Estimating the Effects of Incentives When Workers Learn about Their Ability

Raicho Bojilov*

November 21, 2010

(Job Market Paper)

Abstract

Employees often learn about their ability while working and this learning influences their decision to stay or quit. I show that only when the interaction between learning and turnover is explicitly accounted for can the effects of incentives be properly measured. In a departure from the existing literature, I introduce learning about ability and turnover in a model of effort choice and estimate the model using a unique dataset from a call center in North Carolina. The effect of incentives on effort is significant but small. The results indicate that turnover is a major channel through which incentives affect average performance. Simulating the estimated model shows that neglecting learning and turnover makes estimates of the effect of incentives on effort twice as big as they should be.

KEYWORDS: Piece Rates, Moral Hazard, Learning, Labor Turnover

JEL CODE: J22, J33, D82, D83, M51, M52

.

*Raicho Bojilov, Department of Economics, Columbia University, 1022 International Affairs Building, Mail Code 3308, 420 West 118th Street, New York, NY 10027. E-mail: rbb2104@columbia.edu. I would like to thank Professors J. Currie, L. Edlund, M. Guadalupe, W. Kopczuk, E. Verhoogen, T. von Wachter, and the other participants in the Applied Microeconomics and Labor Seminar at Columbia University for their helpful comments and suggestions. I am grateful to Professors B. MacLeod, L. Munasinghe, and B. O'Flaherty for their advice on this project. I am greatly indebted to Professors P.A. Chiappori and B. Salanié for their continuous support and guidance in this and my other research work.

1 Introduction

Learning about ability, turnover, and unobservable effort choice are defining features of many work environments. In a departure from the existing literature, I consider a model of employment dynamics that incorporates all three. Using unique data from a call center in North Carolina, I apply the model to investigate how learning about ability affects the empirical analysis of the effect of incentives. Furthermore, the model allows me to explore the channels through which pay incentives affect average productivity (performance) and in turn profits.

The data used in this study are ideally suited for the purposes of estimating the effect of incentives on performance; they contain an objective measure of individual performance (defined as output per hour), a compensation policy based on piece rates, and variation in the pay policy. I observe that steeper incentives are associated with higher performance and that persistent differences in individual performance are driven by differences in ability, which reflects the quality of the employer-employee match. Moreover, I find evidence that employees learn about the quality of the match in the course of the employment relation. Their posterior beliefs are largely responsible for their decision to stay or quit and the interaction between incentives and turnover appears to be crucial to evaluating the impact of incentives on individual welfare and profits.

The fact that individual performance affects labor turnover, while turnover determines what performance data are observed posits a serious econometric challenge. In the absence of learning about ability, a popular approach to address this challenge is to introduce individual fixed effects and estimate the observational equation on the subsample of employees who stay for the duration of the study. If the set of unobserved productivity effects remains fixed over time, it is then possible to estimate the time-varying elements of the observational equation. This is the approach adopted in Lazear (2000) and Bandiera, Barankay, and Rasul (2005) (hereafter BBR). I show that this approach, referred to in the rest of the paper as the fixed effects approach, is not appropriate when workers learn about their ability. The main idea behind this result can be illustrated by an example. Consider two workers, Alice and Bob, of identical ability who observe a sequence of two identical productivity signals, a good and a bad one. The difference between them is that Alice receives the good signal first, while Bob receives the bad signal first. Their payoff is equal to the realized signal and they also can quit after the first signal and accept the realization of a random offer that is independent from the signals. When the two know their ability, their probabilities of quitting after the first signal are equal. However, when they learn about their ability, Bob is more likely to quit than Alice. Thus, there are more “Alices” than “Bobs” among the workers who stay and changes in performance are driven by the decision to stay or quit. A failure to control for this econometric implication of learning leads to biased estimates.

Estimating the effects of pay incentives when workers learn about their ability is hard in general. However, if the stochastic technology is additively separable in effort, tenure, and individual productivity, I show how it can be done. These technology restrictions play a crucial role in my empirical work: I use them to develop and estimate a model of effort choice and turnover that nests as a special case the hypothesis of learning about ability. In this way, I obtain valid estimates of the effects of incentives. Furthermore, the estimation of the model allows me to recover the distribution of ability and trace how it evolves over time and across different pay regimes. The bias from neglecting Bayesian learning and attrition can be considerable: simulating the estimated model, I show that the fixed effects approach overestimates the effect of incentives by a factor of two.

