Learning-by-Doing Versus Learning About Match Quality: Can We Tell Them Apart?

Éva Nagypál

First version: December, 1999
This version: September, 2004

Understanding the accumulation of match-specific capital is crucial in shedding light on the reasons for the prevalence of long-term employment relationships and on the welfare consequences of turnover in the labor market. One of the most important source of match-specific capital is human capital acquired through match-specific learning. Such learning can take on two distinct forms. In the first case, workers accumulate match-specific human capital through learning-by-doing. In the second case, a worker and a firm in an employment relationship learn about the quality of the match over time, thereby acquiring valuable information. I construct a structural model that embeds these two learning explanations. I show that it is possible to distinguish the two explanations given turnover data on employing firms coupled with data on workers. Such data now exist in the form of matched employer-employee data sets. I use a matched French data set to estimate the structural model using the Efficient Method of Moments, a simulation-based estimation method. I find that, while learning-by-doing could be present in the first few months of an employment relationship, learning about match quality dominates at tenures above half a year. This finding has important consequences for economists’ understanding of the sources of match-specific capital and for the desirability of policies that alter the incentives for turnover for workers of different tenure.

JEL codes: J41, J63, C51

Keywords: Learning, specific capital, structural estimation, simulated method of moments

---

1I am greatly indebted to Robert E. Hall and Thomas J. Sargent for their invaluable comments and encouragement. Luca Guerrieri, Ken Judd, Dirk Krueger, Hanno Lustig, Thomas MaCurdy, Beatrix Paal, John Pencavel, Harald Uhlig, Ed Vytlacil and participants at the Stanford macro lunch, the Stanford macro reading group, the Society of Economics Dynamics Conference and the European Economic Association Congress have given me helpful comments. I would like to thank Denis Fouëtè for his hospitality during my stay at CREST-INSEE and Francis Kramarz for providing me with the data. Financial support from the Alfred P. Sloan Foundation, the Lynde and Harry Bradley Foundation and the Stanford Graduate Research Opportunities Fund is gratefully acknowledged. All errors are mine.

2Department of Economics, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208-2600; nagypal@northwestern.edu.
1 Introduction

The accumulation of match-specific capital in employment relationships has long been of interest to economists studying labor markets. This is because the amount of match-specific capital present in an employment relationship determines the welfare consequences of turnover in the labor market, and thereby influences how economists evaluate the effect of aggregate shocks on labor market outcomes or the effect of labor market policies.

While match-specific physical capital may be present in many employment relationships, the most important source of match-specific capital is human capital acquired through match-specific learning. Such learning can be of two distinct types: the type that increases the expected productivity of a given employment match over time, which I call learning-by-doing, and the type that allows the partners in the match to form more and more precise estimates of the underlying fixed productivity (quality) of the match, which I refer to as learning about match quality. In the presence of learning by doing, as time on the job increases, the worker accumulates more and more job-specific expertise and hence becomes more productive. In the presence of learning about match quality, the source of match-specific capital is the information that accumulates over time about the quality of the match, which is valuable since it allows the partners to assess more and more precisely the payoffs from continuing the existing match.\(^3\)

This paper tackles the problem of distinguishing between these two learning processes empirically. This is an important task, since the two learning processes have different implications for the source of productivity growth with tenure, for employment policies that induce firms to alter their employment of workers of different tenure, and for the effect of downturns on workers of different tenure and skill.

Most studies attempting to understand the accumulation of match-specific capital focus on wage growth with tenure on the job (see, for example, Topel (1991)). It is difficult, however, to disentangle the effect of general and specific human capital on wages, which explains why there is no agreement in the literature as to how important tenure is in explaining wage growth. While most studies find that wages increase with tenure even after taking into account the effect of experience, the estimates of the wage increase attributable to match-specific capital vary widely. For this reason, in this study I focus on the impact of match-specific capital on turnover as opposed to its effect on wages. There is substantial empirical evidence suggesting that, with the possible exception of the first few months of employment, the hazard rate of employment termination declines with tenure on the job.\(^4\)

This alone is not enough to distinguish the two processes, however, because both learning explanations can explain this fact, since both lead to an increase in expected match-specific capital that shields workers from separation (Mortensen (1988)). In the case of learning-by-doing, increased expertise shields higher-tenure workers from separations, while, in the case of learning about match quality, jobs of longer tenure are less likely to terminate, since the reason they have not already terminated is that the match quality turned out to be favorable.

\(^{3}\)For the sake of conceptual clarity, I treat these two processes as completely independent of each other. It is possible for them to interact, however, for instance, if the worker is learning about the speed at which she learns new tasks on the job. Conceptually, it is still the case that learning-by-doing leads to increases in expected productivity for a given worker, while learning about match quality leads to declining uncertainty about a constant, but unknown, productivity and therefore allows for better and better selection of continuing matches.

\(^{4}\)Studies documenting this finding include Parsons (1978), Mincer and Jovanovic (1981), Topel and Ward (1992), and Farber (1994).
for employment.5

The key insight of this paper is that the two learning processes have different implications regarding separations in response to firm-level shocks under natural assumptions about these shocks. This difference is what I use to distinguish the two types of learning. The intuition behind the use of firm-level shocks to identify the two types of learning is the following. In the presence of learning-by-doing, match-specific capital accumulates over time and is always an increasing function of tenure. Hence, when an unfavorable shock occurs at the firm level, the matches most vulnerable to termination are those of the shortest tenure. In the case of learning about match quality, however, there are workers with low match-specific capital at all tenure levels, who can become the victims of declining demand. This is because there are two effects that determine the amount of match-specific capital. On the one hand, match-specific capital increases with tenure on the job, since average quality increases with tenure as low-quality workers are terminated. On the other hand, match-specific capital declines with tenure as the option value of employment declines. This means that there will be some workers with close to zero match-specific capital even after many months of tenure due to their declining option value. These are the workers who are believed to be just good enough to keep around in normal times, but who are the first ones terminated when an adverse shock hits. Hence, in the case of learning about match quality, there are workers at all tenure levels who are vulnerable to termination when an adverse shock hits.

Figure 1 confirms this intuition. Panel (a) graphs the hazard rate implied by the model developed in this paper when only learning-by-doing is present. The solid line is the hazard rate of separation at a firm experiencing good times, while the dashed line is the hazard rate of separation at a firm experiencing difficult times. The effect of an unfavorable firm-specific shock is to increase the hazard rate at low tenure levels by a much greater proportion than at high tenure levels, implying that it is mostly low-tenure workers that fall prey to adverse shocks at the firm. Panel (b) graphs the same two hazard rate functions implied by the model when only learning about match quality is present. In this case, the effect of an unfavorable firm-specific shock is to increase the hazard rate at all tenure levels by roughly the same proportion, implying that the adverse effects of a bad shock can be felt at all tenure levels. In fact, in the case of learning-by-doing, the effect of an unfavorable shock on the hazard rate goes to zero as tenure increases, while, in the case of learning about match quality, the proportional effect of an unfavorable shock remains stable and above zero. It is this difference in the effect of an unfavorable firm-specific shock that I use to distinguish between the two explanations. Figure 1 also demonstrates that distinguishing between the

5The only other work, to my knowledge, that attempts to distinguish between learning-by-doing and learning about match quality is Farber (1994,1999). He infers the importance of learning-by-doing versus learning about match quality by inspecting the hazard rate function. These inferences are not very satisfactory, since they rely heavily on estimates of the hazard rate during the first few months of tenure, which are sensitive to measurement problems due to the difficulty in acquiring high-quality data on very short spells of employment.
two explanations without considering the effect of firm-specific shocks on the hazard rate function is not possible, since that would amount to distinguishing between the dashed lines in panel (a) and (b), which are qualitatively very similar.\footnote{Quantitatively they differ somewhat, but that is simply the result of the particular parameter values that I chose to generate these graphs.}

To formalize this intuition, I construct a structural model that includes both learning explanations and that has firm-level shocks affecting employment relationships in the economy. Given the presence of match-specific surplus due to learning in my study, it is necessary to use a framework that allows for the determination of turnover in the presence of such surplus. So I study the accumulation of match-specific capital in a search and matching framework introduced by Diamond (1982) and Mortensen (1982). In this framework separations are bilaterally efficient; that is separations only take place when the joint surplus of the match falls below zero. This provides a useful benchmark, since it allows one to isolate and study the allocation of workers across matches without explicitly tackling the issue of bilateral wage contracting.

The key assumption that is necessary for identification is that firm-level shocks affect low- and high-tenure workers symmetrically. In other words, when an adverse demand shock hits the firm, the derived demand for the labor of low- and high-tenure workers is affected the same way. While it is possible to relax this assumption and still get identification under the less restrictive assumption of monotone effects of firm-level shocks across tenure, the assumption adopted is a natural one. I will discuss its validity further after presenting my results.

An additional important contribution of this paper is that it successfully applies a structural estimation method based on the Efficient Method of Moments (EMM) using a novel, albeit natural, set of moments that arise in a dynamic discrete choice model. I use this type of simulation-based estimation method because the likelihood function is prohibitively difficult to calculate due to the lack of data on conditioning variables. EMM is a special case of the Simulated Method of Moments. To choose the moments of interest, EMM relies on the use of an auxiliary model that provides a good statistical description of the data in the dimension addressed by the structural model. EMM then chooses the parameters of the structural model so that observations simulated using the structural model have the same statistical properties as the observed data, as captured by the auxiliary model.

The reason to carry out the structural estimation is two-fold. First, I successfully implement the Efficient Method of Moments in a dynamic discrete choice framework, which is a novel exercise and has rarely been done before. Second, the structural parameters allow me to simulate the effects of employment protection policies affecting the incentives of firms to employ workers of different tenure, an exercise that I undertake in Nagypál (2002).

I use a French matched data set which provides information on workers and their employers.\footnote{Matched employee-employer data sets have become available recently, mostly for European countries. They have received considerable attention, since they promise to enable researchers to analyze questions that could not previously be answered. A description of these data sets and recent work using them can be found in Abowd and Kramarz (1999) or Haltiwanger, Lane, Spletzer, Theuwes, and Troske (2000).} As can be seen from Figure 1, data on job durations and worker characteristics alone do not allow for a distinction between the two explanations. Hence the availability of this data set is key to this study. Despite its clear advantages, the data set also has limitations, since there are many observations on workers for whom there is no information on the employing firm. In fact, despite the presence of close to 250,000 workers in the original
data set on workers, there exists matched data for only about 9,000 workers, which in turn
means that I observe only about 500 separations in the matched data set. Moreover, due to
a filtering problem that is present because of the way the data are collected, short spells are
underrepresented in the data and there are only 18 separations that I observe in the matched
data set that take place at less than six months of tenure and an additional 45 that take
place between six and twelve months of tenure.

The results of my estimation imply that learning about match quality dominates, albeit
the data limitations need to be kept in mind when interpreting this result, since they imply
that the data are not well-suited to understanding the learning process at very short tenures.
I find that the effect of an unfavorable firm-specific shock (captured by an increase in the
appropriately defined separation rate) on the hazard rate of employment termination is
significantly different from zero at all tenure levels up to 10 years and shows no sign of
declining to zero as tenure increases. In fact, the hypothesis that the effect is the same at
all tenure levels cannot be rejected. The qualitative implication of the model that learning
about match quality has to be the more important component at tenures over six months
is validated by the formal structural estimation results of EMM. My results imply that
job-specific learning-by-doing, if present, is a process that takes place quickly in the first
few months of tenure, leading to moderate match-specific capital accumulation. Learning
about match quality, on the contrary, is a lengthy process leading to substantial increases in
match-specific capital with tenure.

To conclude this introduction, I will highlight how this paper relates to the substantial
literature in labor economics attempting to distinguish between the effects of true duration
dependence and pure heterogeneity on the hazard rate. One concise summary of the is-
issues of identification in the presence of duration dependence and heterogeneity is Heckman
(1991). He shows that it is not possible to distinguish between duration dependence and heterogeneity without additional functional form or distributional assumptions. At first
glance, it may seem that learning-by-doing is a way to capture duration dependence and
learning about match quality is a form of heterogeneity. This is not the case, however. It
is certainly true learning-by-doing is not distinguishable from pure heterogeneity, since both
imply that there is an upward drift in match-specific capital and it is only with additional
assumptions on the form of this upward drift that one can distinguish between them. Once
one adds learning about match quality, however, the situation is different. Learning about
match quality has heterogeneity at its core, since, without heterogeneity in match quality,
the learning process would be trivial. The dynamics of learning, however, adds a new chan-
nel, since the dynamic learning process implies that, each period, there is a loss of option
value, implying a downward drift in match-specific capital. This means that learning about
match quality causes a downward drift in match-specific capital together with the upward

8Heterogeneity can be introduced in many ways. The most common is to assume that there is het-
erogeneity among matches in their propensity to terminate. The oldest interpretation of this explanation
asserts that some workers are simply more likely to leave jobs than others, see e.g. Blumen, Kogen, and
McCarthy (1955). Heterogeneity can also be the outcome of a model where workers search for jobs from a
given quality distribution, as in Burdett (1978) or Jovanovic (1979a). In such a model, if a worker draws a
good job from the quality distribution, the probability of leaving this job is low, since the probability of being
offered an even better job is low. I abstract away from such a form of heterogeneity, since it does not have
substantial impact on the identification problem at hand. It is possible to extend my model to accommodate
such heterogeneity in the propensity to separate, but doing so would introduce additional computational
complexity.

9A more recent review is van den Berg (2000).
drift caused by the presence of heterogeneity, while the learning-by-doing process simply causes an upward drift in match-specific capital. It is exactly this difference that I exploit to distinguish the two processes.

Section 2 describes the model and characterizes its equilibrium. Section 3 presents some representative simulations to contrast the predictions of the model in the case when there is only learning-by-doing present to those in the case when there is only learning about match quality present. Section 4 describes the estimation method, the Efficient Method of Moments. Section 5 introduces the data used in the estimation. Section 6 presents the estimation results and Section 7 discusses these further and analyzes their robustness. Supporting results and discussions that are not vital to the argument are presented in Appendices.

2 The Model

Prior to describing the model environment in detail, I sketch its main elements and explain how its different elements interact to provide a framework in which identification of learning-by-doing versus learning about match quality is possible. Readers wishing to skip the technical details can proceed immediately afterwards to Section 3.

I model an economy populated by many ex-ante identical workers and firms. Any single firm employs many workers. For any two workers, working at the same firm simply means that they experience the same firm-level shocks. Otherwise, they work independently, not affecting each other’s output. In other words, labor is the only input into production, and there are constant returns to scale.

The output of a worker is affected by two learning mechanisms. First, there is learning-by-doing, which is a process by which the expected output of a worker deterministically grows with tenure. Realized output is allowed to differ from expected output, but it is expected output that determines separation decisions. I assume that the expected product of a worker is an increasing and asymptotically concave function of tenure that converges to a finite level of maximum productivity. This means that the productivity gains due to learning-by-doing decline and converge to zero over time. The particular functional form that I use is based on the micro-founded learning-by-doing model of Jovanovic and Nyarko (1995). This functional form allows one to express expected output in terms of three parameters, the extent of initial uncertainty about how to perform different tasks on the job, $\sigma_\gamma$, the noisiness of the subsequent signals that allow the worker to learn how to perform the tasks, $\sigma_y$, and the number of tasks, $N$, which captures the complexity of a job. The potential for learning-by-doing increases in all three of these variables. Second, each match has a distinct quality that is unknown at the time the match is formed. Parties learn about the quality as the employment relationship progresses. This learning process is affected by two variables in the model, $\sigma_\mu$, the dispersion of quality across matches, and $\sigma_x$, the dispersion of the signals that the partners in the match receive about match quality. The speed of learning about match quality declines in both of these variables, thereby increasing the potential for accumulation of job-specific capital due to learning about match quality. Both of these learning processes imply that expected match-specific capital increases with tenure, either in terms of increased productivity due to learning-by-doing, or in the form of knowledge about the match quality. This, in turn, implies that the model is able to capture the empirical regularity that the hazard rate of employment termination declines with tenure.

