THE RATE OF LEARNING-BY-DOING: ESTIMATES FROM A SEARCH-MATCHING MODEL

JULIEN PRAT*
Institute for Economic Analysis (CSIC)

SUMMARY
We construct and estimate by maximum likelihood a job search model where wages are set by Nash bargaining and idiosyncratic productivity follows a geometric Brownian motion. The proposed framework enables us to endogenize job destruction and to estimate the rate of learning-by-doing. Although the range of the observations is not independent of the parameters, we establish that the estimators satisfy asymptotic normality. The structural model is estimated using Current Population Survey data on accepted wages and employment durations. We show that it accurately captures the joint distribution of wages and job spells. We find that the rate of learning-by-doing has an important positive effect on aggregate output and a small impact on employment. Copyright © 2009 John Wiley & Sons, Ltd.

1. INTRODUCTION
Over the last few years, search-matching models have become the subject of renewed scrutiny. This spurt of interest is largely motivated by the difficulty to reconcile both micro and macro features of the data with the theory. Central to the issue is the difference between market productivity and the value of leisure. Costain and Reiter (2008) or Hagedorn and Manovskii (2008) show that, by setting it close enough to zero, one can ensure that unemployment and posted vacancies fluctuate as much in the Mortensen-Pissarides model as in the data. This solution, however, leads to a large discrepancy between predicted and observed wage inequality. Hornstein et al. (2007) illustrate this tension in a calibration exercise, finding that parameter values matching business cycle fluctuations greatly underestimate the degree of wage dispersion. Accordingly, recent research highlights the need to carefully analyze the micro predictions of the Mortensen–Pissarides framework.

This paper contributes to this project by estimating a search-matching model where workers accumulate job-specific skills through learning-by-doing (hereafter LBD). This feature allows us to control for the effect of job tenure on the wage distribution; hence it lowers the amount of residual wage dispersion and consequently the cost of being unemployed. We find that LBD reduces the gap between the value of non-market activity derived from our structural estimation and the one needed to fit business cycle fluctuations, but only marginally so. Our analysis therefore concurs with the findings in Hornstein et al. (2007).

The rate of LBD cannot be estimated in a purely deterministic set-up because the lower bound of the wage distribution increases with tenure when workers regularly progress along the learning curve. Given that some highly senior workers earn a wage close to their reservation wage, the

* Correspondence to: Julien Prat, IAE, Campus UAB, 08193 Bellaterra, Barcelona, Spain.
E-mail: julien.prat@iae.csic.es

Copyright © 2009 John Wiley & Sons, Ltd.
estimated rate of LBD necessarily collapses to zero in a deterministic set-up. This is why one needs to introduce some noise into the learning process. A possibility is to allow for measurement errors. The drawback of this approach is that it ignores the interaction between job destruction and the rate of LBD. We assume instead that jobs are affected by idiosyncratic productivity shocks. The addition of this realistic feature allows us to endogenize job separations so that our structural model takes into account the positive influence of LBD on job retention.\footnote{The positive effect of skill accumulation on job retention is commonly invoked to justify training programs. An important example is provided by the official goals of the Workforce Investment Act of 1998. For an econometric assessment of the relationship between job stability and training programs, see Winter-Ebmer and Zweimüller (1996).}

Our analytical framework is therefore closely related to the canonical Mortensen and Pissarides (1994) model (hereafter MP94) with endogenous separations. We assume that productivity and thus human capital are purely match-specific. We also assume that firms and workers cannot commit so that wages are set by Nash bargaining. We extend MP94 by letting initial productivities differ across jobs and by allowing workers to accumulate specific human capital. We also replace Poisson processes with Brownian motions in order to capture the high persistence of earnings shocks. Our model therefore builds on the framework described in Prat (2006). The contribution of this paper consists in structurally estimating the model on micro data. The analysis shows that the proposed extension of the MP94 model is able to match the joint distribution of wages and job spells, and identifies the parameter values required to achieve a good fit.

Given the considerable influence of the MP94 model, it may seem surprising that it has not yet been structurally estimated. A likely explanation is that doing so raises several challenges. First of all, endogenous separations greatly complicate the derivation of the likelihood function because one has to deduce all the sample paths that breach the reservation productivity. We show how this problem can be solved through the introduction of geometric Brownian motions. We therefore assume that log-wages follow a random walk with a deterministic trend. This specification accords well with the high persistence of earning shocks. Yet it substantially simplifies the canonical ARMA decomposition of the individual earnings process since it only retains its martingale component, disregarding transitory shocks.\footnote{An extensive empirical literature documents the accuracy of ARMA representations for earning dynamics. See, among others, the seminal paper by Lillard and Willis (1978) and the contribution by Topel and Ward (1992). More recently, Buhai and Teulings (2005) have used the PSID dataset to substantiate the geometric Brownian specification considered in this paper.} This simplification yields substantial analytical gains, allowing us to solve for equilibrium in closed-form, and thus to estimate the model by full-information methods.\footnote{Lopes de Melo (2006) also uses stochastic calculus to estimate an equilibrium search model. Although his model shares similarities to the one considered here, he focuses on the learning process about the quality of the firm-worker match.}

Hence, although earnings processes of the ARMA-type are more accurate, we propose geometric Brownian motions as a useful first-order approximation.

The other main technical difficulty is due to the non-standard properties of the likelihood function: the reservation wage and consequently the range of the data is a function of the estimated parameters. This peculiarity of job search models is well known (see, for example, Flinn and Heckman, 1982). In order to circumvent this problem, they proposed to evaluate the likelihood function in two steps. Unfortunately, we cannot use their methodology because endogenous job destruction implies that workers and firms separate as soon as the match productivity breaches the reservation threshold. As a result, the lowest reported wage is no longer a super-consistent estimator because the density at the reservation wage is equal to zero. We are nonetheless able to establish the asymptotic property of the estimators. We adapt to our set-up the proof by Greene (1980) that, in order to establish asymptotic normality, standard regularity conditions need not
be satisfied when the likelihood function is equal to zero at the parameter-dependent boundary. Hence endogenous separation actually simplifies the analysis since it enables us to estimate the likelihood function as if it were standard.

After having derived the equilibrium of the economy and characterized the likelihood function, we estimate the model using data from the January 2004 supplement of the Current Population Survey. The supplement contains data on accepted wages and employment durations for a supposedly random sample of the US labor force. Its representative dimension accords well with the macro orientation of our model and especially with our focus on the aggregate wage distribution.

We do not estimate the model on panel data for the three following reasons. First, given the relatively stylized structure of the model, we prefer to restrict our analysis to cross-sectional patterns and leave a more thorough inspection of wage dynamics to extensions with general human capital and on-the-job search. Second, macro panel data for the US economy are not readily available. The relatively small size of the PSID makes it difficult to accurately estimate the wage distribution among job entrants, while linking households over time using the CPS data leads to a significant proportion of mismatches with potential self-selection issues. Lastly, deriving the likelihood function for cross-sectional data turns out to be a comprehensive task since we have to characterize the wage distribution conditional on job tenure. Thus our estimation procedure lays the ground for future empirical research on either panel or cross-sectional data.

We restrict our attention to workers without tertiary education because the estimates do not capture the accumulation of general human capital, which is known to be much more significant for skilled workers. The estimation procedure returns estimates for the rate of LBD of around 2% per year. We assess the ability of the model to fit the joint distribution of wages and job spells and find that it reproduces the data surprisingly well given its parsimonious specification. Then we use the estimates to characterize the impact of the rate of LBD. We show that it shifts to the right the wage distribution and significantly increases its dispersion.

1.1. Related Literature

To the best of our knowledge, this paper is the first to structurally estimate the MP94 model. This gap in the literature is explained by the fact that jobs’ outputs follow stochastic paths in MP94, while estimable search models are typically based on the premise that productivities remain constant through time. As a result, early structural models generated flat wage profiles. Only recently has the empirical literature begun to address the observed pattern of wage dynamics.

4 Indeed, the number of job entrants in the PSID is below 60 in some months. We use the wage distribution among job entrants to control for the job ladder effect in Section 5.1.

5 Madrian and Lefgren (2000) propose several criteria to generate longitudinal data using CPS surveys. They find that, when trying to match individuals from two March CPS surveys, the naive approach based on household numbers leads to a merge rate of about 71%, which is far from the merge rate of 100% that would obtain in the absence of mortality and migration. They investigate more elaborated algorithms but reach the conclusion that none of them can merge individuals who should merge without also merging individuals who should not. This difficulty is compounded by the fact that, as a longitudinal panel, the CPS is of very short duration.

6 Dustmann and Meghir (2005) document that the acquisition of general skills is important for skilled workers, whereas unskilled workers benefit primarily from being attached to a particular firm.

An influential approach proposed by Postel-Vinay and Robin (2002) consider that workers can bring potential employers into Bertrand competition. Employer competition generates upward-sloping wage profiles because it enables workers to gradually appropriate the output of their jobs. Two recent papers build on this wage-setting rule and combine it with human capital accumulation. Bagger et al. (2006) assume that wages are defined as piece rate contracts whose values are determined using the sequential auction model of Postel-Vinay and Robin (2002). Yamaguchi (2006) augments the sequential auction framework with bargaining as in Cahuc et al. (2006).

We focus instead on the Nash bargaining rule prevailing in the standard theory of unemployment so that wages follow changes in productivity. In that respect, our approach is more closely related to the paper by Nagypál (2007), which studies a model with both LBD and learning about match quality. Her analysis aims at disentangling the contributions of these two mechanisms. Her focus is quite different from ours: whereas Nagypál (2007) analyzes in great detail the hazard rate of job separation and the wage profile, we put greater emphasis on the aggregate wage distribution.

The additional features included in these three papers greatly complicate the analysis. This is why they all rely on simulation techniques to estimate their structural models. To the contrary, the framework proposed in this paper can be solved analytically and estimated by maximum likelihood. This reflects our focus on the relatively stylized but nevertheless very influential MP94 model.

Lastly, this paper is naturally connected to the large body of empirical research using Mincer equations to evaluate the rate of LBD. On the one hand, our model is too stylized to contribute to the debate about the relative importance of job tenure versus experience since it does not include general human capital. On the other hand, our structural framework allows us to quantify the aggregate impact of LBD. We find that it has a significantly positive effect on aggregate output but a small effect on employment.

1.2. Structure of the Paper

The rest of the paper is organized as follows. Section 2 discusses the model set-up and characterizes the equilibrium. The econometric procedure and the asymptotic properties of the estimates are detailed in Section 3. Section 4 describes the data and discusses the estimation results. Section 5 assesses the robustness of the estimates. In section 6, we introduce an aggregate matching function to close the model and evaluate the impact of LBD on the equilibrium. Section 7 concludes and the Appendix contains the proofs of the propositions.

2. THE MODEL

We consider a labor market with search frictions where jobs’ output is subject to random fluctuations. The set-up differs in three respects from the one proposed by Mortensen and Pissarides (1994): first, we allow initial productivities to vary across jobs; second, we assume that output follows a geometric Brownian motion; and finally we introduce LBD. Given that the CPS does

---

8 The related literature is too extensive to be comprehensively reported. An arbitrary sample includes the seminal paper by Jovanovic and Mincer (1981) and more recent contributions by Altonji and Shakotko (1987), Topel (1991), Altonji and Williams (2005), and Dustmann and Meghir (2005).
not contain information on the number of posted vacancies, the data will not allow us to estimate the parameters of the matching function. Thus we take as given the rate of contact between firms and job seekers, and postpone the introduction of the aggregate matching function to Section 6.