My work contributes to several strands of the literature, most directly to the empirical literature on pay incentives. The importance of pay incentives for performance has been recognized at least since Taylor (1911). In the last two decades, McMillan, Whalley, and Zhy (1989) provide evidence that 75% of the increase in agricultural productivity in China from 1978 to 1984 can be attributed to the introduction of a responsibility system which allows communes to retain some profits. Kahn and Sherer (1990) document the wide spread use of pay incentives at white-collar office jobs and show that better evaluations are achieved by workers who face steeper incentives. Fernie and Metcalf (1996) study how different forms of compensation affect performance among British jockeys and find that the jockeys employed on fixed compensation perform worse than those who receive prizes when they win. Finally, Lemieux, MacLeod, and Parent (2009) find that since the 1970s a growing proportion of US firms have conditioned pay on performance and that this development contributed to a growing income inequality.

The availability of both performance data and a known compensation policy offers a number of advantages. Most importantly, researchers do not have to rely on strong assumptions to form a link between observed compensation and unobserved performance; knowing the compensation policy allows for a direct test for the effect of pay incentives on effort based on observed performance. The research potential of personnel records has been explored in a number of recent papers. Lazear (2000) considers the effect of switching from an hourly wage to a piece rate on the productivity of installers of windshields. He shows that as a result of the change, average productivity increases by 35% . However, Lazear cautions that about one third of the change can be attributed to selection at entry: the change in the pay regime attracted more qualified employees. To control for the effect of nonrandom attrition on the characteristics of the workforce at different tenure horizons, he employs the fixed effects approach discussed above. The same approach is also used in Lazear and Shaw (2009) that recovers monthly tenure-performance profiles. The fixed effects approach is employed in Bandiera, Barankay, and Rasul (2005) (hereafter BBR) where the authors study the interaction between pay incentives and social preferences, BBR (2007) which considers the effect of pay incentives for managers on performance of subordinates, and BBR (2009) that analyzes the effect of pay incentives on team formation. In this context, the benefits from the experimental environment in Shearer and Paarsch (1999), Shearer (2004), and Shearer and Paarsch (2009) become evident: control over the environment allows them to focus exclusively on the effect of incentives on effort choice. Unfortunately, their framework does not allow for the study of how turnover affects profits. Moreover, economists seldom can fully control the employment environment. My work contributes to the literature on incentives effects in two ways. First, for a family of stochastic technologies, it shows how to estimate the effect of incentives on effort in the presence of learning about ability and attrition. Second, while there is no evidence for selection at entry, the results indicate that turnover is a major channel through which pay incentives affect average performance and in turn profits.

The paper also relates to another strand of the empirical literature that can be traced back to the theoretical work on search by inspection and wage rigidity in Jovanovic (1979) and Harris and Holmstrom (1982). The availability only of compensation data has posited a major challenge to related empirical work. Chiappori, Salanié, and Valentin (1999) address this problem by exploring the testable implications of Bayesian learning and downward rigidity on the dynamics of compensation series. Since turnover is close to nonexistent in their data, their estimates do not suffer from the econometric problems discussed here. Yet, even the dynamics of compensation series are of limited help in distinguishing between learning about match quality and learning-by-doing for a large number of models: Mortensen (1988) shows that these forms of learning impose the same testable implications on the dynamics of compensation data. This property caused insurmountable identification problems to empirical work in the past.[1] For example, in their paper Gibbons, Katz, Lemieux, and Parent (2005) are forced to assume away tenure effects in order to estimate the quality of industry-specific matching. Nagypal (2007) proposes an alternative approach to identification based on estimating a structural model of learning-by-doing and learning about match quality. However, she imposes very strong functional form restrictions and abstracts away from incentive effects. The availability of performance data, along with the estimation approach that controls for learning about ability and turnover, allows me to study key features of the stochastic technology up to a constant. As a result, in this paper I can test for Bayesian learning, characterize its dynamics, and explore how pay incentives affect learning outcomes.