The component that allows me to distinguish between learning-by-doing and learning about match quality is the existence of firm-level shocks. I model these shocks as price
shocks; hence, they enter multiplicatively in the determination of revenue. These shocks are persistent within a single firm, but are independent across firms. Independence, together with a large number of firms, means that there is no aggregate uncertainty in the model. Persistence, on the other hand, means that a low realization of the firm-level shock leads to future realizations that are also expected to be low. This makes the expected future revenue from employing any worker lower. This, in turn, induces firms and workers to rethink whether continuing employment is profitable. Workers with insufficient job-specific capital are released and they look for jobs at other firms (whose productivity is the same on average at all times due to the lack of aggregate shocks). When learning-by-doing is dominant, workers with insufficient job-specific capital are always workers of lower tenure, and they are the ones released when unfavorable firm-level shocks hit. When learning about match quality is dominant, there are workers with insufficient job-specific capital at all tenure levels. This is because there are two effects that determine the amount of job-specific capital. On the one hand, job-specific capital increases with tenure on the job, since average quality increases with tenure as low quality workers are terminated. On the other hand, job-specific capital declines with tenure as the option value of employment declines. Which of these two effects dominates is a function of the match quality.

Firm-level shocks, the key variable of the analysis allowing identification, are generally unobservable. I capture these shocks in the data and in the model by looking at employment changes in firms due to layoffs and quits. In particular, I look at the endogenous separation rate, which is defined as the ratio of workers that are laid off or who quit during a quarter to the total number of workers at the beginning of the quarter. Presumably, at times of hardship at the firm, more workers quit or are laid off. I do not distinguish between layoffs and quits, since many quits are induced by the threat of layoff. Moreover, in the efficient separations framework used in the structural model there is no meaningful distinction between layoffs and quits, because all separations are based on mutual agreement. My choice of this endogenous separation rate to capture firm-level shocks is driven by the availability of data. Its use is validated by the fact that it shows substantial correlation with firm-level shocks in simulations.

The endogenous separation rate is a variable that can be easily matched using the structural model. This is where the assumption that a single firm employs many workers is crucial. In a one worker – one firm setup it would not be possible to construct a meaningful endogenous separation rate variable. Using a many worker – one firm setup, however, makes the likelihood function of the model much more complex. In particular, it turns out that conditioning the hazard rate of separation on the endogenous separation rate makes the likelihood function of the model intractable. There are two reasons for this, one related to the lack of data and the other to the complexity of the likelihood function even in the presence of the desired data. First, the endogenous separation rate is a function of the histories of all the workers that work at a particular firm, and these histories are not observable in the data. Second, these workers are ex-post heterogeneous due to the worker-specific learning processes that they undertake, so even with observable histories on all workers, recovering the belief of workers about the quality of their match is not possible exactly and can be done only in expectation using high-order integrals. The intractability of the likelihood function that arises from these two sources means that I need to turn to simulation-based methods to estimate the model.
2.1 The Environment

The economy is populated by a continuum of infinitely-lived workers, ex-ante identical, of measure one. A worker has to be matched to a firm in order to be able to produce output, which means that firms have some unmodelled input that is essential for production. There is a continuum of firms of measure $\alpha$.

2.1.1 Production Technology

The model has three key components: learning about match quality, job-specific learning-by-doing, and firm-level shocks. I interpret these shocks as price shocks, but they could equally well be firm productivity shocks. Below, I describe each component.

Let the output of a worker $\tau$ periods after the formation of the match be

\[ q_{\tau} = x_{\tau} h(\varepsilon_{\tau}). \]

Here $x_{\tau}$ is worker productivity at tenure $\tau$. $x_{\tau}$ is distributed normally with mean $\mu$ and variance $\sigma^2_{x}$, where $\{x_{\tau}\}_{\tau=1,\ldots}$ are identically distributed across tenure and independent both across tenure and across workers. $\mu$ is the quality of the particular employment match. It is completely match specific, and is observed neither by the worker nor by the firm at the time the match is formed. The dispersion of worker productivity around its unknown mean, $\sigma_{x}$, is common knowledge. When a firm hires a worker, the match quality $\mu$ characterizing that particular match is drawn from a normal distribution $N(\mu, \sigma^2_{\mu})$. The distribution is the same for all matches and is common knowledge, but the particular realization of $\mu$ is unknown. Hence the worker and firm learn about the unknown match quality by observing production outcomes. This is the learning about match quality component of the model and is a discrete-time version of the learning model introduced in Jovanovic (1979b).

The function $h(.)$ in (1) represents the learning-by-doing component of the model. The error term $\varepsilon_{\tau}$ represents a vector of learning-by-doing errors of length $N$, $[\varepsilon_{1\tau}, \varepsilon_{2\tau}, \ldots, \varepsilon_{N\tau}]$. The elements of $\varepsilon_{\tau}$ are identically and independently distributed according to $N(0, \sigma^2_{\tau-1})$. In particular, I choose the functional form for $h(\cdot)$ to be

\[ h(\varepsilon_{\tau}) = \prod_{i=1}^{N} (A - \varepsilon_{i\tau}^2) \]

and $\sigma^2(\cdot)$ to be

\[ \sigma^2(\tau - 1) = \frac{\sigma^2_{3} \sigma^2_{y}}{(\tau - 1) \sigma^2_{\gamma} + \sigma^2_{y}} + \sigma^2_{y}. \]

These functional forms arise naturally in the micro-founded model of learning-by-doing introduced by Jovanovic and Nyarko (1995), which I discuss in more detail in Appendix A. Given the functional form in Equation (2) and the assumption of i.i.d. learning-by-doing shocks, it follows that

\[ E_{\tau-1} [h(\varepsilon_{\tau})] = (A - \sigma^2(\tau - 1))^N, \]

which is a strictly increasing and asymptotically concave function of $\tau$ that converges to $(A - \sigma^2_{y})^N$.

In period $t$, the output produced by the workers of firm $n$ is sold at price $p_{nt}$. (Notice that I use $\tau$ to denote tenure of a particular match and $t$ to denote time period.) $p_{nt}$ follows a finite
state first-order Markov process, i.e., \( p_{nt} \in \mathcal{P} = \{p_1, \ldots, p_M\} \), where \( p_1 < p_2 < \ldots < p_M \). The transition matrix describing this Markov process is \( \Pi \) with elements \( \pi_{ij} \). The price process is assumed to be persistent, which means that the cumulative probability \( \sum_{i=1}^{j} \pi_{ij} \) is decreasing in \( j \). Moreover, the price process is such that it has a unique invariant distribution, denoted by the vector \( \hat{\pi} \). The price processes of different firms are identically distributed and independent of each other, which, together with the assumption of a continuum of firms, means that there is no aggregate uncertainty in this economy, and that in any period the distribution of firms across price states is \( \hat{\pi} \). Additionally, each period any match dissolves for exogenous reasons with probability \( \delta \). This ensures that workers do not all end up in very productive matches over time where there is no threat of separation. It also allows me to make the distinction between exogenous separations and endogenous separations in the model just as I do in the data, the latter being the ones that occur due to a decision to separate made by the agents in an employment relationship.

Timing within a period is as follows. In period \( t \), agents in employment relationships of length \( \tau \) undertake production. At the end of the period, sale price \( p_{nt} \), output \( q_\tau \), and the vector of learning-by-doing errors \( \varepsilon_\tau \) are observed. Note that, given the functional form for output, this means that productivity \( x_\tau \) can be inferred. At the end of the period, exogenous separations take place. If the match does not end due to exogenous reasons, then the agents make decisions whether to continue the match or to separate based on the observation of productivity, learning-by-doing errors, and price up to time \( t \) and tenure \( \tau \) (denoted by \( x_1^\tau \), \( \varepsilon_1^\tau \) and \( p_{-\infty} \)). The decision is made by comparing the joint value of the agents’ outside options with the value of continuing the employment relationship. Moreover, I assume that, if the two parties are indifferent between separation and continuation, then they continue the relationship.

### 2.1.2 Evolution of Beliefs

Agents have rational expectations, which implies that, at the formation of the match, their initial beliefs regarding the quality of a match is the known distribution \( N(\bar{\mu}, \sigma^2_\mu) \). The evolution of beliefs is governed by Bayes’ law. Given normal priors and normal signals, this implies that the posterior beliefs are also normally distributed.

Let the posterior belief of the agents about the match quality \( \mu \), after having observed \( \tau \) signals, be \( N(\hat{\mu}_\tau, \hat{\sigma}^2_\mu) \). Given Bayesian updating,

\[
\hat{\sigma}^2_\mu = \frac{\sigma^2_\mu \sigma^2_x}{\tau \sigma^2_\mu + \sigma^2_x},
\]

\[
\hat{\mu}_\tau = \hat{\sigma}^2_\mu \left( \frac{\hat{\mu}_{\tau-1}}{\hat{\sigma}^2_{\mu_{\tau-1}}} + \frac{x_\tau}{\sigma^2_x} \right),
\]

Notice that the posterior variance, \( \hat{\sigma}^2_\mu \), and the variance of the learning-by-doing errors, \( \hat{\sigma}^2(\tau) \), are both deterministic functions of \( \tau \), hence \( \tau \) is a sufficient statistic for these variances.\(^{11}\)

\(^{10}\)Of course, the caveats discussed in Judd (1985) regarding measurability apply here, too.\(^{11}\)This is not the case for other distributional assumptions, which is why the choice of normal distributions is very convenient.
2.1.3 Preferences

The labor supply of the workers is perfectly elastic at wage $w$, where $w$ is the alternative value of a worker’s time. This means that workers capture none of the surplus when in an employment relationship. Given that all separations are bilaterally efficient in the model, in the sense that separations only take place when the joint outside option of the parties exceeds the value of continuing the match, this assumption does not influence the decisions to separate, since that is independent of the surplus sharing rule. In fact, as I show below, the model is isomorphic to one in which workers capture a positive fraction of the surplus in a Nash bargaining setting, hence the assumption of zero surplus share is innocuous.

Both firms and workers are maximizing their expected wealth, which is just the discounted sum of their revenues, meaning that both firms and workers are risk-neutral. The common discount factor is $\beta$. In an employment relationship, firm and employee make decisions jointly and maximize the surplus of the production unit. This is equivalent to the firm making decisions unilaterally, since the worker is indifferent between being employed and being unemployed.

2.1.4 Hiring

At the end of period $t$ there are $v_{nt}$ vacancies opened by firm $n$. The total number of vacancies is $v_t = \int_0^t v_{nt}dn$. There are also $u_t$ unemployed workers. The matching function determines the number of matches created at the end of a given period and takes the simple form:

\[(7) \quad m_t = \min(v_t, u_t),\]

where $m_t$ is the number of new matches formed.\(^{12}\) Firms open vacancies taking into account two kinds of costs associated with hiring. First, there is a constant $c_0$ cost per vacancy that the firm has to pay in order to open a vacancy. Second, there is a cost $c(e_{nt})$ that firm $n$ incurs after hiring, where $e_{nt}$ is the number of newly employed workers at firm $n$. $c(.)$ is continuous, increasing, and strictly convex. This cost function represents the fact that it is relatively more costly per worker (in terms of some unmodelled organizational capital) to hire more workers in a single period.\(^{13}\) Finally, it is important to highlight that I assume undirected search. If workers were allowed to direct their search towards particular firms with observable past prices, the analysis would become much more complex. It also allows me to abstract from on-the-job search, since workers in the model do not have an incentive to undertake such search.

\(^{12}\)Notice that the timing assumption is such that all matches are created at the end of the period and they become productive in the subsequent period. Thus it is possible for workers to go from one match to the next without experiencing unemployment, thus there are job-to-job transitions in the model.

\(^{13}\)This assumption allows me to keep the size of any one firm bounded in an environment characterized by a production technology linear in labor input. If the hiring costs were linear in the number of new employees, the firm would try to hire no employees if the value of hiring a single employee was negative or zero, or would try to hire as many employees as possible if the value of hiring a single worker was positive. In this second case, workers would always find employment, there would be no unemployment in the model, and workers would be rationed. Employment dynamics in this case would become more cumbersome to simulate without adding realism to the model. Assuming convex hiring costs is the simplest way to keep employment in any one firm bounded even when the probability of filling a vacancy is one. It could also be achieved by assuming that the labor supply function is upward-sloping, that the product demand function is downward-sloping, or that there is another factor of production. Any of these other alternatives would imply complex interactions between matches of different workers reducing the simplicity and tractability of the model that is necessary to be able to carry out the simulation-based estimation.
2.2 Equilibrium

The economy is in a stationary equilibrium when the following conditions are satisfied:

- Agents in period \( t \) in existing matches make continuation decisions \( \{d^\tau\} \) in order to maximize the surplus of the relationship, where \( \{d^\tau\} \) is an adapted process with respect to \( \mathcal{F}_\tau = \sigma(x^\tau_1, \epsilon^\tau_1, p^\tau_{-\infty}, \bar{\mu}_0, \bar{\sigma}_0) \), where \( \bar{\mu}_0 \) and \( \bar{\sigma}_0 \) summarize the prior beliefs of the agents about \( \mu \).
- Agents have rational expectations: \( \bar{\mu}_0 = \mu \) and \( \bar{\sigma}_0 = \sigma_\mu \).
- Firms choose the number of vacancies \( v_{nt} \) in each period in order to maximize the discounted sum of their revenues.
- The distribution of workers across price and belief states at the end of the period, \( l(p, \bar{\mu}_\tau, \tau), p \in \mathcal{P}, \tau = 0, 1, \ldots \), and the state of unemployment is consistent with the optimal decisions of the agents in the model and is constant.

As I show below, the optimal policies are unique, which implies that this equilibrium exists and is unique.

2.2.1 Separation Decisions

Each period, agents in matches decide whether to continue their match or to separate. They base this decision on their belief about the match quality \( \mu \), the length of the relationship, and on the price faced by the firm during the last period. As I noted earlier, given the assumption of normality, the number of signals received is a sufficient statistic for the posterior variance, \( \bar{\sigma}_\mu \). The state space at the end of period \( t \) after \( \tau \) period of employment thus consists of \( p_t, \bar{\mu}_\tau, \) and \( \tau \), implying that cutoff beliefs can be represented as the value of \( \bar{\mu} \) at \( \tau \) for each price realization below which separations take place.

Given the option of quitting and taking a known unemployment value \( U \) and vacancy value \( F \) (to be derived below), the Bellman equation describing the sequential decision problem of the agents at the end of period \( t \) is

\[
V(p_t, \bar{\mu}_\tau, \tau) = \max \left\{ U + F + \sum_{j=1}^{M} \pi(p_j | p_t) \left[p_j \bar{\mu}_\tau (A - \sigma^2(\tau))^N + \beta (\delta(U + F) + (1 - \delta)V(p_j, \bar{\mu}_{\tau+1}, \tau + 1)) \right] \right\}.
\]

The first term in the parentheses represents the value of separating, while the second term is the value of continuing the match in the different price states weighted with the probability of reaching that price state. This has two parts, the expected revenue next period and the continuation value, which takes into account the fact that the match dissolves at the end of the next period for exogenous reasons with probability \( \delta \).

Given Bayesian updating, the posterior belief converges asymptotically to the truth. Hence \( \bar{\mu}_\tau \rightarrow \mu \) and \( \bar{\sigma}_\mu \rightarrow 0 \) as \( \tau \rightarrow \infty \). Asymptotically then,

\[
V(p, \mu) = \max \left\{ U + F + \sum_{j=1}^{M} \pi(p_j | p) \left[p_j \mu (A - \sigma^2)N + \beta (\delta(U + F) + (1 - \delta)V(p_j, \mu)) \right] \right\}.
\]
For a given $\mu$, the above is a system of $M$ equations in $V(p, \mu)$, $p \in \mathcal{P}$, that can be solved analytically. For details see Appendix B. Using the asymptotic value function as approximation of the value function at some very high tenure level, the value function can be approximated by iterating backwards.

From the value function, I can then derive the optimal separation decision $d(p_t, \bar{\mu}_\tau, \tau)$ for any $p_t \in \mathcal{P}$, $\bar{\mu}_\tau$ and $\tau = 0, 1, \ldots$. $d(\cdot)$ is 1 if the firm and worker decide to separate and 0 otherwise.

### 2.2.2 Hiring Decisions

In any period, the decision of firm $n$ to open vacancies is a static decision, since the number of hires in a period does not have dynamic consequences in terms of the hiring or production of future periods. Hence, when deciding how many vacancies to open at the end of period $t$, firm $n$ having faced price $p_t$ solves the following problem:

$$\max_{v_n \in \mathcal{N}} \sum_{e_n = 0}^{v_n} \left( \frac{v_n}{e_n} \right)^{\lambda e_n (1 - \lambda)^{v_n - e_n}} \{ e_n (V(p_t, \bar{\mu}, 0) - U - F) - c(e_n) \} - c_0 v_n, \tag{10}$$

where $e_n$ is the number of newly hired workers by firm $n$. $e_n$ has a binomial $B(v_n, \lambda)$ distribution, where $\lambda$ is the probability that a particular vacancy is filled, where, given the matching function, $\lambda = \min \left( \frac{u}{v}, 1 \right)$.