2.1. The Production Process

Consider a market in which homogeneous workers, who live forever, are either employed or looking for a job. Each competitive firm has one job which can be either filled or vacant. Firms use only labor to produce a unique multi-purpose good. When an unemployed worker meets a firm with a vacant job, they sample a positive output for their match. The initial productivity is a random draw from the exogenous distribution $F(\cdot)$, which is assumed to be continuously differentiable. In the remainder of the paper, we refer to $F(\cdot)$ as the distribution of job offers.

Both parties instantaneously observe the initial productivity. The firm can decide whether or not to make a job offer. If the firm ‘passes’ on the applicant, it does not incur any specific cost for doing so and continues to keep its vacancy open to other workers. Similarly, the worker can choose to refuse the job offer if he prefers to search for a better opportunity.

In the case where both parties decide to match, they immediately start to produce and output begins to fluctuate. We do not consider aggregate shocks, so that stochastic fluctuations are uncorrelated across jobs. The stochastic process that changes the idiosyncratic output is a geometric Brownian motion. Thus its law of motion is given by

$$\text{d}P_t^i = P_t^i(\xi - \sigma \text{d}B_t^i)$$

where $\text{d}B_t^i$ is the increment of a standard Brownian motion. The subscript $i$ indexes jobs. In the remainder of the text we will neglect it when not necessary. According to (1), the expected output at time $t + T$ of a job with current output $P_t$ is equal to $P_t e^{\xi T}$. Hence $\xi$ is the rate at which productivity on-the-job increases. The acquired skills are purely job-specific, so that workers become identical when they return to the unemployment pool. The parameter $\sigma$ measures the variance of the sample paths: the higher it is, the faster output fluctuates.

We also introduce an exogenous source of uncertainty such that jobs are forced out of business when hit by random shocks which arrive at the Poisson rate $\delta$. These shocks capture two type of separations: those that are not motivated by purely economic reasons such as retirement or partner reallocation, as well as those that follow from discontinuous and strongly negative shocks in profitability. Aside from their structural interpretation, exogenous separations are also needed for the following technical reason: given that Brownian motions are non-stationary processes, they do not converge to a well-defined long-run distribution. In the absence of exogenous shocks, the model is not well behaved since the aggregate wage distribution does not admit an ergodic representation.

The introduction of Brownian motions contrasts with the standard practice of considering Poisson processes. Whereas Brownian motions have continuous sample paths, Poisson processes are by definition discontinuous. It is explained in Prat (2006) why Brownian motions deliver more

---

9 Solely productivity shocks are considered in MP94 because Poisson processes have a constant probability of breaching the reservation threshold. This somewhat unrealistic prediction ensures that the aggregation procedure is always consistent. Note, however, that it is possible to recover our distinction in the original MP94 model. One simply has to interpret exogenous shocks as discrete jumps from current to extremely low productivity.
J. PRAT

accurate predictions about the hazard rate of job separation and the shape of the wage distribution. It is also shown in Prat (2006) how most of the statistics of interest can be derived in closed form using stochastic calculus. As we will see in Section 3, the convenient analytical properties of Brownian motions are crucial for the empirical implementation of the model.

We also assume that workers do not receive alternative job offers while employed. Thus we do not consider on-the-job search, so that trade in the labor market is completely separated from production. This restriction is imposed for technical reasons as it is notoriously involved to combine idiosyncratic uncertainty with on-the-job-search. These difficulties partly explain why empirical models of employers competition typically assume away idiosyncratic uncertainty. We make the converse assumption and leave to further research the task of devising a comprehensive framework.

2.2. Optimal Job Separation

Because trading in the labor market is a costly process, matched pairs have to share a quasi-rent. We assume a Nash bargaining rule whereby each party obtains a constant share of the job’s surplus at each point in time. The rent of each party is defined as the difference between the asset value obtained by participating in the match and the disagreement outcome of continued search. Since the two rents remain proportional, it cannot be the case that one is positive and the other negative. In other words, workers and firms always separate by common agreement.

Let \( U \) denote the steady-state expected value of search by an unemployed worker. The search effort costs the worker \( s \) and he meets at the flow rate \( \lambda \) a firm with an open vacancy. The contact leads to a match if the initial output drawn from the distribution of job offers \( F(\cdot) \) is at least as great as the reservation output \( R \). Under the assumption that workers are risk-neutral and that they discount the future at rate \( r \), \( U \) satisfies the following equation:

\[
    rU = -s + \lambda \int_{R}^{+\infty} \beta S(P)dF(P)
\]

where \( \beta \) denotes the worker’s bargaining power. As opposed to the labor force whose size is fixed and normalized to one, new firms enter the market until arbitrage opportunities are exhausted. Thus free-entry ensures that the firm’s outside option is equal to zero and the total surplus of the match can be decomposed in the following way:

\[
    rS(P_t) = P_t - rU - \delta S(P_t) + \frac{E}{d\delta}[dS(P_t)]
\]

where it is assumed that firms discount the future at the same rate than workers. The term \( \delta S(P_t) \) corresponds to the loss incurred by both parties when the job is hit by an exogenous destruction shock. The surplus also evolves through time due to output fluctuations. In the deterministic case,

\[^{10}\text{Nagypál (2005) establishes that the wage distribution cannot be expressed in closed form when workers search on-the-job and uncertainty is modelled using a diffusion process.}\]

\[^{11}\text{The Nash bargaining solution assumes away the difficulty of relating wages to job-specific human capital. As explained in Felli and Harris (1996), wages increase with human capital to the extent that workers are able to appropriate some of the return. As specific human capital enhances the worker’s productivity only in its current working place, it is not clear why the worker should receive any of the return on it. We do not address this issue and instead follow the typical practice of considering that each party receives a fixed share of the expected surplus at any point in time.}\]
one can immediately solve for $S(P_t)$ by combining equations (2) and (3). In the stochastic case, we have to solve the partial differential equation satisfied by $S(P_t)$.

**Proposition 1** The expected surplus of a match with current output $P$ and reservation output $R$ is given by

$$S(P,R) = \frac{P}{r + \delta - \zeta} - \left( \frac{1}{r + \delta} \right) rU - \left[ \frac{R}{r + \delta - \zeta} - \left( \frac{1}{r + \delta} \right) rU \right] \left( \frac{P}{R} \right)^\alpha$$  \hspace{1cm} (4)

where $\alpha$ is the negative root of the following quadratic equation

$$\alpha^2 \sigma^2 + \alpha \left( \zeta - \frac{\sigma^2}{2} \right) - r - \delta = 0.$$

One can solve for the optimal reservation output using a standard first-order condition with respect to $R$. The resulting solution is homogeneous of degree zero in $P$, so that $R$ is identical across matches, as one should expect. Its optimal value is given by

$$R = \left( \frac{\alpha}{\alpha - 1} \right) \left( \frac{r + \delta - \zeta}{r + \delta} \right) rU$$  \hspace{1cm} (5)

From the definition of $\alpha$, it is easily seen that $R$ is upper-bounded by the opportunity cost of employment $rU$. Within the neighborhood of $R$, output is too low to cover costs but the job might turn profitable again. This is why the worker and the firm procrastinate up to the point where the value of waiting equals the operational losses. To the contrary, there is no labor hoarding when productivity is constant. Then the reservation output is equal to the opportunity cost of employment.

**2.3. The Equilibrium**

This section characterizes the equilibrium rate of unemployment and the joint distribution of job spells and wages. It will be shown in Section 3 that these statistics have closed-form solutions when the distribution of job offers is lognormal. For the moment we keep the analysis as general as possible by not imposing any parametric assumption on $F(\cdot)$. First of all, we notice that the Nash-bargaining problem is satisfied if and only if wages are such that

$$w(P_t) = \beta P_t + (1 - \beta)rU$$  \hspace{1cm} (6)

12 As in deterministic search models, when differentiating the surplus with respect to the reservation output $R$, we treat the opportunity cost of employment $rU$ as exogenously given. Thus we do not differentiate the lower bound of the integral in equation (2). This is because the worker’s outside option does not depend on the optimal stopping rule in the current match.

A standard first-order condition holds because the optimal reservation output is independent of current productivity. This property has been established by McDonald and Siegel (1986) for parabolic differential equations, such as the one satisfied by $S(P)$ (see equation (16)). A heuristic derivation of this result can be found in the companion paper Prat (2006). Note that one can also use the standard smooth-pasting condition, as shown in the proof of Proposition 1.

---

Copyright © 2009 John Wiley & Sons, Ltd.  
*J. Appl. Econ.* (2009)  
DOI: 10.1002/jea
Since wages follow from outputs by a location transformation, the discussion can be restricted to the output distribution. The derivations are based on the premise that the labor market is in steady state so that job flows are constant and balance at all time.\textsuperscript{13} The statistics of interest are derived using a progressive approach: starting from the most informative one, namely the joint distribution of job spells and output, we aggregate it step by step in order to obtain the rate of unemployment.

**Proposition 2** The measure of jobs with output $x$ and tenure $T$ is given by

$$\nu(x, T) = u\lambda \left( \int_{R}^{+\infty} \psi(x, T; P)dF(P) \right)$$

where $u$ denotes the rate of unemployment. For $x \in [R, +\infty)$, the function $\psi(x, T; P)$ is

$$\psi(x, T; P) = \left( \frac{e^{-\beta T}}{x} \right) e^{-\frac{1}{2} \left( \frac{\ln(x) - \ln(P) - \alpha T}{\sigma \sqrt{T}} \right)^2} - \left( \frac{R}{P} \right) \frac{2\mu}{\sigma^2} e^{-\frac{1}{2} \left( \frac{\ln(x) + \ln(P) - 2\ln(R) - \mu T}{\sigma \sqrt{T}} \right)^2}$$

where $\mu = \zeta - \frac{\sigma^2}{2}$.\textsuperscript{14}

The first term on the right-hand side of (7) measures the number of contacts between job seekers and firms. The function $\psi(x, T; P)$ is the conditional measure of jobs with current output $x$ and tenure $T$ given initial output $P$. From the set of sample paths starting from $P$ and reaching $x$ after tenure $T$, it deduces all those that breach the separation threshold $R$.\textsuperscript{14} Since initial productivities are drawn from the distribution of job offers, the unconditional measure is obtained integrating $\psi(x, T; P)$ with respect to $F(P)$. The integration is performed from $R$ up to infinity because contacts lead to matches solely when $P$ is above $R$. The measure of jobs with a given current output is readily obtained from (7) after having integrated tenure from 0 up to infinity. The following proposition shows that the resulting integral can be expressed analytically.

**Proposition 3** The measure of jobs with output $x$ is given by

$$\nu(x) = u\lambda \left( \int_{R}^{+\infty} \varphi(x; P)dF(P) \right)$$

\textsuperscript{13} Although conventional for obvious technical reasons, the steady-state assumption is actually quite restrictive, especially because it is imposed over a large number of years. We refer to Jolivet et al. (2006) for empirical evidence in its favor.\textsuperscript{14} Note that an econometrician who observes the workers’ wages at different points in time of their jobs spells could use equation (8) to compute the likelihood of their sample paths.

