tag{10}$$

Then substituting (6), (7) and (10) into (8) straightforwardly yields:

$$G(w|\varepsilon, p) = \left(\frac{1 + \kappa_1 \overline{F}(p)}{1 + \kappa_1 \overline{F}\left[q(\varepsilon, w, p)\right]} \right)^2. \tag{11}$$

Equation (10) implies that, under the model's assumptions, the distribution of individual heterogeneity within the firms is independent of their types. Nothing thus prevents the formation of highly dissimilar pairs (low ε, high p, or low p, high ε) if profitable to both the firm and the worker. This results from the assumption the value of non market time is $b\varepsilon$ with identical b for all workers. Then, all operating firms must have $p > b$ and all possible matches generate a positive surplus and there will always exist a wage acceptable for every worker-firm pair. Finally, given that match productivities are of the multiplicative form $p\varepsilon$ it does not happen that p beats p' for some ε's and p' beats p for some others. So on-the-job search causes no distortion in the conditional wage distribution.

This result deserves some comments. First, one should note that it doesn't rule out assortative matching of workers and firms in a general sense: remember that we are considering a labor market for workers with identical observed characteristics. Going back to the labor market as a whole, it may very well be the case that the within firm distributions of *observed* individual heterogeneity (which defines the marketed profession) vary significantly across firms. Our model merely predicts the absence of sorting *once observed worker heterogeneity is controlled for.*

This result finds some empirical support. The somewhat limited available evidence about the correlation between worker and firm productive heterogeneity components indeed shows that the degree of sorting is in any case small, controlling for observed worker heterogeneity. Abowd, Kramarz and Margolis (1999) estimate a correlation between firm and worker effects of 0.08 in the French DADS panel (order-dependent estimation of the correlation between α and ϕ in table VI), and Abowd, Finer and Kramarz (1999) find essentially 0 using the Washington

State UI data.

Steady-state earnings distribution. The preceding results have an immediate consequence: the steady-state earnings distribution, i.e. the cross-sectional equilibrium distribution of wages in the population of employees, is the distribution of $\phi(\varepsilon, q, p)$ where ε, p, q are three random variables such that

1. ε is independent of (p, q),
2. the cdf of the marginal distribution of ε is H on $[\varepsilon_{\min}, \varepsilon_{\max}]$,
3. the cdf of the marginal distribution of p is L on $[p_{\min}, p_{\max}]$, and
4. the cdf of the conditional distribution of q given p is $\widetilde{G}$ on $\{b\} \cup [p_{\min}, p]$ such that

$$\begin{aligned} \widetilde{G}(q) &= G\left(\phi(\varepsilon, q, p)|\varepsilon, p\right) \\ &= \frac{\left[1+\kappa_1 \overline{F}(p)\right]^2}{\left[1+\kappa_1 \overline{F}(q)\right]^2} \end{aligned}$$

   for all $q \in \{b\} \cup [p_{\min}, p]$.

A very simple algorithm can thus be designed to generate random draws from the steady-state distribution of earnings:

1. Draw ε from distribution with pdf h;
2. Independently draw p from pdf ℓ;
3. Draw $q = \max(q_1, q_2)$ independently of ε with q_i, $i = 1, 2$, such that:
   (a) $q_i = b$ with probability $\frac{1+\kappa_1 \overline{F}(p)}{1+\kappa_1} = \frac{1}{1+\kappa_1 L(p)}$;
   (b) and with probability $1 - \frac{1+\kappa_1 \overline{F}(p)}{1+\kappa_1} = \frac{\kappa_1 L(p)}{1+\kappa_1 L(p)}$, q_i is a draw from the conditional distribution of productivities truncated above at p, i.e. with density: $\ell(q)/L(p)$ at $q \in [p_{\min}, p]$.

3 Application

In this section we develop an estimation procedure for the preceding model in the case $U(w) = \ln w$. The reason for this choice of specification is that, as will clearly appear below, it readily delivers nice log-wage equations for the empirical applications; but more general specifications, like the Box-Cox transform (CRRA utility function), could also be used. Because both worker abilities and firm technological parameters are unobserved to the econometrician, one cannot identify the location parameter for the respective distributions of both $\ln \varepsilon$ and $\ln p$. We therefore add the normalization assumption that $E \ln \varepsilon = \int_{\varepsilon_{\min}}^{\varepsilon_{\max}} \ln \varepsilon \, h(\varepsilon) d\varepsilon = 0$.

We use the same matched employer-employee wage data as Abowd, Kramarz and Margolis (1999). We start by describing the data, then the estimation procedure and the results.

3.1 The DADS data

The "Déclarations Annuelles des Données Sociales" dataset is a large collection of matched employer-employee information collected by the Income Division of the French Statistical Institute INSEE (*Institut National de la Statistique et des Etudes Economiques — Division des Revenus*). The data are based on a mandatory employer report of the gross earnings of each salaried employee of the private sector subject to French payroll taxes. (See Abowd, Kramarz and Margolis (1999) for a complete description of the DADS data.)

We use two datasets. The first sample follows all individuals employed on January 1st, 1996 at firms installed in the region *Ile-de-France* (greater Paris), who were born in October of even-numbered years. Our extract runs from 1996 through 1998, our last available survey. We have deliberately selected a much shorter period than is available because we want to find out whether it is possible to estimate our structural model on a homogeneous period of the business cycle. It would have been very hard indeed to defend the assumption of time-invariant parameters had we been using a much longer panel like the one used by Abowd *et al.* (1999). Each observation corresponds to a unique firm-individual-year combination. The observation includes an identifier that corresponds to the employee and an identifier that corresponds to the establishment. For each observation, we have information on the number of days during the calendar year the individual worked at the establishment, as well as the full-time/part-time/intermittent/at-home work-status of the employee. Each observation also includes, in addition to the variables listed above, the sex, month, year and place of birth, occupation, total net nominal earnings during the year and annualized gross nominal earnings during the year for the individual, as well as the location—region, *département* ("district"), and town[13]—and industry of the employing establishment. There is no information on workers' education in the data and we could hardly use the Census data to get information on education as Abowd, Kramarz and Margolis did because the last available Census was done in 1990 and many of the laborers of 1996-1998 were still in school in 1990.

Beside this (already pretty huge) sample of workers (about 120,000 workers) that we follow over three years, we also have access for each of these three years to the exhaustive data on employers' reports of all salaried workers of the *Ile-de-France* region. The exhaustive panel cannot be constructed because the individual indices were dropped from the exhaustive data by INSEE for confidentiality reasons. We nevertheless use the exhaustive data to compute in 1997 for all establishments around Paris having to report to the French Tax Administration the mean sizes, mean log-wages and log-wage variances at the establishment level, for seven categories of occupations (managers, administrative staff, manual workers, etc.). In order to limit measurement error and to produce distributions which are as close as possible to steady-state, we have chosen to retain only those employees working full-year (not necessarily full-time) in 1997. We thus get rid of short-term employment contracts which might respond to a different behavioral logic than the one governing the theory put forward in this paper. This provides a cross-sectional sample of firm data involving a total of just over 3 million workers. In order to reduce the computational cost of the non parametric estimations we are going to use, we round

[13] The *région* is Ile-de-France, and it comprises 8 *départements* (Paris, Seine et Marne, Yvelines, Essonne, Hauts de Seine, Seine-Saint-Denis, Val de Marne, Val d'Oise).

mean log wage values after the third digit to create repeated observations. We also trim one percent of the data at the two extremes of the cross-worker mean log wage distribution to get rid of exceptionally low or high wage values.

3.2 Descriptive analysis of the data

We start the descriptive analysis with a look at worker mobility patterns. The panel sample provides individual wage bills reported by the employers on a yearly basis. We know for example that worker i was employed by establishment j during d days in 1996 within a time interval beginning this day of 1996 and ending that day of 1996. A trajectory featuring an employer change may be such that the end of one employment spell does not coincide with the beginning of the next one, and a worker may also leave the panel before the end of the recording period. There is no way of knowing the status of the worker during such periods not covered by a wage statement. He/she may have permanently or temporarily quit participating, or be unemployed, or have found a job in the Public Sector, or have started up his/her own business. In the estimation, we shall interpret temporary attrition as resulting from layoffs and permanent attrition as resulting from either layoffs or retirements. Moreover, we arbitrarily define a *job-to-job mobility* as an employer change with an intervening unemployment spell of less than 15 days.

Table 1 reports some statistics about worker mobility. It shows that, depending on the occupational category, 42 to 55 per cent of the workers stayed in the same job over the entire recording period of 3 years, while only 5 to 23 per cent changed jobs without passing through a period of unemployment. Job-to-job mobility therefore appears to be rather limited in this period, which corresponds to the end of a recession, in spite of the fact that job-to-job mobility (and worker mobility in general) is usually found much more substantial around Paris than in the rest of France. Concerning the mobility between employment and non-employment, the sample mean employment duration (which is censored at 3 years) is close to 2 years for all worker categories, while the median of that same duration (not reported here) is above 3 years for all categories. The sample mean duration of non-employment lies between 12 and 14 months, while its median (not reported here) is close to one year for all categories.

<Table 1 about here>

To reassure ourselves that it is legitimate to consider the sole region *Ile-de-France* as a self-contained labor market, we can look at cross-regional worker mobility. Looking at the sequence of employer locations for all workers in the panel, we find that only 4.7 per cent of them leave *Ile-de-France* during the recording period. Cross-regional mobility is therefore extremely limited over the period considered, and we can safely ignore it.

Finally, we may want to look at the stability of our occupational categorization of workers. We use the loosest available classification (next to pooling all workers together in a single class), which contains 7 categories (see various Tables). Looking at how impermeable those categories are, we found that in total, 81.3 per cent of the workers do not change category over the recording period, and close to 4 per cent change more than once. A more detailed look at those mobility patterns showed that the mobility is notably due to skilled white collar becoming executives, and unskilled blue collar becoming skilled blue collar.

We now turn to a description of wage mobility. Table 2 displays some information about the wage changes experienced by workers after their first recorded job-to-job mobility. The nominal wages available in the data were deflated using the Consumer Price Index (+1.23% in 1996 and +0.7% in 1997). The reported statistics include medians and 5 selected points of the cdf of wage changes in the relevant population of workers. We see on that Table that, even though the median wage variation after a job-to-job mobility is practically always positive, between 36 and 55 per cent of workers changing jobs do it at the price of a wage decrease. This observation confirms our initial feeling that it was important to model a wage setting mechanism allowing for such wage cuts due to job changes.

<Table 2 about here>

Table 3 reports similar information about the wage changes experienced between January 1, 1996 and December 31, 1997 for workers who held the same job over this period. Indeed, we may have several wages recorded for the same individual in the same firm-establishment if the worker stays employed by one firm for more than one year. Unfortunately, there is no way to know exactly at which moment he/she experienced a wage increase if the daily wage reported one year is greater than the one reported the year before. As the Table shows, it frequently happens (around 30 per cent of the times, depending on worker categories) that real wages decrease from one year to the next even when the worker has not changed employers. Obviously, our model cannot deliver such downward wage changes. They may reflect fluctuations of bonuses with the firm's activity since there is no way of separating contractual wages from bonuses which in some cases may be a non negligible share of salaries. Wage changes may also reflect occupation changes within the same establishment and compensating differentials. These wage fluctuations could be captured in the model in an *ad hoc* way by a pure idiosyncratic shock. Nevertheless, we prefer to estimate the structural model as it was laid out in the preceding sections at the price of a lack of fit because our main goal here is precisely to evaluate the capacity of the structural model to reproduce the main features of the dynamics of wages. Incorporating productivity fluctuations into the model is certainly not a straightforward extension, as we know that it generates endogenous job destruction (see e.g. Mortensen and Pissarides, 1994).

<Table 3 about here>

3.3 Estimation procedure

The discrete nature of the data, the fact that all individual wage records are aggregated within each calendar year, implies a complicated censoring of the continuous-time trajectories generated by the theoretical model, which can hardly be described analytically in a likelihood framework. We bypass this difficulty by proceeding to separate estimations of the various equilibrium parameters of the model.

The multi-step estimation procedure relies on the assumption that there exists an observable firm variable y that is (strictly) positively related to p. We assume (and we shall check on this assumption) that we can take the within-firm average log-wage $E(\ln w|p)$ for y, i.e.

$$\begin{aligned} y &\equiv E(\ln w|p) \\ &= \ln p - \left[1+\kappa_1 \overline{F}(p)\right]^2 \cdot \int_b^p \frac{1+(1-\sigma)\kappa_1 \overline{F}(q)}{\left[1+\kappa_1 \overline{F}(q)\right]^2} \frac{dq}{q} \\ &= \ln p - \left[1+\kappa_1 \overline{F}(p)\right]^2 \cdot \int_{p_{\min}}^p \frac{1+(1-\sigma)\kappa_1 \overline{F}(q)}{\left[1+\kappa_1 \overline{F}(q)\right]^2} \frac{dq}{q} \\ &\quad - \left[1+\kappa_1 \overline{F}(p)\right]^2 \cdot \frac{1+(1-\sigma)\kappa_1}{(1+\kappa_1)^2} \left[\ln p_{\min} - \ln b\right], \end{aligned} \tag{12}$$

where $\sigma = \frac{\rho}{\rho+\delta+\mu}$ and $\kappa_1 = \frac{\lambda_1}{\delta+\mu}$. (See Appendix B for a proof.) [14]

We proceed in the five following steps: The first step uses the cross-section of establishment data to provide a non-parametric estimate of the cdf of the distribution of y in the worker population, $L(p(y)) \equiv Z(y)$ (say). Z is the cdf of the distribution of workers across firms as characterized by their mean log wage payment y.

The second step uses the individual panel together with the preceding estimate of Z to provide an estimate of transition rate parameters δ, μ, λ_0 and λ_1. We maximize the likelihood of the duration of the first employment spell, the occurrence of an earnings change, and the following out-of-sample spell (if there exists one), conditional on employers' y's.