#### Proposition 1

For any $\{ \bar{\mu}, \sigma_\mu, \sigma_x, \sigma_\gamma, \sigma_y, N, \mathcal{P}, \Pi, \delta, \beta, w, c(\cdot), c_0 \}$ there exists $\hat{\alpha} > 0$ such that, for any $\alpha \leq \hat{\alpha}$, in equilibrium $\lambda = 1$.

**Proof:** See Appendix C.

In the remainder, I will restrict my attention to a value of $\alpha > 0$ such that $\lambda = 1$, which always exists according to Proposition 1. Also, it is clear that the number of new hires of firm $n$ is solely the function of last period’s price faced by the firm, i.e., $v_n = v(p_t)$.

Finally, I need to determine the value of a vacancy and that of unemployment. The value of the vacancy $F$ is bid down to zero, since firms are free to open vacancies. In addition, since, each period, the worker receives $w$, either as wage or as the alternative value of time or unemployment compensation when unemployed, the value of unemployment is simply

$$U = \frac{w}{1 - \beta}. \tag{11}$$

While the above determination of $U$ makes simulating the model easier, one can show that this economy is isomorphic to one in which workers capture a positive fraction of the surplus in a Nash bargaining setting.

#### Proposition 2

Given an economy $\{ \bar{\mu}, \sigma_\mu, \sigma_x, \sigma_\gamma, \sigma_y, N, \mathcal{P}, \Pi, \alpha, \delta, \beta, w^\kappa, c^\kappa(\cdot), c_0^\kappa \}$ with a positive worker share of the surplus, $\kappa$, such that $0 < \kappa < 1$ there exists an equivalent economy $\{ \bar{\mu}, \sigma_\mu, \sigma_x, \sigma_\gamma, \sigma_y, N, \mathcal{P}, \Pi, \alpha, \delta, \beta, w^0, c^0(\cdot), c_0^0 \}$ with zero worker share of the surplus, where equivalence means that the optimal policies and the distribution of workers across states is the same in the stationary equilibria of the two economies.

**Proof:** See Appendix C.
2.3 Likelihood Function

The derivation of the likelihood function is a task that I relegate to Appendix D. It requires first the derivation of the equilibrium distribution of workers, which then is used to derive the likelihood function. In the derivation, I need to consider that the firm-specific shocks are not observable. But, as I argued above, the distinction between learning-by-doing and learning about match quality is based on the different responses to firm-specific shocks under the two explanations. In order to preserve identification, I need a variable that is both correlated with the firm-specific shocks and is observable. This variable is the endogenous separation rate: the ratio of workers laid off or who quit to the total number of workers in a period. As I will show using simulations, this ratio is negatively correlated with the firm-specific shocks, as intuition would suggest.

Hence, the problem becomes that of calculating the likelihood of separation given the endogenous separation rate and potentially other observable variables. I find that this likelihood function is computationally intractable. This is because the probability of separation at a firm with given endogenous separation rate depends on the history of all the ex-post heterogeneous coworkers in the same firm, since the endogenous separation rate depends on these histories. I do not observe these histories. Moreover, even if I observed these histories, this would not be enough information to determine the distribution of beliefs at the firm, which is what determines the endogenous separation rate together with the firm-level shock. Hence estimating the above structural model by maximizing the likelihood function directly is not a feasible alternative. This leads me to consider simulation-based estimation methods since the model is relatively easy to simulate.

3 Some Representative Simulations

Prior to estimation, I present results for representative simulations in this section in order to demonstrate the qualitative difference in the predictions of learning-by-doing and those of learning about match quality. I solve the above model numerically for the simple case when there are only two prices, i.e., \( p \in P = \{p_l, p_h\} \). I choose a symmetric transition matrix

\[
\left[ \begin{array}{cc}
\rho & 1 - \rho \\
1 - \rho & \rho \\
\end{array} \right].
\]

(12)

I approximate \( V(p, \bar{\mu}_\tau, \tau) \) in Equation (8) using piece-wise linear functions in \( \bar{\mu}_\tau \) with 501 equi-distanced nodes separately for each \( p \in P \) and each \( \tau \). Piece-wise linear functions are chosen since the kink introduced by the maximum operator makes the use of smooth approximations inappropriate. I equate the value function at tenure \( \tau_{\text{max}} \) to the limit of the value function in Equation (9), then solve backwards. I choose \( \tau_{\text{max}} = 500 \) by finding the lowest value of \( \tau_{\text{max}} \) for which the optimal policy at tenure 0 does not change with the increase of \( \tau_{\text{max}} \).

Given the value function, the optimal policy function can be derived. The optimal policy function determines, for each value of the price, the cutoff belief or cutoff match quality as a function of tenure. For any tenure level, above this cutoff the worker is further employed, below it she is terminated. As I have argued above, in the case of normal conjugate priors, this cutoff belief is simply the value of \( \bar{\mu} \) at each \( \tau \) below which separations take place.
I use the optimal policy function to simulate 300,000 employment spells in a firm starting in period 5000 to eliminate the effect of initial conditions. Simulating a single firm is sufficient since, in the stationary equilibrium of the model, the distribution of worker-tenure profiles across firms is the same as the distribution of worker-tenure profiles in a single firm over time. From the simulation, I derive the endogenous separation rate at the firm in each period.

For the parameters of the structural model, I use the values in Table 1 that are chosen so that the qualitative properties of the model can be highlighted. Notice that choosing the constant $A$ in the learning-by-doing function, the average match quality, $\bar{\mu}$, and the price in the low state, $p_l$, to be equal to unity is simply convenient normalization. $\beta$, the discount factor, is set to 0.99, implying that a period in the model is roughly equal to a month. $\delta$, the rate at which exogenous separations take place, is set to 0.003, a relatively low value, implying that 0.3% of workers separate from their employer each month for exogenous reasons. The cost parameters are chosen so that firms hire five workers in the low-price state and ten workers in the high-price state.

I consider two polar cases. First, I consider the case when there is only learning-by-doing in the model. What this means is that, while there is dispersion in the match quality across workers, the quality of the match is observed by both agents immediately upon meeting. In other words, this is the case of learning-by-doing with pure heterogeneity present. (Note that the Bellman equations are altered slightly in this case: there is no updating of $\tilde{\mu}_\tau$, since $\tilde{\mu}_\tau = \mu \forall \tau$.) Let the dispersion of match quality be $\sigma_{\mu} = 0.4$, and the parameters determining the extent of learning-by-doing be $\sigma_{\gamma} = 0.4$, $\sigma_{y} = 0.4$, and $N = 5$. I set the wage to $w = (\rho p_h + (1 - \rho)p_l)(\bar{\mu} - \sigma_{\mu})(A - \sigma_{y}^2)^N$. Such a choice means that asymptotically the cutoff match quality in the high price state is $\bar{\mu} - \sigma_{\mu}$. This can also be seen from Figure 2, where the cutoff match quality converges to $\bar{\mu} - \sigma_{\mu} = 0.6$ as tenure increases.

Panel (a) of Figure 2 shows the optimal policy function for the different values of the price. The optimal policy function entails determining the cutoff match quality above which employment is continued (recall that the match quality is known in this case from the start). The cutoff match quality decreases with tenure. The firm is willing to employ lower- and lower-quality workers, since they accumulate experience which compensates for their lower match quality. Not surprisingly, the cutoff for the high price is much lower than for the low price. This simply reflects the fact that, when the price is high, it is worthwhile for the firm to continue employment relationships of less productive workers, while in times of low demand and low price, these workers are terminated. This is the “cleansing effect” of downturns, which, in this model, are firm specific. Panel (a) of Figure 3 shows the endogenous separation rate for 1000 periods between period 5000 and 6000 (to eliminate the effect of initial conditions), and shows that endogenous separations take place in bursts following negative price shocks. This turnover pattern is not unlike the one that is observable in firm-level turnover data, where separations are very concentrated in time (Hamermesh (1989)).

The correlation between the endogenous separation rate and the price shock is -0.119, showing that the use of the endogenous separation rate to capture firm-level shocks is justified. Even more striking, the correlation between the endogenous separation rate and a variable that takes on the value -1 when a low price shock “hits” (when the price process switches from $p_h$ to $p_l$) and zero otherwise is -0.738.

---

14 I start out the firm with zero employment. By period 5000 the employment of the firm always reaches its stationary level.
15 Other values of the parameters give the same qualitative but different quantitative results.
Next, assume that there is only learning about match quality in the model. This means that there is no learning-by-doing, so output each period is \( q_t = x_t \). Let the dispersion of match quality be \( \sigma_{\mu} = 0.4 \) and the dispersion of productivity around its mean be \( \sigma_x = 0.6 \). The wage is set the same way as in the case of learning-by-doing only.

Panel (b) of Figure 2 shows the optimal policy function for the different values of the price. At the beginning of any employment relationship, rational expectations imply that the agents have the average \( \bar{\mu} = 1 \) belief regarding the match quality. As signals regarding the match quality are observed, the belief of the agents changes. The upward slope in the cutoff belief is the result of the fact that the option value of employment decreases with tenure as uncertainty is resolved. The fact that, in the case of a low price, the agents with belief \( \tilde{\mu} = 0.7 \) at ten months of tenure continue the employment relationship, while with the same belief at twenty months of tenure they terminate reflects the fact that given belief \( \tilde{\mu} = 0.7 \), the probability of the actual productivity being much higher is smaller at twenty months than at ten months of tenure. The cutoff for the high price is much lower than for the low price, and this is for the same reason as in the case of learning-by-doing only. Panel (b) of Figure 3 shows the endogenous separation rate and shows that endogenous separations, here too, take place in bursts following negative price shocks.

The correlation between the endogenous separation rate and the price shock is -0.257. Also, the correlation between the endogenous separation rate and the variable that takes on the value -1 when a low price shock “hits” and zero otherwise is -0.730. This shows, once again, that the use of the endogenous separation rate is justified.

Next, I use the simulated observations to estimate a discrete-time piecewise logit hazard model with the endogenous separation rate included as an explanatory variable. The hazard rate takes the form in (13) discussed in the next section. Since I am simulating the observations, I do not have a restriction on the sample size that I can use, thus the grid \( \tau \), which determines the intervals over which the hazard rate function is the same, can be chosen to be very fine. In fact, I choose the finest grid possible, \( \tau = \{1, 2, ..., 120\} \).

In Figure 4, I plot the estimated hazard rate predicted for the two cases. In panel (a) I plot the hazard rate function at a firm that has an endogenous separation rate that is half the average and at a firm that has an endogenous separation rate that is twice the average using simulated observations for the case when there is only learning-by-doing present in the structural model. By comparing the two hazard rates, I can evaluate the effect of an increased separation rate on the hazard rate of workers of different tenure. The proportional effect of an increased endogenous separation rate declines to zero as tenure increases. In panel (b), in turn, I plot the hazard rate function for the same two values of the endogenous separation rate using simulated observations for the case when there is only learning about match quality present in the structural model. In this case, the proportional effect of an increased endogenous separation rate is the same across tenure. This figure demonstrates that the effect on the hazard rate of an increased endogenous separation rate is qualitatively very different in the two cases, which is what allows me to distinguish the two learning processes in the structural estimation.
4 Estimation

I use the Efficient Method of Moments (EMM) (Smith (1993), Gallant and Tauchen (1996)) to estimate the structural model. Simulation-based methods are very useful in situations where a complex structural model or lack of data on conditioning variables leads to an intractable likelihood function. The basic idea of such methods is to choose the parameters of the structural model so as to get similar properties for observed endogenous variables and simulated endogenous variables. The methods differ in how they make the notion of similar properties more precise. EMM does it through the use of an auxiliary model. This auxiliary model is chosen in such a way that it provides a good statistical description of the data along the dimensions of interest for the structural model. The parameters of the structural model are then chosen to match the statistical properties of the data, as captured by the auxiliary model.

The main steps of EMM are as follows. First, an auxiliary model with easily estimable parameters is chosen. Next, these parameters are estimated by maximizing the criterion function of the auxiliary model using the observed data. \( \phi \), the parameter vector of the structural model, is then chosen to minimize the distance, in a sense that will be made precise below, between the score of the auxiliary model calculated using the observed data (which is zero by construction) and the score calculated using data simulated for a given value of \( \phi \).

The choice of the auxiliary model is, of course, a crucial step in this estimation procedure. There are two approaches to a good choice. The auxiliary model can be a model that is close to the initial model or is an approximation of it. Another approach is to choose a descriptive auxiliary model with a large number of parameters that captures the data’s statistical properties along the dimensions addressed by the structural model often in an economically meaningful way. I follow this second approach and choose a discrete-time hazard model as the auxiliary model. As explanatory variables, I include piece-wise constant terms to capture non-parametrically the variation in the hazard rate across tenure and the endogenous separation rate at the firm interacted with piece-wise constant terms, once again to allow for non-parametric variation in the effect of the separation rate on the hazard rate across tenure. It is this time variation that I use to capture the difference between learning-by-doing and learning about match quality.

Furthermore, I assume that the parameters describing the learning process are the same for all workers. This assumption is necessitated by the relatively small size of the sample used, an issue that I return to in Section 7.

I model the hazard rate as a piecewise logit function. Assume that there is a grid on the integers, \( \tau = \{\tau_0, ..., \tau_n\} \) such that on interval \( m, \{\tau_{m-1} + 1, \tau_m\} \), the hazard rate function

---

\(^{16}\)This method and other simulation-based methods are discussed in Gourieroux and Monfort (1996) and Tauchen (1997).

\(^{17}\)Efficiency of the EMM estimator has only been rigorously established for auxiliary models generated by a particular semi-non-parametric method discussed in Gallant and Long (1997). The auxiliary model that I use is not generated by this method. I use the term EMM, nonetheless, based on the conjecture of Tauchen (1997) that the efficiency result holds for other classes of auxiliary models, too.

\(^{18}\)In duration analysis, the most commonly used models are continuous time hazard models. In this application, however, the existence of a large number of ties — same tenure at separation for different individuals, a zero probability event in continuous time models — makes the use of continuous time models inappropriate.
takes the logit form:

\[ h(\tau, s; \eta) = \frac{\exp(\eta_m + \eta_{n+m}s)}{1 + \exp(\eta_m + \eta_{n+m}s)}, \]

where \( s \) is the endogenous separation rate at the employing firm.