Copyright © 2009 John Wiley & Sons, Ltd.  
*J. Appl. Econ.* (2009)

DOI: 10.1002/jae
The rate of learning-by-doing

where the function \( \varphi(x; P) \) is defined for \( x \in [R, +\infty) \) and is equal to

\[
\varphi(x; P) = \begin{cases} 
    P^{-1} \left( \frac{x}{P} \right)^{\frac{\mu - \gamma}{\sigma^2}} \left( 1 - \frac{R}{P} \right)^{\frac{2\gamma}{\sigma^2}} ; & \text{if } x > P \\
    P^{-1} \left( \frac{x}{P} \right)^{\frac{\mu + \gamma}{\sigma^2}} \left( 1 - \frac{R}{x} \right)^{\frac{2\gamma}{\sigma^2}} ; & \text{if } x \in [R, P] 
\end{cases}
\]

(10)

with \( \gamma = \sqrt{\mu^2 + 2\delta \sigma^2} \).

In a similar way, the aggregate rate of employment follows integrating equation (9) from \( R \) up to infinity. Again the calculation leads to a closed-form expression that is given in Proposition 4. The expression is reminiscent of the equilibrium rate of unemployment under certainty. Actually, when uncertainty vanishes so that \( \sigma \) goes to zero, the term \( \frac{R}{P} \sigma^2 \) also converges to zero. Thus the expression of \( u \) converges to the standard solution under certainty.

**Proposition 4** The equilibrium rate of unemployment is equal to

\[
u = \frac{\delta}{\delta + \lambda \int_{R}^{+\infty} \left( 1 - \frac{R}{P} \right)^{\frac{\mu + \gamma}{\sigma^2}} dF(P)}
\]

(11)

In this section we have derived the statistics that will be useful for the econometric estimation. The next section details the econometric procedure and analyzes the property of the likelihood function.

### 3. ESTIMATION PROCEDURE

We now discuss how to estimate the model’s parameters. Our structural framework leads to similar identification problems to the deterministic search model. We therefore adopt the identification strategy laid out by Flinn and Heckman (1982). First, the truncation of the wage offer distribution implies that the model is fundamentally underidentified. We have to impose a parametric restriction on the distribution of job offers \( F(\cdot) \) such that it belongs to the class of distribution functions which can be recovered from truncated observations. Section 3.3 discusses this problem in further detail.

The searching costs \( s \) and discount rate \( r \) are not individually identified because they both enter the likelihood function only through their impact on \( R \). For this reason, we treat the reservation output \( R \) as if it were a parameter and estimate it directly. This may seem inconsistent with the fact that \( R \) is not an exogenous primitive of the model but rather an endogenous variable. Thus one may wonder whether we are actually discarding the information contained in the optimal stopping
rule (5). However, when all parameter estimates have been obtained, equation (5) remains central to the estimation procedure as it allows us to ‘back out’ the locus relating $r$ and $s$. Thus one should think of equation (5) as a generalization of the reservation rule in deterministic settings, the only difference being the additional term $\alpha/(\alpha - 1)$ which captures the value of waiting. Given that the point estimates of the identified parameters pin down the value of $\alpha$, we can follow, without loss of generality, the same procedure as Flinn and Heckman (1982).

Lastly, the bargaining power $\beta$ has to be fixed prior to the estimation. Although $\beta$ is theoretically identified due to the highly nonlinear likelihood function, trials show that its estimates often do not converge to an interior solution. In the absence of information about firms’ profits, it is not surprising that our dataset does not allow us to recover both sizes and allocations of job surpluses. This well-known difficulty is now gradually overcome by research based on matched employer–employees data (see Cahuc et al., 2006). Given the one-sided nature of the CPS data, we stick to the usual practice of assuming symmetric bargaining and then perform some robustness tests with respect to $\beta$.

### 3.1. The Likelihood Function

Following these preliminary steps, the likelihood of the sample can be expressed as a function of the remaining set of parameters. We slightly restrict the generality of the problem by assuming that the distribution of job offers $F(\cdot)$ can be completely parametrized in terms of a finite-dimensional vector $\Omega$ so that the set of estimated parameters $\Theta = \{\xi, \delta, \sigma, \Omega, R, \lambda\}$.

The likelihood of the sample is computed as follows. Let $Y$ denote the set of observations, so that $Y \equiv \{y^1, y^2, \ldots, y^n\}$, where $n$ is the total number of workers in the sample. The individual observations are defined using four variables: $e^i$, $w^i$, $T^i$, $\tau^i$. Let $e^i$ denote an indicator function which takes value 0 when worker $i$ is currently unemployed, 1 when he is employed but fails to report his tenure, and 2 when worker $i$ is employed and reports both his wage and job spell. The variables $w^i$ and $T^i$ are the current hourly wage and job tenure. In the case where worker $i$ fails to report the length of his job spell, i.e. $e^i = 1$, $T^i$ is obviously ignored. If worker $i$ is currently searching for a job, i.e. $e^i = 0$, $y^i$ is set equal to the unemployment duration $\tau^i$. The likelihood function is therefore made of three distinct components. The individual contribution of a job searcher is equal to the density associated with an ongoing unemployment spell of length $\tau$ conditional on unemployment times the probability of observing an unemployed worker:

\[
\int_{+\infty}^{+\infty} \left(1 - \frac{R}{P} \frac{e^{\chi + \tau}}{\sigma^2} \right) \, dF(P)
\]

Note that equation (5) also implies that estimating the reservation $R$ is formally equivalent to estimating the opportunity cost of employment $\tau U$.
THE RATE OF LEARNING-BY-DOING

where \( F(P) \equiv 1 - F(P) \). The likelihood of observing an employed worker paid wage \( w \) is given by \( \nu(x(w)) \). The expression can be further decomposed reinserting (11) into (9) to obtain

\[
    f(e = 1, w) = \nu(x(w)) = \frac{\delta \lambda \left( \int_{R}^{+\infty} \phi(x(w); P)dF(P) \right)}{\delta + \lambda \int_{R}^{+\infty} \left( 1 - \left( \frac{R}{P} \right)^{\mu+y} \right) dF(P)}
\]

Note that output is defined as a function of the observed wage. Its implicit value follows from combining (5) with (6). Similarly, the joint likelihood of observing a worker paid wage \( w \) with a job tenure equal to \( T \) is given by

\[
    f(e = 2, w, T) = \nu(x(w), T) = \frac{\delta \lambda \left( \int_{R}^{+\infty} \psi(x(w), T; P)dF(P) \right)}{\delta + \lambda \int_{R}^{+\infty} \left( 1 - \left( \frac{R}{P} \right)^{\mu+y} \right) dF(P)}
\]

Putting together these three components, the log-likelihood for the observed sample reads

\[
    \ln L(\Theta, Y) = n \left[ \ln \lambda + \ln \delta - \ln \left( \delta + \lambda \int_{R}^{+\infty} \left( 1 - \left( \frac{R}{P} \right)^{\mu+y} \right) dF(P) \right) \right] \\
    + n_S \ln \bar{F}(R) - \lambda \bar{F}(R) \sum_{i \in S} \tau_i + \sum_{i \in W} \ln \left( \int_{R}^{+\infty} \phi(x(w_i); P)dF(P) \right) \\
    + \sum_{i \in H} \ln \left( \int_{R}^{+\infty} \psi(x(w_i), T_i; P)dF(P) \right)
\]

where \( n_S \) is the number and \( S \) is the set of indices of job searchers in the sample, \( W \) is the set of indices of employees who only report their current wages and \( H \) is the set of indices of employees who report both wages and job spells. Although the parameters \( \xi \) and \( \Omega \) do not appear in the analytical expression of the likelihood function, they are implicitly identified: \( \xi \) determines the values of \( \mu, \gamma, \phi(\cdot) \) and \( \psi(\cdot) \), while the parametric vector \( \Omega \) obviously influences \( F(\cdot) \). Note that, as discussed before, the reservation output is treated as a primitive parameter of the model.

### 3.2. Asymptotic Properties

The likelihood function is continuously differentiable and its parameters belong to a compact support. Yet it does not satisfy all the standard requirements for a well-behaved likelihood function since the support of the distribution is a function of the parameters. Furthermore, the density functions \( f(e, w(R)) \) and \( f(e, w(R), T) \) are both equal to zero at the reservation wage. Hence the lowest reported wage is not a super-consistent estimator of the reservation wage, so that we
have to rely on a different estimation method from the two-step approach proposed by Flinn and Heckman (1982).

Our problem bears similarities to the estimation of optimal production frontiers. Optimal frontiers models also imply that the range of observations changes with the parameters being estimated. Moreover, they share with our model the additional implication that agents are never exactly on the optimal frontier.16 As firms cannot perfectly counteract random perturbations, they remain within the neighborhood of the optimal combination of inputs without ever achieving it perfectly. Given that the estimation of optimal frontiers is one of the most popular areas of applied econometrics, great attention has been devoted to the econometric solutions for this kind of problem. In an influential paper, Greene (1980) showed that when the density at the parameter-dependent boundary is equal to zero standard regularity conditions need not be satisfied in order to produce standard asymptotic distribution results. We adapt Greene’s proof to our set-up and establish that the estimators satisfy asymptotic normality under standard requirements.

**Proposition 5** Suppose that: (i) the parameter space $\Gamma$ is compact and contains an open neighborhood of the true value $\Theta_0$ of the population parameter; (ii) the distribution of job offers $F(P)$ is continuously differentiable. Then the maximum likelihood estimator

$$\hat{\Theta} = \arg \max_{\Theta \in \Gamma} \ln L(\Theta, Y)$$

converges in probability to $\Theta_0$ so that $\sqrt{n}(\hat{\Theta} - \Theta_0) \rightarrow d N(0, H^{-1}JH^{-1})$, where $H$ is the Hessian of the likelihood function and $J$ is the information matrix.

The proof of Proposition 5 relies on the fact that $f(1, w(R))$ and $f(2, w(R), T)$ are both equal to zero. This property justifies interchanging the order of integration and differentiation so that the asymptotic property of the estimators can be characterized by linear approximation. Our problem is slightly less standard than the one considered by Greene (1980) because the derivatives of the density functions with respect to $\Theta$ are not equal to zero when evaluated at the reservation output. Thus interchanging the order of integration and differentiation is justified solely for the first derivative. This is why the hessian matrix $H$ is not equal to $-J$, so that the asymptotic covariance matrix cannot be simplified and set equal to $J^{-1}$. In any case, as explained in Newey and McFadden (1994), the information matrix equality is not essential to asymptotic normality. The only complication is technical, as reflected by the intricate form of the asymptotic variance.

### 3.3. Lognormal Distribution of Job Offers

We have characterized the estimation procedure for general distributions of job offers. The econometric implementation of the model requires to narrow the analysis to a particular family of distributions. We hereafter assume that $F(\cdot)$ is lognormal. Lognormal distributions are commonly considered because they satisfy the ‘recoverability condition’ defined by Flinn and Heckman

---

16 This feature directly follows from the structure of the stochastic process and does not depend on whether one assumes that separations occur below or at the reservation threshold. This is because sample paths are not of bounded variation: for any small time interval $[t, t + \epsilon]$, future outputs will be such that $P_t < P_t$ for some $t \in [t, t + \epsilon]$. In other words, whenever the worker is indifferent between staying and separating, he will strictly prefer to separate ‘an instant later’. It follows by continuity that the left limit of the density at the reservation threshold must be equal to zero.