The third step uses the firm data on mean log wages y together with the preceding estimates of $\kappa_1 = \lambda_1/(\delta+\mu)$ and Z to provide a semi-parametric estimate $\widehat{p}(y;\sigma)$ of p given y (the inverse

[14] How reasonable is this assumption? It is easy to show by differentiation of equation (12) that $E(\ln w|p)$ is locally increasing at p if and only if

$$\frac{dE(\ln w|p)}{dp} = -(1-\sigma)\kappa_1 \frac{\overline{F}(p)}{p} + 2\kappa_1 f(p)\left[1+\kappa_1 \overline{F}(p)\right] \cdot \int_b^p \frac{1+(1-\sigma)\kappa_1 \overline{F}(q)}{\left[1+\kappa_1 \overline{F}(q)\right]^2} \frac{dq}{q} > 0.$$

If there is a non-monotonicity problem, it can thus only be at the left end of the support of p (the negative contribution to the positivity of $dE(\ln w|p)/dp$ is proportional to a decreasing function of p: $\overline{F}(p)/p$). In particular, for $p_{\min}$,

$$\left.\frac{dE(\ln w|p)}{dp}\right|_{p=p_{\min}} > 0 \Longleftrightarrow p_{\min} f(p_{\min}) > \frac{1}{2} \cdot \frac{(1-\sigma)\kappa_1}{1+(1-\sigma)\kappa_1} \cdot \frac{1}{\ln p_{\min} - \ln b},$$

which implies that the left-tail of the sampling distribution of p must not be too thin. It will be verified if workers are sufficiently myopic (σ large), or if $(\ln p_{\min} - \ln b)$ is large enough, or if on-the-job turnover is limited (κ_1 small).

of $p \mapsto y = E(\ln w|p))$ for any value of $\sigma = \rho/(\rho + \delta + \mu)$ in $[0, 1]$.

The fourth step uses the firm data on the conditional log wage variance given mean log wage y to provide an estimate of σ and of the variance of the log of individual abilities $\ln \varepsilon$.

The fifth step uses the cross-section of individual wages (previously used to compute firm sizes, firm mean log wages and within-firm log wage variances) to provide a non-parametric (up to parameters κ_1 and σ) estimate of the distribution of individual log abilities $\ln \varepsilon$ (when $V \ln \varepsilon$ is estimated positive).

Each step uses the results of the preceding one. Estimation errors are thus passed on and this should be taken care of properly. Now, the complexity of the whole procedure renders the computation of appropriate standard errors intractable. Fortunately, the huge sizes of the samples we use for inference in this work legitimate the claim that neither efficiency nor asymptotic standard errors are a problem we have to worry about.[15]

We now describe each estimation step in detail.

Step 1: Non parametric estimation of the sampling distribution $F(p(\cdot))$. As will become clear below, knowledge of the sampling distribution of firms F is essential in every step of the estimation. We therefore need to construct an empirical counterpart of F before we move on to the estimation of the rest of the model.

F is given by equation (6) as a function of the flow parameter κ_1 and the cdf of firm types (mpl's) in the population of workers, i.e. L. The fact that we do not observe p is therefore *a priori* problematic. However, looking at equation (6), we realize that the only thing that matters in the definition of $F(p)$ is the *ranking* of the type-p firm in the population of workers.

[15] Another reason why we prefer to use this multi-stage estimation procedure instead of a more efficient synthetic, GMM-type estimation method is that we want to control which data is used to identify which parameter. For example, parameter κ_1 contributes to the likelihood of both durations and cross-sectional wages. We certainly do not want κ_1 to be used to help fit the wage data if it is structurally determining the point process of employment transitions. The theory tells us that there exists a mechanism which induces a determination of wage dispersion by κ_1. We claim that in such a case the right way of testing the specification is to verify after estimating κ_1 from durations whether such an estimate of κ_1 also provides a good fit to the wage data, and not to proceed to an efficient GMM specification test which does not allow to recognize the directions of causality implied by the theory.

Therefore, the cdf in the population of workers of any observed variable that is in a one-to-one relationship with the p's can be used as an empirical counterpart of $L(p)$.

Provided that p is related to y through an increasing function $p(y)$, the cdf of average log earnings (denoted by $Z(y)$) in the population of workers equals the cdf of firm types in that same population, i.e. $Z(y) = L(p(y))$, and the sampling distribution $F(\cdot)$ can be redefined from equation (6) as

$$1 + \kappa_1 \overline{F}(p(y)) = \frac{1+\kappa_1}{1+\kappa_1 Z(y)}. \tag{13}$$

From a cross-sectional sample of observations on firms' mean log wage costs, $y_1, ..., y_N$, we estimate $Z(y)$ by integration of a normal kernel density estimator.[16]

Step 2: Estimation of the transition parameters δ, μ, λ_0 and λ_1. The recording period starts at time 0 (namely January 1st, 1996) and ends at time T (namely December 31st, 1998). All the N sampled individuals are employed at the beginning of the observation period. Define d_{i1} as the length of individual i's first employment spell, i.e. the amount of time this individual stays at his/her first employer. If the spell ends before the end of the recording period T, and if it is not immediately followed by another employment spell in a different establishment (job-to-job transition), let d_{i2} denote the length of the period spent out of the survey (in unemployment, inactivity, self-employment or the Public Sector) before a possible reentry. An individual initially present in the panel may therefore be in one of the following four situations:

1. The first employment spell is censored: $d_{i1} = T$;
2. The first employment spell is not censored ($d_{i1} < T$), and ends with a job-to-job transition: $d_{i2} = 0$;

[16] In the summation of mean log wages y_j we weight each observation y_j by the number of workers employed by an establishment with mean log wage y_j. We also work with a "firm"-dataset of reasonable size by rounding first mean log wages to the third digit and "collapsing" all establishments with the same value of y_j, as the theory suggests we should do.

3. The first employment spell is not censored ($d_{i1} < T$), does not end with a job-to-job transition, and the subsequent attrition period is censored: $d_{i2} = T - d_{i1}$;

4. The first employment spell is not censored ($d_{i1} < T$), does not end with a job-to-job transition, and the subsequent attrition period is not censored: $0 < d_{i2} < T - d_{i1}$.

Moreover, as was already mentioned, wages do not vary continuously over time and the administrative data give no clue as to exactly when promotions take place. Under the model's assumption, however, yearly wages cannot decline unless the worker changes employers. Then, if two subsequent yearly wage declarations by the same employer for the same worker significantly differ from one year to the next, then it must be that at least one contact was made by the worker of an alternative employer which was productive enough for his/her current employer to grant the employee a wage rise. Now, let n_i be the number of recorded wage rises within the period of time d_{i1}. If $d_{i1} \leq 1$, then $n_i = 0$ with probability one; if $1 < d_{i1} \leq 2$, then n_i is either 0 or 1; if $d_{i1} > 2$ then n_i can be either 0, 1 or 2; etc... It is difficult to derive the distribution of n_i (whatever conditional on) when $d_{i1} > 2$. Fortunately, it is rather easy to calculate the probability of $n_i = 0$ given $d_{i1} = d$ and given the employer's type in the first spell p_i. It is the expected value of the probability of $n_i = 0$ given d_{i1} and p_i and given the unobserved worker type-ε and initial wage w (i.e. at the onset of the recording period), that is the expected value of $\exp\left\{-\lambda_1 \left[\overline{F}\left(q\left(\varepsilon, w, p\right)\right) - \overline{F}\left(p\right)\right] d\right\}$ with respect to ε and w.

We estimate δ, μ, λ_0 and λ_1 by maximizing the likelihood of the N observations $(d_{i1}, \mathbf{1}\{n_i = 0\}, d_{i2};\ i = 1, ..., N)$ conditional on the observed indicator of the first employer's type '(the average log earnings y_i).[17] Detailed derivation of the likelihood is an algebraically rather tedious, although not particularly difficult exercise. It is carried out in Appendix C.

Step 3: Estimation of p and $\ln p_{\min} - \ln b$ given y and ρ. Steps 1 and 2 provide estimates of distribution Z and the transition parameters (in particular $\kappa_1 = \frac{\lambda_1}{\delta+\mu}$). From now on, we shall

[17] Note that we could use more in- and out-of-sample spells than the first two.

thus consider Z and κ_1 known. We then construct an estimator of the marginal productivity of labor (p_j) for each firm j from its observed mean log wage y_j as follows. We start with the expression (12) of a given firm type's mean wage. Substituting $p(y)$ for p in expression (12) we have:

$$\ln p(y) = y + \left[1 + \kappa_1 \overline{F}\left(p(y)\right)\right]^2 \cdot \int_b^{p(y)} \frac{1}{1 + \kappa_1 \overline{F}\left(x\right)} \cdot \left[1 - \sigma \frac{\kappa_1 \overline{F}\left(x\right)}{1 + \kappa_1 \overline{F}\left(x\right)}\right] \frac{dx}{x}. \tag{14}$$

Differentiating once w.r.t. y, using equality (13) to substitute $Z(y)$ for $F(p(y))$, we get, after some rearrangements:

$$\frac{2}{(1-\sigma)} \cdot \frac{\overline{Z}'\left(y\right)}{\overline{Z}\left(y\right)} \cdot \left[\ln p\left(y\right) - y\right] + \left[\frac{p'(y)}{p(y)} - 1\right] = -\frac{1+\kappa_1}{\kappa_1 \overline{Z}\left(y\right)} \cdot \frac{1 - \sigma \frac{\kappa_1}{1+\kappa_1} \overline{Z}\left(y\right)}{1-\sigma}. \tag{15}$$

At the maximum observed value of y, say $y_{\max} = E(\ln w | p_{\max})$, for which $\overline{Z}\left(y_{\max}\right) = 0$, the above equation implies the following initial condition:[18]

$$\ln p\left(y_{\max}\right) = y_{\max} + \frac{1+\kappa_1}{2\kappa_1 Z'\left(y_{\max}\right)}. \tag{16}$$

Given this initial condition, equation (15) then solves as:

$$\ln p(y) = y + \frac{1+\kappa_1}{\kappa_1} \overline{Z}\left(y\right)^{-\frac{2}{1-\sigma}} \cdot \int_y^{y_{\max}} \frac{\overline{Z}\left(t\right)^{\frac{1+\sigma}{1-\sigma}}}{1-\sigma} \cdot \left[1 - \sigma \frac{\kappa_1 \overline{Z}\left(x\right)}{1+\kappa_1}\right] dx. \tag{17}$$

Equation (17) can be used to predict a value for p given a value of y for any given value of $\sigma = \frac{\rho}{\rho+\delta+\mu}$.

That equation (14) implies the following relation between $p_{\min}$, $y_{\min} = E(\ln w | p_{\min})$ and $\ln b$:

$$\ln p_{\min} = y_{\min} + (\ln p_{\min} - \ln b)(1 + \kappa_1(1-\sigma)).$$

(Set $y = y_{\min}$ and $p(y) = p_{\min}$ in (14)). One can thus deduce an estimate of $\ln b$ from the observed minimum mean log wage $\widehat{y}_{\min}$ using $p\left(\widehat{y}_{\min}\right)$ to estimate $p_{\min}$ (conditional on σ).

Now this procedure offers no guaranty of a good fit of observed mean log wages $y_1, ..., y_N$ when one uses formula (14) to predict y_j, $i = 1, ..., N$, from $p(y_j)$. This happens in particular

[18] Which holds true only if $p'(y)$ is not infinite at $y_{\max}$, or equivalently if $f(p_{\max})$ is non zero. But it must be the case that $f(p_{\max}) \neq 0$, otherwise the type $p_{\max}$ firms would employ no worker.

if the estimated function $p(y)$ is not exactly everywhere monotonic or if measurement errors affect the recorded minimum mean log wage $\widehat{y}_{\min}$.[19] One can alternatively regress

$$\ln p(y_j) - y_j - \left[1+\kappa_1 \overline{F}\left(p(y_j)\right)\right]^2 \cdot \int_{p_{\min}}^{p(y_j)} \frac{1+(1-\sigma)\kappa_1 \overline{F}\left(q\right)}{\left[1+\kappa_1 \overline{F}\left(q\right)\right]^2} \frac{dq}{q}$$

against $\frac{1+(1-\sigma)\kappa_1}{(1+\kappa_1)^2}\left[1+\kappa_1 \overline{F}\left(p(y_j)\right)\right]^2$ to obtain an estimate of $\ln p_{\min} - \ln b$. We use weighted OLS, weighing each firm observation by its size in order to maximize the fit to worker data rather than firm data.

Although we have used this second method in practice, we have also checked in the application that the two procedures give very similar results.

Step 4: Estimation of ρ and the variance of the log of individual abilities from a cross-section of wages. The preceding step allows to estimate p given y up to a predefined value of ρ (say $p(y;\rho)$ to emphasize the dependence of $p(y)$ on ρ). The parameter ρ remains to be estimated, together with the parameters of the distribution of individual heterogeneity ε[20] and the opportunity cost of employment b. We use the second-order moments $V(\ln w|p)$ to provide an estimation of these three parameters.[21] Specifically, we obtain consistent estimates of $V\ln\varepsilon$ and ρ using weighted non-linear least squares, by regressing the within-firm empirical variance of the set of log-wages in any firm on the firm's mean log-wage y. We use the firm sizes (the number of employees of a given occupation) as regression weights to take into account the variable precision of each firm's empirical log-wage variance (of order one over the sample size). Practically, we re-run step 3 over for each value of σ in a grid of step size 1% over $[0,1]$. We select the value of σ which yields the minimal value of the weighted standard deviation of

[19] In various experiments it occurred that $p(y)$ was even decreasing in the lowest range of y.

[20] As we mentioned at the beginning of this Section, the entire estimation is conducted under the normalization assumption that the mean of $\ln\varepsilon$ is equal to zero. We therefore only focus on the estimation of the variance of $\ln\varepsilon$, which turns out to be of particular interest for the application of our model to the decomposition of log-wage variance. The complete non-parametric estimation of the distribution of $\ln\varepsilon$ is left to the next paragraph.

[21] In the sequel, we shall refer to the conditional variance $V(\ln w|p)$ as the *within-firm log wage variance*. This may sound somewhat abusive, since what $V(\ln w|p)$ really measures is the variance of log wages in *all* firms with productivity p. Since for the sake of computational tractability we have estimated the p's by rounding the mean log wages of firms to the third digit, it necessarily turns out that somewhat different firms—particularly in terms of their size—have equal p's. This conditional heterogeneity is implicitly ignored when we speak of the conditional variance as the within-firm variance.

the difference between the observed within-firm log wage variance and the predicted one. A consistent estimate of $V \ln \varepsilon$ is obtained by computing the weighted mean of this difference.[22] The specific form of the theoretical within-firm log wage variance is derived in Appendix D.