Let there be \( I \) monthly observations \( \{y_i\}_{i=1}^I \) with tenure less than or equal to \( \tau_n \) on \( J \leq I \) individuals. (I leave the hazard rate for tenure greater than \( \tau_n \) unspecified.) Each observation is of the form \( y_i = (\tau_i, f_i, s_i) \), where \( \tau_i \) is the tenure of the worker in the period of observation, \( f_i \) is 1 if the worker separates from her employer at the end of tenure \( \tau_i \) and is zero otherwise, and \( s_i \) is the endogenous separation rate at the employing firm in the given period of observation. Each individual \( j \) contributes \( \tau_j^e - \tau_j^b + 1 \) monthly observations, where \( \tau_j^b \) is the tenure at which the individual comes under observation and \( \tau_j^e \) is the tenure at which she is last observed. The probability of a single observation \( y_i \) is

\[ P(y_i; \eta) = f_i h(\tau_i, s_i; \eta) + (1-f_i)(1 - h(\tau_i, s_i; \eta)). \]

Hence the likelihood function given \( I \) monthly observations \( y_I = \{y_i\}_{i=1}^I \) is

\[ L(y_I; \eta) = \prod_{i=1}^I P(y_i; \eta) = \prod_{i=1}^I [f_i h(\tau_i, s_i; \eta) + (1-f_i)(1 - h(\tau_i, s_i; \eta))]. \]

Since, as I describe below, EMM relies on the use of gradients to determine how good the fit of a particular set of simulated data is, I need to normalize the likelihood function to make its gradients comparable across sets of simulated data. First, since the gradients are defined separately for each interval, and I wish to give the same importance to all intervals, it is not observations but intervals that need to be given equal weight. Without this normalization the estimation method would put almost all the weight on matching the coefficient for the interval with the most number of observations. This is clearly not desirable if one wants to match coefficients across all intervals as well as possible. To address this issue, I normalize the likelihood function for each interval by the number of observations that fall into that interval. Second, I need to normalize the separation rate by the average separation rate across the simulated observations. Without this normalization lower levels of the endogenous separation rate would directly imply a smaller value for the objective function to be minimized, something that is not desirable. The normalized hazard rate function is then

\[ h_N(\tau, s; \eta) = \frac{\exp(\eta_m + \eta_{n+m}s/\bar{s})}{1 + \exp(\eta_m + \eta_{n+m}s/\bar{s})}, \]

where the average endogenous separation rate in the sample is \( \bar{s} = \frac{\sum_{i=1}^I s_i}{I} \). The normalized likelihood function is

\[ L_N(y_I; \eta) = \prod_{m=1}^n \left( \prod_{i=I_{m-1}+1}^{I_m} P_N(y_i; \eta) \right)^{\frac{1}{I_{m-I_{m-1}}}} = \prod_{m=1}^n \left( \prod_{i=I_{m-1}+1}^{I_m} [f_i h_N(\tau_i, s_i; \eta) + (1-f_i)(1 - h_N(\tau_i, s_i; \eta)) \right)^{\frac{1}{I_{m-I_{m-1}}}}, \]
where \( I_m \) is the number of observations that fall into intervals 1 through \( m \), so that \( I_0 = 0 \) and \( I_n = I \). This normalization simply scales the estimates, so that

\[
(\text{arg max } L(y_1; \eta))_m = (\text{arg max } L_N(y_1; \eta))_m \quad m = 1, \ldots, n,
\]

and

\[
\bar{s} (\text{arg max } L(y_1; \eta))_{n+m} = (\text{arg max } L_N(y_1; \eta))_{n+m} \quad m = 1, \ldots, n.
\]

The corresponding normalized log-likelihood function is

\[
L_N(y_1; \eta) = \sum_{m=1}^{n} \frac{1}{I_m - I_{m-1}} \sum_{i=I_{m-1}+1}^{I_m} [f_i h_N(\tau_i; s_i; \eta) + (1 - f_i)(1 - h_N(\tau_i; s_i; \eta))].
\]

The maximum likelihood estimate of \( \eta \) is

\[
\hat{\eta} = \text{arg max}_{\eta} L_N(y_1; \eta).
\]

By construction \( \hat{\eta} \) is such that it sets the score of the likelihood function equal to zero:

\[
\frac{\partial L_N}{\partial \eta}(y_1; \hat{\eta}) = 0.
\]

Using this observation, the parameters of the structural model are chosen so as to make the score of the auxiliary model evaluated at \( \hat{\eta} \) given the simulated observations as close as possible to zero. Hence \( \phi \) is chosen such that

\[
\hat{\phi}_S(\Omega) = \text{arg min}_{\phi} \frac{\partial L_N}{\partial \eta'}(y_S(\phi); \hat{\eta}) \Omega \frac{\partial L_N}{\partial \eta}(y_S(\phi); \hat{\eta}),
\]

where \( y_S(\phi) \) is a matrix of \( S \) simulated observations of the endogenous variables using the structural model and a value \( \phi \) for the parameter vector of the structural model. \( \Omega \) is a symmetric nonnegative definite matrix weighting the elements of the score vector.

**Proposition 3** (Special case of Propositions 4.2 and 4.3 of Gouriéroux and Monfort (1996).) Under regularity assumptions (see Appendix E), the EMM estimator \( \hat{\phi}_S(\Omega) \) is consistent, asymptotically normal when \( \hat{S} = S/I \) is fixed and \( I \) goes to \( \infty \):

\[
\sqrt{I}(\hat{\phi}_S(\Omega) - \phi_0) \overset{d}{\to} N\left(0, W(\hat{S}, \Omega)\right),
\]

where

\[
W(\hat{S}, \Omega) = \left(1 + \frac{1}{\hat{S}}\right) [A_0' \Omega A_0]^{-1} A_0' \Omega B_0 \Omega A_0 [A_0' \Omega A_0]^{-1},
\]

\[
A_0 = \frac{\partial^2 L_\infty}{\partial \eta \partial \phi'}(\phi_0; b(\phi_0)),
\]

and

\[
B_0 = \lim_{I} V_0 \left(\sqrt{I} \frac{\partial L_N}{\partial \eta}(y_1; b(\phi_0))\right).
\]

Moreover, the optimal weighting matrix is \( \Omega^* = B_0^{-1} \). In this case the variance-covariance matrix simplifies to

\[
W(\hat{S}, \Omega^*) = \left(1 + \frac{1}{\hat{S}}\right) [A_0' B_0^{-1} A_0]^{-1}.
\]

For a definition of \( L_\infty \) and \( b(.) \) see Appendix E.
Notice that the matrix $A_0$ can be consistently estimated using simulations:

$$p \lim_S \frac{\partial^2 L_S}{\partial \eta \partial \phi}(y_S(\hat{\phi}_S(\Omega^*)); \hat{\eta}) = A_0.$$  

(29)

The matrix $B_0$ can be consistently estimated using the Hessian evaluated at the observed data:

$$p \lim_I \frac{\partial^2 L_N(y_I; \hat{\eta})}{\partial \eta' \partial \eta} = B_0.$$  

(30)

Given Equations (29) and (30), the optimal weighting matrix $B_0^{-1}$ and the variance-covariance matrix in Equation (28) can be consistently estimated. Furthermore, in terms of global specification tests, the following result is useful:

**Proposition 4** (Proposition 6 of Gouriéroux, Monfort, and Renault (1993).) The statistic based on the minimized objective function

$$\frac{\hat{S}I}{1 + S} \frac{\partial L_S}{\partial \eta}(y_S(\hat{\phi}_S(\Omega^*)); \hat{\eta})\Omega \frac{\partial L_S}{\partial \eta}(y_S(\hat{\phi}_S(\Omega^*)); \hat{\eta})$$  

(31)

has a $\chi^2$ distribution asymptotically with $\dim(\eta) - \dim(\phi)$ degrees of freedom.

5 Data

The data I use in this paper come from two sources. The first one is the French labor force survey (*Enquête Emploi*, hereafter EE), collected by the *Institut National de la Statistique et des Études Économiques* (INSEE). This is a rotating panel survey of a sample of 60,000 households in which one third of the sample is replaced each year. A given household is surveyed in March in three consecutive years. I use the surveys from 1991 to 1998. The EE contains standard demographic and labor market data. Moreover, it contains the identifier of the establishment that employs the interviewed individual, where INSEE can determine this identifier reliably. This identifier allows me to match the EE with the second data source, a data set on establishments.

This second data source is the *Déclaration Mensuelle des Mouvements de Main-d’Oeuvre* (DMMO) which is an administrative record of workers’ movements within French private establishments that employ at least 50 workers. It contains data on the beginning-of-month and end-of-month employment of the establishment and the number of entries and exits during the month. Furthermore, exits are broken down by reason; hence, I can distinguish between quits, economic and other layoffs, end of fixed term contracts, end of trial periods, retirement, exit to military service, transfer out, and other exits. I construct the endogenous separation rate variable as the sum of quits and layoffs divided by employment at the beginning of the period. The use of the endogenous separation rate allows me to have a measure that is independent of the average size of firms, which varies considerably across firms in the data, but which is constant across firms in my stationary model. Moreover, the use of total separations would be inappropriate since that would include workers on fixed-term contracts. These workers are excluded from both the labor force data and the data on separations from the firm for reasons explained in Section 6.

I construct employment history data by looking at two consecutive observations on the same individual. I do this using information in the EE that allows me to calculate how long
an individual has been in a particular labor market state at the time of the survey. The details can be found in Appendix F. In the 1991 through 1998 EE there are 297,191 spells consisting of two consecutive observations on an individual where the individual works on at least one of the observation dates. Of these, after deletions, I construct a sample consisting of 91,544 monthly observations of the form \( y_i = (\tau_i, f_i, s_i) \), where \( \tau_i \) is the tenure of the worker in the month of observation, \( f_i \) is unity if the worker separates from her employer at the end of tenure \( \tau_i \) and is zero otherwise, and \( s_i \) is the endogenous separation rate at the employing establishment in the given quarter of observation. The sample represents 9,420 individuals. Of course, this means that many individuals who are in the EE are not represented in my sample. A large number of deletions are due to the fact that an individual does not have a corresponding establishment identifier in the EE, or even if the identifier exists, it is not present in the DMMO. To examine the extent to which the deletions bias the remaining sample, I report in Table 7 the average characteristics of all 248,068 workers in the EE who are employed when they are first observed and those of the 9,420 workers in the matched sample that I use. We can see that, compared to the full sample, workers in the matched sample are less likely to be female or to work part time, are three and a half years younger on average, have 0.4 more years of schooling on average, and have a salary that is 3% higher than average. While most of these differences are not large, they should be kept in mind when interpreting the results. I have also estimated a discrete-time piece-wise constant hazard model without establishment variables for the 91,544 observations used and for the complete set of 1,188,434 observations that could be created the same way as the included observations when matching with establishment level data was not required. The two functions are plotted in Figure 5. It can be seen that at all tenures the hazard rate function is significantly lower for the final sample with establishment variables implying that the observations for which establishment variables exist in the DMMO tend to be for individuals in less fragile employment relationships. This observation is very important to keep in mind when interpreting the estimation results in Section 6.

There are two sources of error in the way I construct the survival time data. First, I assume that the job held in March of the first year ends when the spell of employment or non-employment starts for the second year. This implies that there is no possibility of multiple job loss within a year, and that employment spells are at times overestimated. Second, survival data are constructed using two consecutive observations in time that are a year apart. This leads to the undersampling of short employment spells that are more likely than longer spells to both start and end between the two observation dates.

To demonstrate the severity of these two sources of error, I do the following simple exercise. I posit a true hazard function \( h(\tau) = \frac{0.04}{\sqrt{\tau} + 3} \). I then simulate labor market histories comprising of employment and unemployment spells using this hazard function and assuming a second hazard function for the hazard of leaving unemployment. Simulations show that the most important determinant of the size of the first source of bias is the expected duration of unemployment spells, and not their particular distribution. This implies that the fact that the hazard rate of leaving unemployment is not constant over the unemployment spell, as has been suggested by a large body of evidence, does not influence the size of the first bias. For the sake of simplicity, therefore, I use a constant hazard rate of leaving unemployment of 0.3 that is chosen to match the expected duration of unemployment. In Figure 6 panel (a) I plot the empirical hazard function (dashed line) when all employment spells are observed. Of course, with a sufficient number of simulations, I expect to get a hazard function that is
equivalent to the posited true hazard function (solid line), which is the case here. The two functions are indeed so close to each other that it is difficult to tell them apart on the graph. In panel (b) of Figure 6, in turn, I plot the empirical hazard function when the data are collected as in the EE, with two consecutive observation dates one year apart and assuming that the spell at the first observation date ends when the spell at the second observation date starts. The empirical hazard function (dashed line) is quite different from the posited true hazard function (solid line) as a result of the bias introduced by the way the data are collected. In particular, the hazard rate at low tenures is understated both because of the undersampling of short spells and because of the overestimated separation tenures.

The severity of this bias means that it is something I need to take into account in the simulations of the structural model, since my aim is to match the statistical properties of the actual data that contain this bias. Hence, in the simulations of the structural model I collect data exactly as I do in the EE, with two consecutive observation dates one year apart and assuming that the spell at the first observation date ends when the spell at the second observation date starts. Of course, this means that simulating employment spells only is not sufficient and I also have to simulate unemployment spells. For this, I need to know the expected duration of unemployment. In the stationary equilibrium of the structural model this duration is determined by the hazard rate of leaving unemployment, which is constant and is a function of $\alpha$, the measure of firms. This means that I can choose freely the hazard rate of leaving unemployment by appropriately choosing $\alpha$.

6 Estimation Results

6.1 Auxiliary Model Using Observed Data

I estimate the auxiliary model in (20). I exclude from the estimation workers on fixed-term contracts both when constructing the sample and when calculating the endogenous separation rate. Workers on fixed-term contracts exhibit a very different pattern than workers on indeterminate-term contracts. In particular, the endogenous separation rate has no explanatory power for workers on fixed-term contracts in explaining the hazard of separation. In fact, the hypothesis that all the coefficients on the endogenous separation rate are equal to zero cannot be rejected for these workers, not even at the 80% level of confidence. This implies that the termination of fixed-term contracts is mostly due reasons other than negative shocks at the firm level, with the most likely reason being simply that these contracts expire and workers who are hired on such contracts fulfill temporary tasks. Given these differences, I exclude observations on workers on fixed-term contracts from the sample. To maintain consistency, I also exclude end of fixed-term contract from the definition of the endogenous separation rate, though it should be noted that their inclusion would not substantially alter the auxiliary model results reported below.

In choosing the grid $\tau$, I take into account two considerations. On the one hand, a finer grid better captures the variation in the hazard rate and in the effect of the endogenous separation rate across tenure. On the other hand, a finer grid decreases the number of observations in each interval $[\tau_{m-1}+1, \tau_m]$, $m = 1, ..., n$, and hence leads to less precise estimates of the parameter vector $\eta$. Taking these considerations into account, I choose a grid that is finer at lower tenure levels and is coarser at higher tenure levels: $\tau = \{0, 6, 12, 20, 30, 50, 80, 120\}$. The number of total observations and of separations per interval is recorded in Table 8. The greatest shortcoming of the data is that there are very few separations at low tenure levels,
in large part due to the filtering problem discussed in the previous section. With only 18 separations that take place at tenure of six months or less, choosing a finer grid does not provide a more accurate statistical description of the data at these tenure levels. This is an issue that influences the estimation of the structural parameters and I will return to it below.

The estimation results for the auxiliary model are reported in Table 2. Recall that $\eta_1$ through $\eta_7$ are the constant terms and $\eta_8$ through $\eta_{14}$ are the coefficients on the endogenous separation rate variable in the kernel of the logit specification for the seven intervals defined by the above grid. I report the point estimates of the coefficients in the logit specification together with asymptotic standard errors. For the constants, I also report the average predicted hazard rate for each interval calculated as the average of the predicted hazard rate over all the observations that fall into the particular interval. For the coefficients on the separation rate I report the marginal effect of a one percentage point increase in the separation rate.\footnote{A one percentage point absolute increase, not a one percentage point increase in the ratio of the separation rate to the average separation rate.} For example, the effect of an increase in the endogenous separation rate by ten percentage points compared to the average at ten months of tenure is to increase the hazard rate of separation from 1.084% to 1.744%, a substantial increase. Figure 7 is another way to present the estimation results. It plots the predicted hazard rate for a worker that works for a firm with an endogenous separation rate that is half the average and for a worker that works for a firm with an endogenous separation rate that is twice the average.

The hazard rate is declining with tenure, except for the first few months of tenure. From Figure 6 it is clear, however, that the initial rise could be the result of the way the data are collected. The effect of the endogenous separation rate on the hazard rate is always positive, as expected, and is significantly different from zero at all tenure levels. Also, it is slightly higher at lower tenure levels.

From the theory of EMM, I know that the auxiliary model has to satisfy two criteria. First, it needs to provide a good statistical description of the data along the dimension of interest for the structural model. Second, it needs to capture the features of the data that allow for the identification of the parameters of the structural model. Here I discuss whether the first criterion is satisfied. Whether the second criterion is satisfied can be gauged by looking at the standard errors on the estimated parameters of the structural model, which I report in the next subsection.

To start with, it is important to note that the specification of the auxiliary model does not include conditioning on observable characteristics of the workers besides their tenure. The reason for this exclusion is that preliminary results show that standard observable characteristics, such as education, gender, marital status, nationality, are all insignificant, even at the 90% level of confidence. The only demographic characteristic that enters significantly at the 95% level of confidence is age for workers above 50 (though not for younger workers), while working part time enters significantly at the 90% level of confidence, but not at the 95% level of confidence. The inclusion of these variables does not noticeably alter the coefficients on the separation rate variable, therefore given their limited significance and the fact that the structural model does not include these variables, I do not include them in the baseline specification. I will return to the issue of age in the robustness checks in Section 7.

The pseudo $R^2$ of the auxiliary model is 0.0322. This is quite low, implying that lot of the variation in the data is not explained by variation in the hazard rate across the intervals and variation in the endogenous separation rate. I will nonetheless argue that the first criterion of
EMM is satisfied, since the auxiliary model captures well the statistical properties of the data along the dimensions of interest for the structural model. As mentioned above, the inclusion of observable characteristics does not help in explaining the variation in the data, so most of the unexplained variation is due to variation in unobservable characteristics. The structural model itself, however, allows for a lot of variation in separation due to such unobservable characteristics through the learning about match quality channel, so a low pseudo $R^2$ can be explained in the realm of the structural model.