In the process of estimating the model, I also recover the returns to months of tenure. This result relates to a large body of literature starting with the empirical findings in the late 1970s and early 1980s, Jovanovic and Mincer (1981) being one of the most cited, of a large seniority or tenure effect on earnings. Abraham and Farber (1987) caution that the empirically observed strong relation between tenure and earnings in many cases is a statistical

[1] Nagypal (2007) provides a comprehensive review of the related literature since 1988.

artifact due to the positive relation between seniority and an omitted variable. Altonji and Shakotko (1987) confirm the findings in Abraham and Farber (1987). However, Topel (1991) finds that a two-step first difference approach applied to the same model and similar data yields substantially higher returns to tenure. In a later work, Altonji and Williams (2005) review the strengths and weaknesses of the applied methods and conclude that, if turnover is driven by learning about match quality, Topel (1991) does not control adequately for attrition bias. In contrast to this literature, I model explicitly the attrition process and estimate the effect of monthly tenure on performance[2]. Furthermore, I show that learning about ability still generates a bias, as discussed in Abraham and Farber (1987), even when compensation is not a function of past performance signals but a simple deterministic rule.

The rest of the paper is organized as follows. Section 2 explores the implications of learning about ability and attrition on the estimation of the effect of incentives and of returns to tenure. Section 3 presents the data used in the empirical work. Section 4 demonstrates how to estimate the effects of incentives in the presence of learning and nonrandom attrition. Section 5 presents the estimated effects of incentives and tenure under the fixed effects approach and under the estimation approach of section 4. The section continues with an investigation of the bias resulting from the inappropriate use of the fixed effects approach. It ends with a set of robustness checks that confirm the choice of technology restrictions and the specification of the attrition process. Section 6 summarizes the main results and concludes with remarks on related research.

2 Bayesian Learning and Nonrandom Attrition

Labor turnover may depend on unobserved productivity parameters. Indeed, this is a central feature of many models, starting with Jovanovic (1979). Such models predict that workers with low values of the unobserved productivity parameters are more likely to quit. If the

[2] Buchinsky et. al. (2009) rely purely on wage data to model expliticly turnover when it is generated through search by inspection, while preserving the structure of the observational equation as in Abraham and Farber (1987).

econometrician ignores such a process of nonrandom attrition and pools all available observations to estimate an equation for, say, performance, the estimated effects of tenure and incentives are biased. As pointed out in the introduction, past research addresses this problem by estimating the observational equation using fixed effects only on the subsample of employees who stay at the firm for the duration of the study. In what follows, I will show that this fixed effects approach yields biased estimates when the workers engage in Bayesian learning about their ability. I discuss the econometric implications of nonrandom attrition and learning in the context of a model that incorporates both effort choice and labor turnover. A central feature of the model is the strong separability of the stochastic technology in effort, ability, and tenure. This restriction is consistent with the properties of the data used in the empirical analysis, and it emphasizes that in the presence of Bayesian learning, even when effort choice does not depend on posterior beliefs and ability can be differenced out, the fixed effects approach yields biased estimates of the effects of incentives.

When applied to the data generated by such a model, fixed effects approach yields biased estimates simply because there is a nonzero correlation between noise in the performance signal and the outside offer. This problem is well understood in the existing literature on selection and attrition. Here I focus on the bias generated by learning about ability and in the rest of the section maintain the assumption that noise in the performance series and the outside offer are independent. If workers learn about their ability only in the course of their employment relation, their separation decisions are based on their posterior beliefs and through them on observed noisy signals. Consequently, the decision to stay is not independent from noise in the performance series.

2.1 Model

The model is a variation on the standard model of search by experience, first introduced in Jovanovic (1979). Each period, workers choose not only whether to stay or quit, but also how much effort to exert. The crucial element in the model is an ability parameter θ_i that is unknown at the time of hiring and both the employee and the employer learn symmetrically about its value over time through a sequence of noisy performance signals y_{it}. At the beginning of the employment relation, the worker has a prior θ_{i1}. At the beginning of each period t, she receives an outside offer ξ_{it}, and decides to stay if the value of continued employment is greater than the outside offer. I assume that if an employee quits, she is never hired again.[3] If the worker decides to stay, she chooses a level of effort l_{it}, that is not observable or verifiable by the firm.[4] Then she receives an output signal y_{it} governed by the following stochastic technology:[5]