Note that the reliability of the thus estimated value of $V \ln \varepsilon$ clearly hinges on how well the model does in predicting the within-firm wage variance. Because apart from this last step, the model was entirely estimated using only first-order moments of the within- and between-firm wage distributions, one may be dubious about the chances that a second-order moment like this be well matched. We shall therefore take a careful look at this particular issue when we expose the corresponding estimation results.

Step 5: Estimation of the density of workers' abilities. If the preceding estimation yields a positive estimate of $V \ln \varepsilon$ then the distribution of individual abilities is non degenerate and one can obtain an estimate of the whole distribution of ε by using the non-parametric deconvolution method of Stefanski and Caroll (1990).

Section 2.3 has demonstrated that the cross-sectional distribution of wages was equal to the distribution of $\phi(\varepsilon, q, p) = \varepsilon\phi(1, q, p)$ with ε and (q, p) independently distributed. It thus follows that the cross-sectional distribution of log wages is equal to the convolution of the cross-sectional distribution of $\ln \varepsilon$ and the cross-sectional distribution of $\ln \phi(1, q, p)$. In practice we proceed as follows: for any wage observation w_i for a worker i working in a set $j(i)$ of firms with same value of mean log wage $y_{j(i)}$ (we aggregate all firms with values of y equal up to 3 digits). For each y_j we compute $p_j = p(y_j)$ using the estimates of the previous steps. The distribution of $p_{j(i)}$ across workers is clearly equal to $\ell(\cdot)$ (neglecting the estimation errors). One then uses the algorithm detailed in section 2.3 to draw a value of q_i for each i and $p_{j(i)}$. At this stage we have an individual sample $(\ln w_i,\ \ln \phi_i,\ i = 1, ..., n)$ of draws of log wages and logged values of

[22] We naturally constrain $V \ln \varepsilon$ to be nonnegative.

$\phi(1, q, p)$. We obtain a estimate of the density of $\ln \varepsilon$ at any point x as

$$\widehat{h}_{\ln \varepsilon}(x) = \frac{1}{2\pi} \int_{-1/\lambda}^{1/\lambda} \chi(t) e^{-itx} dt, \tag{18}$$

where $\chi(t)$ is the ratio of the empirical characteristic functions of w_j and ϕ_j:

$$\chi(t) = \frac{\frac{1}{n} \sum_{j=1}^{n} \exp\left(it \ln w_j\right)}{\frac{1}{n} \sum_{j=1}^{n} \exp\left(it \ln \phi_j\right)},$$

and where the bandwidth λ is obtained as a zero of

$$I(\lambda) = \frac{n}{n-1} \cdot \left[2 - (n+1) \left| \frac{1}{n} \sum_{j=1}^{n} \exp \frac{i \ln w_j}{\lambda} \right|^2 \right].$$

We also estimate the cdf of $\ln \varepsilon$ to simulate cross-sections of log wages $\ln \widehat{w} = \ln \varepsilon + \ln \phi(1, q, p)$. It is sufficient for that to integrate e^{-itx} in (18) with respect to x:[23]

$$\widehat{H}_{\ln \varepsilon}(x) = -\frac{1}{2\pi} \int_{-1/\lambda}^{0^-} \chi(t) \cdot \frac{e^{-itx}}{it} dt - \frac{1}{2\pi} \int_{0^+}^{1/\lambda} \chi(t) \cdot \frac{e^{-itx}}{it} dt + \widehat{H}_{\ln \varepsilon}(\ln \varepsilon_{\min}).$$

To draw random values of $\ln \varepsilon$ in distribution $\widehat{H}_{\ln \varepsilon}$ we draw uniform numbers in $[0, 1]$ and transform them by the inverse of $\widehat{H}_{\ln \varepsilon}$.

3.4 Estimation results

3.4.1 Transition rates

We first report the estimated transition parameters in Table 4 below. Layoffs and reemployment rates vary with skills as expected. Layoffs occur on average every 10 to 15 years and unemployment lasts between 6-8 months. Attrition is a rare event (once every 65 years for unskilled blue collars who display the highest rate!). Surprisingly, the arrival rate of alternative offers vary relatively little with the worker category. On an average, employees are solicited by 'poachers' every 16-19 months.

<Table 4 about here>

[23] It is straightforward to show that the integrals are well defined in the vicinity of $t = 0$.

3.4.2 Productivity estimates

Knowledge of the transition parameters for each category of labor allows us to apply the estimator (17) derived in Paragraph 3.3 to our cross section of firm data from which the mean log wage earned by each category of worker within each firm is available.

The resulting estimated mpl's are plotted on Figure 1 against the corresponding mean log wage ($y = E(\ln w|p)$). The vertical lines indicates the 10th, 25th, 50th, 75th and 90th percentiles of the distribution of y or p in the population of workers. Estimates of $\ln p_{\min}$ and $\ln b$ are given in the first two columns of Table 5. Our initial assumption that b is always less than $p_{\min}$, which implies that any type of firm can potentially hire any type of worker, holds true in the data.

We first check that labor productivity is an increasing function of mean log wages. Then looking at how the 45° line divides the area below the productivity curve, one sees that the profit share of value-added is not a monotonic function of labor productivity, except for the two categories of lowest-skilled workers. Finally, the slope of the productivity curve is particularly steep at the right tail of the distribution. This happens because mean log wages are particularly dispersed in the upper part of the distribution, which in turn implies very small values of the density and correspondingly high productivity estimates (see equation (16) to see why).

<Figure 1 about here>

3.4.3 Discount rates

The next step is to estimate the discount rate ρ using the marginal productivities estimates obtained above (which are conditional on ρ) and following the procedure detailed in Paragraph 3.3. The results are gathered in Table 5. Column 4 shows the estimated values of the discount rate (and column 3 the estimate of σ). In general, workers show a strong impatience rate, yet increasingly strong as the amount of 'sophistication' incorporated in the profession decreases. By and large, we estimate that the first four categories discount between 40 and 50% of each additional year, sales and services employees 60% and unskilled blue collars 80%. These high discount rate values might reflect the fact that workers are more risk averse than is implied by the log utility assumption. More risk averse agents would be less likely to accept to trade income today for higher income prospects tomorrow, which is exactly what a greater discount rate also implies. It is therefore difficult to empirically distinguish the degree of concavity of the utility function from the amount of time discounting.

<Table 5 about here>

3.4.4 Within-firm log-wage variance

Estimating $V \ln \varepsilon$ so as to minimize the distance between the variance of actual wages and the predicted variance, we compute an estimate of the within-firm variance that we compare to the observed one on Figure 2. It is first plain clear that the data are heteroskedastic and that the conditional log wage variance appears to be an increasing function of mean log wages. This is *per se* a very interesting result, the implications of which we shall discuss at length in the next paragraph. For now, it is clear that the model definitely picks the right overall correlation *and* the right magnitudes.

That the conditional log wage variance shows an increasing trend against $E(\ln w|p)$ for all categories of labor is not an unexpected result. One indeed typically expects the distribution of wages in large-p firms (or, equivalently, in firms with large mean wages) to be more dispersed than the distribution of wages in smaller firms both because they offer lower wages to unemployed workers, due to increased monopsony power, and because they can poach the employees of the less productive firms by offering them higher income prospects. That the magnitudes are also correct is a remarkable result if one remembers how few free parameters were estimated to fit within-firm variances (ρ and $V \ln \varepsilon$), all other parameters being estimated so as to provide a perfect fit to within-firm mean log wages (the infinite dimensional function $p(y)$ and b).

Nonetheless, the predicted conditional variance shows undulations which do not exist in the data and tends to overshoot its target for high p's (especially for the last four worker categories). This might indicate that the assumption of independence between ε and p is too strong. A model allowing for some extent of *ex ante* worker selection by firms and predicting that the within-firm variance of worker abilities is a hump-shaped function of firm productivity would certainly provide a better fit.

<Figure 2 about here>

3.4.5 Individual log-wage variance decomposition

A decomposition of the total variance of workers' log wages arises naturally from our model. This decomposition is into three components: between-firm variance (firm effect), the variance of log abilities (individual effect) and the within-firm residual reflecting market frictions. Specifically, we write

$$V(\ln w) = EV(\ln w|p) + VE(\ln w|p) \tag{19}$$

$$= \underbrace{V \ln \varepsilon}_{\substack{\text{Individual} \\ \text{effect}}} + \underbrace{(EV(\ln w|p) - V \ln \varepsilon)}_{\substack{\text{Effect of market} \\ \text{frictions}}} + \underbrace{VE(\ln w|p)}_{\substack{\text{Firm} \\ \text{effect}}}. \tag{20}$$

(See equation (37) in Appendix D for a precise computation of each term.)

<Table 6 about here>

The log-wage variance decomposition is reported in Table 6. We obtain a remarkable result: individual ability differences explain about 50% of the log wage variance for managers and engineers, 20% for workers with lower executive functions, about 15% for technicians and technical supervisors and virtually nothing for the other categories. It therefore seems that the more sophisticated the profession is, the more difficult it is to predict the efficiency of a worker given his observable attributes. To put it differently, the more skill-intensive an occupation is, the more heterogeneous is the category of workers who can apply to it. At the bottom of the skill hierarchy, manual workers and employees are rather homogeneous as far as productive efficiency is concerned.

Another interesting result is that once the person effect has been removed, firm effects and search frictions explain approximately identical parts of the residual variance.

As a matter of comparing our results to those of previous contributions, again we should cite Abowd *et al* (1999), and Abowd and Kramarz (1999), who use the same data as we do and find over the whole sample, controlling for observed skill characteristics, that the person effect accounts for more or less 50% of total log wage variance. Even though we ran separate estimations for each skill category, our results make it clear that the average weight of the person effect over the whole sample is by far less than a half. Why this discrepancy? We believe heteroskedasticity is the crux here. What equation (20) tells us is that the expectation decomposition of $\ln w$ into the sum of a worker effect $\ln \varepsilon$ and a firm effect $E(\ln w|p)$ does not fully account for the contribution of firm heterogeneity to wage dispersion, as it omits the conditional heteroskedasticity which appears in the data (see Figure 2) and is generated by our model from the wage mobility induced by job search. As was discussed and observed in the previous paragraph, large p-firms have more dispersed conditional wage distributions than small p-firms. On the other hand, due to the cross-sectional orthogonality of firm and worker heterogeneity parameters, worker heterogeneity contributes to wage dispersion in the same way at all firms. Assuming homoskedasticity, as Abowd *et al* do, amounts to attributing all the within-firm wage variance component to worker heterogeneity and therefore results in an overestimation of the contribution of worker heterogeneity to wage dispersion.

3.4.6 Cross-sectional earnings distributions

All the parameters of the model being now estimated, we can simulate the model and compare the actual distributions of earnings to the predicted ones. Figure 3 provides, for each of the seven professions we consider, the graphs of the quantile functions for the distribution of individual (log) wages and the distribution of $\phi(1, q, p)$ when (p, q) is distributed as explained in paragraph

2.3 of the theoretical part of the paper. The last distribution is the distribution predicted by the model when there is no dispersion of abilities. Otherwise, the distribution of log wages is equal to the convolution of the distribution of $\phi(1,q,p)$ with the distribution of $\ln\varepsilon$, i.e. H. Figure 4 plots the deconvolution results using the method described in paragraph 3.3 for the first four categories.[24] There is not much to say about them except that, as it should be given the preceding estimates of $V\ln\varepsilon$, the distribution for the first group of workers is flatter than that for the second group which is itself flatter than the last two. Note that the right tail of the distribution of $\ln\varepsilon$ is also thicker for the first group.

<Figure 4 about here>

There is more to say about Figure 3. First, we observe a discontinuity in the quantile function for $\phi(1,q,p)$ which is entirely due to the gap between b and $p_{\min}$. It is plain clear that $\ln p_{\min} - \ln b$ is far too large for the distribution of wages offered to former unemployed to mix with the distribution of wages obtained from on-the-job search. The data seem to require heterogeneity in b as well as heterogeneity in ε to mix the lower part of the distribution of predicted wages et provide a better fit. We leave this extension to further work.[25]

Second, one sees why we estimate no ability dispersion for low skilled workers. The model with no worker heterogeneity works quite well to explain the dispersion of log earnings in this case. For skilled manual workers, the fit is good in the upper part of the distribution but bad in the lower part because of the wide wedge between $\ln p_{\min}$ and $\ln b$.

For the first four categories of more skilled workers, the actual distribution of wages dominates the predicted one (with not worker heterogeneity) in the upper part of the distribution. This demonstrates the necessity of allowing for heterogeneous abilities, which we now do. As was explained in paragraph 3.3, the deconvolution method can also deliver the cdf's of $\ln\varepsilon$,

[24] Our attempts at retrieving a non-degenerate distribution of $\ln\varepsilon$ for the remaining three categories were unsuccessful, as expected given that their estimated variance of $\ln\varepsilon$ was 0.

[25] Heterogeneity in b is present in the theoretical model that we constructed in a previous paper Postel-Vinay and Robin (1999).

from which we can get random draws of worker abilities and thus simulate a complete cross-section of wages following the algorithm described in paragraph 2.3 (in fact, all we have to do at this point is add a cross section of $\ln \varepsilon$'s randomly selected from H to the cross section of $\phi(1, q, p)$'s already simulated). The predicted cross-worker log-wage densities and cdf's are plotted together with the observed ones on Figure 5. We see that the fit is almost perfect, except at the left end of the wage distribution. This again points to the need of some heterogeneity in the workers' 'at-home' productivity parameters, b.

<Figure 5 about here>

3.4.7 Recruiting effort, productivity and firm size

As we argued when exposing the basic assumptions of our theoretical model, our specialization of an unconstrained 'sampling density' $f(\cdot)$ and its relationship to that of firm types in the population of firms $\gamma(\cdot)$ potentially carries some information about the process through which firms and workers are matched. More precisely, we saw that the sampling weights $f(p)/\gamma(p)$ of firms by workers in the search process could be interpreted as the average flow of 'help-wanted ads' or 'job vacancies' posted by type p firms per unit time. Broadly speaking, those sampling weights provide a measure of the average effort put into hiring by type p firms.