The relevant question to address then is whether the auxiliary model provides a good description of the data along the dimension of interest, i.e. whether it provides a good statistical description of the variation in the hazard rate and of the impact of the endogenous separation rate on this hazard rate at different tenures. To do this I perform several tests to see if the separation rate variables are indeed significant and if the auxiliary model captures all the relevant variation in the sense of outperforming a more parameter intensive specification. First, I perform a likelihood ratio test of the hypothesis that the hazard rate is the same and constant across all intervals. The value of the likelihood ratio test is 199.64 with thirteen degrees of freedom, which rejects the hypothesis at the 95% level of confidence. Next, I perform a likelihood ratio test of the hypothesis that the separation rate coefficients are all zero. The value of the likelihood ratio test is 105.76 with seven degrees of freedom, meaning that this hypothesis is also rejected at the 95% level of confidence. On the contrary, the hypothesis that all the coefficients on the separation rate are the same cannot be rejected, not even at the 50% level of confidence, since the likelihood ratio test in this case is 2.11 with six degrees of freedom. This means that while the separation rate variables have significant explanatory power, they are not significantly different from each other. This observation is crucial to keep in my mind when interpreting the estimation results for the structural model. Next, I test whether a finer grid would improve the fit of the model. Choosing a twice as fine grid, the likelihood ratio test does not reject the hypothesis that the coarser grid is appropriate at the 95% level of confidence (the value of the likelihood ratio test is 24.56 with 14 degrees of freedom.) Therefore, I conclude that, despite the low pseudo $R^2$, the auxiliary model gives a good statistical description of the data along the dimensions addressed by the structural model, which is the first criterion for the successful use of EMM.

6.2 Parameters of the Structural Model

Next, I turn to the estimation of the parameters of the structural model. As described in Section 4, this is done by finding the parameter vector $\phi$ that minimizes the EMM criterion function in Equation (23). For the weighting matrix, I use a consistent estimate of the optimal weighting matrix $B_0^{-1}$, as calculated from Equation (30). In terms of simulating observations, I proceed as in Section 3, except that now I collect the simulated observations in the same way as the data in the EE are collected. I do this to replicate the bias introduced by the data collection method, as explained in Section 5. In particular, given a set of simulated employment spells, I take a random permutation of them, insert simulated unemployment spells between two consecutive employment spells, and treat the derived data as data on a single infinitely-lived individual. Then I collect data in the same way as it is collected in the EE. The stationary nature of the model implies that this is equivalent to collecting filtered data on many individuals over a shorter period of time.

The limit of the EMM criterion function as the number of simulations goes to infinity is continuous in $\phi$. The EMM criterion function is not continuous in $\phi$ for a finite number
of simulations, however, causing difficulty in its minimization. This is because a subset of the endogenous variables is discrete, meaning that \( f \) can take on the value of zero or one depending on whether separation takes place or not. Thus, due to a perturbation in the parameter vector \( \phi \), \( f \) changes by a discrete amount. In other words, as the parameter vector is perturbed, a particular simulated spell might switch from separation to continuation, causing a discontinuous change in the EMM criterion function as long as the number of simulated spells is finite. So for any numerical implementation, the EMM criterion function is discontinuous, even though the extent of the jaggedness of the function declines as I increase the number of simulations. To demonstrate this, in Figure 8 I plot the EMM criterion function evaluated at \( \delta = 0.003, 0.00301, \ldots, 0.004 \) keeping the other parameters constant for different number of simulated monthly observations. The lowest number of observations is 3 million, the highest is 40 million. To put these numbers in perspective, 3 million observations means roughly 400,000 observations per interval, of which roughly 0.5%, or 2000, represent separations. Thus 3 million observations represents 2000 “relevant” observations per interval, thus 3 million is not an especially high number of observations. Figure 8 demonstrates two facts. First, the extent of jaggedness declines with the number of observations. In fact, the relative error in the EMM criterion function compared to its smoothed counterpart is 3.14% for 3 million observations, 1.57% for 10 million observations, 1.14% for 20 million observations and 0.66% for 40 million observations.\(^{20}\) Second, there is substantial deviation from the asymptotic value of the EMM criterion function for lower number of observations. It is only with 20 million observations that the EMM criterion function remains stable when the number of observations is increased.

The computational problem is thus that of minimizing a discontinuous function. The discontinuity makes the use of derivative-based methods inappropriate. Hence, I turn to methods that use only function evaluations. One further difficulty is computational time. A single evaluation of the EMM criterion function with the simulation of 3 million monthly observations takes around 15 seconds to perform on a Pentium III 600 Mhz machine using a computationally efficient Fortran code. This means that simply evaluating the function on a grid of five points in each of ten directions (to estimate ten parameters) would take over four years!

Given normalization \( p_l = 1, A = 1, \) and \( \bar{\mu} = 1 \), the parameters of the structural model are the following: \( \alpha, \beta, \sigma_\delta, \sigma_x, \sigma_\gamma, \sigma_y, w, N, c_0 \), the \( M(M - 1) \) free elements of the transition matrix \( \Pi \), the \( M - 1 \) elements of the price vector, and the parameters of the cost function \( c(.) \). The computational complexity of the problem requires me, however, to set some parameters of the structural model a priori, and estimate only the key parameters. I thus only estimate the parameters determining the extent of learning and of turnover, \( \delta, \sigma_x, \sigma_\mu, \sigma_y, \sigma_\gamma, N, \) and \( w \).

The remaining parameters are set in the following way. I assume that the price process follows the simplest process possible, a two-state Markov process. For the price shocks to be relevant, it is necessary that being in the low-price state means substantial revenue loss compared to being in the high-price state. For this to be the case, two conditions have to be met. The prices in the two states need to differ substantially, and the probability of transition between the two states has to be low. Hence, I assume that the price process can take on the values \( p_l = 1 \) and \( p_h = 2 \), and that the probability of remaining in the same state is 0.95 which means that prices change every twenty periods on average. The discount

\(^{20}\)The smoothing of the series is performed using a Hodrick-Prescott filter with a smoothing parameter of 14400.
factor $\beta$ is set equal to 0.99, implying that a period corresponds roughly to one month in the model. The measure of firms $\alpha$ is set so that the probability of finding a job in a period is 0.3, which roughly corresponds to the monthly job finding rate in France. Again, recall from Section 5 that it is the expected duration of the unemployment spell and not its particular distribution that has the largest impact on the size of the bias introduced by filtering, so, for simplicity, I do not consider the fact that job finding probability changes with the duration of unemployment. Finally, the cost parameters are set so that firms hire five workers in the low price state and ten workers in the high price state. This means that hiring is profitable even in bad times, which is underscored by the fact that firms hire workers even in recessions.

To find the vector $(\delta, \sigma_x, \sigma_\mu, \sigma_y, \sigma_\gamma, N, w)$ that solves the minimization problem in Equation (23) I use a hybrid method. First, I evaluate the EMM criterion function on a grid. I start with a relatively low number of simulations, and then refine the function evaluations only for points that perform well in the first round. This allows me to have quite precise estimates of the EMM criterion function at the relevant points without having to estimate it precisely at all points on the initial grid. To do this, I execute the following algorithm:

1. Set $k = 1$ and choose $S_1$.
2. Choose an initial grid $(\phi^i_1)^{T_i=1}$.
3. Evaluate the EMM criterion function on this grid using $S_k$ simulated observations. Let $(\phi^k_i, E^k_i)$ indicate the points where the function has been evaluated.
4. Put $E^{k*} = \min_i E^k_i$, and let $\phi^{k*}$ denote the corresponding argument.
5. Evaluate the EMM criterion function at $\phi^{k*}$ $J$ times, with the random number generator starting at a different value at each time to evaluate the variation in the EMM criterion function that is due to the fact that simulated data are used. Let $(E^{k*_j})_{j=1}^J$ indicate the $J$ function values.
6. Put $\sigma_k = \text{stdev}((E^{k*_j})_{j=1}^J)$.
7. Build grid $k + 1$ out of the points for which the following relationship holds:

$$(32) \quad E^k_i < E^{k*} + Z\sigma_k.$$ 

8. Choose $S_k > S_{k-1}$.
9. If $k \geq K$ then stop; otherwise, let $k = k + 1$ and go to Step 3.

The initial grid that I choose has dimensions $4 \times 10 \times 10 \times 5 \times 3 \times 3 \times 3$. I choose $K = 3$, $S_1 = 3$ million, $S_2 = 10$ million, $S_3 = 40$ million, $J = 100$, and $Z = 5$. The resulting final grid has 75 points with EMM-criterion function values between $8.92 \times 10^{-5}$ and $11.24 \times 10^{-5}$.

Second, I use these final grid points as starting values in a simplex algorithm. In evaluating the EMM criterion function, I use $S = 40$ million so that the function is evaluated precisely. I reparameterize $\sigma_\gamma$ and $w$ to make the simplex procedure more stable. In particular, I estimate $\hat{\sigma}_\gamma = \frac{\sigma_\gamma}{\sigma_y}$ and $\hat{w}$, where $w = (\rho_{ph} + (1 - \rho)p_l)(\bar{\mu} - \hat{w}\sigma_{\mu})(A - \sigma_y^2)^N$. Also,

---

21I decide how many grid points to use along each dimension based on preliminary results that indicate the sensitivity of the function to the different parameters.
I do not restrict \( N \) to take on integer values even though its interpretation in the model is number of tasks.

The final estimates of the parameter vector are displayed in Table 3 with asymptotic standard errors in parentheses. The asymptotic variance-covariance matrix is calculated using Equation (28), where \( A_0 \) is calculated using Equation (29). To calculate the derivative in \( A_0 \), I use a finite-difference method with \( S = 40 \) million and a perturbation of 0.005\% around the estimated values. The parameters describing the learning about match quality component of the model, namely \( \sigma_x \), the dispersion of productivity, and \( \sigma_\mu \), the dispersion of match quality, are quite precisely estimated and are significantly different from zero. Of the parameters describing the learning-by-doing component, namely \( \sigma_\gamma \), the extent of initial uncertainty about how to perform different tasks on the job, \( \sigma_y \), the noisiness of the subsequent signals that allow the worker to learn how to perform the tasks, and \( N \), the number of tasks, only \( N \) is significantly different from zero at the 95\% level of confidence. In fact, \( \sigma_\gamma \) is very imprecisely estimated. This is because at the estimated value of \( \sigma_y \) and \( N \) the learning-by-doing process is very quick, so the value of \( \sigma_\gamma \) could only be determined by looking at the change in the coefficients during the first few months of tenure. Such change is not allowed in the auxiliary model, however, since it imposes the restriction that the hazard rate function is the same for the first six months of tenure. As I explained above, this restriction is necessitated by the fact that there are only 18 separations falling into the interval between one and six months, partly due to the small number of overall observations, but mostly due to the filtering problem described in Section 5. This means that the data at my disposal are not informative about the speed of learning at very low tenures.

Despite this problem, the statistic in Proposition 4 is equal to 7.518, which is below the \( \chi^2 \) critical value of 14.067 at the 95\% level of confidence with seven degrees of freedom, meaning that the hypothesis that the model is well specified cannot be rejected.

7 Discussion of results and robustness

To check how well the above estimates perform, I have simulated 40 million data points using the estimated structural parameter vector and estimated the auxiliary model using this simulated data set. The results and their comparison with the coefficients estimated from the observed data are displayed in Table 4 and in Figure 9. The estimates of the coefficients of the auxiliary model using the simulated data set are close to the ones that are estimated using the observed data. This supports the claim that the structural model is well specified and does a good job of matching the statistical characteristics of the observed data as captured by the auxiliary model. The average relative difference between the constants is 2.6\%, while the average relative difference between the coefficients on the separation rate is 9.9\%. In fact, the coefficients estimated using the simulated data always fall into the 95\% confidence intervals, and in most cases are much closer to the coefficients estimated using the observed data. The coefficients on the separation rate estimated from the simulated data set match the level of the coefficients estimated from the observed data well, but do not match their slightly declining nature. I know from Section 6.1, however, that this is not a robust feature, since the hypothesis that all the separation rate coefficients are the same cannot be rejected, not even at the 50\% level of confidence.

The policy function evaluated at the final estimates is displayed in Figure 10. The cutoff belief is upward sloping as in the case of learning about match quality, implying that learning about match quality is the dominant component. In Figure 11, I plot the increase in average
output due to the learning about match quality component, the increase due to the learning-by-doing component, and the total increase. Panel (a) plots the average quality of surviving matches as a function of tenure. The average quality is increasing for a long period of time, implying that the learning about match quality is a slow process, in fact, the half life of the standard deviation of the belief about match quality is almost one year. The slowness of this process also implies that the potential for the accumulation of match-specific capital is large, since the accumulated information is very valuable. If the relationship were to end, the agents would have to invest a long time to learn about the quality of a new match. Panel (b) plots the learning-by-doing component of output as a function of tenure. Given the lack of data at low tenure levels discussed above, only limited conclusions can be drawn from the structural estimates regarding learning-by-doing. Given that the impact of the separation rate does not vanish with tenure, however, it is possible to conclude that job-specific learning-by-doing takes place quickly as output increases quickly initially, but then reaches a plateau after just a few months. Here, at the estimated values of the structural parameters, the half life of the standard deviation of the belief about the learning-by-doing component is less than one month. The speed of this process implies that the potential for the accumulation of match-specific capital is likely to be small, since going through the same learning process at another firm would take a short time. Also note that given the imprecise estimate of \( \sigma_\gamma \), it is not possible to reject the hypothesis that in fact there is no job-specific learning-by-doing at all. Panel (c) demonstrates that learning-by-doing contributes to the increase only in the first few months of tenure, while learning about match quality contributes at all tenure levels up to ten years.

All these results imply that, albeit there is a possibility that there is substantial job-specific learning-by-doing taking place during the first few months of tenure, learning about match quality is the dominant learning component after the first six months. This is not surprising, since, as I showed in Section 3, it is learning about match quality that is able to match the key observation in the data that the coefficients on the separation rate do not go to zero as tenure increases. Were learning-by-doing the key component at longer tenures, there would be a significant decline in the coefficients on the separation rate with tenure.

Notice that the presence of learning about match quality leads to a roughly 30\% increase in output over a ten year horizon, which under a constant sharing rule would translate into a 30\% increase in wages. This is in line with the estimates of Topel (1991) who finds that wage increases attributable to match-specific components are on the order of 30\% at ten years of tenure. Of course, this increase cannot be interpreted as “true” return to tenure, since the learning about match quality component simply reflects the fact that better quality matches are more likely to lead to long-term employment leading to positive selection.

It is important to keep in mind that all the conclusions in this paper are about job-specific learning-by-doing. There could be general learning-by-doing that is not related to the particular match, the magnitude of which is not determined by the model. A richer specification would allow for such general learning-by-doing. In its presence, adverse firm-level shocks would induce workers with more general learning capital (experience) to seek employment at firms in better conditions. The presence of such general learning-by-doing could be differentiated from the match-specific learning considered here by stratifying the sample by experience. Unfortunately, such stratification is not very meaningful in the present application, given the small sample size. Other interesting extensions that are beyond the scope of this paper include further breakdown of the specificity of knowledge to job- vs. firm-specific and industry-specific vs. general knowledge; and the study of the assignment of
workers to different tasks within a firm based on the accumulated knowledge of the worker and about the worker.