Copyright © 2009 John Wiley & Sons, Ltd.  
*J. Appl. Econ.* (2009)  
DOI: 10.1002/jae
(1982), meaning that their location and scale parameters can be recovered from truncated observations. The class of functions which satisfy the ‘recoverability condition’ also encompasses, among others, gamma distributions.\textsuperscript{17} Lognormality is eventually justified by its good fit of the data. In our case this assumption has a more crucial role. Given the intricate expression of the likelihood function, there is little hope to derive it in closed form. Solely when initial productivities are lognormally distributed in the population does the likelihood function admit an analytical expression so that approximation errors due to numerical integrations can be avoided.\textsuperscript{18} Given its length, we do not include the expression of \( L(\Theta) \) in the body of the paper.

**Proposition 6** Under the assumption that the initial productivities are drawn from a lognormal distribution, so that

\[
dF(P) = e^{-\frac{1}{2} \left( \frac{\ln P - \Sigma}{\xi} \right)^2} \frac{1}{P \xi \sqrt{2\pi}} dP,
\]

the likelihood functions \( L(\Theta) \) has a closed-form solution. The resulting expression is reported in the Appendix.

### 4. EMPIRICAL RESULTS

#### 4.1. Data

We estimate the model using cross-sectional data. Whereas most surveys systematically ask unemployed workers to report the time they have been searching for a job, employees are rarely asked the length of their job spells. As a result, data on job durations are scarcer than data on unemployment durations. A notable exception is the January/February supplement of the Current Population Survey. The CPS is structured as a rotating panel with 4 months of participation, 8 months without interviews, and 4 more months of participation after which the household is taken out of the panel. In January or February, the current wages and job spells of the Outgoing Rotation Groups\textsuperscript{19} are collected. More precisely, employees are asked the following question: ‘How long have you been working continuously for your present employer?’

The job tenure supplement therefore provides data on both job spells and wages for a supposedly random sample of one-quarter from the January 2004 CPS. We use data on males and females with an age between 20 and 65 years. Since our model does not include a state of non-participation to the labor market, we have restricted our sample to individuals who indicated that they were currently employed or actively searching for a job. For the same reason, we have excluded individuals observed as self-employed, working part time or employed in the non-civilian labor

\textsuperscript{17} See Flinn (2006) for a careful discussion of the class of functions which satisfy the ‘recoverability condition’.

\textsuperscript{18} The derivation of an analytical expression for the likelihood function is made possible by the fact that geometric Brownian motions are also lognormally distributed.

\textsuperscript{19} The Outgoing Rotation Groups are composed of households who answered their outgoing interview, that is, their 4th and 8th interview. We restrict our attention to the Outgoing Rotation Groups because usual weekly hours/earning questions are asked only at households in their outgoing interviews. In our case, the Outgoing Rotation Groups include respondents that entered the panel in October 2002 or 2003.
force. After excluding observations with missing wage data, we restricted the sample to workers with a high school graduation diploma or less. Finally we have trimmed the subsample by excluding observations below the bottom percentile of the wage distribution. The trimming is particularly important for the estimation of the reservation wage since it allows us to avoid implausibly low estimates due to measurement errors.\footnote{The trimming of the data is especially useful given the nature of the observations. For the workers who are not paid on an hourly basis, we have divided their gross weekly wage by their usual hours of work per week in order to impute their hourly wage. This computation interacts potential measurement errors and for some observations leads to implausibly low hourly wages.}

Descriptive statistics are reported in Table I. It contains statistics for jobs with a reported tenure below 1 year. Their wage distribution will be close to the estimated distribution of job offers because the estimation procedure approximates the latter using observations with short job spells. This points to potential bias since the distribution of wages among job entrants has a lower mean than the distribution of wages among workers with less than 1 year of tenure.\footnote{We are able to identify individuals entering the employment pool because the CPS tenure supplements ask workers whether they were employed a year ago. An alternative approach consists in checking the employment status of the individuals in the December 2003 file. This methodology has the advantage of identifying workers who entered employment between December 2003 and January 2004. However, as discussed in Madrian and Lefgren (2000), matching CPS}

Table I. Descriptive statistics

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td>40.475</td>
<td>39.841</td>
<td>41.358</td>
</tr>
<tr>
<td></td>
<td>(11.403)</td>
<td>(11.316)</td>
<td>(11.467)</td>
</tr>
<tr>
<td><strong>Working week (hours)</strong></td>
<td>41.411</td>
<td>42.324</td>
<td>40.163</td>
</tr>
<tr>
<td></td>
<td>(5.750)</td>
<td>(6.720)</td>
<td>(3.723)</td>
</tr>
<tr>
<td><strong>Average spells (months)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>4.537</td>
<td>4.438</td>
<td>4.704</td>
</tr>
<tr>
<td></td>
<td>(5.011)</td>
<td>(4.956)</td>
<td>(5.117)</td>
</tr>
<tr>
<td>Job</td>
<td>92.586</td>
<td>96.773</td>
<td>86.601</td>
</tr>
<tr>
<td></td>
<td>(97.160)</td>
<td>(101.926)</td>
<td>(89.604)</td>
</tr>
<tr>
<td><strong>Hourly wage</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All jobs</td>
<td>13.777</td>
<td>14.932</td>
<td>12.199</td>
</tr>
<tr>
<td></td>
<td>(6.982)</td>
<td>(7.458)</td>
<td>(5.922)</td>
</tr>
<tr>
<td>Job spell &lt; 1 year</td>
<td>11.829</td>
<td>12.801</td>
<td>10.394</td>
</tr>
<tr>
<td></td>
<td>(6.308)</td>
<td>(6.843)</td>
<td>(5.109)</td>
</tr>
<tr>
<td>Entrants</td>
<td>10.301</td>
<td>11.059</td>
<td>9.439</td>
</tr>
<tr>
<td></td>
<td>(4.486)</td>
<td>(5.140)</td>
<td>(3.436)</td>
</tr>
<tr>
<td>Number of observations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labor market position</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>413</td>
<td>260</td>
<td>153</td>
</tr>
<tr>
<td>Employed</td>
<td>4336</td>
<td>2504</td>
<td>1832</td>
</tr>
<tr>
<td>Reported job spells</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total number</td>
<td>3676</td>
<td>2163</td>
<td>1513</td>
</tr>
<tr>
<td>Job spell &lt; 1 year</td>
<td>535</td>
<td>319</td>
<td>216</td>
</tr>
<tr>
<td>Entrants</td>
<td>154</td>
<td>82</td>
<td>72</td>
</tr>
<tr>
<td>Sample size</td>
<td>4749</td>
<td>2764</td>
<td>1985</td>
</tr>
</tbody>
</table>

Note: Standard deviations in parentheses.
the data is easily explained by on-the-job search as employees select offers which are above their current wage. Given that our model excludes on-the-job search, the job-ladder effect is ignored and consequently our estimates of the location and scale parameters of the distribution of job offers will be biased upwards. As shown in Section 5, where we analyze the robustness of the estimation procedure, this leads to downward biases for the estimates of the rate of LBD.

The non-parametric kernel estimates of the three wage densities are reported in Figure 1. As expected, the distribution of wages among job entrants is located to the left and exhibits slightly less dispersion than the distribution of wages among workers with less than 1 year of tenure. On the contrary, the dispersion of the aggregate wage distribution is much higher. This feature fits the model well since Brownian motions are diffusion processes, so that their distributions become increasingly dispersed as time elapses.

4.2. Estimates

The estimated parameters and their standard deviations are reported in Table II. We also estimate the deterministic model using the procedure devised by Flinn and Heckman (1982). Table II makes clear that their approach is nested into the one proposed in this paper. Note that, in addition to the rate of LBD $\xi$ and the variance parameter $\sigma$, the model also allows us to estimate the value and standard deviation of the reservation wage $w_r$. Conversely, the value reported for the deterministic model corresponds to the lowest wage in the sample so that its standard deviation is not defined.
Table II. Model estimates

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\beta = 0.5$</th>
<th>$\beta = 0.4$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stochastic</td>
<td>Deterministic</td>
</tr>
<tr>
<td>$w_r$</td>
<td>3.76</td>
<td>4.11</td>
</tr>
<tr>
<td></td>
<td>(0.180)</td>
<td>(—)</td>
</tr>
<tr>
<td>$\xi$</td>
<td>0.0204</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td></td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.0350</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td></td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.133</td>
<td>0.133</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>$\Sigma$</td>
<td>2.86</td>
<td>3.01</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>$\xi$</td>
<td>0.488</td>
<td>0.526</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>1.95</td>
<td>1.95</td>
</tr>
<tr>
<td></td>
<td>(0.073)</td>
<td>(0.081)</td>
</tr>
<tr>
<td>$\ln(L)$</td>
<td>−28,874</td>
<td>−29,045</td>
</tr>
</tbody>
</table>


The estimates of $\Sigma$ and $\xi$ imply that the mean and dispersion of the distributions of job offers are higher in the deterministic model than in the stochastic model. This result is intuitive since the deterministic model is based on the premise that the distribution of job offers and the aggregate distribution are one and the same. Conversely, the estimation of the stochastic model sets the parameters $\Sigma$ and $\xi$ so as to fit the distribution of wages among jobs with a short tenure and so yields smaller values for both parameters.

The exogenous rate of job destruction $\delta$ is rather similar. This is somewhat surprising since the stochastic model generates endogenous separations. Instead, one might expect that the rate of exogenous job destruction would be significantly lower in a stochastic environment. For the very small estimate of the variance parameter, however, endogenous separation is a marginal phenomenon. Accordingly, the separation rates are quite similar in both models, with an estimated average length of a job close to 90 months.

The estimated values of $\lambda$ imply that job searchers receive a job offer every 6 months. Although the model overestimates the average unemployment duration, the predicted unemployment rate is equal to 6.4%, whereas its value in the data is 8.7%. These opposite biases suggest that the sample does not completely satisfy the stationarity assumption, most probably because of a significant and recent entry of workers into the labor force.

We now turn our attention to the estimates that are specific to the stochastic model. First, we note that the variance parameter $\sigma$ is quite low. With a standard deviation close to 3.5%, the model predicts that the sample paths are nearly deterministic. Nonetheless, the estimated value of $\sigma$ is highly significant. This is because one cannot set $\sigma$ to zero and at the same time estimate $\xi$. In a deterministic environment the lower bound of the joint distribution is an increasing function

---

23 This finding contrasts with the result in Nagypal (2007), where exogenous reasons account for at most 50% of all separations.

24 The rather high rate of unemployment is explained by the fact that the sample contains solely workers with a high school graduation diploma or less.
of tenure and consequently the likelihood of observing a worker with a seniority equal to $T$ and a wage inferior to $e^{T \bar{w}_r}$ is zero. Given that the sample contains such observations as long as $\zeta$ significantly differs from zero, the deterministic model collapses to the case where the job offers and aggregate distributions are indistinguishable. Thus there is a fundamental link between the introduction of uncertainty and the estimation of the LBD rate, the former being necessary to implement the latter.

The estimated rate of LBD is close to 2%. This translates into a smaller wage growth of around 1.75% per year because the constant outside option accounts for a substantial share of wages. The model predicts that 10 years of tenure raises the average wage by about 18.8%, so that our estimate of the cumulative returns to tenure lies in the range separating the low returns obtained by Altonji and Williams (2005) from the high returns obtained by Topel (1991). Note that, given the low value of the variance, non-random selection in who acquires seniority is not a source of important bias.