Formally, an expression of $f(p)/\gamma(p)$ is readily available from equation (7):

$$\frac{f(p)}{\gamma(p)} = \frac{\left[1 + \kappa_1 \overline{F}(p)\right]^2}{1 + \kappa_1} \cdot \frac{\ell(p)}{\gamma(p)}. \tag{21}$$

The densities $\ell(p)$ and $\gamma(p)$ of firm types respectively in the populations of workers and firms are estimated using a normal kernel. The estimated sampling weights are plotted against p on Figure 6, together with the mean firm sizes $\ell(p)/\gamma(p)$, in log-coordinates.[26]

<Figure 6 about here>

[26] It is important to note at this point that there is conditional heterogeneity in firm sizes within each firm type p. As not all type p firms have equal sizes, the thus estimated hiring efforts and firm sizes are conditional mean values.

The most obvious result, which is robust across all categories of labor is that the sampling weights decrease with productivity: more productive firms devote less effort to hiring, which naturally makes them less efficient in contacting potential new employees. On the other hand, since they are also more attractive to workers, they are more efficient in retaining their employees and attracting the workers that they do contact. Those two counteracting forces sum up to a non monotonic effect on mean firm size, which is generally a hump-shaped function of firm type: low-p firms do not fully compensate their lack of competitiveness in the Bertrand game by their higher recruiting effort, while high-p firms are not among the largest in spite of their attractiveness because they contact too few workers.

Those results bring about a comment. The common usage in the job search literature is to assume a particular 'matching technology' that precisely connects the sampling distribution to the distribution of firm types. Two extreme benchmark cases are the assumption of *random matching* (all firms have an equal probability of being sampled, implying $f(p) = \gamma(p)$; see Burdett and Mortensen, 1998, among others), and that of *balanced matching* (the probability of being sampled is proportional to firm size, implying $f(p) = \ell(p)$; see Burdett and Vishwanath, 1988)[27]. We stand somewhere in between those two extremes, which are a priori both encompassed by our more general assumption.

Given our estimated relationship between firm hiring efforts and sizes, we may find it interesting to assess which one of the above two assumptions is closest to our more general model's predictions. Mean firm sizes are plotted against hiring efforts on Figure 7 (again in log-coordinates). Even though the graph shows stark cross-firm differences in the amount of effort put forth for hiring, those differences are not in a monotonic relationship with size. We thus clearly reject both assumptions of random and balanced matching,[28] and rather plead in

[27] A slightly different approach also exists, which consists in allowing firms to decide upon an endogenous 'search effort' that increases their visibility. This idea is borrowed from matching models (see e.g. Pissarides, 1990).

[28] Note that balanced matching in a strict sense—i.e. $f(p) = \ell(p)$—is obviously incompatible with equation (21). This is because, as can be seen from equation (5), balanced matching is in fact incompatible with non trivial productivity dispersion. This is the translation into our model of the original findings of Burdett and

favor of differentiated search efforts put forth by the various firm type—and even *within* each firm type, given the conditional heterogeneity of firm sizes. A deeper look into the 'job vacancy posting' behavior of firms is on our research agenda.

<Figure 7 about here>

3.4.8 Dynamic simulations

The last thing that we do in this paper is to proceed to the most severe specification test we could think of, which is looking at how good (or bad) the model is at predicting wage mobility along the line of Tables 2 and 3 that we have already commented. Tables 7 and 8 display the results of a dynamic simulation of 10,000 trajectories for each professional category. The major discrepancy between actual and simulated data is that the model does not do well (to say the least) in predicting downward wage mobility. We produce rather good upward wage mobility predictions for workers changing employers (the last two columns of Tables 2 and 7 are quite close). Yet, the simulations are clearly not as good for those workers holding the same job over the one year simulation period since we predict too few downward and upward wage changes.

<Table 7 about here>

<Table 8 about here>

We believe that these differences are genuine. They are not (or not only) the sign of an inefficient estimation method, mostly based on cross-sectional analysis. On the contrary, we view the fact that we make as little use of the panel as possible to estimate the model as an advantage, allowing for rigorous specification testing (even if a precise statistical test is missing). We think that these results reflect the lack of idiosyncratic shocks on labor productivity in the model. About 15 to 20% of earnings are bonuses which are indexed on the firms' performances. They can naturally (partly) be explained by moral hazard considerations but they are also

Vishwanath (1988), that the non degenerate wage dispersion result of Burdett and Mortensen (1998) collapses under balanced matching.

likely to reflect fluctuations in firms' outcomes. Yet another item to add to the research agenda is thus an extension of the model to allow for idiosyncratic productivity shocks, maybe along the lines of Mortensen and Pissarides (1994).

4 Concluding remarks

The main contribution of this paper is an investigation of the properties of the distribution of wages within an equilibrium job search model with on-the-job search, using matched employer and employee data. The theoretical model features heterogeneous productivity attributes for both firms and workers, and an original wage setting mechanism that departs from the conventional alternative assumptions of wage posting or wage bargaining. The model fits the data well and provides new results about the decomposition of log-wage variance. Our structural model delivers a clear economic interpretation for three components of the log wage variance: a firm effect, a person effect and an effect of labor market frictions. Taking this explicit account of the role of labor market frictions in the determination of wage dispersion, we find that the share of wage variance explained by the heterogeneity in unobserved worker productive ability is in general much smaller than what is found in earlier analyses of the same panel. The weight of the labor market friction effect in total wage variance approximately equals that of the pure firm effect and varies from 22 to 50 p. cent of total wage variance, depending on the labor category considered.

The model's success at fitting the data and passing specification tests is overall satisfactory. A few problems nonetheless persist, which are suggestive of leads that should perhaps be pursued in future research. One problem is the less-than-perfect fit to the observed within-firm log wage variance (see Figure 2 and paragraph 3.4.4). We believe this could be fixed by allowing for differences in the within-firm distributions in personal abilities, i.e. in other words, for some amount of sorting. Tighter links between the job search approach and the job assignment literature are surely worthwhile looking at. Secondly, the last paragraph pointed to a general lack of downward wage mobility in the individual wage paths simulated by the model. A possible remedy for this would probably be to incorporate isiosyncratic productivity shocks into the picture, as is done in the matching literature (Mortensen and Pissarides, 1994). This addition would bear the additional advantage of endogenizing layoffs. Finally, an important shortcoming of the model is that it keeps the "vacancy posting" behavior of firms exogenous, even though it draws a non trivial picture of this behavior. Going into the direction of an endogenous recruitment effort, as is done in Mortensen's (1998, 1999) attempts at merging the job search/wage posting and the job matching literatures will certainly provide interesting new insights on this central issue.

References

[1] Abowd J. M. and F. Kramarz (1999) "Inter-industry and Firm-size Wage Differentials in France and the United States", mimeo.

[2] Abowd J. M., F. Kramarz and D. Margolis (1999), "High Wage Workers and high Wage Firms", *Econometrica*, **67** (2), 251-334.

[3] Abowd J. M., H. Finer and F. Kramarz (1999), "Individual and Firm Heterogeneity in Compensation", in Haltiwanger, Lane, Spletzer, Theeuwes, eds, *The Creation and Analysis of Employer-Employee Matched Data*, North-Holland.

[4] Albrecht, J. W. and B. Axell (1984), "An Equilibrium Model of Search Unemployment", *Journal of Political Economy*, **92** , 824–840.

[5] Bontemps, C., J.-M. Robin and G. J. Van den Berg (1999), "An Empirical Equilibrium Job Search Model with Search on the Job and Heterogeneous Workers and Firms", *International Economic Review*, **40** (4), 1039-75.

[6] Bontemps, C., J.-M. Robin and G. J. Van den Berg (2000), "Equilibrium Search with Continuous Productivity Dispersion: Theory and Non-Parametric Estimation", *International Economic Review*, **41** (2), 305 -358.

[7] Burdett, K. and K. L. Judd (1983), "Equilibrium Price Dispersion", *Econometrica*, **51**, 955-970.

[8] Burdett, K. and D. T. Mortensen (1998) "Wage Differentials, Employer Size and Unemployment", *International Economic Review*, **39**, 257-73.

[9] Burdett, K. and T. Vishwanath (1988), "Balanced Matching and Labor Market Equilibrium", *Journal of Political Economy*, **96** (5), 1048-1065.

[10] Burdett, K. and R. Wright (1998), "Two-Sided Search with Nontransferable Utility", *Review of Economic Dynamics*, **1** (1), 220-245.

[11] Dey, M. S. and C. J. Flinn (2000) "An Equilibrium Model of Health Insurance Provision and Wage Determination", mimeo, University of Chicago and New York University.

[12] Eckstein, Z. and K. I. Wolpin (1990) "Estimating a Market Equilibrium Search Model from Panel Data on Individuals", *Econometrica*, **58** (5), 1001-44.

[13] Heckman, J. J. and B. E. Honoré (1990), "The Empirical Contents of the Roy Model", *Econometrica*, **58** (5), 1121-49.

[14] Heckman, J. J. and G. L. Sedlacek (1985), "Heterogeneity, Aggregation and Market Wage Functions: An Empirical Model of Self-selection in the Labor Market", *Journal of Political Economy*, **93** (6), 1077-1125.

[15] Heckman, J. J. and J. Scheinkman (1987), "The Importance of Bundling in a Gorman-Lancaster Model of Earnings", *Review of Economic Studies*, **54** (2), 243-55.

[16] Mortensen, D. T. (1998), "Equilibrium Unemployment with Wage Posting: Burdett-Mortensen Meet Pissarides", WP 98-14, CLS, University of Aarhus.

[17] Mortensen, D. T. (1999), "Modelling Matched Job-Worker Flow", mimeo.

[18] Mortensen, D. T. and C. A. Pissarides (1994) "Job Creation and Job Destruction in the Theory of Unemployment", *Review of Economic Studies*, **61**, 397-415.

[19] Mortensen, D. T. and C. A. Pissarides (1998) "New Developments in Models of Search in the Labor Market", mimeo, Northwestern University and LSE.

[20] Moscarini, G. (2000) "Excess Worker Reallocation", forthcoming *Review of Economic Studies.*

[21] Pissarides, C. A. (1990) *Equilibrium Unemployment Theory*, Oxford: Basil Blackwell.

[22] Postel-Vinay, F. and J.-M. Robin (1999) "The Distribution of Earnings in an Equilibrium Search Model with State-Dependent Offers and Counter-Offers", INRA-LEA WP. No. 98.04.

[23] Robin, J.-M. and S. Roux (1998) "Random or Balanced Matching: An Equilibrium Search Model with Endogenous Capital and Two-Sided Search", CREST-INSEE WP. No. 98.38.

[24] Roy, A. D. (1951), "Some Thoughts on the Distribution of Earnings", *Oxford Economic Papers (New Series)*, **3**, 135-46.

[25] Sattinger, M. (1993) "Assignment Models of the Distribution of Earnings", *Journal of Economic Literature*, **31** (2), 831-80.

[26] Stefanski, L. and R. J. Caroll (1990), "Deconvoluting Kernel Density Estimators", *Statistics*, **21**, 169-184.

[27] Van den Berg, G. J. and G. Ridder (1998), "An Empirical Equilibrium Search Model of the Labor Market", *Econometrica*, **66** (5), 1183-222.

Appendix

A Equilibrium wage determination

In this appendix we derive the precise form of equilibrium wages $\phi(\varepsilon, p, p')$.

The first step is to compute the value functions $V_0(\cdot)$ and $V(\cdot)$. Since offers accrue to unemployed workers at rate λ_0, $V_0(\varepsilon)$ solves the following Bellman equation:

$$(\rho + \mu + \lambda_0) \cdot V_0(\varepsilon) = U(\varepsilon b) + \lambda_0 \cdot E_F \{V(\varepsilon, \phi_0(\varepsilon, X), X)\},$$

where E_F is the expectation operator with respect to a variable X which has distribution F. Using definition (1) to replace $V(\varepsilon, \phi_0(\varepsilon, p), p)$ by $V_0(\varepsilon)$ in the latter equation then shows that:

$$V_0(\varepsilon) = \frac{U(\varepsilon b)}{\rho + \mu}. \tag{22}$$

We thus find that an unemployed worker's expected lifetime utility depends on his personal ability ε only through the amount of output he produces when engaged in home production, εb. This naturally results from the fact that their first employer is able to appropriate the entire surplus generated by the match until the worker gets his first outside offer. The only income the employer originally has to compensate the worker for is εb.

Now turning to employed workers, consider a type-ε worker employed at a type-p firm and receiving a wage $w \leq \varepsilon p$. This worker is hit by outside offers from competing firms at rate λ_1. If the offer stems from a firm with mpl p' such that $\phi(\varepsilon, p', p) \leq w$, then the challenging firm is obviously less attractive to the worker than his current employer since it cannot even offer him his current wage. The worker thus rejects the offer and continues his current employment relationship at an unchanged wage rate. Now if the offer stems from a type-$p' < p$ firm such that $w < \phi(\varepsilon, p', p) \leq \varepsilon p$, then the offer is matched by p, in which case the challenging firm p' will not be able to attract the worker but the incumbent employer will have to grant the worker a raise—up to $\phi(\varepsilon, p', p)$—to retain him from accepting the other firm's offer. This leaves the worker with a lifetime utility of $V(\varepsilon, \varepsilon p', p')$. Finally, if the offer originates from a firm more productive than p, then the worker eventually accepts the outside offer and goes working at the type-p' firm for a wage $\phi(\varepsilon, p, p')$ and a utility $V(\varepsilon, \varepsilon p, p)$.

For a given worker type-ε and a given mpl p, define the threshold mpl $q(\varepsilon, w, p)$ by $\phi(\varepsilon, q(\varepsilon, w, p), p) = w$, so that $\phi(\varepsilon, p', p) \leq w$ if $p' \leq q(\varepsilon, w, p)$. Contacts with firms less productive than $q(\varepsilon, w, p)$ end up not causing any wage increase because the current employer (with a technology yielding productivity p) can outbid such a challenging firm by offering a wage *lower* than w. Since in addition layoffs and deaths still occur at respective rates δ and μ, we may now write the Bellman equation solved by the value function $V(\varepsilon, w, p)$:

$$\begin{aligned}\left[\rho+\delta+\mu+\lambda_1\overline{F}\left(q\left(\varepsilon,w,p\right)\right)\right]\cdot V\left(\varepsilon,w,p\right) &= U(w) \\ &+\lambda_1\left[F\left(p\right)-F\left(q\left(\varepsilon,w,p\right)\right)\right]\cdot E_F\left\{V\left(\varepsilon,\varepsilon X,X\right)|q\left(\varepsilon,w,p\right)\leq X\leq p\right\} \\ &+\lambda_1\overline{F}\left(p\right)\cdot V\left(\varepsilon,\varepsilon p,p\right)+\delta V_0\left(\varepsilon\right). \quad (23)\end{aligned}$$

Imposing $w = \varepsilon p$ in the latter relationship, we easily get:

$$V\left(\varepsilon,\varepsilon p,p\right)=\frac{U(\varepsilon p)+\delta V_0\left(\varepsilon\right)}{\rho+\delta+\mu}. \quad (24)$$

Note that this expression is independent of the particular form of the unemployment value $V_0(\varepsilon)$.