$$y_{it} = \theta_i + g(t) + l_{it} + \varepsilon_{it} \tag{1}$$

where ε_{it} is iid over time and across individuals, θ_i is independent of the error process, and $g(t)$ represents the accumulation of firm-specific knowledge.[6] After observing the output signal y_{it} the worker is paid $w_{it} = \alpha_{it} + \beta_{it} y_{it}$, according to a linear compensation regime $R_{it} = (\alpha_{it}, \beta_{it})'$. Regime R_{it} is said to be more generous than regime R'_{it} , $R_{it} > R'_{it}$, if both $\alpha_{it} > \alpha'_{it}$ and $\beta_{it} > \beta'_{it}$. For the sake of exposition in this section, workers are assumed to be risk-neutral with a utility function $u(R_{it}, l_{it}, y_{it}) = \alpha_{it} + \beta_{it} y_{it} - \psi(l_{it})$.[7] The piece rate is taken as given from the perspective of the worker and changes in it are modelled as unforeseen shocks. In this way, I abstract away the issue of forming expectations, which is outside the scope of this paper.[8] The cdf of the posterior belief θ_{it} is a function of the cdf of the initial prior θ_{i1} and the noisy signals about θ_i up to period t, $\{y_{ik} - g(k) - l(R_{ik})\}_{k=1}^{t}$. Let μ_{it} denote the mean of the posterior belief. Since θ_i and l_{it} enter additively in the utility function, optimal effort

[3] This assumption is not restrictive to the empirical work in the second half of the paper: only five out of 675 employees are rehired after quitting. The second spells of employment are dropped out.

[4] See Malcomson and MacLeod (1992) for a discussion on the implications of these assumptions.

[5] Since both the actual and estimated performance are always greater that 0, the restriction $y_{it} \geq 0$ never binds.

[6] Jovanovic and Nyarko (1994) provide an alternative specification for the accumulation of firm-specific knowledge, which is sometimes refered to as learning-by-doing. However, the data do not support the prediction of their model that the variance of individual performance declines over time. Note that the model also assumes that the accumulation of knowledge does not depend on past or present effort.

[7] Prendergast (2002) discusses in great detail the relation between risk-aversion and optimal performance pay. However, for the purposes of this paper, the investigation the effect of a change in incentives on effort, the assumption of risk-neutrality is not restrictive.

[8] While at first sight this assumption may appear restrictive, Kanemoto and MacLeod (1992) show that when firms learn about individual ability the existence of an outside option for workers disciplines firms to keep piece rates fixed even after beliefs are updated. This issue is discussed further in section 4.

choice does not depend on θ_{it} and is function only of R_{it}, $l(R_{it})$. Define

$$G(\theta_{it}, R_{it}, t) = \max_{l_{it}} E_{\theta_{it},\xi}(u(R_{it}, l_{it}, y_i)) + \delta \int \int \max[\xi_{it}, G(\theta_{it+1}, R_{it}, t+1)]\, dF_{\theta_{it+1}} dF_{\xi},$$

where $E_{\theta_{it},\xi}$ indicates that expectation is taken with respect to the cdf of θ_{it} and ξ. $F_{\theta_{it+1}}$ is the transitional distribution over the set of possible beliefs at the beginning of period $t+1$ based on the information at t and F_{ξ} is the cdf of outside offers[9]. An employee decides to stay if

$$G(\theta_{it}, R_{it}, t) > \xi_{it} \tag{2}$$

and quits otherwise. Consequently, performance for period T is observed only if (2) holds for $t = 2, ..., T$. The assumptions of the model, the existence of a solution to the worker's problem and its characterization are presented in Appendix A. Under assumptions AA1 to AA4 in the appendix, if R_{it} is more generous than R'_{it} optimal effort $l(R_{it}) > l(R'_{it})$ and $G(\theta_{it}, R_{it}, t) > G(\theta_{it}, R'_{it}, t)$. Furthermore, the value of continued employment increases in θ_{it} in the sense of the likelihood ratio property.

Alternatively, the worker may know the value of her θ_i at the time of hiring. This is a special case of the model presented above, referred to in the rest of the paper as the case of known ability. Here the prior belief is a degenerate distribution centered at the true value. Performance in period T is observed if for all $t = 2, ..., T$

$$G(\theta_i, R_{it}, t) > \xi_{it} \tag{3}$$

Again, if R_{it} is more generous than R'_{it}, optimal effort $l(R_{it}) > l(R'_{it})$ and $G(\theta_i, R_{it}, t) > G(\theta_i, R'_{it}, t)$, while $G(\theta_i, R_{it}, t)$ increases in θ_i. While on the surface the two cases appear very similar, they have very different econometric implications explored below.

[9] For simplicity of exposition, I take the outside offer as fixed over time. The model taken to the data, however, allows for more general specification of the outside offer. See section 4 for more details.