Table II also contains the estimates when $\beta$ is equal to 0.4 instead of 0.5. A bargaining power of 0.4 is in the range of the estimates obtained by Flinn (2006) using wage-share information for CPS sample members between 16 and 24 years of age. Of all the estimates, the most sensitive to changes in $\beta$ are the parameters determining the shape of the distribution of job offers. A lower value of $\beta$ implies that the worker receives a smaller share of the job’s output, so the observed dispersion of wages must result from a higher degree of productivity dispersion. This is why both $\Sigma$ and $\xi$ substantially increase.

Conversely, the estimate of $\zeta$ is quite robust to variations in bargaining power. It slightly increases because workers must learn at a higher rate to benefit from a given pay raise.

We now consider the ability of the model to fit the sample information. Of most interest to our analysis are its predictions about the joint distribution of wages and job spells. The data for jobs with a tenure below 1, 5 and 10 years as well as the aggregate distribution are reported against their simulated counterparts in Figure 2. The figures illustrate the ability of the model to fit the gradual increase in dispersion of the cross-sectional distributions. The model matches accurately the right tails of the distributions. This a classical test for models of wage dispersion due to the ‘heavy tail’ property of the data. As explained in Prat (2006), the cross-sectional distribution aggregates underlying distributions with right tails of Pareto functional form. It is therefore not surprising that the model easily fits the wage distribution at high quantiles.

We also report in each panel the $p$-value of the $\chi^2$ test for the goodness-of-fit of the simulated distributions. Equality of the observed and simulated distributions is accepted for jobs below 5 and 10 years of tenure and marginally rejected (at the 5% level) for the aggregate distribution. On the other hand, the estimated distribution among workers with less than 1 year of tenure fails the test. Careful inspection of the upper-left panel shows that its mean is slightly higher than in the data.

---

25 An alternative strategy is to introduce measurement errors. As discussed in the Introduction, it has the drawback of neglecting the interactions between job separation and LBD. Furthermore, the likelihood function with normally distributed measurement errors cannot be expressed in closed form. Thus, although this alternative strategy is arguably more stylized, it has to be implemented through numerical integration.

26 On the other hand, given that our structural model ignores returns to labor market experience and their positive correlation with job tenure, our estimates are likely to be affected by a positive bias. This may explain why we find higher returns than Altonji and Williams (2005) in spite of the positive non-random selection bias affecting their OLS estimates.

27 Note that the dispersion of the distribution of job offers does not vary with the estimated values for $\sigma$, $\zeta$ and $\delta$. This is why the decrease in $\beta$ has to be compensated by an increase in the dispersion $\xi$ and average value $\Sigma$ of productivities in new jobs.
Figure 2. Wage densities at different job spells. This figure is available in color online at www.interscience.wiley.com/journal/jae

This implies that the simulated distributions tend to be a little less responsive to changes in tenure. The next section proposes an alternative estimation procedure which reduces this discrepancy and evaluates its impact on $\zeta$.

Since we have excluded on-the-job-search, one may wonder whether the fit of the wage distribution is achieved at the expense of the turnover process. To address this potential concern, we report in Figure 3 the actual distribution of job spells together with the structural estimation. While not as convincing as for wages, the simulation is still reasonably close to the data. Given that endogenous separations are quite rare, the stochastic and deterministic models have very similar predictions about the distribution of job spells. Thus the hazard rate of job separation is almost flat. This result is somewhat disappointing since it is shown in Prat (2006) that the framework estimated in this paper has the potential to fit the hump-shaped hazard rate of job separation that is observed in the data. Unfortunately, this would require to set the idiosyncratic variance $\sigma$ to a higher value than the one resulting from the estimation.

Copyright © 2009 John Wiley & Sons, Ltd.

J. Appl. Econ. (2009)
DOI: 10.1002/jae
5. ROBUSTNESS

This section assesses the robustness of the estimation procedure. First, we focus on the biases induced by the exclusion of on-the-job search. Then we analyze the role played by exogenous separations.

5.1. On-the-Job Search

A full treatment of on-the-job search would require a comprehensive model to be devised. Given that such a project is beyond the scope of this paper, we try to partly control for the impact of the job ladder effect on the estimation of the distribution of job offers. By focusing on employees who were unemployed a year ago, we can infer the actual wage distribution among job entrants. Thus we can set $\Omega = \{\Sigma, \xi\}$ so as to fit the entrants distribution for every given value of the vector of remaining parameters $\Theta_1 \equiv \{\xi, \delta, \sigma, R, \lambda\}$ and then maximize the sample likelihood with respect to $\Theta_1$.

The resulting estimators are not 'two-step' estimators as the optimal parametric vector $\Omega = \{\Sigma, \xi\}$ changes with $\Theta_1$. In other words, we cannot estimate $\Omega$ independently because the mapping between wages and productivity depends on $\Theta_1$. Instead we define $\Omega(\Theta_1, Y_1)$ as a function of $\Theta_1$ and of the subset of observations $Y_1 \subseteq Y$ used to infer the shape of the distribution of job offers.

---

28 See Christensen et al. (2005) for an early example of such an approach.
The procedure shares some formal similarities with concentrating the likelihood function but it is substantially different since we do not set $\Omega(\theta_1, Y_1)$ to maximize the likelihood of the sample but instead to approximate available information about the distribution of job offers. This approach is particularly justified if one suspects that the model is somehow misspecified. It is well known that, in such cases, full-information estimation can be a source of significant bias. By constraining $\Omega$ to fit the entrants distribution and imposing its value afterwards, we are able to reduce the size of the bias.

We call the values resulting from this procedure restricted estimates. They are reported in Table III for three distinct restrictions where $Y_1$ contains either: (i) the wages of job entrants that were unemployed a year ago, as documented in the January 2004 tenure supplement; (ii) the wages of job entrants that were unemployed a month ago, as inferred from matched monthly files; (iii) the wages of workers with less than 1 year of tenure. In the first two cases, the experiment gives us a sense of how estimates are biased down by the job-ladder effect. As expected, $\Sigma$ and $\xi$ significantly decrease. This leads to a noticeable increase in $\zeta$: given that the wage distribution among job entrants has a lower mean than the one resulting from the full estimation, the fit of the sample information is achieved through a higher rate of LBD. The estimated value of 3.4% for $\zeta$

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\beta = 0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_r$</td>
<td>3.79</td>
</tr>
<tr>
<td></td>
<td>(0.398)</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.132</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
</tr>
<tr>
<td>$\Sigma$</td>
<td>2.67</td>
</tr>
<tr>
<td></td>
<td>(0.037)</td>
</tr>
<tr>
<td>$\xi$</td>
<td>0.477</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>1.94</td>
</tr>
<tr>
<td></td>
<td>(0.082)</td>
</tr>
<tr>
<td>$\ln L$</td>
<td>$-29,009$</td>
</tr>
</tbody>
</table>

The restricted estimates are equivalent to concentrating the likelihood function so that it coincides with the full estimation procedure.

Since the properties used in the proof of Proposition 5 are satisfied, the estimators are still consistent and normally distributed, although not efficient. The asymptotic variance of $\tilde{\theta}$ is derived using the delta method so that

$$\text{var}(\tilde{\theta}) = \left[ \frac{\partial \Omega(\theta_1, Y_1)}{\partial \theta_1^T} \right]_{\theta_1 = \hat{\theta}_1} (H^{-1}JH^{-1}) \left[ \frac{\partial \Omega(\theta_1, Y_1)}{\partial \theta_1^T} \right]_{\theta_1 = \hat{\theta}_1}$$

The matched files are taken from the dataset constructed by Christian Haefke, Marcus Sonntag and Thijsvan Rens. See their working paper (Haefke et al., 2008) for a thorough description of the matching procedure.

THE RATE OF LEARNING-BY-DOING

should be interpreted as an upper bound since the job-ladder effect is excluded from the distribution of job offers but not from the aggregate distribution. Accordingly, a complete evaluation would require explicit modelling of on-the-job search and the restricted estimation suggests that this would lead to substantial adjustments of the estimates.

In the second case $\Omega(\theta_1, Y_1)$ fits the distribution of wages among workers with less than 1 year of tenure. As can be seen from Figure 2, this restriction is motivated by the fact that the full estimation uses $\Omega$ to improve the fit of the aggregate distribution at the cost of slightly overestimating the mean of the wage distribution among jobs with a tenure below 1 year. Controlling for this bias leads to a small but noticeable increase in $\zeta$.

5.2. Exogenous vs. Endogenous Separations

A surprising outcome of the estimation procedure is the low value of the variance parameter $\sigma$. Consequently, most separations occur because of exogenous shocks. One may wonder whether that result is due to the difficulty of disentangling exogenous from endogenous separations. In order to address this concern, we reestimate the model while fixing *ex ante* the rate $\delta$ at which destruction shocks arrive.

As explained in Section 2.1, the equilibrium is not well defined when $\delta$ is equal to zero. We therefore apply a less drastic restriction and assume that $\delta = 0.05$ instead of 0.13. Thus jobs last on average 20 years in the absence of endogenously induced separations. This seems to be a reasonable lower bound for the influence of non-economical factors such as taste shocks, reallocation or retirement. We also consider intermediate cases where $\delta = 0.09$. For both restrictions, we report in Table IV the estimates derived using the full information procedure and the restricted estimates based on the wage distribution among job entrants.

Perhaps surprisingly, lowering the value of $\delta$ does not affect strongly the value of the variance parameter $\sigma$. Accordingly, the total rate of job destruction falls, which explains why the contact rate $\lambda$ drops significantly. In other words, the restricted model underestimates the amount of worker

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\delta = 0.05$</th>
<th>$\delta = 0.09$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_r$</td>
<td>3.91</td>
<td>3.74</td>
</tr>
<tr>
<td>$\xi$</td>
<td>0.0163</td>
<td>0.0196</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.0389</td>
<td>0.033</td>
</tr>
<tr>
<td>$\Xi$</td>
<td>2.86</td>
<td>2.66</td>
</tr>
<tr>
<td>$\bar{\xi}$</td>
<td>0.494</td>
<td>0.488</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>1.21</td>
<td>1.61</td>
</tr>
<tr>
<td>$\ln L$</td>
<td>-30.329</td>
<td>-29.149</td>
</tr>
</tbody>
</table>

Table IV. Estimates with fixed $\delta$

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\delta = 0.05$</th>
<th>$\delta = 0.09$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_r$</td>
<td>3.77</td>
<td>3.74</td>
</tr>
<tr>
<td>$\xi$</td>
<td>0.0298</td>
<td>0.0196</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.052</td>
<td>0.033</td>
</tr>
<tr>
<td>$\Xi$</td>
<td>2.65</td>
<td>2.66</td>
</tr>
<tr>
<td>$\bar{\xi}$</td>
<td>0.485</td>
<td>0.488</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>1.22</td>
<td>1.61</td>
</tr>
<tr>
<td>$\ln L$</td>
<td>-30.483</td>
<td>-29.280</td>
</tr>
</tbody>
</table>

Note: Discount rate: $r = 5\%$, $\beta = .5$. Standard errors in parenthesis.
turnover. Similar conclusions hold when the estimation is based on the wage distribution among job entrants.

The new optimal values of the likelihood function show that treating $\delta$ as a fixed parameter significantly worsens the fit of the model. Much of the loss in explanatory power is due to the lower accuracies of the estimated rates of separation and job finding. In order to alleviate this tension, the estimation procedure lowers the rate of LBD $\zeta$ so as to raise the incidence of job destruction. But, as for $\sigma$, the adjustment is rather small.