Plugging this back into (23), replacing the expectation term by its expression and integrating by parts, we finally get a definition of $V(\cdot)$:

$$\left(\rho+\delta+\mu\right)\cdot V\left(\varepsilon,w,p\right)=U(w)+\delta V_0\left(\varepsilon\right)+\frac{\lambda_1\varepsilon}{\rho+\delta+\mu}\cdot\int_{q(\varepsilon,w,p)}^{p}\overline{F}\left(x\right)U'(\varepsilon x)dx. \quad (25)$$

We can now derive expressions of the reservation wages $\phi_0(\cdot)$ and $\phi(\cdot)$, as well as the threshold mpl $q(\cdot)$. We begin with the latter for a given productivity p and a given worker type-ε. Using (24) and (25) together with the fact that, by definition, $V(\varepsilon, w, p) = V(\varepsilon, \varepsilon q(\varepsilon, w, p), q(\varepsilon, w, p))$, we get an implicit definition of $q(\varepsilon, w, p)$:

$$U(\varepsilon q\left(\varepsilon,w,p\right))-\frac{\lambda_1}{\rho+\delta+\mu}\cdot\int_{q(\varepsilon,w,p)}^{p}\overline{F}\left(x\right)\varepsilon U'(\varepsilon x)dx=U\left(w\right). \quad (26)$$

Note that, as intuition suggests, (26) shows that $q(\varepsilon, \varepsilon p, p) = p$. Now consider a pair of firm types $p \leq p'$. Substituting $\phi(\varepsilon, p, p')$ for w in (26), using the fact that $q(\varepsilon, \phi(\varepsilon, p, p'), p') = p$, and rearranging terms, we get:

$$U\left(\phi\left(\varepsilon,p,p'\right)\right)=U(\varepsilon p)-\frac{\lambda_1}{\rho+\delta+\mu}\cdot\int_{p}^{p'}\overline{F}\left(x\right)\varepsilon U'(\varepsilon x)dx. \quad (27)$$

We now turn to the unemployed workers' reservation wages $\phi_0(\cdot)$, which are defined by the equality (1). Replacing w by $\phi_0(\varepsilon, p)$ in (23) and noticing that $q(\varepsilon, \phi_0(\varepsilon, p), p) = b$,[29] we get for any given ε:

$$\phi_0\left(\varepsilon,p\right)=\phi\left(\varepsilon,b,p\right)=U^{-1}\left(U(\varepsilon b)-\frac{\lambda_1}{\rho+\delta+\mu}\cdot\int_{b}^{p}\overline{F}\left(x\right)\varepsilon U'(\varepsilon x)\,dx\right). \quad (28)$$

[29] This is shown by the definitions of $q(\cdot)$ and $\phi_0(\cdot)$:

$$V_0\left(\varepsilon\right)=V\left(\varepsilon,\phi_0\left(\varepsilon,p\right),p\right)=V\left(\varepsilon,\varepsilon q\left(\cdot\right),q\left(\cdot\right)\right),$$

which implies from (22) and (24) that $q(\varepsilon, \phi_0(\varepsilon, p), p) = b$.

B Computation of $E[T(w)|p]$ for any integrable function $T(w)$

The lowest paid type-ε worker in a type-p firm is one that has just been hired, therefore earning $\phi_0(\varepsilon, p)$, while the highest-paid type-ε worker in that firm earns his marginal productivity εp. Having thus defined the support of the within-firm earnings distribution of type ε workers for any type-p firm, we can readily show that for any integrable function $T(w)$,

$$\begin{aligned} E[T(w)|p] &= \int_{\varepsilon_{\min}}^{\varepsilon_{\max}} \left(\int_{\phi_0(\varepsilon,p)}^{\varepsilon p} T(w) \cdot G(dw|\varepsilon, p) + T(\phi_0(\varepsilon,p)) \cdot G(\phi_0(\varepsilon,p)|\varepsilon,p) \right) h(\varepsilon) d\varepsilon \\ &= \left[1+\kappa_1 \overline{F}(p)\right]^2 \cdot \left\{ \frac{1}{(1+\kappa_1)^2} \int_{\varepsilon_{\min}}^{\varepsilon_{\max}} T(\phi_0(\varepsilon,p)) h(\varepsilon) d\varepsilon \right. \\ &\quad \left. + \int_{p_{\min}}^{p} \left[\int_{\varepsilon_{\min}}^{\varepsilon_{\max}} T(\phi(\varepsilon,q,p)) h(\varepsilon) d\varepsilon \right] \cdot \frac{2\kappa_1 f(q)}{\left[1+\kappa_1 \overline{F}(q)\right]^3} dq \right\} \\ &= \int_{\varepsilon_{\min}}^{\varepsilon_{\max}} T(\varepsilon p) h(\varepsilon) d\varepsilon \\ &\quad - \left[1+\kappa_1 \overline{F}(p)\right]^2 \cdot \int_{p_{\min}}^{p} \left[\int_{\varepsilon_{\min}}^{\varepsilon_{\max}} \frac{T'(\phi(\varepsilon,q,p))}{U'(\phi(\varepsilon,q,p))} \varepsilon U'(\varepsilon q) h(\varepsilon) d\varepsilon \right] \cdot \frac{1+(1-\sigma)\kappa_1 \overline{F}(q)}{\left[1+\kappa_1 \overline{F}(q)\right]^2} dq. \end{aligned}$$

The first equality follows from the definition of $G(w|\varepsilon, p)$ as

$$G(w|\varepsilon,p) = \frac{\left[1+\kappa_1 \overline{F}(p)\right]^2}{\left[1+\kappa_1 \overline{F}(q(\varepsilon,w,p))\right]^2}$$

yielding

$$G'(w|\varepsilon,p) = \left[1+\kappa_1 \overline{F}(p)\right]^2 \cdot h(\varepsilon) \cdot \frac{2\kappa_1 f(q)}{\left[1+\kappa_1 \overline{F}(q)\right]^3} \cdot \frac{\partial q(\varepsilon,w,p)}{\partial w} dw.$$

The second equality is obtained by integration by part, computing the partial derivative of $\phi(\varepsilon, q, p)$ wrt w from (27) as

$$U'(\phi(\varepsilon,q,p)) \cdot \frac{\partial \phi(\varepsilon,q,p)}{\partial w} = \varepsilon U'(\varepsilon q) \cdot \left[1+\kappa_1(1-\sigma)\overline{F}(q)\right].$$

Equation (12) follows when $T(w) = U(w) = \ln w$.

C Derivation of the likelihood in estimation step 2

Let ℓ_i designate the contribution of individual i to the likelihood of the N observations. We can factorize ℓ_i into the product of two components: ℓ_{i1} which is the likelihood of (d_{i1}, d_{i2}) given y_i, and ℓ_{i2} which is the probability of $\mathbf{1}\{n_i = 0\}$ given y_i and d_{i1}.

We begin with ℓ_{i2}. As is explained in the main text, the probability of $n_i = 0$ given $d_{i1} = d$ and the employer's type in the first spell p_i is the expected value of $\exp\left\{-\lambda_1 \left[\overline{F}(q(\varepsilon,w,p)) - \overline{F}(p)\right] d\right\}$ with respect to ε and w:

$$
\begin{aligned}
& \Pr\{n_i = 0 \mid d_{i1} = d, p_i = p\} \\
= & \int_{\varepsilon_{\min}}^{\varepsilon_{\max}} \left(\int_{\phi_0(\varepsilon,p)}^{\varepsilon p} e^{-\lambda_1 \left[\overline{F}(q(\varepsilon,w,p)) - \overline{F}(p)\right] d} \cdot G(dw|\varepsilon, p) + e^{-\lambda_1 F(p) d} \cdot G(\phi_0(\varepsilon,p)|\varepsilon,p) \right) d\varepsilon \\
= & 1 - \left[1 + \kappa_1 \overline{F}(p)\right]^2 \cdot \int_{p_{\min}}^{p} \frac{\lambda_1 f(q) d}{\left[1 + \kappa_1 \overline{F}(q)\right]^2} \cdot e^{-\lambda_1 \left[\overline{F}(q) - \overline{F}(p)\right] d} dq \\
= & \frac{\left[\delta + \mu + \lambda_1 \overline{F}(p)\right]^2}{\left[\delta + \mu + \lambda_1\right]^2} \cdot e^{-\lambda_1 F(p)} + \operatorname{Ei}\left(-\left[\delta + \mu + \lambda_1 \overline{F}(p)\right] d\right) - \operatorname{Ei}\left(-\left[\delta + \mu + \lambda_1\right] d\right),
\end{aligned} \tag{29}
$$

after integrating by parts and making appropriate changes of variables, and where Ei is the Exponential Integral function ($\operatorname{Ei}(u) = \int_{-\infty}^{u} \frac{e^x}{x} dx$ or $\int_u^{\infty} \frac{e^{-x}}{x} dx = -\operatorname{Ei}(-u)$).

We further need to observe the Poisson exit rate out of a firm of given mpl p_i, which equals $\Delta(p_i) = \delta + \mu + \lambda_1 \overline{F}(p_i)$ (see above paragraph 2.3). Using the estimator (13) constructed in paragraph 3.3, $\Delta(p_i)$ rewrites as a function of the observed average log earnings:

$$
\Delta(y_i) = (\delta + \mu) \cdot \frac{1 + \kappa_1}{1 + \kappa_1 Z(y_i)}.
$$

Since $Z(y_i)$ is recorded for all of the N firms corresponding to the N employment spells d_{i1}, we can use those observations in the likelihood derived below.

Using the last equation together with (29), we come up with an expression of ℓ_{i2}:

$$
\begin{aligned}
\ell_{i2} = & 1 - \mathbf{1}\{n_i = 0\} - [1 - 2 \cdot \mathbf{1}\{n_i = 0\}] \\
& \times \left\{ \frac{e^{-\lambda_1 F(p)}}{[1 + \kappa_1 Z(y_i)]^2} + \operatorname{Ei}\left(-\frac{(\delta + \mu)(1 + \kappa_1) d}{1 + \kappa_1 Z(y_i)}\right) - \operatorname{Ei}\left(-(\delta + \mu)(1 + \kappa_1) d\right) \right\}.
\end{aligned}
$$

We now turn to ℓ_{i1}, which has different expressions depending on worker i's particular history.

1. **First employment spell censored.** Given the Poisson exit rate out of a job derived above, the probability that an employment spell at firm i last longer than T is given by:

$$
\begin{aligned}
\ell_{i1} & = e^{-\Delta(y_i) \cdot T} \\
& = \exp\left[-\frac{(\delta + \mu)(1 + \kappa_1)}{1 + \kappa_1 Z(y_i)} T\right].
\end{aligned}
$$

2. **Job-to-job transition after the first employment spell.**[30] Here we know that the first job spell has a duration of exactly d_{i1}, an event that has probability $\Delta(y_i) \cdot \exp\{-\Delta(y_i) \cdot d_{i1}\}$. We also know that the transition is made directly toward another job, which has conditional probability

[30] Recall that we arbitrarily define a job-to-job transition as an employer change with an intervening unemployment spell of less than 15 days. This convention can be varied within a reasonable range without dramatically affecting the estimates.

$\lambda_1 \overline{F}(p_i) / \Delta(p_i)$. The probability of observing such a transition is therefore:

$$\begin{aligned}\ell_{i1} &= \lambda_1 \overline{F}(p_i)\, e^{-\Delta(y_i)\cdot d_{i1}} \\ &= \frac{(\delta+\mu)\kappa_1 \overline{Z}(y_i)}{1+\kappa_1 Z(y_i)} \cdot \exp\left[-\frac{(\delta+\mu)(1+\kappa_1)}{1+\kappa_1 Z(y_i)} d_{i1}\right].\end{aligned}$$

3. **Permanent exit from the sample.** Again here the probability of observing a first job spell of length d_{i1} equals $\Delta(y_i) \cdot \exp\{-\Delta(y_i) \cdot d_{i1}\}$. Now since the subsequent spell is censored, there is no way we can know for sure whether the worker has permanently left the labor force or just experiences a protracted period of unemployment. The conditional probability that worker i's initial exit from the sample corresponds to a 'death' is $\mu/\Delta(y_i)$. Similarly, this exit is the result of a layoff with probability $\delta/\Delta(y_i)$. In the latter case, however, the fact that worker i does not re-enter the panel before date T can be caused either by this worker's 'death' occurring before he/she finds a new job, or by this worker not dying before T but simply experiencing a protracted unemployment spell. Overall, the conditional probability of not seeing worker i reappear in the sample before date T, given a transition at date d_{i1} is given by:

$$\frac{\mu}{\Delta(y_i)} + \frac{\delta}{\Delta(y_i)} \cdot \left[\int_{d_{i1}}^{T} \mu e^{-\mu x} \cdot e^{-\lambda_0 x} dx + e^{-(\mu+\lambda_0)\cdot T}\right].$$

The contribution to the likelihood of an observation like case 3 is the product of the above two probabilities:

$$\begin{aligned}\ell_{i1} &= \left[\mu \frac{\delta+\mu+\lambda_0}{\mu+\lambda_0} + \frac{\delta\lambda_0}{\mu+\lambda_0} \cdot e^{-(\mu+\lambda_0)\cdot T}\right] \cdot e^{-(\delta+\mu)\cdot \frac{1+\kappa_1}{1+\kappa_1 Z(y_i)} \cdot d_{i1}} \\ &= \left[\delta+\mu-\delta\lambda_0 \frac{1-e^{-(\mu+\lambda_0)\cdot T}}{\mu+\lambda_0}\right] \cdot \exp\left[\frac{(\delta+\mu)(1+\kappa_1)}{1+\kappa_1 Z(y_i)} d_{i1}\right]\end{aligned}$$

4. **Job-to-unemployment transition followed by a reentry.** Once again the probability of observing a first job spell of length d_{i1} equals $\Delta(y_i)\cdot\exp\{-\Delta(y_i) \cdot d_{i1}\}$. Concerning the subsequent spell, we know in this case that it can only be an unemployment spell of exact length d_{i2}. The conditional probability of such a spell is

$$\frac{\delta}{\Delta(y_i)} \cdot \lambda_0 e^{-(\mu+\lambda_0)\cdot d_{i2}},$$

which implies a contribution to the likelihood expressed as:

$$\ell_{i1} = \delta\lambda_0 \exp - \left[\frac{(\delta+\mu)(1+\kappa_1)}{1+\kappa_1 Z(y_i)} d_{i1} - (\mu+\lambda_0)\, d_{i2}\right].$$

The complete likelihood of the N observations ($\ell_i = \ell_{i1} \times \ell_{i2}$) can thus be written as a function of the sole transition parameters δ, μ, λ_0, and κ_1.