2.2 Estimating the Effect of Tenure

Suppose for the moment that the principal interest of the econometrician lies in the estimation of returns to tenure. Assume that the observational equation is defined by (1) and that the piece rate remains the same across individuals and over time. In the case of known ability, i knows θ_i at the time of hiring and performance is observed in period T if (3) holds for $t = 2, ..., T$. If ξ_{it} is iid over time, and is independent of the error process for the observational equation, the decision to stay and the errors $\{\varepsilon_{ik}\}_{k=1}^{T-1}$ are independent, so for $t = 1, ..., T-1$

$$E\left[\varepsilon_{it} | \; min\left[\{G(\theta_i, R_{it}, k) + \xi_{ik}\}_{k=1}^{T-1}\right] > 0\right] = 0.$$

Thus, estimating (1) using fixed effects or first differences, subject to (3) for all $t = 2, .., T-1$, yields unbiased estimates of the effect of tenure on performance in the first T periods.

Next, consider the case of Bayesian learning, in which i learns the value of θ_i over time. The expected value of the disturbance term ε_{it} for $t = 1, ..., T-1$

$$E\left[\varepsilon_{it} \middle| \; min\left\{\{G(\theta_{ik}, R_{it}, k) + \xi_{ik}\}_{k=1}^{T-1}\right\} > 0 \;\right] \tag{4}$$

is generally different from zero. Since θ_{it} is an increasing function of $\{\varepsilon_{ik}\}_{k=1}^{t}$,[10] the conditions for staying define implicitly left truncations of the unconditional distribution of ε_{it}. To illustrate, assume that the distribution of ε_{it} is log-concave. Then (4) is positive and increases in $min\left\{\{G(\theta_{ik}, R_{it}, k) + \xi_{ik}\}_{k=1}^{T-1}\right\}$. Intuitively, for any θ_i if employee i stayed at least T periods, then in each period t before T she could not have been excessively unlucky in her draws of ε_{it}. What "excessively" means depends on the structural parameters of the model, in particular on θ_i. Thus, estimating the observational equation (1) using fixed effects or first differences, subject to (2) for $t = 2, ..., T$ yields biased estimates of the effect of tenure on performance in the first T periods.

Diagram 1A illustrates the point that Bayesian learning imposes a left truncation on the

[10] More details can be found in Appendix A.

distribution of observed signals for $T = 2$. For example, the fact that individual i stays for at least 3 periods indicates that given θ_i she must have been sufficiently lucky in both the first and second period. However, whether the conditional expectation of noise for the first period or the second period is larger depends on the structural parameters of the model, in particular θ_i. The former case is presented on Diagram 1B and the latter on Diagram 1C. Since the truncation thresholds decrease in θ_i, the "learning" bias declines with θ_i. If for the majority of workers the conditional expectation of noise in $t = 1$ is smaller than its counterpart in $t = 2$, as shown on Panel C, then the fixed effects approach overestimates the effect of tenure on performance. However, if the majority of the conditional expectations trend downwards as shown on Panel B , the fixed effects approach underestimates the effect of tenure. Consequently, the direction and magnitude of the "learning" bias depend in a complicated fashion on tenure, ability, the realized errors, and the other structural parameters.

2.3 Estimating the Effect of Incentives

The analysis in the preceding paragraphs assumed that the piece rate does not vary over time or across individuals. In what follows, I relax this assumption. Suppose that regime 1 is more generous than regime 2, and consider the case when the pay regimes are introduced sequentially, first regime 1 and then regime 2, and suppose that θ_i is known at the time of hiring. The fixed effects approach again yields unbiased estimates of the effect of incentives, since the same set of workers are exposed to both pay regimes. However, when Bayesian learning takes place, the estimated incentives effect is biased upwards, as illustrated on Diagram 1B. Suppose that a worker switches from regime regime 1 to regime 2 in the second period and that the econometrician conditions on staying for at least two periods. Since the employee stays for more than one period, the conditional expectation of noise for $t = 1$ is positive, while for $t = 2$ it is zero. As a result, the fixed effects approach overestimates the effect of incentives. The same argument extends to more than two periods. Suppose that beliefs converge to the true value of θ_i quickly and that regime 1 is in place during these crucial periods. Then, the average of the conditional expectations of noise under the initial regime 1 is positive, while under regime 2 close to zero.[11]

2.4 Implications

The preceding paragraphs indicate that the fixed effects approach yields biased estimates of both tenure and incentives effect when posterior beliefs drive separation decisions. This result follows from the fact that the noise from the signals affect posterior beliefs, which in turn determine individual actions, such as the decision to stay or quit. At a higher level of generality, Bayesian learning introduces dependency on tenure in the observed series of signals through the separation decisions. These observations leave the econometrician with a choice to model attrition explicitly or to estimate the treatment and tenure effects on the set of agents who have a probability of quitting close to zero. The first approach utilizes all available data, but at the price of imposing strong assumptions on the attrition and performance processes. In what follows, I adopt the second approach and investigate the appropriateness of the associated assumptions in section 5.4.