It should not be too surprising that restricting $\delta$ is mostly to the detriment of how the model fits the turnover process. Given the cross-sectional nature of the data and the fact that more than 90% of the workers belong to the employment pool, the estimation procedure puts a greater weight on wage dispersion than on job turnover. This is why the two key parameters $\zeta$ and $\sigma$ turn out to be relatively robust. On a less positive note, these results suggest that there exists a tension between the structural model and the data that prevents productivity shocks from exerting too strong an influence. Future research with a greater focus on wage dynamics should determine whether this is due to our specification of the stochastic process.

6. THE IMPACT OF LEARNING-BY-DOING

In this section we introduce an aggregate matching function to close the model and evaluate the impact of LBD on labor market outcomes. We assume that the matching process is similar to the one described in Pissarides (2000). Since the matching function has become the workhorse for the study of equilibrium unemployment, the exposition can be brief. Firms post vacancies that are randomly matched and incur a flow cost equal to $c$. The number of job matches per unit of time is a function of the number of vacancies and job seekers. When the aggregate matching function is homogeneous of degree one, the rate at which a vacancy meets a worker depends only on the unemployment rate $u$ and on the ratio $v$ of vacant jobs divided by the size of the labor force. The transition rate for vacancies is given by a function $q(\phi)$ where the labor market tightness parameter $\phi$ denotes the vacancy–unemployment ratio. Similarly, jobs seekers meet firms at the rate $\theta q(\phi)$. As opposed to the labor force whose size is fixed and normalized to one, new firms enter the market until arbitrage opportunities are exhausted. Therefore the Free-Entry condition is given by

$$c = q(\phi) \int_R^{+\infty} (1 - \beta)S(P)dF(P)$$

(14)

Similarly we can replace in (2) the exogenous contact rate $\lambda$ by $\theta q(\phi)$ to obtain

$$rU = -s + \theta q(\phi) \int_R^{+\infty} \beta S(P)dF(P)$$

Reinserting (14) into the previous equation allows us to solve for the asset value of being unemployed as a function of labor market tightness;

$$rU = -s + c\theta \left( \frac{\beta}{1 - \beta} \right)$$
The job’s surplus follows from replacing the previous equation in (3). Accordingly the **Optimal Separation rule** is such that

$$ R = \left( \frac{\alpha}{\alpha - 1} \right) \left( \frac{r + \delta - \zeta}{r + \delta} \right) \left( -s + c\theta \left( \frac{\beta}{1 - \beta} \right) \right) $$

(15)

The equilibrium values of the two endogenous variables \( \theta \) and \( R \) are determined by the equilibrium conditions (14) and (15).\(^{32}\) Since the aggregate matching function defines a one-to-one mapping between \( \lambda \) and \( \theta \), a parametric assumption allows us to retrieve the values of the labor market tightness and searching costs using the estimates reported in the previous section. As is common in the literature, we assume that the matching function is Cobb–Douglas. We further restrict our attention to the case where the allocation is efficient and consequently use the ‘Hosios condition’ to set the elasticity of the matching function equal to \( \beta \), so that \( q(\theta) = \theta^{-1/2} \).

Table V contains the implied costs of search and equilibrium labor market tightness for the deterministic and stochastic models as well as for the restricted estimates. The high values of equilibrium tightness are unreasonable if interpreted as the ratio of vacancies to job seekers. Thus one should interpret \( \theta \) as measuring the ratio of recruitment effort to search effort. The relatively high searching costs are required to offset the important surpluses of the matches in the right tail of the distribution of job offers.

By keeping the values of the environmental parameters constant and varying the rate of LBD, we can simulate its impact on labor market outcomes. The results are reported in Figure 4. The upper-left panel shows the effect on aggregate wage distribution. Not surprisingly, a higher rate of LBD increases the mass in the right tail. This does not necessarily lead to higher inequality, however, because the left tail of the wage distribution is truncated by the increase in the reservation wage. The ratio of standard deviation to average wage reported in the upper-right panel shows that the latter effect dominates when the rate of LBD is close to zero.

Unemployment rate as a function of \( \zeta \) is reported in the lower-left panel. As expected, the function is decreasing but its elasticity is close to zero. To understand why, it is useful to recall that the opportunity cost of employment \( rU \) is equal to \(-s + c\theta)/(1 - \beta)\). This implies that, for the estimated value of the recruitment costs \( c \), the impact of \( \theta \) on the worker’s outside option is

<table>
<thead>
<tr>
<th><strong>Table V. Point estimates of remaining variables</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tightness</strong></td>
</tr>
<tr>
<td>( \theta )</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Flow costs of search</strong></td>
</tr>
<tr>
<td>( s )</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>( c )</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

**Note:** Standard errors in parentheses.

\(^{32}\) We refer the reader to Prat (2006) for a proof of the existence and uniqueness of the equilibrium when the distribution of job offers \( F(\cdot) \) is degenerate.
amplified by more than an order of magnitude. Small adjustments of the vacancy–unemployment ratio have drastic effects on wages, which explains why employment remains remarkably stable.

The rigidity of the unemployment rate and the flexibility of wages are even more pronounced in the deterministic model. This is because the model without LBD does not control for the effect of job tenure on wage dispersion. As a result, the opportunity cost of employment is lower in the deterministic model. Introducing LBD reduces the gap between the value of non-market activity derived from structural estimations and the one required to fit business cycle fluctuations of the unemployment rate. However, the adjustment remains rather modest, so that the conclusion reached by Hornstein et al. (2007) continues to hold in our set-up.

For the same reasons the estimates imply that labor market policies, such as employment subsidies or unemployment benefits, affect mostly wages and leave employment nearly unchanged. The predictions of the deterministic model are similar with an even stronger employment rigidity. Accordingly, the model suggests that policies aimed at reducing the rate of unemployment should focus on lowering both recruitment and search costs.

The lower-right panel contains a plot of the aggregate output as a function of the LBD rate. We normalize aggregate output to one when $\xi = 0$, for ease of interpretation. The model predicts that an increase of the LBD rate from 0 to 4% raises aggregate output by around 45%. These gains arise due to three reinforcing effects: (i) the direct impact of LBD obviously leads to a higher average output for a given job spell; (ii) the increase of the reservation wage implies that, ceteris

---

Figure 4. Effect of LBD on the equilibrium. This figure is available in color online at www.interscience.wiley.com/journal/jae

33 On the other hand, the predictions of the structural models starkly differ from those of calibrated models based on the set of parameters proposed by Shimer (2005).
THE RATE OF LEARNING-BY-DOING

paribus, ongoing job relationships have a higher average productivity; (iii) the higher rate of employment mechanically raises aggregate output. Decomposing the relative importance of these effects shows that the first one accounts for nearly 98% of the total gains. The increase in the reservation productivity explains 1.8% of the total gains, while the slight increase in employment accounts for the remaining 0.2%.

Before concluding, we would like to discuss two ways in which the simulation exercise could be improved. First, the results are partly driven by the low estimates for the variance parameter \( \sigma \), since higher volatility would magnify the impact of the reservation wage on job destruction and thus unemployment. One might suspect that estimating the model on panel data would lead to higher values for \( \sigma \). Hence the model’s implication might differ when estimated on a data source with richer information about wage dynamics. Second, our structural model does not take into account prevalent labor market regulations such as firing costs and employment subsidies. A recent paper by Silva (2008) illustrates how our framework could be used to evaluate the impact of labor market institutions. She extends the set-up by introducing a minimum wage and severance payments, before estimating the resulting model using data on employment histories from Chile. She finds that severance payments are more efficient than minimum wages in reducing distortions due to low bargaining power for workers.

7. CONCLUSION

It has been shown in this paper how the production side of the Mortensen–Pissarides model with endogenous job separation can be estimated by maximum likelihood using cross-sectional data. The analysis establishes that the parsimoniously specified model convincingly fits the joint distribution of wages and job spells once LBD is taken into account. A concrete contribution of the analysis is to identify the rate of LBD in an equilibrium set-up, whereas the estimates available in the literature are typically based on ‘reduced-form’ estimations. Taking into account the accumulation of job-specific skills raises the opportunity cost of employment. Yet the increase is too small to noticeably augment the elasticity of the unemployment rate. For example, we find that the rate of LBD has a significantly positive effect on aggregate output but a small impact on employment.

As we have deliberately tipped the balance in favor of tractability over realism, the model lends itself to several theoretical extensions. We conclude by briefly discussing some of these. The most obvious refinement would be to introduce general human capital. Although not so demanding at the conceptual level, this extension will come at the cost of closed-form solutions. More promising is the introduction of on-the-job search since it would connect the model with the burgeoning econometric literature based on employers’ competition. Until recently, uncertainty and on-the-job search have been considered in isolation. But, as attested by a series of recent papers (Bagger et al., 2006; Yamaguchi, 2006), the importance of combining both dimensions is now widely recognized. Such a research project raises serious technical challenges. For the moment, these structural models treat job separations as exogenous and available estimates are based on indirect inference methods. This paper suggests that stochastic calculus helps to alleviate some of the difficulties. Finally, we also hope that the derivations of the asymptotic properties of the estimators will be of some interest to researchers working in areas other than labor economics since our result can be applied to a wide class of models with endogenous exit.
ACKNOWLEDGEMENTS

I am grateful to Giuseppe Bertola, Christopher Flinn, Neil Foster, Giuseppe Moscarini and three anonymous referees for their helpful comments. I am indebted to Christian Haefke for providing me with useful data. I also thank conference participants at the SED 2006, ESEM 2006, NOeG 2006, and at the universities of Vienna and Tübingen for their comments and discussions. The author acknowledges the support of the University of Vienna, of the Institute for Economic Analysis (CSIC), of the Barcelona GSE and of the Government of Catalonia.

REFERENCES

THE RATE OF LEARNING-BY-DOING


APPENDIX

Proof of Proposition 1

We assume that $R$ does not depend on current output. The Bellman equation satisfied by the surplus within the continuation region then follows by Ito’s Lemma:

$$(r + \delta)S(P^i_t, R) = P^i_t - rU + \zeta P^i_t S_1(P^i_t, R) + \frac{(P^i_t \sigma)^2}{2} S_{11}(P^i_t, R)$$

(16)

where number subscripts denote the partial derivatives of the function. It is well known that the general solution of this partial differential equation is of the form

$$S(P^i_t, R) = C(R) \left( \frac{P^i_t}{R} \right)^{\alpha} + D(R) \left( \frac{P^i_t}{R} \right)^{\eta} + E_{P_t} \left[ \int_{t}^{T} e^{-(r+\delta)(t-\tau)}(P^i_\tau - rU) d\tau \right]$$

(17)

where $C(R)$ and $D(R)$ are some constants of integration which do not depend on the current state $P^i_t$, while $\alpha$ and $\eta$ are respectively the negative and positive roots of the quadratic equation

$$\frac{\sigma^2}{2} \chi(\chi - 1) + \chi \zeta - r - \delta = 0$$

The values of $C(R)$ and $D(R)$ are pinned down by two boundary conditions. First, when $P^i_t$ diverges to infinity, the values of the option to separate converges to zero. This implies that we can set $D(R)$ equal to zero. Otherwise, the term $D(R)(P^i_t/R)^\eta$ would diverge to infinity with $P^i_t$ because the root $\eta$ is superior to zero. But this would contradict the fact that the value of the option to separate should become negligible. The second boundary, $\lim_{P^i_t \to R} S(P^i_t, R) = 0$, follows from the definition of the reservation threshold $R$. One can easily verify that the solution proposed in (4) satisfies the differential equation and boundary conditions. The optimal reservation productivity is set so as to maximize the surplus. Since current revenues are independent of the reservation productivity, it can be shown that $\delta S(P^i_t, R)/\delta R = 0$ when

$$\left. \frac{\partial S(P^i_t, R)}{\partial P^i_t} \right|_{P^i_t = R} = 0$$

(18)

Equation (18) is commonly referred to as the smooth-pasting condition. Its solution reads

$$R = \left( \frac{\alpha}{\alpha - 1} \right) \left( \frac{r + \delta - \zeta}{r + \delta} \right) rU$$

(19)

which is equivalent to (5) and verifies our guess that $R$ is independent of $P^i_t$.