D Within-firm log wage variance

Using the result of Appendix B, simple calculations show that

$$E(\ln w|\varepsilon,p) = \ln\varepsilon + m_1(p) \tag{30}$$

where

$$\begin{aligned} m_1(p) &= E(\ln w|p) \\ &= \ln p - \left[1+\kappa_1\overline{F}(p)\right]^2 \cdot \int_b^p \frac{1+\kappa_1(1-\sigma)\overline{F}(q)}{\left[1+\kappa_1\overline{F}(q)\right]^2}\cdot\frac{dq}{q}. \end{aligned} \tag{31}$$

Likewise,

$$E\left((\ln w)^2|\varepsilon,p\right) = (\ln\varepsilon)^2 + 2\ln\varepsilon\cdot m_1(p) + m_2(p) \tag{32}$$

where

$$\begin{aligned} m_2(p) &= E\left((\ln w)^2|p\right) - V\ln\varepsilon \\ &= (\ln p)^2 - \left[1+\kappa_1\overline{F}(p)\right]^2 \cdot \int_b^p 2\ln\phi(1,q,p)\cdot\frac{1+\kappa_1(1-\sigma)\overline{F}(q)}{\left[1+\kappa_1\overline{F}(q)\right]^2}\cdot\frac{dq}{q}, \end{aligned} \tag{33}$$

using the result of appendix B.

It thus follows from (30) and (32) that

$$V(\ln w|\varepsilon,p) = m_2(p) - m_1(p)^2, \tag{34}$$

which is independent of ε in the absence of assortative matching (in which large-p firms would be more likely to hire large-ε workers), and that

$$V(\ln w|p) = V\ln\varepsilon + m_2(p) - m_1(p)^2. \tag{35}$$

What we have to do at this point is to construct $m_1(p)$ and $m_2(p)$.

The construction of $m_1(p)$ is straightforward: the only thing we need to do is to split the integral in (31) into two over the intervals $[b, p_{\min}]$ (over which $\overline{F}(\cdot)\equiv 1$) and $[p_{\min}, p]$ to obtain:

$$m_1(p) = \ln p - \left[1+\kappa_1\overline{F}(p)\right]^2\cdot\int_{p_{\min}}^p \frac{1+\kappa_1(1-\sigma)\overline{F}(q)}{\left[1+\kappa_1\overline{F}(q)\right]^2}\cdot\frac{dq}{q} - \left[1+\kappa_1\overline{F}(p)\right]^2\cdot\frac{1+\kappa_1(1-\sigma)}{(1+\kappa_1)^2}\cdot(\ln p_{\min}-\ln b). \tag{36}$$

From equation (30) and our normalization assumption $E_H(\ln\varepsilon)=0$, $m_1(p)$ has the average log wage $E(\ln w|p)$ in any particular type-p firm as an immediate empirical counterpart.

The construction of $m_2(p)$ proceeds as follows. First, using the definition (27) of $\phi(\cdot)$, $m_2(p)$ can be rewritten from (33) as

$$\begin{aligned}
m_2(p) &= (\ln p)^2 - 2\left[1+\kappa_1\overline{F}(p)\right]^2 \cdot \int_b^p \left(\ln q - \kappa_1(1-\sigma)\int_q^p \overline{F}(x)\frac{dx}{x}\right)\cdot \frac{1+\kappa_1(1-\sigma)\overline{F}(q)}{\left[1+\kappa_1\overline{F}(q)\right]^2}\cdot\frac{dq}{q} \\
&= (\ln p)^2 - 2\left[1+\kappa_1\overline{F}(p)\right]^2 \cdot \left\{\int_b^p \ln q \cdot \frac{1+\kappa_1(1-\sigma)\overline{F}(q)}{\left[1+\kappa_1\overline{F}(q)\right]^2}\cdot\frac{dq}{q} - \kappa_1(1-\sigma)\int_b^p \overline{F}(q)\cdot\frac{\ln q - \Omega_1(q)}{\left[1+\kappa_1\overline{F}(q)\right]^2}\cdot\frac{dq}{q}\right\} \\
&= (\ln p)^2 - 2\left[1+\kappa_1\overline{F}(p)\right]^2 \cdot \left\{\ln p\cdot\frac{\ln p - m_1(p)}{\left[1+\kappa_1\overline{F}(p)\right]^2} - \int_b^p \left[1+\kappa_1(1-\sigma)\overline{F}(q)\right]\cdot\frac{\ln q - m_1(q)}{\left[1+\kappa_1\overline{F}(q)\right]^2}\cdot\frac{dq}{q}\right\} \\
&= 2\ln p\cdot m_1(p) - (\ln p)^2 + 2\left[1+\kappa_1\overline{F}(p)\right]^2\cdot\int_b^p \left[1+\kappa_1(1-\sigma)\overline{F}(q)\right]\cdot\frac{\ln q - m_1(q)}{\left[1+\kappa_1\overline{F}(q)\right]^2}\cdot\frac{dq}{q},
\end{aligned}$$

where the second and third lines above are deduced from the first using integrations by parts and the definition (31) of $m_1(p)$. Finally, splitting the integral in the last line into two over the intervals $[b, p_{\min}]$ (over which $\overline{F}(\cdot)\equiv 1$ and $\ln q - m_1(q) = (1+\kappa_1(1-\sigma))(\ln q - \ln b)$) and $[p_{\min}, p]$ yields:

$$\begin{aligned}
m_2(p) = 2\ln p\cdot m_1(p) - (\ln p)^2 + 2\left[1+\kappa_1\overline{F}(p)\right]^2\cdot\int_{p_{\min}}^p \left[1+\kappa_1(1-\sigma)\overline{F}(q)\right]\cdot\frac{\ln q - m_1(q)}{\left[1+\kappa_1\overline{F}(q)\right]^2}\cdot\frac{dq}{q} \\
+\left[1+\kappa_1\overline{F}(p)\right]^2\cdot\left(\frac{1+\kappa_1(1-\sigma)}{1+\kappa_1}\right)^2\cdot(\ln p_{\min} - \ln b)^2.
\end{aligned}$$

It thus follows that and the within-firm log-wage variance can be written as

$$\begin{aligned}
V(\ln w|p) = 2\ln p\cdot m_1(p) - (\ln p)^2 + 2\left[1+\kappa_1\overline{F}(p)\right]^2\cdot\int_{p_{\min}}^p \left[1+\kappa_1(1-\sigma)\overline{F}(q)\right]\cdot\frac{\ln q - m_1(q)}{\left[1+\kappa_1\overline{F}(q)\right]^2}\cdot\frac{dq}{q} \\
+(\ln p_{\min} - \ln b)^2\cdot\left(\frac{1+\kappa_1(1-\sigma)}{1+\kappa_1}\right)^2\left[1+\kappa_1\overline{F}(p)\right]^2 - m_1(p)^2 + V\ln\varepsilon. \quad (37)
\end{aligned}$$

QED.

Occupation	Number of indiv. trajectories	Percentage with no recorded mobility	Percentage whose first recorded mobility is from job... ... to-job	... to-out of sample	sample mean unemployment spell duration	sample mean employment spell duration
Executives, managers and engineers	22,757	46.2%	23.4%	30.4%	0.96 yrs	2.09 yrs
Supervisors, administrative and sales	14,977	48.1%	19.3%	32.5%	1.16 yrs	2.11 yrs
Technical supervisors and technicians	7,448	55.5%	16.0%	28.6%	1.07 yrs	2.28 yrs
Administrative support	14,903	54.3%	8.2%	37.5%	1.30 yrs	2.23 yrs
Skilled manual workers	12,557	55.9%	5.2%	38.9%	1.16 yrs	2.28 yrs
Sales and service workers	5,926	45.1%	5.5%	49.4%	1.28 yrs	2.06 yrs
Unskilled manual workers	4,416	42.5%	7.0%	50.5%	1.29 yrs	1.98 yrs

Table 1: Descriptive analysis of worker mobility

Occupation	Nb obs.	Median Δlog wage	% obs such that Δlog wage ≤ -0.10	-0.05	0	0.05	0.10
Executives, managers and engineers	5,335	3.1%	23.6	28.5	38.1	55.1	65.4
Supervisors, administrative and sales	2,893	3.7%	21.6	27.1	36.6	54.3	65.2
Technical supervisors and technicians	1,190	3.8%	14.0	20.2	32.2	55.5	67.3
Administrative support	1,222	2.2%	21.5	28.7	40.7	60.5	69.2
Skilled manual workers	657	0.5%	33.2	37.7	49.2	62.3	72.0
Sales and service workers	326	1.4%	31.3	37.7	45.1	58.0	67.5
Unskilled manual workers	310	-1.3%	33.5	42.9	54.5	63.4	72.3

Table 2: Variation in real wage after first recorded job-to-job mobility
(i.e. with less than 15 days work interruption) in 96-98

Occupation	Nb obs.	Median Δlog wage	% obs such that Δlog wage ≤ -0.10	-0.05	0	0.05	0.10
Executives, managers and engineers	16,102	2.7%	6.6	11.3	28.5	64.4	80.0
Supervisors, administrative and sales	15,592	2.6%	7.9	12.9	28.6	65.2	81.1
Technical supervisors and technicians	5,644	2.5%	6.6	11.9	29.6	68.1	85.0
Administrative support	11,105	2.2%	7.9	12.4	30.0	69.8	84.2
Skilled manual workers	9,747	1.9%	7.9	15.0	34.9	69.5	85.1
Sales and service workers	4,192	2.5%	7.4	12.8	31.4	64.5	79.1
Unskilled manual workers	2,847	2.2%	7.7	14.6	32.9	66.4	81.9

Table 3: Variation in real wage between 01/01/96 and 31/12/97
when holding the same job over this period

	Parameter				
Occupation	δ	μ	λ_0	λ_1	κ_1
Executives, managers and engineers	0.0776 (0.0009)	0.0070 (0.0005)	2.104 (0.063)	0.643 (0.009)	7.61 (0.14)
Supervisors, administrative and sales	0.0859 (0.0014)	0.0065 (0.0007)	1.956 (0.081)	0.666 (0.015)	7.21 (0.21)
Technical supervisors and technicians	0.0686 (0.0016)	0.0042 (0.0008)	2.055 (0.137)	0.646 (0.021)	8.87 (0.37)
Administrative support	0.0932 (0.0020)	0.0085 (0.0011)	1.678 (0.078)	0.737 (0.026)	7.24 (0.32)
Skilled manual workers	0.0886 (0.0020)	0.0082 (0.0012)	1.499 (0.071)	0.685 (0.027)	7.07 (0.35)
Sales and service workers	0.1016 (0.0031)	0.0045 (0.0016)	1.486 (0.097)	0.716 (0.038)	6.75 (0.44)
Unskilled manual workers	0.0989 (0.0036)	0.0153 (0.0020)	1.529 (0.099)	0.666 (0.038)	5.84 (0.41)

Table 4: Estimated transition parameters
(annual values, standard errors in parentheses)

Occupation	$\ln b$	$\ln p_{\min}$	σ	ρ
Executives, managers and engineers	4.47	4.66	0.860	0.52 (40.5% annual)
Supervisors, administrative and sales	3.87	4.11	0.890	0.75 (52.7% annual)
Technical supervisors and technicians	3.94	4.11	0.910	0.74 (52.1% annual)
Administrative support	3.67	3.80	0.890	0.82 (56.0% annual
Skilled manual workers	3.69	3.84	0.875	0.68 (49.2% annual)
Sales and service workers	3.50	3.58	0.900	0.95 (61.5% annual)
Unskilled manual workers	3.51	3.62	0.935	1.64 (80.7% annual)

Table 5: Estimation of the parameters of the productivity function

Occupation	Nobs.	Mean log wage: $E(\ln w)$	Total log wage variance: $V(\ln w)$	Firm effect: $VE(\ln w\|p)$		Search friction effect: $EV(\ln w\|p) - V\ln\varepsilon$		Person effect: $V\ln\varepsilon$	
				Value	% of $V(\ln w)$	Value	% of $V(\ln w)$	Value	% of $V(\ln w)$
Executives, managers and engineers	647,674	4.78	0.159	0.038	23.9%	0.035	22.0%	0.086	54.1%
Supervisors, administrative and sales	571,646	4.25	0.119	0.043	36.1%	0.052	43.7%	0.024	20.2%
Technical supervisors and technicians	260,926	4.29	0.070	0.029	41.4%	0.031	44.3%	0.010	14.3%
Administrative support	553,657	3.97	0.076	0.031	40.8%	0.035	46.1%	0.010	13.2%
Skilled manual workers	482,350	4.00	0.065	0.034	52.3%	0.031	47.7%	0	0%
Sales and service workers	267,535	3.72	0.048	0.023	47.9%	0.025	52.1%	0	0%
Unskilled manual workers	228,253	3.76	0.050	0.027	54.0%	0.023	46.0%	0	0%