3 Data

This section presents the environment of the empirical study and the descriptive analysis. The data set has several features that make it comparable to the data sets used by Bandiera, Barankay, and Rasul, as well as the data sets used by Lazear and Shaw. It contains a clean performance measure and three piece rates that were implemented in a way that allows to identify each one's effect on performance. However, what makes it particularly appealing is the presence of considerable turnover, consistently above 50% during the first six months of employment. The descriptive analysis indicates that neither a pure moral hazard model nor a model of pure learning about match quality can account for the observed data patterns.

[11] Matters are complicated further by sample size and the horizon of the study. Convergence of beliefs to the true value of the unobserved parameters is achieved for the model presented in this paper. Since the issues is not central to the argument, no proofs are provided. An excellent treatment of the issue can be found in Easley and Kiefer (1988) and Aghion, Bolton, Julien and Harris (1991).

3.1 Context

The data are collected at a call center in North Carolina owned and operated by a multinational company. The call center collects outstanding debt and fees on behalf of cable TV companies, which ensures a stable demand for its services. An automated switchboard operator allocates inbound and outbound calls, so that the longest weighting customer is matched with the longest weighting operator. Employees rotate their work stations on a daily basis. In its recent history, the call center suffered from low average productivity and high labor turnover.

As part of its reorganization plans, the central management implemented a piece rate (regime 1) as a pilot project to evaluate the consequences of switching from hourly wage to a piece rate across all of its call centers. This regime change was implemented at the beginning of January 2005. The piece rate was a linear function of the performance metric, the number of calls per hour that end with collection of the outstanding debt. Importantly, one's pay did not depend on the performance of others; in theory there may be competition among the employees for calls, but in practice this possibility is ruled out by the chronic shortage of workers at the call center. The firm experienced difficulties attracting candidates to fill in vacancies, so the management hired virtually all candidates during its monthly hiring rounds. The central management was concerned that the company was paying "too much," so it implemented a new piece rate for the newly-hired employees in June 2005 (regime 2). Relative to regime 1, regime 2 offered a lower base pay, decreased the slope of the piece rate for those with performance less than 3.8 call per hour, and increased the slope of the piece rate for those with performance greater than 3.8 call per hour (regime 2). All previously hired employees continued to be paid according to regime 1. Since the central management was worried about possible negative effects of the piece rate on the quality of service, it changed the pay regime yet again in November 2005. The new regime 3 had two components: all employees were paid according to the pay schedule of regime 2, but in addition employees had to meet certain minimum quality standards of service to qualify for the piece rate. Twenty per cent of one's calls were randomly monitored and the quality of service was rated on a scale from 0 to 100. An employee who did not meet the minimum quality standard was relegated to an hourly wage equal to the base pay of the piece rate. Since 99% of performance lies between 1.05 and 3.8, regimes 2 and 3 effectively lowered incentives relative to regime 1. Diagram 2 shows a time line for the implementation of the three regimes and Table 2 some descriptive statistics of interest.

3.2 Descriptive Analysis

The call center experienced high turnover rates under all pay regimes: more than 50% of all employees under regime 1 quit within the first six months of employment, while under regimes 2 and 3 the turnover for the first six months approached 67%. There also appears to be a noisy downward trend in the separation rates as tenure increases. This noisiness is probably due to the small sample size, but it also suggests that separation decisions depend to a large extent on individual-specific factors. Table 1 reports the average performance for the first six months of employment across regimes. Again, as one may expect, the average performance under regime 1 is higher than its counterparts for regimes 2 and 3. Furthermore, the average performance on the subset of workers who stay for at least six months is higher than the simple average, suggesting that poor performers quit.