---

34 Since $P^i_t$ grows at rate $\zeta$ and the opportunity cost of employment remains constant through time, the integral on the right-hand side of (17) is equal to $\frac{P^i_t}{r + \delta - \zeta} - \frac{rU}{r + \delta}$, as stated in Proposition 1. The decomposition in equation (17) shows that the value of the job adds the option to separate to the discounted stream of revenues given by the integral.

35 See Merton (1973, p. 171).
THE RATE OF LEARNING-BY-DOING

Proof of Proposition 2

Consider a match \( i \) that is created at date \( t \). We define the stopping time \( \tau^i_1 \) as the time of arrival of the first exogenous destruction shock and

\[
\tau^i_2 = \min \{ \tau > t : P^i_\tau = R \}
\]

Thus \( \tau^i_1 \) is the first time at which the job would have been endogenously destroyed. Hence, job \( i \) is operational at time \( t + T \) if and only if \( \tau^i_1 \) and \( \tau^i_2 \) are both superior to \( t + T \). As destruction shocks and idiosyncratic fluctuations are independent, it follows by complementarity that

\[
\begin{align*}
\Pr & [ P^i_{t+T} \in A \cap \tau^i_2 > t + T \cap \tau^i_1 > t + T | P^i_t ] \\
& = (\Pr [ P^i_{t+T} \in A | P^i_t ] - \Pr [ P^i_{t+T} \in A \cap \tau^i_2 \leq t + T | P^i_t ] ) \times \Pr [ \tau^i_1 > t + T ]
\end{align*}
\]

(20)

where the Borel set \( A \subset (R, +\infty) \). These probabilities are more easily computed considering \( \ln(P^i_{t+T}) \) since it is a standard Brownian motion. Thus

\[
\Pr [ \ln(P^i_{t+T}) \in A | P^i_t ] = \int_A e^{-\frac{1}{2} \left( \frac{\ln(x) - \ln(P^i_t) - \mu T}{\sigma \sqrt{T}} \right)^2} \frac{1}{\sqrt{2\pi}} \, d\ln(x)
\]

(21)

where \( \mu = \xi - \frac{\sigma^2}{2} \) is the trend of \( \ln(P^i_t) \). When \( \mu \) is equal to zero, the expression of \( \Pr [ \ln(P^i_{t+T}) \in A \cap \tau^i_2 \leq t + T | P^i_t ] \) is easily obtained from the reflection principle. The general expression is derived in Harrison (1985) through a change of measures:

\[
\Pr [ \ln(P^i_{t+T}) \in A \cap \tau^i_2 \leq t + T | P^i_t ] = \int_A \left( \frac{R}{P_t} \right)^{2\mu} e^{-\frac{1}{2} \left( \frac{\ln(x) + \ln(P^i_t) - 2 \ln(R) - \mu T}{\sigma \sqrt{T}} \right)^2} \frac{1}{\sqrt{2\pi}} \, d\ln(x)
\]

(22)

Expressing the densities (21) and (22) in terms of \( x \) instead of \( \ln(x) \), substituting the resulting expressions into (20) and multiplying by \( \Pr [ \tau^i_1 > t + T ] = e^{-\beta T} \) implies that for all \( x > R \)

\[
\int_A \psi(x, T; P) \, dx = \Pr [ P^i_{t+T} \in A \cap \tau^i_2 > t + T \cap \tau^i_1 > t + T | P^i_t = P ]
\]

\[
= \int_A e^{-\beta T} \left( \frac{1}{x} \right) \left( e^{-\frac{1}{2} \left( \frac{\ln(x) - \ln(P^i_t) - \mu T}{\sigma \sqrt{T}} \right)^2} - \left( \frac{R}{P} \right)^{2\mu} e^{-\frac{1}{2} \left( \frac{\ln(x) + \ln(P^i_t) - 2 \ln(R) - \mu T}{\sigma \sqrt{T}} \right)^2} \right) \, dx
\]

(23)

According to Bayes’ rule, the unconditional mass of jobs is given by

\[
\Pr [ P^i_{t+T} \in A \cap \tau^i_2 > t + T \cap \tau^i_1 > t + T ] = \Pr [ P^i_{t+T} \in A \cap \tau^i_2 > t + T \cap \tau^i_1 > t + T | P^i_t \in B ] \times \Pr [ P^i_t \in B ]
\]

where the Borel set $B \subset (R, +\infty)$. Remember that we are considering a match which has been created at date $t$. Therefore, under the assumption according to which the initial output $P_i^t$ is drawn from $F(\cdot)$, $Pr[P_i^t \in B] = \int_R dF(P)/(1 - F(R))$ and so the unconditional mass of jobs reads

$$Pr\{P_{i+T}^t \in A \cap t_1 > t + T \cap t_1 > t + T\} = \int_A \left( \frac{\int^{+\infty}_R \psi(x, T; P)dF(P)}{1 - F(R)} \right) dx$$

Finally the measure $\nu(x, T)$ is given by the unconditional mass of jobs multiplied by the rate of job creation. According to the stationarity assumption, the job creation flow is constant and equal to $u\lambda(1 - F(R))$, which yields the expression in Proposition 2.

**Proof of Proposition 3**

By definition, the mass of jobs with current output equal to $x$ is given by

$$\nu(x) \equiv \int_0^{+\infty} \nu(x, T)dT = u\lambda \int_0^{+\infty} \left( \int_R^{+\infty} \psi(x, T; P)dF(P) \right) dT$$

The equality follows from the fact that $\psi(x, T; P)$ is the mass of jobs with output $x$ and tenure $T$ conditional on the initial productivity $P$. Reversing the order of integration allows us to find an analytical solution for $\nu(x)$. Some algebra yields

$$\psi(x, T; P) = P^{-1} \left( \frac{x}{P} \right)^{\mu + \nu - 1} \left( \frac{e^{-\frac{1}{2} \left( \frac{\ln(x) - \ln(P) + \gamma T}{\sigma \sqrt{T}} \right)^2}}{\sigma \sqrt{2\pi T}} \right)$$

where $\gamma = \sqrt{\mu^2 + 2\lambda \sigma^2}$. Using the result in Leland and Toft (1997) according to which for positive values of $x$

$$\int_0^T e^{\frac{-1}{2} \left( \frac{\ln(x) + \gamma T}{\sigma \sqrt{T}} \right)^2} d\tau = \left( \frac{1}{\gamma} \right) (\Phi \left( \frac{-\ln(x) - \gamma T}{\sigma \sqrt{T}} \right) + x \frac{2\nu}{\sigma^2} \Phi \left( \frac{-\ln(x) + \gamma T}{\sigma \sqrt{T}} \right))$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function, we obtain

$$\lim_{t \to +\infty} \int_0^T e^{\frac{-1}{2} \left( \frac{\ln(x) + \gamma T}{\sigma \sqrt{T}} \right)^2} d\tau = \frac{x}{\gamma} \frac{2\nu}{\sigma^2}$$

Using this limit to integrate $\psi(x; P) \equiv \int_0^{+\infty} \psi(x, T; P) dT$ and ensuring that the integration is always performed over positive values, yields the expression of $\nu(x)$ reported in Proposition 3.
THE RATE OF LEARNING-BY-DOING

Proof of Proposition 4

Given that the size of the labor force has been normalized to one, the rate of employment is equal to the integral of \( \nu(x) \) from \( R \) up to infinity. Thus

\[
1 - u = \int_{R}^{+\infty} \nu(x) dx
\]

\[
= u\lambda \int_{R}^{+\infty} \left( \int_{R}^{+\infty} \varphi(x; P) dF(P) \right) dx = u\lambda \int_{R}^{+\infty} \left( \int_{R}^{+\infty} \varphi(x; P) dx \right) dF(P)
\]

Integrating \( \varphi(x; P) \) with respect to \( x \) yields

\[
\int_{R}^{+\infty} \varphi(x; P) dx = \left( \frac{1}{\beta} \right) \left( 1 - \left( \frac{R}{\beta} \right)^{\frac{\mu+\nu}{\sigma}} \right)
\]

The expression of the unemployment rate \( u \) is immediately obtained reinserting this solution into the previous equation and simplifying.

Proof of Proposition 5

For tractability, let us focus on the analytically simpler case where all employed workers fail to report their job spells, so that \( e \in [0, 1] \) and \( y = \{e, w, \tau\} \). Then the density function is given by

\[
f(y, \Theta) = (1 - e)\lambda F(R)e^{-\lambda F(R)T(R; \Theta)} + e[\nu(x(w); \Theta)]
\]

Since \( F(\cdot) \) is continuously differentiable, the density function \( f(y, \Theta) \) and consequently the likelihood function \( L(y, \Theta) \) are also continuously differentiable. Differentiating the density function yields

\[
\frac{\partial}{\partial \Theta} \int f(y, \Theta) dy = \frac{\partial}{\partial \Theta} \left[ \sum_{e=0}^{1} (1 - e)u(\Theta) \int_{0}^{+\infty} \lambda F(R)e^{-\lambda F(R)T(R; \Theta)} d\tau + e \int_{R(\Theta)}^{+\infty} \nu(x(w); \Theta) dx \right]
\]

\[
= \sum_{e=0}^{1} (1 - e) \frac{\partial u(\Theta)}{\partial \Theta} + e \frac{\partial}{\partial \Theta} \int_{R(\Theta)}^{+\infty} \nu(x(w); \Theta) dx = 0
\]

(24)

where the second equality follows from \( \int_{0}^{+\infty} \lambda F(R)e^{-\lambda F(R)T(R; \Theta)} d\tau = 1 \), for all \( R > 0 \); while the last equality holds true by definition. The second term of the sum is non-standard because the lower bound of the integral varies with the set of parameters \( \Theta \). Leibnitz’s rule implies that

\[
\int_{R(\Theta)}^{+\infty} \frac{\partial \nu(x(w); \Theta)}{\partial \Theta} dx = \frac{\partial}{\partial \Theta} \int_{R(\Theta)}^{+\infty} \nu(x(w); \Theta) dx + \nu(R(\Theta); \Theta) \frac{dR(\Theta)}{d\Theta}
\]

\[
= \frac{\partial}{\partial \Theta} \int_{R(\Theta)}^{+\infty} \nu(x(w); \Theta) dx = - \frac{\partial u(\Theta)}{\partial \Theta}
\]