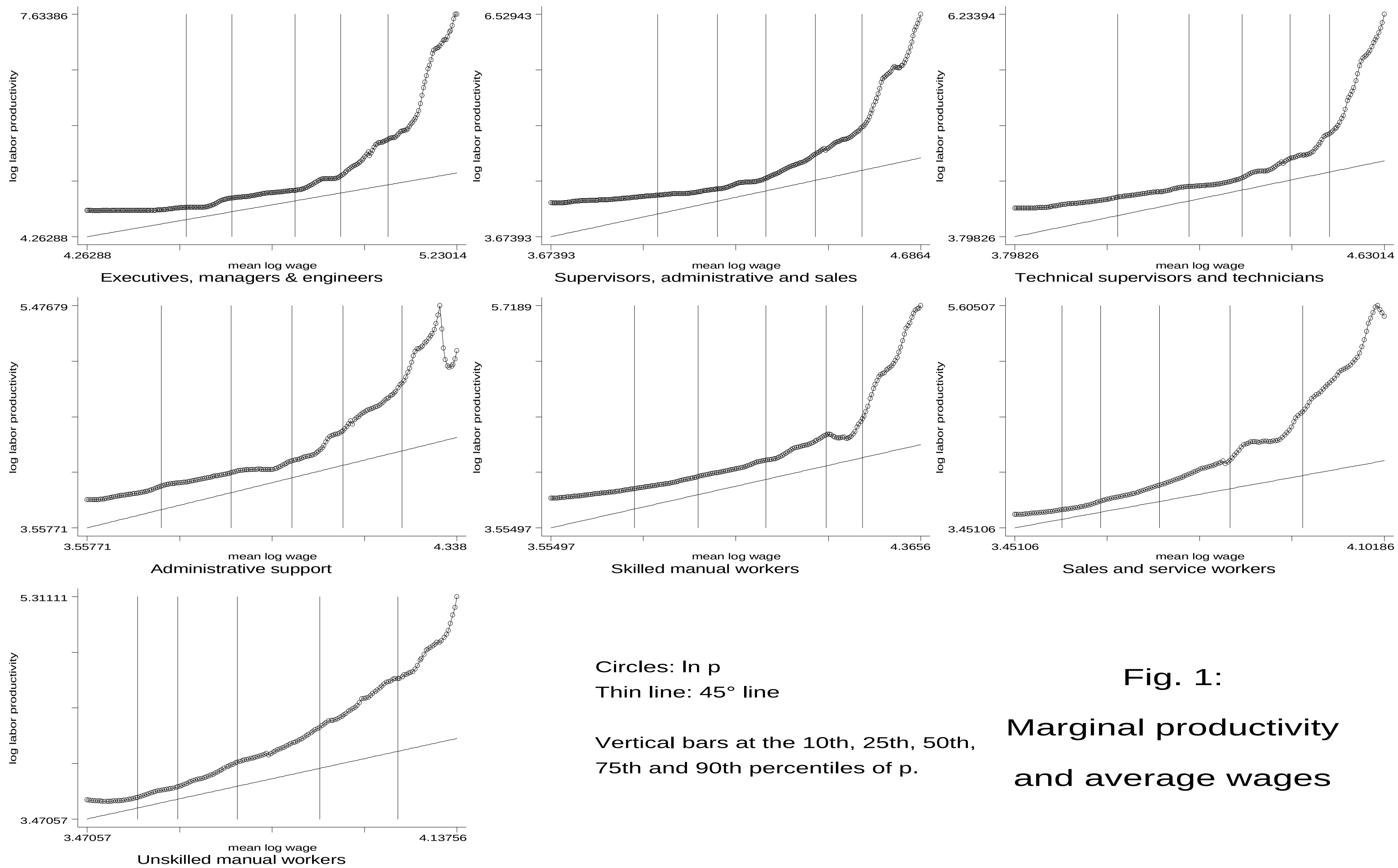
Table 6: Log wage variance decomposition

Occupation	Median Δlog wage	% obs such that Δlog wage $\leq$ -0.10	-0.05	0	0.05	0.10
Executives, managers and engineers	3.7%	1.2	10.8	32.7	53.0	65.3
Supervisors, administrative and sales	5.0%	0.2	3.1	22.3	50.0	64.8
Technical supervisors and technicians	5.8%	0	3.3	19.7	46.8	64.2
Administrative support	5.3%	0.3	3.6	24.4	48.9	64.0
Skilled manual workers	5.4%	0.3	4.6	24.2	48.9	64.5
Sales and service workers	4.0%	0	0.9	22.2	54.9	71.0
Unskilled manual workers	4.5%	0	0.4	21.8	51.9	67.3

Table 7: Dynamic simulation–Variation in real wage after first recorded job-to-job mobility

Occupation	Median Δlog wage	% obs such that Δlog wage $\leq$ -0.10	-0.05	0	0.05	0.10
Executives, managers and engineers	0	0	0	86.3	95.3	97.4
Supervisors, administrative and sales	0	0	0	86.3	95.1	97.5
Technical supervisors and technicians	0	0	0	87.4	96.0	98.3
Administrative support	0	0	0	84.9	94.3	97.2
Skilled manual workers	0	0	0	85.9	94.4	97.0
Sales and service workers	0	0	0	84.6	95.3	97.7
Unskilled manual workers	0	0	0	83.9	94.4	97.2

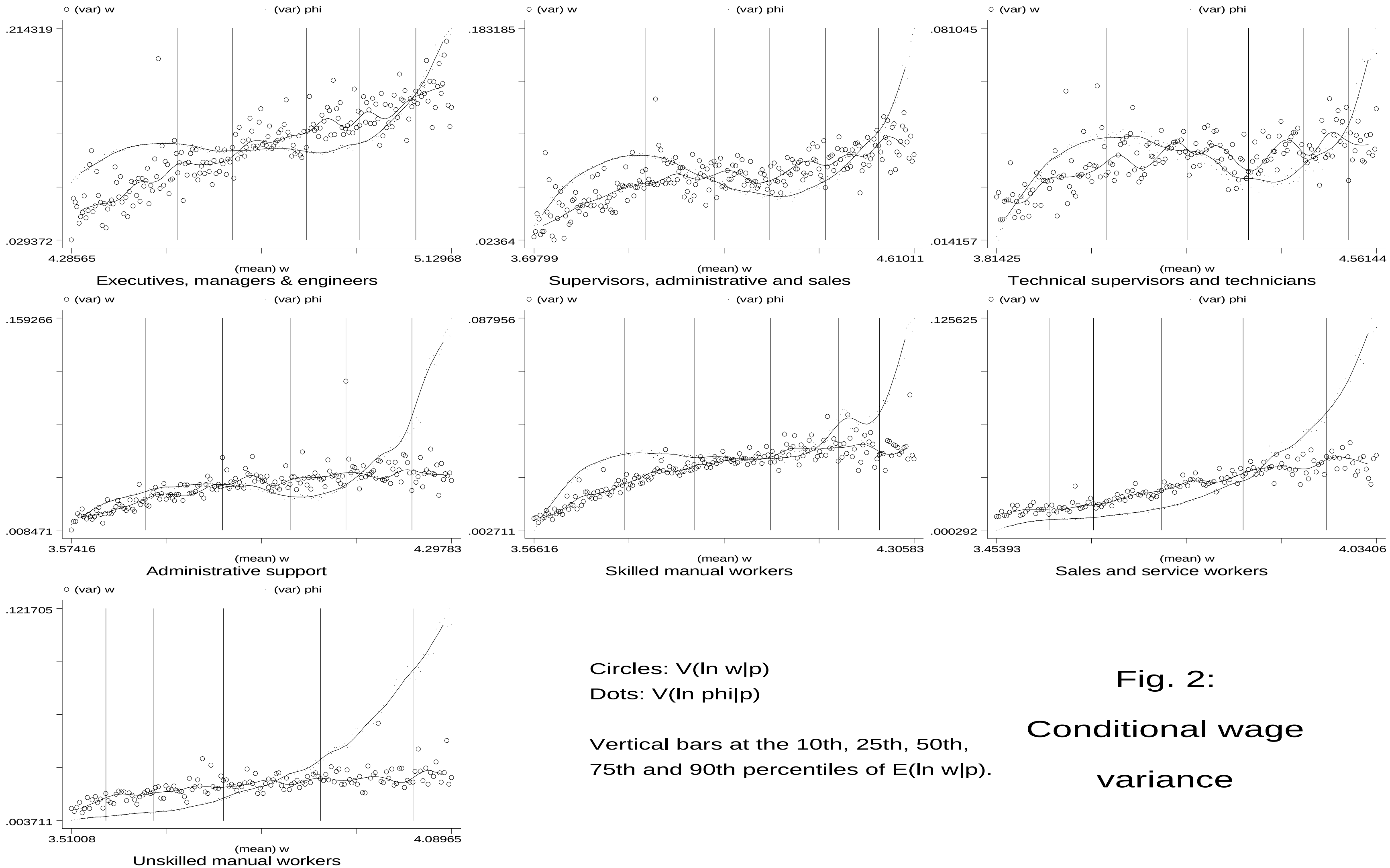
Table 8: Dynamic simulation–Yearly variation in real wage when holding the same job over this period



Circles: ln p
Thin line: 45° line

Vertical bars at the 10th, 25th, 50th,
75th and 90th percentiles of p.

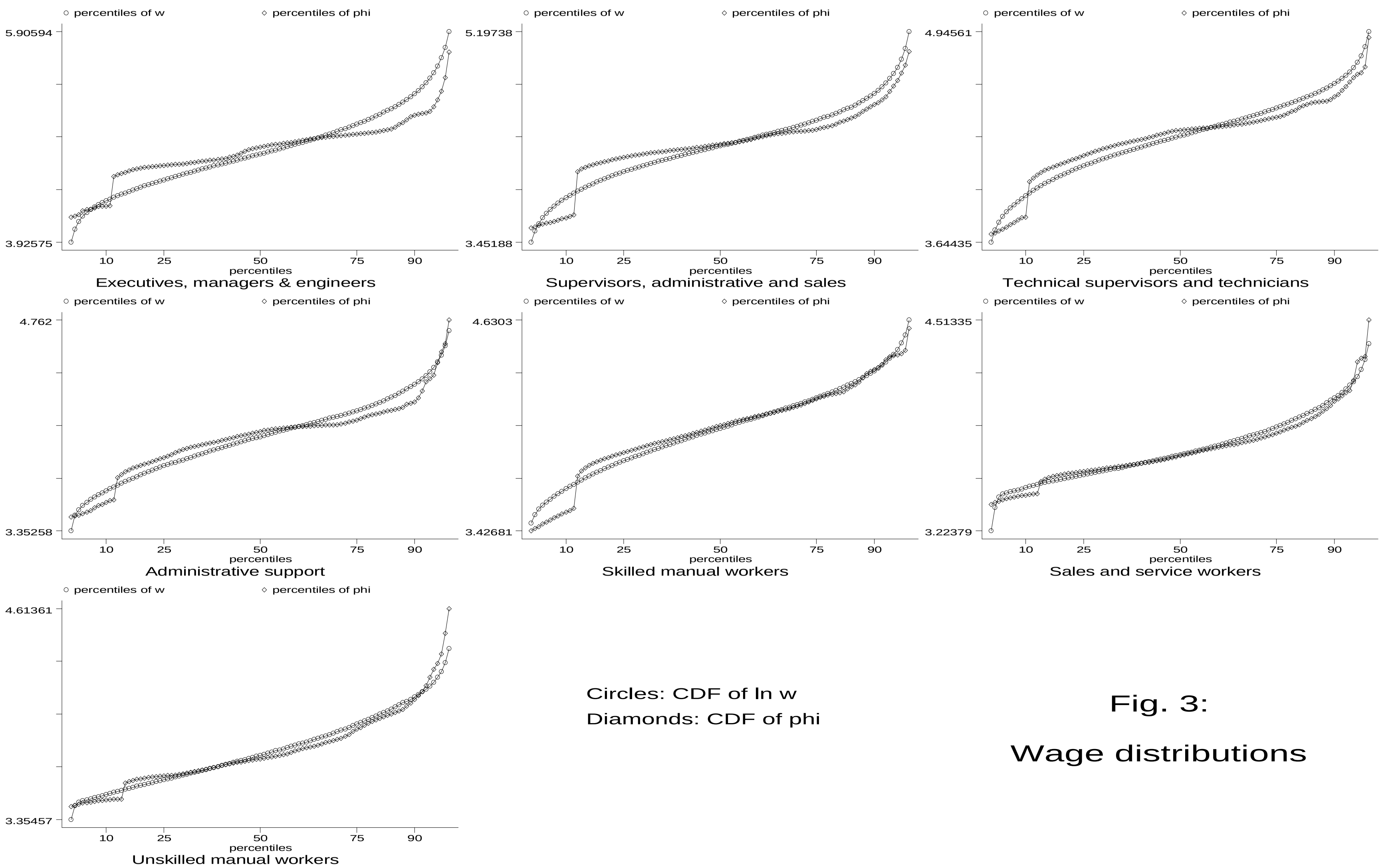
Fig. 1:
Marginal productivity
and average wages



Circles: V(ln w|p)
Dots: V(ln phi|p)

Vertical bars at the 10th, 25th, 50th, 75th and 90th percentiles of E(ln w|p).