Figure 11 presents evidence for persistent differences in performance across individuals that are consistent with the existence of unobserved individual productivity effects. The figure plots average performance in periods 2 to 5 conditional on the performance quartile in the first month of employment. If there were no persistent differences in the productivity of employees, performance in months 2 to 5 would be the same across the initial performance quartiles. This hypothesis is not supported by the data: the workers in the top initial quartile have consistently higher performance in periods 2 to 5 than their counterparts in the other three quartiles. Furthermore, average performance for the employees in each initial quartile increases over time: for all quartiles, the difference between average performance in periods 1 and 5 is statistically significant at 5 %. Finally, performance does not seem to be "fanning out" over time.

This evidence suggests that steep pay incentives lead to high performance; that attrition appears to be nonrandom, since workers with higher performance are more likely to stay; that individual-specific effects are present, but performance does not "fan out" over time; and finally that workers accumulate experience or knowledge in the course of their first six months of employment.

4 Estimation

This section starts by introducing the attrition model taken to the data. One crucial implication of the functional form restriction on the technology is that posterior beliefs do not affect effort choice which simplifies considerably the estimation of the effects of incentives. The rest of the section is devoted to presenting how to estimate the model using MLE. The validity of the technology restriction is discussed in section 5.4.

4.1 Performance and Attrition Equations

The estimated model is identical to the one presented in section 2, except that the observational and attrition equations include also all covariates observable at period t, X_{it}. I do not specify a utility function explicitly and estimate G flexibly to accommodate a number of variations on the basic model. For example, the outside offer may vary with the accumulated firm-specific knowledge if that knowledge is exportable[12], it may vary with some of the observed covariates, or it may depend on an unobserved individual effect. Without imposing structure on the utility function, it is possible to identify only the effect of a change in pay incentives relative to the benchmark regime 1, since one cannot distinguish between effort under the initial regime 1 and the mean of the distribution of θ_i. Nevertheless, such a specification is sufficient to test whether incentives affect performance. The downside is that many parameters in the attrition equations do not have a clear interpretation in relation to the underlying model. I also assume that employees take the piece rate as given and do

[12] See Mortensen (1988) for a discussion of the conditions on a time-varying outside offer that ensure a well-behaved solution of the worker's problem.

not expect it to change. Given that the average tenure at the firm is around 3.5 months, an employee could have realistically expected that the same regime would last for the duration of her employment spell.

The model is estimated under the following additional distributional assumption.

Assumption A (i) *$\varepsilon_{it} \sim N\left(0, \sigma_{\varepsilon}^{2}\right)$ and $\xi_{it} \sim N\left(0, \sigma_{\xi}^{2}\right)$ are iid across tenure horizons and individuals, independent from the rest of the covariates.* (ii) *$\theta_{i} \sim N\left(0, \sigma_{\theta}^{2}\right)$ is iid over time and across individuals and is independent from the rest of the covariates.*

These assumptions impose strong restrictions: they rule out temporary but persistent health or family shocks. The plausibility of these assumptions is evaluated in section 5.4 through some simple post-estimation tests. Under the assumption above, the posterior belief θ_{it} is also normally distributed for all $t : \theta_{it} \sim N\left(\mu_{it}, \sigma_{t}^{2}\right)$, where for $t > 1$

$$\mu_{it} = (1 - K_{t})\mu_{it-1} + K_{t}(y_{it-1} - l\left(R_{it}\right) - g\left(t-1\right)) \tag{5}$$

$$\sigma_{t}^{2} = \frac{\sigma_{\varepsilon}^{2}\sigma_{\theta}^{2}}{\sigma_{\theta}^{2}\left(t-1\right) + \sigma_{\varepsilon}^{2}}$$

$$K_{t} = \frac{\sigma_{\theta}^{2}}{\sigma_{\theta}^{2}\left(t-1\right) + \sigma_{\varepsilon}^{2}}$$

Thus, Bayesian updating becomes quite tractable. In particular, the precision of beliefs depends only on t, so the average of the demeaned past signals is a sufficient statistic to characterize posterior beliefs. The first equation in (5) can be rewritten as

$$\mu_{it} = k\left(t\right). \left(\frac{1}{t-1}\sum_{k}^{t-1}\left(y_{ik} - l\left(R_{it}\right) - g\left(k\right) - m\left(X_{it}\right)\right)\right) + \left(1 - k\left(t\right)\right)\mu_{i1}.$$

The function $k\left(t\right)$ represents the precision that the worker attaches to the average demeaned past performance as a signal about her θ. Since $k\left(t\right)$ increases over time, she attaches greater and greater weight to the average of the demeaned past performance and less to the mean of the initial belief. Under the assumption of a common prior, this discussion implies that the average of demeaned past performance is a sufficient statistic for posterior beliefs and their effect on the decision to stay or quit.