The structural model of this paper assumes that all shocks are firm-specific and there are no aggregate shocks. This is a reasonable assumption given that less than 1% of the variation in the endogenous separation rate is explained by year dummies. There is substantial variation in the average endogenous separation rate across the different years, however, with the average endogenous separation rate varying from a low of 1.66% in 1996 to a high of 2.66% in 1991. While incorporating aggregate shocks into the structural model is beyond the scope of this paper, it is instructive to understand how allowing for variation in the average level of separation alters the auxiliary model results. Recall that the normalized hazard rate function in the baseline specification is

\[ h_N(\tau, s; \eta) = \frac{\exp(\eta_m + \eta_{n+m}s/\bar{s})}{1 + \exp(\eta_m + \eta_{n+m}s/\bar{s})}, \]

where \( \bar{s} \) is the average of the endogenous separation rate in the sample. One way to allow for the presence of aggregate variation in the data is to consider the alternative specification for the normalized hazard rate function

\[ h_N(\tau, s, t; \eta) = \frac{\exp(\eta_m + \eta_{n+m}s/\bar{s}_t)}{1 + \exp(\eta_m + \eta_{n+m}s/\bar{s}_t)}, \]

where \( t \) is the year of the observation and \( \bar{s}_t \) is the average of the endogenous separation rate across observations in year \( t \). The estimated parameters of the auxiliary model using this alternative specification are shown in the second column of Table 5. Comparing with the baseline estimates of the first column, we can see that the estimated coefficients are very similar. In fact, the hypothesis that the coefficients on the separation rate are equal to the baseline estimates cannot be rejected, not even at the 5% level of confidence, since the Wald test statistic in this case is 2.45 with seven degrees of freedom.

Another robustness check is to allow for firm heterogeneity to be present in the data that is not present in the structural model. One way to allow for the presence of such heterogeneity is to consider the alternative specification for the normalized hazard rate function

\[ h_N(\tau, s, k; \eta) = \frac{\exp(\eta_m + \eta_{n+m}s/\bar{s}_k)}{1 + \exp(\eta_m + \eta_{n+m}s/\bar{s}_k)}, \]

where \( k \) is the identifier of the employing firm and \( \bar{s}_k \) is the average of the endogenous separation rate across observations for firm \( k \). The estimated parameters of the auxiliary model using this alternative specification are shown in the third column of Table 5. Comparing with the baseline estimates of the first column, we can see that the coefficients are similar, though they have higher standard errors, and two of the coefficient estimates on the separation rate are negative, though not at all significant. More importantly, the hypothesis that the coefficients on the separation rate are equal to the baseline estimates cannot be rejected, as the Wald test statistic in this case is 7.66 with seven degrees of freedom. Thus, allowing for aggregate variation or firm heterogeneity in the level of the endogenous separation rate does not significantly alter the results.

A crucial question is whether the findings in the data displayed in the auxiliary model are indeed a result of the effects studied in the structural model, or are driven by other possible explanations not incorporated in the model. In particular, the structural model does not explicitly incorporate labor market institutions prevalent in the French economy.
that might affect the separation decision of workers and firms. As most of these institutions, such as employment protection and unions, tend to protect high-tenure workers more than low-tenure workers,\footnote{Incorporating employment protection via a separation cost that does not change with tenure is trivial in the model, since it simply affects the outside option of the agents that is constant in the model. In fact, the model is isomorphic to one with such separation costs.} the failure to incorporate them biases my estimates towards learning-by-doing, which only supports the finding that learning about match quality is dominant at longer tenures. The reason for the direction of the bias is that protection of high-tenure workers acts the same way as the learning-by-doing component: it increases with tenure the incentive to continue the relationship, and it does so in a strictly increasing fashion.

The single labor-market institution that favors low-tenure workers over high-tenure ones is early retirement. It is possible that, given the prevalence of early retirement in France, the fact that even high-tenure workers are vulnerable to termination is simply a result of the fact that workers close to retirement have higher tenure on average. In fact, as I mentioned above, age increases the hazard rate of separation significantly for workers above 50 years of age, while there is no such effect for younger workers. To study this issue, I re-estimate the auxiliary model using only those observations where the worker is 50 years old or younger at the time he or she comes under observation. The results are reported in Table 6 and show that the effect of an increased endogenous separation rate is effectively the same for workers below the age of 50 as in the baseline case. The average hazard rate drops at higher tenure levels compared to the baseline case, which is to be expected since older workers of high tenure are much more likely to take early retirement. Comparing with the baseline estimates, I find that the hypothesis that the coefficients on the separation rate are equal to the baseline estimates cannot be rejected at any level of confidence, since the Wald test statistic in this case is 0.30 with seven degrees of freedom.

Another alternative is that the vulnerability of high tenure workers to separation is due to rapid technological change within the firm. It is possible, in principle, to study this issue by stratifying the sample according to the industry to which the employing firm belongs to and asking the question whether industries with different speed of technological change differences in the vulnerability of high tenure workers. Again, such stratification is not very meaningful in the present application, given the small sample size, hence the study of this issue is beyond the scope of this study and is left for future research.

8 Conclusion

One important source of match-specific capital is learning on the job, which can take on two distinct forms: learning-by-doing and learning about match quality. While these are conceptually distinct processes, distinguishing between them empirically has not been achieved in past work. This study distinguishes between these two explanations by building on the insight that the two explanations have different implications for how firm-specific demand shocks affect the distribution of tenure among a firm’s displaced workers. The estimation results imply that learning about match quality is the dominant explanation at tenures above six months. Learning about match quality is a process that takes place slowly and can become an important source of match-specific capital, while job-specific learning-by-doing takes place swiftly in the initial first few months and leads to no substantial accumulation of match-specific capital.
There are several caveats to keep in mind. First, the computational complexity of the problem does not allow me to estimate all the parameters of the structural model. Instead, I fix some parameters and estimate only the key parameters guiding learning-by-doing and learning about match quality. I believe that this does not have significant effects on my findings. Learning about match quality is the more important component due to the very robust finding in the data that the effect of an unfavorable firm-level shock on the hazard rate of employment termination does not die out as tenure increases. The robustness of this finding lends confidence to the results of the structural estimation. Second, the sample used comprises of workers at relatively large establishments in France in the private sector. I have shown that the employment experience of these workers is quite different from that of the average French worker, since they work in employment relationships that are overall less fragile. Third, the demanding data requirements of the model mean that only relatively few observations in the matched data set contain all the necessary information for estimation. I have close to 100,000 monthly observations, which is not a large sample considering that a separation, the event of interest, takes place with less than one percent probability. Fourth, the assumption of worker and firm homogeneity is one that has been necessitated by the computational complexity of the problem and by the relatively small sample size, but it is most likely a restrictive one. It is possible to extend the structural model to contain both worker and firm heterogeneity. With the wider availability of matched data sets, it will be possible to estimate such an extended version of the model. Hence, the question how the contribution of the two explanations varies across heterogeneous groups of workers (education groups, industries, etc.) could be addressed. Answers to this question will shed light on the different labor market experiences of these groups. For example, as shown in Nagypál (2000), variation in the importance of learning about match quality among groups can lead to unemployment rate differences among them.

There are many implications of my finding. Beyond contributing to economists’ understanding of job-specific learning in employment relationships and providing a guideline for appropriately modelling such learning, this study has important implications for labor market policies that alter the incentives of firms to substitute between workers of different tenure. In a companion paper (Nagypál (2002)), I study the effects of the introduction of employment protection policies that raise the costs of dismissal. Macroeconomists studying these policies have thus far focused primarily on their effect on the flexibility of firms in adjusting their employment. They have found that, while qualitatively the imposition of these costs leads to a reduction in welfare, quantitatively they do not seem to have a large impact. In the presence of match-specific learning, however, these policies have an additional important effect on average productivity by influencing the distribution of productivity across employed workers. Moreover, the magnitude of this productivity effect is very different depending on whether match-specific capital accumulates due to learning-by-doing or due to learning about match quality. In the companion paper I show that in the presence of learning about match quality, the imposition of dismissal costs leads to a much larger reduction in average productivity than if only learning-by-doing is present. Moreover, I show that even if the first-best policy of removing these dismissal costs completely is not available, removing dismissal costs for low-tenure workers through the liberalization of the use of fixed-term contracts can undo most of the above productivity loss when learning about match quality dominates, while there is no such positive effect when learning-by-doing dominates.
References


Appendix A: The learning-by-doing model

I follow Jovanovic and Nyarko (1995) in modelling learning-by-doing as a dial-setting problem. Each period, the worker sets a dial. The farther away her dial-setting is from the best dial setting, the lower her output. Besides being unknown to the worker, the best dial setting changes over time. The time variation in the best dial setting captures the idea that workers perform different tasks over time; for example, a sales manager is faced with different clients or a researcher with different problems. The best dial setting, however, has a component that is initially unknown, but is constant across time. For example, clients have similar needs, or problems at hand have similar characteristics. At the end of each period, the worker observes what the best dial setting was for that period. This allows her to make inferences about the constant component which, in turn, makes the prediction of the next period’s best dial setting easier, and the worker becomes more productive. Learning-by-doing is affected by three variables in the model, $\sigma_\gamma$, $\sigma_y$, and $N$. In terms of the dial setting analogy, $\sigma_\gamma$ is the dispersion of the constant component in the best dial setting across matches, $\sigma_y$ is the dispersion of the best dial setting around its constant, but unknown mean and $N$ is the number of tasks the worker carries out. The potential for productivity growth increases in all three of these variables.

Formally, let the output of a worker during the $\tau$th period of employment be

\[(36) \quad h(\epsilon_\tau) = h(\mid y_\tau - z_\tau \mid),\]

where $y_\tau$ is a vector of length $N$ of random variables that describe for period $\tau$ the target decisions the worker should take regarding $N$ independent tasks. In the dial setting analogy this is the best dial setting. $y_\tau$ is distributed normally according to $N(\gamma_i, \sigma_y^2)$. $\gamma$ is an unknown vector of constants $\{\gamma_i\}_{i=1,...,N}$ — the unknown constant component of the best dial setting — that determines the distribution of $\{y_\tau\}_{i=1,...,N}$, which are identically distributed across tenure and independent across tasks, tenure, and workers. The fact that for any $i$ the $\{y_\tau\}_{\tau=1,...}$ are stochastic but are drawn from the same distribution captures the idea that the task a worker carries out varies over time but has a common component captured by the common mean. At the beginning of the match, each element of $\gamma$ is drawn from the normal distribution $N(\bar{\gamma}, \sigma_\gamma^2)$. This distribution is the same for all matches and is common knowledge, but the particular realization of $\gamma$ is unknown. Workers learn about the constant component of the best dial setting by observing past realizations of the best dial setting. $z_\tau$, in turn, is a vector of production decisions — the actual dial setting — that the worker has to take at the beginning of period $\tau$ prior to observing $y_\tau$ at the end of period $\tau$. The difference between $z_\tau$ and $y_\tau$ (the error $\epsilon_\tau = \mid y_\tau - z_\tau \mid$) leads to a decline in output captured by the function $h(.)$. This function determines output taking into account the loss due to the imperfect knowledge of the target decision, i.e., $\partial h(.) / \partial \epsilon_i < 0 \ \forall i = 1, ..., N$. This modelling of learning-by-doing has been developed by Jovanovic and Nyarko.

Since agents are assumed to have rational expectations, their initial belief regarding each common component of the target dial setting is the known distribution $N(\bar{\gamma}, \sigma_\gamma^2)$. Moreover, let the posterior belief of the agents about the constant $\gamma_i$, after having observed $\tau$ signals,
be \( N(\tilde{\gamma}_{i\tau}, \tilde{\sigma}^2_{i\tau}) \). From Bayes’ rule

\[
\begin{equation}
\tilde{\sigma}^2_{i\tau} = \tilde{\sigma}^2_{i\tau} = \frac{\sigma^2_i \sigma^2_y}{\tau \sigma^2_i + \sigma^2_y} \quad \forall i = 1, \ldots, N,
\end{equation}
\]

and

\[
\begin{equation}
\tilde{\gamma}_{i\tau} = \tilde{\sigma}^2_{i\tau} \left( \frac{\tilde{\gamma}_{i\tau-1}}{\tilde{\sigma}^2_{i\tau-1}} + \frac{y_{i\tau}}{\sigma^2_y} \right) \quad \forall i = 1, \ldots, N.
\end{equation}
\]

When determining the optimal choice of \( z_{i\tau} \), notice that the worker is solving a static optimization problem. This is because the choice of \( z_{i\tau} \) does not affect the learning dynamics of the agents, since the signals are independent of \( z_{i\tau} \). In other words, the agents in the model are passive learners: they cannot influence the arrival of signals. Relaxing this feature is a very interesting extension of the model, but it goes beyond the scope of the exercise in this paper. I choose the function \( h(\cdot) \) mapping the mistakes into output to be of the form

\[
\begin{equation}
h(\varepsilon_{i\tau}) = \prod_{i=1}^N \left( A - \tilde{\sigma}^2_{i\tau-1} - \sigma^2_y - (E_{\tau-1}(y_{i\tau}) - z_{i\tau})^2 \right).
\end{equation}
\]

This functional form is compelling, since it states that mistakes lead to output loss at an increasing rate, and it allows the derivation of a particularly simple choice of \( z_{i\tau} \).

The agents solve the following maximization problem each period before observing \( y_{\tau} \):

\[
\begin{equation}
\max_{z_{i\tau}} E_{\tau-1}[h_{i\tau}].
\end{equation}
\]

Using the independence of signals one can show that

\[
\begin{equation}
E_{\tau-1}[h_{i\tau}] = \prod_{i=1}^N \left( A - \tilde{\sigma}^2_{i\tau-1} - \sigma^2_y - (E_{\tau-1}(y_{i\tau}) - z_{i\tau})^2 \right).
\end{equation}
\]

Thus, the optimal production decision is clearly \( z_{i\tau} = E_{\tau-1}(y_{i\tau}) \). Given this optimal production decision, \( E_{\tau-1}[h_{i\tau}] = \prod_{i=1}^N (A - \tilde{\sigma}^2_{i\tau-1} - \sigma^2_y) = (A - \tilde{\sigma}^2_{\tau-1} - \sigma^2_y)^N \), which is exactly the form for the learning-by-doing function used in Equation (4).

**Appendix B: Solving the asymptotic Bellman equation**

Let the \( M \)-state price process be represented by the vector \( P \) and its transition matrix by \( \Pi \). If the price process is persistent, then the optimal policy for a given \( \mu \) will take on the following form: separate in the lowest \( m \) price states and continue for higher prices. Assume that separation is optimal in the first \( m \) states. Partition the matrix \( \Pi \) the following way

\[
\begin{equation}
\Pi = \begin{bmatrix}
\Pi_{11} & \Pi_{12} \\
\Pi_{21} & \Pi_{22}
\end{bmatrix},
\end{equation}
\]

and the vector \( V(\mu) \) the following way

\[
\begin{equation}
V(\mu) = \begin{bmatrix}
V(p_1, \mu) \\
\vdots \\
V(p_M, \mu)
\end{bmatrix} = \begin{bmatrix}
V_1(\mu) \\
V_2(\mu)
\end{bmatrix},
\end{equation}
\]

where \( \Pi_{11} \) is a \( m \times m \) matrix and \( V_1(\mu) \) is a vector of length \( m \).
Then, using the fact that $F = 0$, the following relationships hold:

\[(43) \quad V_1(\mu) = U_1m,\]

where $1_m$ is a vector of 1’s of length $m$, and

\[(44) \quad V_2(\mu) = [\Pi_{21} \Pi_{22}]P \mu (A - \sigma_y^2)^N + \beta (\delta U_1m - (1 - \delta)\Pi_{21}U_1m + (1 - \delta)\Pi_{22}V_2(\mu)),\]

from which it follows that

\[(45) \quad V_2(\mu) = (I - \beta(1 - \delta)\Pi_{22})^{-1}
\begin{bmatrix}
(\Pi_{21} \Pi_{22})P \mu (A - \sigma_y^2)^N + \beta U_1m - (1 - \delta)\Pi_{21}U_1m
\end{bmatrix}.\]

The asymptotic Bellman equation then can be easily solved for any $\mu$ by finding the lowest $m$, such that $V_2(\mu) \geq U_1m$.

\[\text{Appendix C: Proofs}\]

**Proof of Proposition 1:** Assume that $\lambda < 1$ for any $\alpha$, contrary to Proposition 1, meaning that all displaced workers find new jobs immediately. Notice that this means no unemployment in any period, hence $u = 0$. Denote the solution to Equation (10) for a given value of $\lambda$ by $v(\lambda)$. $v(\lambda)$ is increasing in $\lambda$. Hence, the total measure of vacancies as a function of $\lambda$, $v(\lambda) = \alpha \sum_{i=1}^M \tilde{\pi}_i v(p_i; \lambda)$, is also increasing in $\lambda$. This means that $v(\lambda) \leq v(1)$ for any $0 \leq \lambda \leq 1$. $v(1)$, however, can be set arbitrarily low by appropriately choosing $\alpha$. Hence, it can be set so that the total number of vacancies is below the flow of workers out of employment in any period which is at least as large as $\delta$ when $u = 0$. This, however, contradicts the assumption that all displaced workers find a new job immediately. This means that for $\alpha$ low enough, the flow of vacancies is lower than the flow out of employment ensuring that every vacancy is filled, thereby verifying Proposition 1.