Fig. 2: Conditional wage variance



Circles: CDF of ln w
Diamonds: CDF of phi

Fig. 3:

Wage distributions

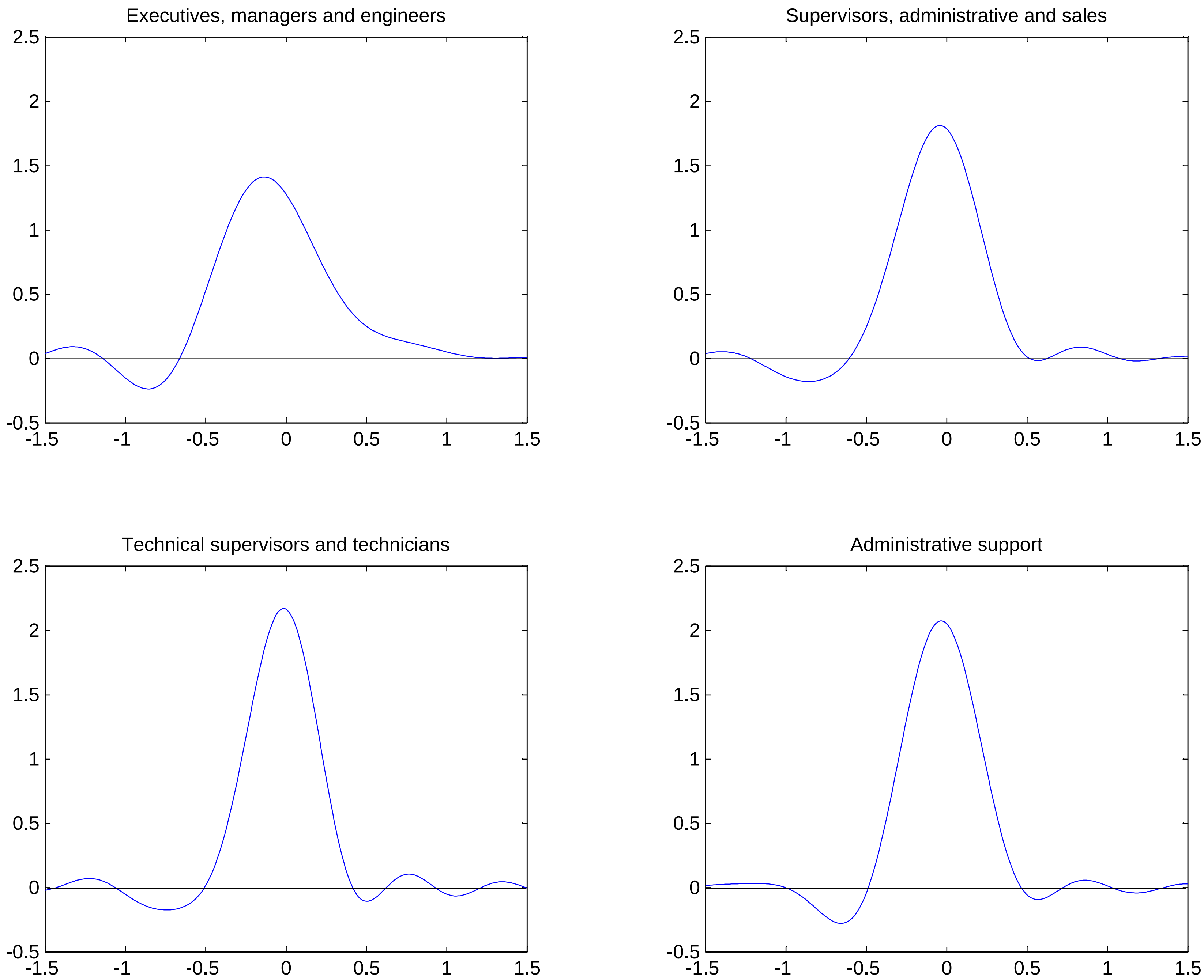
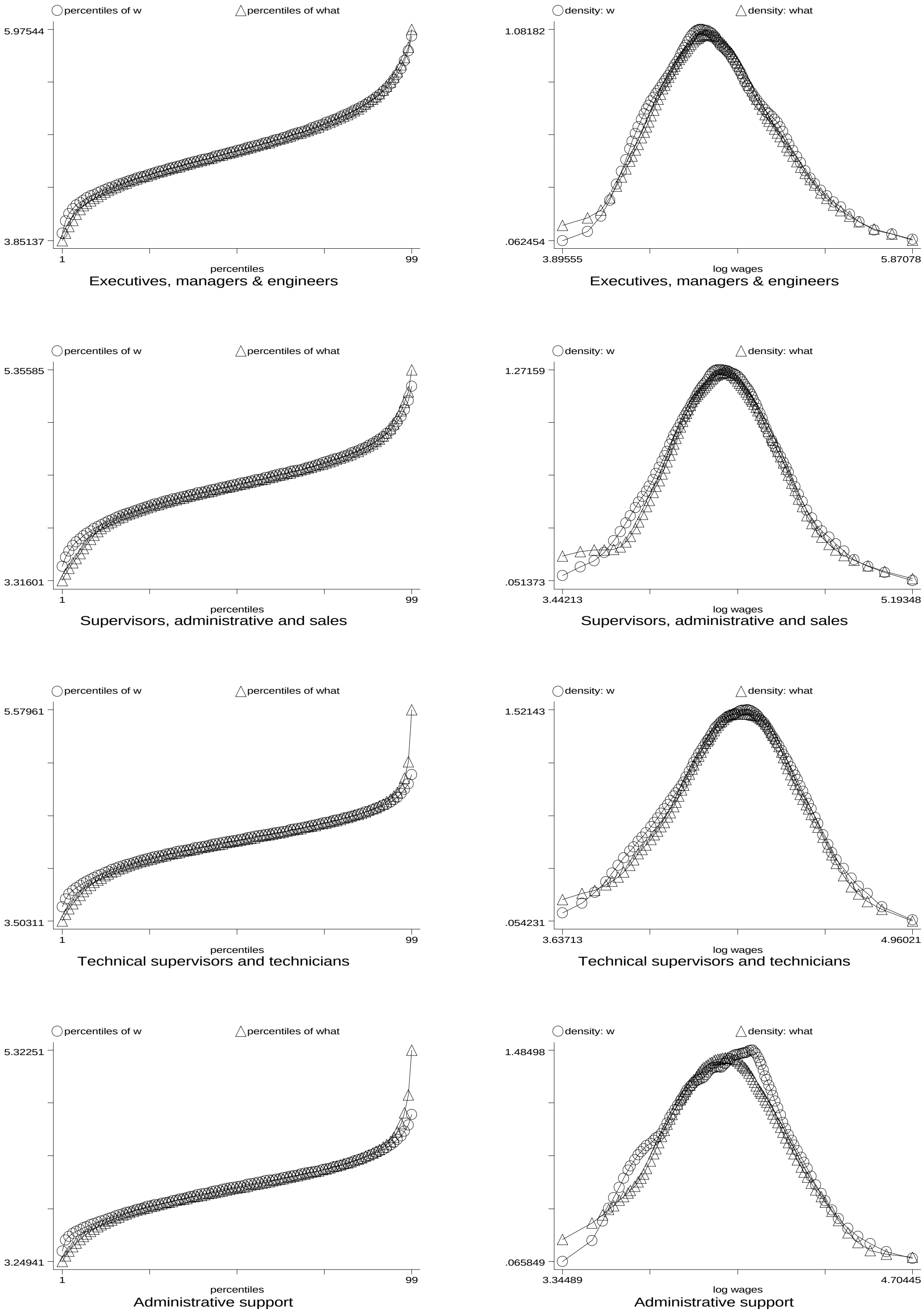
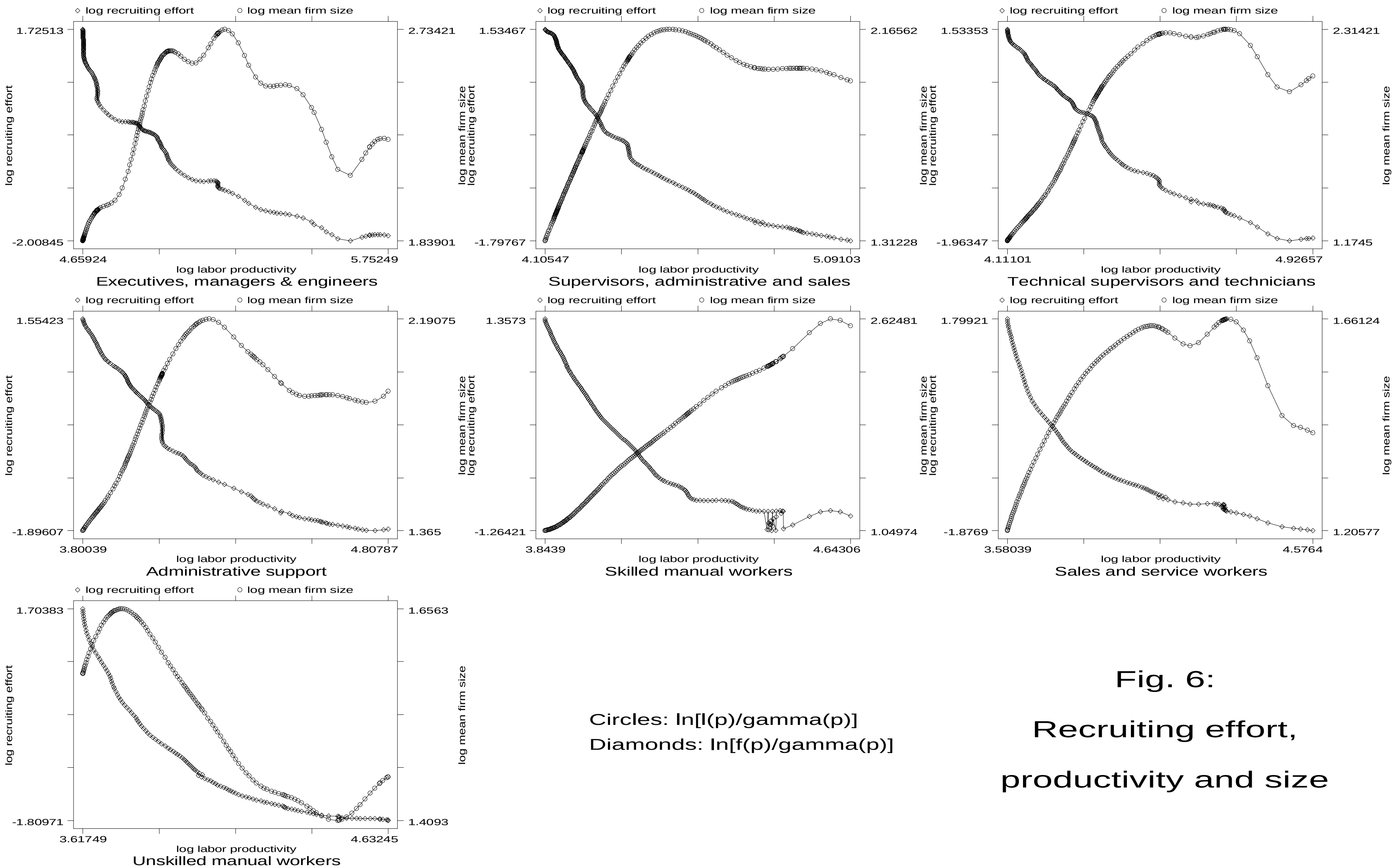


Fig. 4: density of ln(")

Fig. 5: Observed and predicted log earnings distributions





Circles: ln[l(p)/gamma(p)]
Diamonds: ln[f(p)/gamma(p)]

Fig. 6:

Recruiting effort,

productivity and size

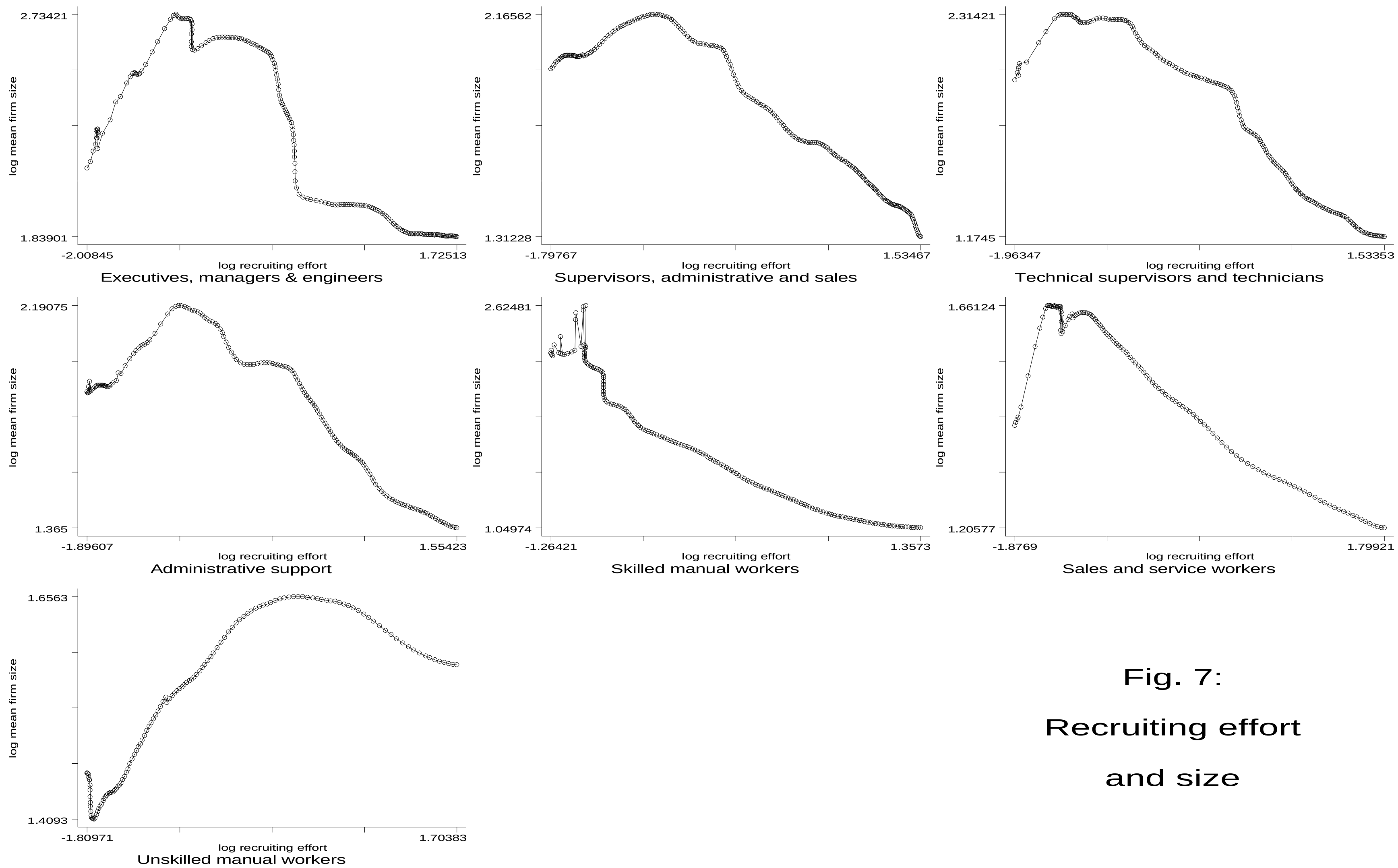


Fig. 7:

Recruiting effort

and size