(25)
The second equality holds true because \( \varphi(R(\Theta); P) = 0 \) for all \( P \geq R \), so that \( \nu(R(\Theta); \Theta) = u\lambda \int_{R(\Theta)}^{\infty} \varphi(R(\Theta); P) dF(P) = 0 \). As explained in the proof of Proposition 4, the last equality follows from the definition of \( \nu(\cdot) \). Combining equations (24) and (25), we find that \( \frac{\partial}{\partial \Theta} \int f(y, \Theta)dy = \int \frac{\partial}{\partial \Theta} f(y, \Theta)dy = 0 \). Accordingly, the order of integration and differentiation can be reversed and the central limit theorem yields

\[
\frac{1}{\sqrt{n}} \left( \sum_{i=1}^{n} \frac{\partial \ln f(y_i, \Theta)}{\partial \Theta} \right) \rightarrow^{d} \mathcal{N}(0, J)
\]

where \( J \) is the information matrix. Since the estimator \( \hat{\Theta} \) is consistent, the law of large numbers implies that

\[
-\frac{1}{n} \left( \sum_{i=1}^{n} \frac{\partial^2 \ln f(y_i, \Theta)}{\partial \Theta \partial \Theta} \right) \rightarrow^{p} H
\]

where \( H \) is the Hessian matrix. Note that the Hessian matrix is not equivalent to the information matrix as

\[
\int_{R(\Theta)}^{\infty} \frac{\partial^2 \nu(x(w); \Theta)}{\partial \Theta \partial \Theta'} dx = \frac{\partial}{\partial \Theta} \int_{R(\Theta)}^{\infty} \frac{\partial \nu(x(w); \Theta)}{\partial \Theta} dx + \frac{\partial \nu(x; \Theta)}{\partial \Theta} \bigg|_{x=R(\Theta)} \frac{dR(\Theta)}{d \Theta}
\]

\[
\neq \frac{\partial}{\partial \Theta} \int_{R(\Theta)}^{\infty} \frac{\partial \nu(x(w); \Theta)}{\partial \Theta} dx
\]

The inequality holds because the derivative of the measure \( \nu(\cdot) \) w.r.t. \( \Theta \) at the reservation output differs from zero.\(^{36}\) Hence \( \int \frac{\partial^2 f(y, \Theta)}{\partial \Theta \partial \Theta'} dy \neq \frac{\partial}{\partial \Theta} \int f(y, \Theta)dy \) and so the order of differentiation and integration cannot be interchanged twice.\(^{37}\) Given that the likelihood function satisfies all the other regularity conditions, asymptotic efficiency and asymptotic normality of the maximum likelihood estimator are established. The proof is easily extended to the general case where a share of employed workers also report their tenure, i.e. \( e \in \{0, 1, 2\} \) and \( y = \{e, w, T, t\} \), noting that \( \psi(R(\Theta), T; P) = 0 \) for all \( P \geq R \) and \( T \geq 0 \).

**Proof of Proposition 6**

The proof of proposition 6 follows by direct calculation. Given that the algebra is tedious, we decompose the solution in several steps. First consider the integral with respect to \( \psi(x, T; P) \). Under the parametric assumption that \( F(P) \) is lognormal, it reads

\[
\int_{R}^{\infty} x\psi(x, T; P)dF(P) e^{-\delta t} = \int_{R}^{\infty} \left( x\psi(x, T; P) e^{-\delta t} \right) \left( e^{-\frac{1}{2} \left( \frac{\ln(P) - \Sigma}{\xi} \right)^2 \frac{\xi}{\xi \sqrt{2\pi}}} \right) d\ln(P)
\]

\(^{36}\)To verify this claim consider, for example, \( \frac{\partial \psi(x; P)}{\partial R_{i=0}} \). It is easily shown that the derivative is negative for all \( P \geq R \).

\(^{37}\)See Greene (1980) for further details on the relationship between the second-order derivative and the information matrix equality.
The rate of learning-by-doing

\[
\begin{align*}
&= \left( -\frac{1}{2} \left( \frac{B(x, T) \left( \xi^2 + \sigma^2 T \right)}{\xi^2 \sigma^2 T} \right) \right) \int_{\mathbb{R}} \left( -\frac{1}{2} \left( \frac{\ln(P) - (A(x, T) \rho^2(T) + \Sigma (1 - \rho^2(T)))}{\sqrt{\xi^2 + \sigma^2 T} \rho(T)(1 - \rho^2(T))} \right)^2 \right) d \ln(P) \\
&= \left( -\frac{1}{2} \left( \frac{D(x, T) \left( \xi^2 + \sigma^2 T \right)}{\xi^2 \sigma^2 T} \right) \right) \int_{\mathbb{R}} \left( -\frac{1}{2} \left( \frac{\ln(P) - (C(x, T) \rho^2(T) + \Sigma (1 - \rho^2(T)))}{\sqrt{\xi^2 + \sigma^2 T} \rho(T)(1 - \rho^2(T))} \right)^2 \right) d \ln(P) \\
&= \left( -\frac{1}{2} \left( \frac{B(x, T) \left( \xi^2 + \sigma^2 T \right)}{\xi^2 \sigma^2 T} \right) \right) \Phi \left( \frac{-\ln(R) + A(x, T) \rho^2(T) + \Sigma (1 - \rho^2(T))}{\sqrt{\xi^2 + \sigma^2 T} \rho(T)(1 - \rho^2(T))} \right) \\
&= \left( -\frac{1}{2} \left( \frac{D(x, T) \left( \xi^2 + \sigma^2 T \right)}{\xi^2 \sigma^2 T} \right) \right) e^{\left( -\frac{2\mu}{\sigma^2} \left( \frac{\xi^2 \sigma^2 T}{\xi^2 + \sigma^2 T} \right) + \left( \frac{C(x, T) \rho^2(T) + \Sigma (1 - \rho^2(T))}{\sqrt{\xi^2 + \sigma^2 T} \rho(T)(1 - \rho^2(T))} \right) \right) + \left( \frac{2\mu}{\sigma^2} \right) \left( \frac{\xi^2 \sigma^2 T}{\xi^2 + \sigma^2 T} \right) + \left( \frac{C(x, T) \rho^2(T) + \Sigma (1 - \rho^2(T))}{\sqrt{\xi^2 + \sigma^2 T} \rho(T)(1 - \rho^2(T))} \right)} \\
&\times \Phi \left( \frac{-\ln(R) + \left( \frac{2\mu}{\sigma^2} \left( \frac{\xi^2 \sigma^2 T}{\xi^2 + \sigma^2 T} \right) + C(x, T) \rho^2(T) + \Sigma (1 - \rho^2(T))}{\sqrt{\xi^2 + \sigma^2 T} \rho(T)(1 - \rho^2(T))} \right)} \right),
\end{align*}
\]

where

\[\rho^2(T) = \frac{\xi^2}{\xi^2 + \sigma^2 T} \]

\[A(x, T) = -\mu T + \ln(x) \quad B(x, T) = \left( -A(x, T) \rho^2(T) + \Sigma (1 - \rho^2(T))^2 \right) \]

\[+ A(x, T)^2 \rho^2(T) + \Sigma^2 (1 - \rho^2(T)) \]

\[C(x, T) = 2 \ln(R) + \mu T - \ln(x) \quad D(x, T) = \left( -C(x, T) \rho^2(T) + \Sigma (1 - \rho^2(T))^2 \right) \]

\[+ C(x, T)^2 \rho^2(T) + \Sigma^2 (1 - \rho^2(T)) \]

Now consider the integral with respect to \(\psi(x; P)\)

\[
\int_{\mathbb{R}} \psi(x; P) dF(P) = \int_{\mathbb{R}} \left( \frac{1}{P} \right)^{2\gamma} \left( \frac{\ln(P)}{\xi \sqrt{2\pi}} \right)^{2\gamma} \left( \frac{1 - \left( \frac{R}{P} \right)^{\gamma} \sigma^2}{\gamma} \right) \left( \frac{1}{\sigma^2} \right) d \ln(P)
\]

\[ + \int_{x}^{\infty} P^{-1} \left( \frac{x}{P} \right) \frac{\mu + \gamma}{\sigma^2} \left( 1 - \left( \frac{R}{\mu + \gamma} \right) \frac{2^{-1} \left( \frac{\ln(P) - \Sigma}{\xi} \right)^2}{\xi \sqrt{2\pi}} \right) d\ln(P) \]

\[ = \left( \frac{\mu - \gamma}{x \sigma^2} \right) e^{\left( \frac{\gamma - \mu}{\sigma^2} \right)^2 \left( \xi^2 \right) + \left( \frac{\gamma - \mu}{\sigma^2} \right) \Sigma} \left( e^{-\ln(x) + \left( \frac{\gamma - \mu}{\sigma^2} \right) \xi^2 + \Sigma} \right) \Phi \left( \frac{\ln(x) - \left( \frac{\gamma - \mu}{\sigma^2} \right) \xi^2 + \Sigma}{\xi} \right) \Phi \left( \frac{\ln(R) - \left( \frac{\gamma - \mu}{\sigma^2} \right) \xi^2 + \Sigma}{\xi} \right) \]

\[ + \left( \frac{\mu + \gamma}{x \sigma^2} \right) e^{\left( \frac{-\gamma - \mu}{\sigma^2} \right)^2 \left( \xi^2 \right) + \left( \frac{-\gamma - \mu}{\sigma^2} \right) \Sigma} \Phi \left( -\ln(x) + \left( \frac{-\gamma - \mu}{\sigma^2} \right) \xi^2 + \Sigma \right) \Phi \left( -\ln(R) + \left( \frac{-\gamma - \mu}{\sigma^2} \right) \xi^2 + \Sigma \right) \]

\[ - \frac{2\gamma}{R \sigma^2} \left( \frac{\mu - \gamma}{x \sigma^2} \right) e^{\left( \frac{\gamma - \mu}{\sigma^2} \right)^2 \left( \xi^2 \right) + \left( \frac{\gamma - \mu}{\sigma^2} \right) \Sigma} \Phi \left( -\ln(R) + \left( \frac{\gamma - \mu}{\sigma^2} \right) \xi^2 + \Sigma \right) \]

Finally consider

\[ \int_{R}^{\infty} \left( 1 - \left( \frac{R}{P} \right) \frac{\mu + \gamma}{\sigma^2} \right) dF(P) = \int_{R}^{\infty} \left( 1 - \left( \frac{R}{P} \right) \frac{\mu + \gamma}{\sigma^2} \right) \left( e^{-\ln(P) + \left( \frac{\mu + \gamma}{\sigma^2} \right) \xi^2 + \Sigma} \right) \phi \left( \frac{\ln(P) - \ln(R) + \left( \frac{\mu + \gamma}{\sigma^2} \right) \xi^2 + \Sigma}{\xi} \right) \]

\[ = F(R) - R \frac{\mu + \gamma}{\sigma^2} e^{\left( \frac{-\mu - \gamma}{\sigma^2} \right)^2 \left( \xi^2 \right) + \left( \frac{-\mu - \gamma}{\sigma^2} \right) \Sigma} \phi \left( -\ln(R) + \left( \frac{-\mu - \gamma}{\sigma^2} \right) \xi^2 + \Sigma \right) \]

The closed-from expression of the likelihood function is obtained inserting (26), (27) and (28) into (12). Notice that the unemployment rate also has an analytical solution

\[ u = \frac{\delta}{\delta + \lambda} \left[ F(R) - R \sqrt{\phi \left( -\ln(R) + \left( \frac{-\mu - \gamma}{\sigma^2} \right) \xi^2 + \Sigma \right)} \right] \]