**Proof of Proposition 2:** Denote variables in the economy where workers receive $\kappa$ and zero fraction of the surplus by the superscript $\kappa$ and 0, respectively. First, note that the result that the value of vacancy is zero is independent of the assumption of zero worker surplus. Denote the solution of the Bellman equation in Equation (8) for a given value of unemployment $U$ by $V(p_t, \bar{\mu}, \tau; U)$. In the economy with $\kappa$ worker share of the surplus the following functional equation has to hold for the value of unemployment:

\[(46) \quad U^\kappa = w^\kappa + \beta \left[ \xi^\kappa \kappa \left( \sum_{i=1}^M \tilde{\pi}_i V(p_i, \bar{\mu}, 0; U^\kappa) - U^\kappa \right) + U^\kappa \right],\]

where $\xi^\kappa$ is the probability of finding a vacancy which is simply $\min \left( 1, \frac{v^\kappa}{w^\kappa} \right)$ given the matching function.

From (46)

\[(47) \quad U^\kappa = \frac{w^\kappa + \beta \xi^\kappa \kappa \sum_{i=1}^M \tilde{\pi}_i V(p_i, \bar{\mu}, 0; U^\kappa)}{1 - \beta + \beta \xi^\kappa \kappa}.\]

Now choose $w^0 = (1 - \beta)U^\kappa$ which implies that $U^\kappa = U^0$. In turn $V(p, \bar{\mu}, 0; U^\kappa) = V(p, \bar{\mu}, 0; U^0)$. Hence, the optimal separation decisions coincide in the two economies. To show that equilibrium outcomes are equivalent, I need to show that the hiring decisions of
firms are the same given proper reparametrization. In an economy where workers get fraction $\kappa$ of the surplus, firms, instead of solving the maximization problem in Equation (10), solve the following problem (taking into account that $F = 0$):

$$\max_{v_n \in \mathbb{N}} \sum_{e_n=0}^{v_n} \left( \frac{v_n}{e_n} \right) \lambda^{e_n} (1 - \lambda)^{v_n-e_n} \{e_n (1 - \kappa)(V(p_i, \bar{\mu}, 0) - U) - c^e(e_n)\} - c^0 v_n.$$

Clearly, by rescaling the cost of hiring by $1 - \kappa$, this maximization problem is the same as that in (10). Hence all optimal decision rules are the same as in an economy where workers get zero surplus. Lastly, since the distribution of workers across states is uniquely determined by the optimal decision rules of the agents, I can conclude that the economies are indeed equivalent. The reparametrization needed to achieve this equivalence is $c^0 = \frac{\kappa}{1 - \kappa}$, $c^0(e) = \frac{\kappa^e(e)}{1 - \kappa}$, and

$$w^0 = \frac{(1 - \beta)(w^\kappa + \beta \xi^e \kappa \sum_{i=1}^{M} \hat{\pi}_i V(p_i, \bar{\mu}, 0; U^\kappa))}{1 - \beta + \beta \xi^e \kappa}.$$

**Appendix D: Derivation of the likelihood function of the model**

**Equilibrium distribution of workers**

The distribution of workers across different price and belief states at the end of a period (after separations and hiring took place) is $l(p, \bar{\mu}_\tau, \tau)$ in a stationary equilibrium and is defined by three sets of equations.

$$l(p, \bar{\mu}_{\tau+1}, \tau + 1) = (1 - \delta)(1 - d(p, \bar{\mu}_{\tau+1}, \tau + 1)) \int \sum_{i=1}^{M} \pi(p \mid p_i) l(p_i, \bar{\mu}_\tau, \tau) f(\bar{\mu}_{\tau+1} \mid \bar{\mu}_\tau) d\bar{\mu}_\tau,$$

$p \in \mathcal{P}$, $\tau = 0, 1, \ldots$. Here $f(\bar{\mu}_{\tau+1} \mid \bar{\mu}_\tau)$ is the density of the posterior mean of $\mu$ after having seen $\tau + 1$ signals, given the posterior belief having seen $\tau$ signals. This equation is simply a transition equation from states after $\tau$ periods of employment into states after $\tau + 1$ periods of employment, taking into account the price transition and the fact that exogenous separations take place with probability $\delta$ and endogenous separations take place with probability $d(.).$

The second set of equations states that the rate at which workers enter into initial states $(p, \bar{\mu}, 0)$, $p \in \mathcal{P}$, depends on the rate at which firms hire workers in different price states and the probability of that price state occurring:

$$\frac{v(p_i) \hat{\pi}_i}{v(p_j) \hat{\pi}_j} = \frac{l(p_i, \bar{\mu}, 0)}{l(p_j, \bar{\mu}, 0)} \quad \forall i \neq j \quad i = 1, \ldots, M \quad j = 1, \ldots, M.$$

Finally, the total measure of employed workers is $1 - u$, where $u$ is the unemployment rate:

$$\sum_{\tau=0}^{\infty} \sum_{i=1}^{M} l(p_i, \bar{\mu}_\tau, \tau) d\bar{\mu}_\tau = 1 - u.$$

In turn, the equilibrium rate of unemployment can be determined by equating the flow out of and into unemployment. The flow out of unemployment is equal to the total number of vacancies opened in each period, since each of these is filled with probability one. This
total number is determined by the measure of firms, \( \alpha \), and the expected number of vacancies opened. The expectation is taken with respect to the distribution of firms across price states, which is simply the invariant price distribution:

\[
(53) \quad v = \alpha \sum_{i=1}^{M} v(p_i) \tilde{\pi}_i.
\]

The flow into unemployment is determined by two components, exogenous and endogenous separations, integrated across the possible states:

\[
(54) \quad \sum_{\tau=0}^{\infty} \int \sum_{i=1}^{M} (\delta + (1 - \delta)d(p_i, \tilde{\mu}_\tau, \tau)) \hat{l}(p_i, \tilde{\mu}_\tau, \tau) d\tilde{\mu}_\tau,
\]

or

\[
(55) \quad \delta(1 - u) + (1 - \delta)(1 - u) \sum_{\tau=0}^{\infty} \int \sum_{i=1}^{M} d(p_i, \tilde{\mu}_\tau, \tau) \hat{l}(p_i, \tilde{\mu}_\tau, \tau) d\tilde{\mu}_\tau,
\]

where \( \hat{l}(p_i, \tilde{\mu}_\tau, \tau) = \frac{\hat{l}(p_i, \tilde{\mu}_\tau, \tau)}{1 - u} \) so that \( \sum_{\tau=0}^{\infty} \int \sum_{i=1}^{M} \hat{l}(p_i, \tilde{\mu}_\tau, \tau) d\tilde{\mu}_\tau = 1 \). From equating (53) and (55), the equilibrium rate of unemployment is

\[
(56) \quad u = 1 - \frac{\alpha \sum_{i=1}^{M} v(p_i) \tilde{\pi}_i}{\delta + (1 - \delta) \sum_{\tau=0}^{\infty} \int \sum_{i=1}^{M} d(p_i, \tilde{\mu}_\tau, \tau) \hat{l}(p_i, \tilde{\mu}_\tau, \tau) d\tilde{\mu}_\tau}.
\]

**Likelihood of separation**

Next, I turn to the determination of the likelihood of separation. The likelihood of separation of a worker in state \( (p_i, \tilde{\mu}_\tau, \tau) \) is simply \( \delta + (1 - \delta)d(p_i, \tilde{\mu}_\tau, \tau) \). The posterior beliefs of the agents are not observable, however, so I cannot condition on them in the derivation of the likelihood function. Hence, I need to calculate the probability of separation at the end of period \( \tau \) for a worker that has been employed for \( \tau \) periods without conditioning on beliefs. In equilibrium this probability is

\[
(57) \quad \delta + (1 - \delta)S(p_i, \tau),
\]

where \( S(p_i, \tau) \) is the probability of separating endogenously, which can be expressed as

\[
(58) \quad S(p_i, \tau) = P(d(p_i, \tilde{\mu}_\tau, \tau) = 1 \mid p_i, \tau) = \int d(p_i, \tilde{\mu}_\tau, \tau) \frac{l(p_i, \tilde{\mu}_\tau, \tau)}{\hat{l}(p_i, \tilde{\mu}, \tau)d\tilde{\mu}} d\tilde{\mu}_\tau.
\]

As I argued in Section 2.3, due to the fact that prices are not observable, I need to be able to calculate the probability of separation given the endogenous separation rate and potentially other observable variables.

Consider a worker of a firm, worker 1, who has been employed for \( \tau \) periods up to time \( t \). Worker 1 works for a firm that employs \( N_t \) workers at the beginning of period \( t \) and that is endogenously separating from \( L_t \) of its workers in period \( t \). Assume for the moment that I not only know the number of employees of the firm, but I also know the tenure distribution of these workers at the end of period \( t \), \( \{\tau_i\}_{i=1}^{N_t} \). What is the probability, then, that the worker of interest separates at the end of period \( t \) conditional on the endogenous separation
rate? I can express this probability given the different price states and then integrate over the conditional probability distribution of the price states:

\[(59) \quad P(d_{t_1} = 1 \mid L_t, N_t, \{\tau_i\}_{i=1}^{N_t}) = \sum_{j=1}^{M} P(d_{t_1} = 1 \mid L_t, N_t, \{\tau_i\}_{i=1}^{N_t}, p_j) P(p_j \mid L_t, N_t, \{\tau_i\}_{i=1}^{N_t}).\]

The conditional probability in the first term in (59) can be expressed as follows:

\[(60) \quad P(d_{t_1} = 1 \mid L_t, N_t, \{\tau_i\}_{i=1}^{N_t}, p_j) = \frac{P(d_{t_1} = 1, L_t, N_t \mid \{\tau_i\}_{i=1}^{N_t}, p_j)}{P(L_t, N_t \mid \{\tau_i\}_{i=1}^{N_t}, p_j)}.

In (60) the numerator can be written as the sum of the probability of two disjoint events, the probability that worker 1 exogenously separates and the firm separates from \(L_t\) of its \(N_{t-1}\) other workers endogenously, and the probability that worker 1 endogenously separates and the firm separates from \(L_{t-1}\) of its \(N_{t-1}\) other workers endogenously:

\[(61) \quad P(d_{t_1} = 1, L_t, N_t \mid \{\tau_i\}_{i=1}^{N_t}, p_j) = \delta P(L_t, N_{t-1} \mid \{\tau_i\}_{i=1}^{N_t}, p_j) + (1 - \delta) S(p_j, \tau_1) P(L_{t-1}, N_{t-1} \mid \{\tau_i\}_{i=1}^{N_t}, p_j).

For any \(L\) and \(N\), I can write down the probability of a firm in state \(p_j\) endogenously separating from \(L\) of its \(N\) workers, given the tenure distribution of workers, by considering all the possible ways the \(L\) endogenously separating workers can be chosen:

\[(62) \quad P(L, N \mid \{\tau_i\}_{i=1}^{N}, p_j) = \sum_{(k_i)_{i=1}^{N}, k_i \in \{0, 1\}} \prod_{i=1}^{N} S(p_j, \tau_i)^{k_i} (1 - S(p_j, \tau_i))^{1-k_i}.

Finally, the second term in (59) can be expressed using Bayes’ law as

\[(63) \quad P(p_j \mid L_t, N_t, \{\tau_i\}_{i=1}^{N_t}) = \frac{P(L_t, N_t \mid \{\tau_i\}_{i=1}^{N_t}, p_j) \hat{\pi}_j}{\sum_{i=1}^{M} P(L_t, N_t \mid \{\tau_i\}_{i=1}^{N_t}, p_i) \hat{\pi}_i}.

Combining these expressions I can derive an expression for \(P(d_{t_1} = 1 \mid L_t, N_t, \{\tau_i\}_{i=1}^{N_t})\), which then I can integrate over the probability distribution of tenure distributions to get \(P(d_{t_1} = 1 \mid L_t, N_t)\). The above derivations should convince the reader that the resulting probability is prohibitively difficult to calculate. There are two reasons for this, one is conceptual and one relates to lack of data. First, even if the tenure distribution of workers at the firm was observable, the distribution of beliefs across workers is never observable, and that is what determines the likelihood of endogenous separation via Equation (58) and thereby the endogenous separation rate for the firm. Second, I do not actually observe the tenure distribution of workers at the employing firm.

**Appendix E: Regularity conditions needed to prove consistency and asymptotic normality of EMM estimator**

To prove consistency, the following five assumptions are needed:

**Assumption 1** *The normalized function \(L_N(y_1(\phi); \eta)\) tends almost surely to a deterministic function \(L_\infty(\phi, \eta)\) uniformly in \((\phi, \eta)\) as \(I \to \infty\).*
Assumption 2 This limit function has a unique maximum with respect to $\eta$: $b(\phi) = \arg \max_{\eta} L_\infty(\phi, \eta)$.

Assumption 3 $L_N$ and $L_\infty$ are differentiable with respect to $\eta$, and $\frac{\partial L_\infty}{\partial \eta}(\phi, \eta) = \lim_I \frac{\partial L_N}{\partial \eta}(y_I(\phi); \eta)$.

Assumption 4 The only solution of the asymptotic first order condition is $b(\phi)$; $\frac{\partial L_\infty}{\partial \eta}(\phi, \eta) = 0 \Rightarrow \eta = b(\phi)$.

Assumption 5 The equation $\eta = b(\phi)$ admits a unique solution in $\phi$.

The function $b(.)$ introduced in Assumption 2 is the binding function which maps the parameters of the structural model to corresponding parameters of the auxiliary model. The assumed invertability of this binding function allows one to prove the consistency of $\hat{\phi}$ by proving the consistency of $\hat{\eta}$, which is a standard problem.

For the proof of asymptotic normality two additional assumptions are needed:

Assumption 6 $p\lim_I - \frac{\partial^2 L_N}{\partial \eta \partial \eta'}(y_I; b(\phi_0)) = - \frac{\partial^2 L_\infty}{\partial \eta \partial \eta'}(\phi_0; b(\phi_0)) = A_0$.

Assumption 7 $\sqrt{I} \frac{\partial L_\infty}{\partial \eta}(y_I; b(\phi_0)) \overset{d}{\to} N[0, B_0]$.

Appendix F: Construction of Data Set

I use four series to calculate employment duration data for individuals. These are:

1. ANCA — the year an employed individual started to work for her current employer,
2. ANCM — the month an employed individual started to work for her current employer,
3. ACTA — the year a non-employed individual stopped to work for her last employer,
4. ACTM — the month a non-employed individual stopped to work for her last employer.

Unfortunately, there are inconsistencies in the data with respect to these four series. There are observations for which all four series exist, even though the ANCA/ANCM series should only exist for those who are employed at the time of the survey, while the ACTA/ACTM series should only exist for the non-employed. I use the ACT7 variable, labor force status in a seven category breakdown, to determine which observations are inconsistent. I drop observations for which there exists an ACTA that is greater than or equal to ANCA even though according to ACT7 the person is employed. I also drop observations for which there exists an ANCA that is greater than or equal to ACTA even though according to ACT7 the person is non-employed.

Next, I sort these observations into one of five cases based on whether the individual works or not at either of the observation dates, and whether she has changed jobs between the first and the second observation date (inferred from the duration of employment at the second date.)

Table 9 shows the disposition of the sample. The surveys of 1991 through 1998 comprise 586,484 individuals. From these I can construct 589,167 observations comprising of two consecutive years, which represent 367,662 individuals. Households are surveyed in three
consecutive years, unless they refuse to participate in the second or third year of the survey or they move to another address in which case no attempt is made to trace them. All members of the household who are at least 15 years old are interviewed.

Of the 589,167 observations, 291,976 (49.6%) are observations in which the individual does not work on either of the observation dates. An additional 7,120 observations (2.4% of the remainder) are excluded due to inconsistencies in the ANCA/ACTA series described above. A further 11,885 observations (4.1% of the remainder) are dropped because I cannot create employment histories due to missing data. Of the remaining observations, 66,596 (23.9%) are on fixed-term contracts, which I exclude from this study since the termination of such contracts is heavily influenced by considerations dictated by the legislation of these contracts that are not present in my model. Finally, 119,476 observations (56.5% of the remainder) have no corresponding establishment identifier for any of the observation dates on which the worker works. This highlights the fact that the collection of establishment identifiers for employed individuals in the EE is far from complete.

These deletions leave me with 92,114 observations comprising of two consecutive years. These represent 1,077,691 monthly observations. Of these, the corresponding establishment identifier can be found in the DMMO for 334,812 monthly observations (31.1%). An establishment identifier is not in the DMMO either because the establishment employs fewer than 50 workers or because the establishment is public. The matching rate of 31.1% underlines the fact that both public and small firm employment in France is substantial. The corresponding quarterly endogenous separation rate can be calculated for 240,491 monthly observations (71.8%). My final sample consists of 91,544 of these 240,491 monthly observations for which tenure is less than or equal to 120 months, or 10 years. This final sample represents 9,420 individuals.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>1.0</td>
</tr>
<tr>
<td>$N$</td>
<td>5.0</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.95</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.99</td>
</tr>
<tr>
<td>$\mu$</td>
<td>1.0</td>
</tr>
<tr>
<td>$p_l$</td>
<td>1.0</td>
</tr>
<tr>
<td>$p_h$</td>
<td>2.0</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.003</td>
</tr>
</tbody>
</table>

Table 2: Auxiliary model estimation results (standard errors in parentheses).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficients</th>
<th>Average hazard rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta_1$</td>
<td>-5.7787 (0.2724)</td>
<td>0.414%</td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>-4.7664 (0.1576)</td>
<td>1.084%</td>
</tr>
<tr>
<td>$\eta_3$</td>
<td>-4.7397 (0.1322)</td>
<td>1.027%</td>
</tr>
<tr>
<td>$\eta_4$</td>
<td>-4.8932 (0.1377)</td>
<td>0.892%</td>
</tr>
<tr>
<td>$\eta_5$</td>
<td>-5.3637 (0.1089)</td>
<td>0.528%</td>
</tr>
<tr>
<td>$\eta_6$</td>
<td>-5.5251 (0.1001)</td>
<td>0.452%</td>
</tr>
<tr>
<td>$\eta_7$</td>
<td>-5.8743 (0.1185)</td>
<td>0.311%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficients</th>
<th>Marginal effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta_8$</td>
<td>0.1394 (0.0405)</td>
<td>0.029%</td>
</tr>
<tr>
<td>$\eta_9$</td>
<td>0.1223 (0.0243)</td>
<td>0.066%</td>
</tr>
<tr>
<td>$\eta_{10}$</td>
<td>0.1136 (0.0353)</td>
<td>0.059%</td>
</tr>
<tr>
<td>$\eta_{11}$</td>
<td>0.1167 (0.0249)</td>
<td>0.052%</td>
</tr>
<tr>
<td>$\eta_{12}$</td>
<td>0.0953 (0.0139)</td>
<td>0.026%</td>
</tr>
<tr>
<td>$\eta_{13}$</td>
<td>0.1139 (0.0164)</td>
<td>0.026%</td>
</tr>
<tr>
<td>$\eta_{14}$</td>
<td>0.1002 (0.0234)</td>
<td>0.016%</td>
</tr>
</tbody>
</table>

Log Likelihood: -2998.6
Table 3: Structural model estimation results (standard errors in parentheses).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>δ</td>
<td>0.00322(0.00169)</td>
</tr>
<tr>
<td>σ_μ</td>
<td>0.6261 (0.2652)</td>
</tr>
<tr>
<td>σ_χ</td>
<td>1.0283 (0.3740)</td>
</tr>
<tr>
<td>σ_γ</td>
<td>0.6016 (7.1750)</td>
</tr>
<tr>
<td>σ_γ</td>
<td>0.3075 (0.1667)</td>
</tr>
<tr>
<td>w</td>
<td>0.5189 (0.4546)</td>
</tr>
<tr>
<td>N</td>
<td>5.0901 (2.1846)</td>
</tr>
</tbody>
</table>

EMM criterion function \( 8.17 \times 10^{-5} \)

Table 4: Comparison of the estimates of the coefficients of the auxiliary model using the observed data and using simulated observations with the estimated structural coefficients.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficients using observed data with 95% conf. intervals</th>
<th>Coefficients using simulations</th>
<th>Relative difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>η_1</td>
<td>-5.779 (-6.310 -5.248)</td>
<td>-6.141</td>
<td>6.3%</td>
</tr>
<tr>
<td>η_2</td>
<td>-4.766 (-5.074 -4.459)</td>
<td>-4.685</td>
<td>1.7%</td>
</tr>
<tr>
<td>η_3</td>
<td>-4.740 (-4.998 -4.482)</td>
<td>-4.671</td>
<td>1.4%</td>
</tr>
<tr>
<td>η_4</td>
<td>-4.893 (-5.162 -4.625)</td>
<td>-5.074</td>
<td>3.7%</td>
</tr>
<tr>
<td>η_5</td>
<td>-5.364 (-5.576 -5.151)</td>
<td>-5.409</td>
<td>0.9%</td>
</tr>
<tr>
<td>η_6</td>
<td>-5.525 (-5.720 -5.330)</td>
<td>-5.637</td>
<td>2.0%</td>
</tr>
<tr>
<td>η_7</td>
<td>-5.874 (-6.105 -5.643)</td>
<td>-5.725</td>
<td>2.5%</td>
</tr>
<tr>
<td>η_8</td>
<td>0.139 (0.060 0.218)</td>
<td>0.117</td>
<td>15.9%</td>
</tr>
<tr>
<td>η_9</td>
<td>0.122 (0.075 0.170)</td>
<td>0.115</td>
<td>5.7%</td>
</tr>
<tr>
<td>η_10</td>
<td>0.114 (0.045 0.182)</td>
<td>0.119</td>
<td>5.1%</td>
</tr>
<tr>
<td>η_11</td>
<td>0.117 (0.068 0.165)</td>
<td>0.123</td>
<td>5.2%</td>
</tr>
<tr>
<td>η_12</td>
<td>0.095 (0.068 0.122)</td>
<td>0.120</td>
<td>26.3%</td>
</tr>
<tr>
<td>η_13</td>
<td>0.114 (0.082 0.146)</td>
<td>0.113</td>
<td>1.0%</td>
</tr>
<tr>
<td>η_14</td>
<td>0.100 (0.055 0.146)</td>
<td>0.090</td>
<td>10.2%</td>
</tr>
</tbody>
</table>
Table 5: Comparison of the estimates of the coefficients of the auxiliary model when allowing for aggregate or firm-level variation in the average endogenous separation rate.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficients in the benchmark case</th>
<th>Coefficients allowing for aggregate variation</th>
<th>Coefficients allowing for firm-level variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta_1$</td>
<td>-5.779 (0.272)</td>
<td>-5.828 (0.279)</td>
<td>-5.298 (0.332)</td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>-4.766 (0.158)</td>
<td>-4.722 (0.155)</td>
<td>-4.732 (0.168)</td>
</tr>
<tr>
<td>$\eta_3$</td>
<td>-4.740 (0.132)</td>
<td>-4.722 (0.130)</td>
<td>-4.558 (0.157)</td>
</tr>
<tr>
<td>$\eta_4$</td>
<td>-4.893 (0.138)</td>
<td>-4.866 (0.136)</td>
<td>-4.955 (0.156)</td>
</tr>
<tr>
<td>$\eta_5$</td>
<td>-5.364 (0.109)</td>
<td>-5.361 (0.109)</td>
<td>-5.390 (0.129)</td>
</tr>
<tr>
<td>$\eta_6$</td>
<td>-5.525 (0.100)</td>
<td>-5.508 (0.099)</td>
<td>-5.531 (0.106)</td>
</tr>
<tr>
<td>$\eta_7$</td>
<td>-5.874 (0.118)</td>
<td>-5.859 (0.117)</td>
<td>-5.924 (0.139)</td>
</tr>
<tr>
<td>$\eta_8$</td>
<td>0.139 (0.040)</td>
<td>0.165 (0.046)</td>
<td>-0.185 (0.276)</td>
</tr>
<tr>
<td>$\eta_9$</td>
<td>0.122 (0.024)</td>
<td>0.108 (0.026)</td>
<td>0.183 (0.063)</td>
</tr>
<tr>
<td>$\eta_{10}$</td>
<td>0.114 (0.035)</td>
<td>0.105 (0.034)</td>
<td>-0.016 (0.107)</td>
</tr>
<tr>
<td>$\eta_{11}$</td>
<td>0.117 (0.025)</td>
<td>0.101 (0.023)</td>
<td>0.205 (0.062)</td>
</tr>
<tr>
<td>$\eta_{12}$</td>
<td>0.095 (0.014)</td>
<td>0.095 (0.015)</td>
<td>0.145 (0.058)</td>
</tr>
<tr>
<td>$\eta_{13}$</td>
<td>0.114 (0.016)</td>
<td>0.101 (0.015)</td>
<td>0.141 (0.034)</td>
</tr>
<tr>
<td>$\eta_{14}$</td>
<td>0.100 (0.023)</td>
<td>0.085 (0.021)</td>
<td>0.146 (0.064)</td>
</tr>
</tbody>
</table>

Table 6: Comparison of the estimates of the coefficients of the auxiliary model when using the whole sample as opposed to when restricting the sample to workers 50 years old or younger.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficients in the benchmark case</th>
<th>Coefficients for workers 50 years old or younger</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta_1$</td>
<td>-5.779 (0.272)</td>
<td>-5.801 (0.280)</td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>-4.766 (0.158)</td>
<td>-4.742 (0.159)</td>
</tr>
<tr>
<td>$\eta_3$</td>
<td>-4.740 (0.132)</td>
<td>-4.730 (0.135)</td>
</tr>
<tr>
<td>$\eta_4$</td>
<td>-4.893 (0.138)</td>
<td>-4.920 (0.144)</td>
</tr>
<tr>
<td>$\eta_5$</td>
<td>-5.364 (0.109)</td>
<td>-5.374 (0.112)</td>
</tr>
<tr>
<td>$\eta_6$</td>
<td>-5.525 (0.100)</td>
<td>-5.641 (0.109)</td>
</tr>
<tr>
<td>$\eta_7$</td>
<td>-5.874 (0.118)</td>
<td>-6.015 (0.132)</td>
</tr>
<tr>
<td>$\eta_8$</td>
<td>0.139 (0.040)</td>
<td>0.146 (0.040)</td>
</tr>
<tr>
<td>$\eta_9$</td>
<td>0.122 (0.024)</td>
<td>0.120 (0.024)</td>
</tr>
<tr>
<td>$\eta_{10}$</td>
<td>0.114 (0.035)</td>
<td>0.114 (0.035)</td>
</tr>
<tr>
<td>$\eta_{11}$</td>
<td>0.117 (0.025)</td>
<td>0.108 (0.027)</td>
</tr>
<tr>
<td>$\eta_{12}$</td>
<td>0.095 (0.014)</td>
<td>0.092 (0.014)</td>
</tr>
<tr>
<td>$\eta_{13}$</td>
<td>0.114 (0.016)</td>
<td>0.119 (0.017)</td>
</tr>
<tr>
<td>$\eta_{14}$</td>
<td>0.100 (0.023)</td>
<td>0.103 (0.023)</td>
</tr>
</tbody>
</table>
Table 7: Comparison of matched sample to complete sample of employed workers from French Labor Force Survey.

<table>
<thead>
<tr>
<th></th>
<th>Matched sample</th>
<th>Complete sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>34.09%</td>
<td>44.23%</td>
</tr>
<tr>
<td>Age</td>
<td>34.88</td>
<td>38.50</td>
</tr>
<tr>
<td>Years of education</td>
<td>10.38</td>
<td>9.97</td>
</tr>
<tr>
<td>Married</td>
<td>60.04%</td>
<td>60.81%</td>
</tr>
<tr>
<td>Monthly salary (in FF)</td>
<td>8682.7</td>
<td>8432.5</td>
</tr>
<tr>
<td>Part time</td>
<td>9.42%</td>
<td>14.28%</td>
</tr>
<tr>
<td>Sample size</td>
<td>9,420</td>
<td>248,068</td>
</tr>
</tbody>
</table>

Table 8: Number of observations and separations by interval.

<table>
<thead>
<tr>
<th>Tenure</th>
<th>Observations</th>
<th>Separations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 to 6 months</td>
<td>4,327</td>
<td>18</td>
</tr>
<tr>
<td>7 to 12 months</td>
<td>4,058</td>
<td>45</td>
</tr>
<tr>
<td>13 to 20 months</td>
<td>6,415</td>
<td>71</td>
</tr>
<tr>
<td>21 to 30 months</td>
<td>8,934</td>
<td>78</td>
</tr>
<tr>
<td>31 to 50 months</td>
<td>17,537</td>
<td>95</td>
</tr>
<tr>
<td>51 to 80 months</td>
<td>24,516</td>
<td>113</td>
</tr>
<tr>
<td>81 to 120 months</td>
<td>25,258</td>
<td>79</td>
</tr>
</tbody>
</table>

Table 9: Disposition of sample.

<table>
<thead>
<tr>
<th>Spell Description</th>
<th>Spells</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two year spells in EE between 1991-1998</td>
<td>589,167</td>
</tr>
<tr>
<td>Deletions</td>
<td></td>
</tr>
<tr>
<td>Not working at any of the two dates</td>
<td>291,976</td>
</tr>
<tr>
<td>Inconsistencies in start of employment/non-employment series</td>
<td>7,120</td>
</tr>
<tr>
<td>Cannot create employment history due to missing data</td>
<td>11,885</td>
</tr>
<tr>
<td>On fixed term contract</td>
<td>66,596</td>
</tr>
<tr>
<td>No firm identifier</td>
<td>119,476</td>
</tr>
<tr>
<td>Total deletions</td>
<td>497,053</td>
</tr>
<tr>
<td>Intermediate sample</td>
<td>92,114</td>
</tr>
<tr>
<td>Monthly observations</td>
<td></td>
</tr>
<tr>
<td>Intermediate sample</td>
<td>1,077,691</td>
</tr>
<tr>
<td>Deletions</td>
<td></td>
</tr>
<tr>
<td>Not in DMMO</td>
<td>742,879</td>
</tr>
<tr>
<td>Cannot calculate endogenous separation rate</td>
<td>94,321</td>
</tr>
<tr>
<td>Tenure greater than 10 years</td>
<td>148,947</td>
</tr>
<tr>
<td>Total deletions</td>
<td>986,147</td>
</tr>
<tr>
<td>Final sample</td>
<td>91,544</td>
</tr>
</tbody>
</table>
Figure 1: Comparison of the hazard rate of employment termination at a firm with a low separation rate (solid line) and at a firm with a high separation rate (dashed line).
Figure 2: Comparison of optimal policy functions in the low-price state (dashed line) and in the high-price state (solid line) as a function of tenure for representative simulations in two polar cases: when there is either only learning-by-doing or only learning about match quality present.
Figure 3: Comparison of the time path of endogenous separation rate in the firm over time for representative simulations in two polar cases: when there is either only learning-by-doing or only learning about match quality present.
Figure 4: Comparison of the effect of an increased endogenous separation rate on the hazard rate of separation for representative simulations in two polar cases: when there is either only learning-by-doing or learning about match quality present. The solid and the dashed lines are the empirical hazard rate of employment termination at a firm with low endogenous separation rate and at a firm with high endogenous separation rate, respectively.
Figure 5: Hazard rate of employment termination in the sample with establishment variables (solid line) and the complete sample without deleting observations due to missing establishment variables (dashed line).
Figure 6: Comparison of empirical hazard rates (dashed line) with the true hazard rate (solid line) with and without the bias introduced by the data collection method.
Figure 7: Predicted hazard rate of employment termination for a worker that works for a firm that has an endogenous separation rate that is half of the average (solid line) and for a worker that works for a firm that has an endogenous separation rate that is twice the average (dashed line).

Figure 8: The EMM criterion function evaluated at $\delta = 0.003, 0.00301, ..., 0.004$ keeping the other parameters constant for different number of simulated observations.
Figure 9: Comparison of the estimates of the coefficients of the auxiliary model using the data (solid line) with 95% confidence interval bands (crossed line) and using simulated observations with the estimated structural coefficients (dashed line).
Figure 10: Optimal cutoff belief in the low-price state (dashed line) and in the high-price state (solid line) evaluated at the final estimates.
(a) Increase in average output due to the learning about match quality component.

(b) Increase in quality unadjusted output due to the learning-by-doing component.

(c) Total increase in output.

Figure 11: Breakdown of increase in total output due to the two learning components.