The two cases of the model in section 2 can be nested within the following general model:

$$y_{it} = \theta_i + l\left(R_{it}\right) + g\left(t\right) + m\left(X_{it}\right) + \varepsilon_{it},$$

$$s_{ik} = 1\left[G\left(\underbrace{\lambda\mu_{ik} + (1-\lambda)\,\theta_i}_{\widetilde{\theta}_{ik}}, R_{ik}, k, X_{ik}\right) - \xi_{ik} > 0\right] \tag{6}$$

where y_{it} is observed if $s_{ik} = 1$ for all $k = 2, ..., t$. If $\lambda = 1$, Bayesian learning is present and employees share a common prior. If $\lambda = 0$, workers know their match quality at the time of hiring. A $\lambda \in (0,1)$ is difficult to interpret, but probably suggests that the initial prior is correlated with ability[13]; finally $\lambda < 0$ is a clear rejection of the model.

4.2 MLE

The crucial difference between attrition and selection models is that $s_{it} = 1$ implies that observing performance in one period implies that performance is observed also in all preceding periods. Thus, $s_{it} = 1$ provides information about the value of θ that affects the estimation of $y_{i\tau}$ across all observed periods. In contrast, selection models assume that the selection process takes place in each period independently. The literature on estimation of attrition models starts with Hausman and Wise (1979) who offer an estimation method based on a full information MLE. The MLE method in this study is similar, but also involves "integrating out" the unobserved effects[14]. The derivation of the likelihood is discussed in more detail in Appendix B. Here I provide only a summary of the main features of the MLE.

Let Θ_1 be the vector of parameters to be estimated and the available information about individual i be W_i conditional on θ_i. The likelihood for individual i conditional on the data

[13] I do not pursue this avenue any further because a nonparametric test discussed later indicates that there is no self-selection at entry as regimes vary; the finding is consistent with the absence of any prior knowlege about ability.

[14] Florens, Heckman, Meghir, and Vytlacil (2008) provide the basis of an alternative approach based on the use of control functions, which does not require distributional assumptions on the noise in the performance signal. However, small sample size and the fact that the MLE fits the data well, as discussed in the following section 6.2, argue in favor of the approach taken in the paper.

and θ_i can be written in a standard way as follows:

$$
\begin{aligned}
& l_i \left(\Theta_1 | \theta_i, W_i\right) \\
& = \left[\prod_{t=1}^{T_i} \left(\varphi \left(\frac{y_{it} - g\left(t\right) - m\left(X_{it}\right) - l\left(R_{it}\right) - \theta_i}{\sigma} \right) \Phi \left(G \left(R_{it}, t, X_{it}, \widetilde{\theta}_{it} \right) \right) \right)^{S_{it}} \right] \\
& \left(1 - \Phi \left(G \left(R_{iT_i}, T_i, X_{iT_i}, \widetilde{\theta}_{iTi} \right) \right) \right)^{1 - S_{iT_i}},
\end{aligned}
$$

where $S_{it} = \prod_{k=1}^{t} s_{ik}$ and T_i is the last period in which i is observed. The interpretation of this expression is quite intuitive. The individual observes a performance signal, updates her belief, and decides whether to stay or quit. If she stays, the econometrician observes her effort choice and separation decisions in the following periods. If she quits, the econometrician does not, so the unobserved performance and separation series are "integrated out" and do not appear in the conditional ML. Since θ_i is not observed, it is integrated out to obtain

$$
l_i \left(\Theta | W_i\right) = \int l_i \left(\Theta_1 | \theta_i, W_i\right) .\varphi(\theta_i | W_i, \Theta_2) d\theta_i,
$$

where Θ_2 is a vector of parameters that define the distribution of θ_i and Θ is a vector that contains all parameters in Θ_1 and Θ_2 . Finally, the log-likelihood is obtained by taking logs and summing over i:

$$
l \left(\Theta | \left\{ W_i \right\}_{i=1}^{N} \right) = \sum_{i=1}^{N} \log l_i \left(\Theta\right)
$$

Note that the model of attrition with Bayesian learning also implies the exclusion restriction that the average of past performances enters the attrition but not performance equations. Under alternative estimation methods, this restriction provides the basis for identification.

5 Results

This section presents the results from estimating the model from section 2 and investigates some alternative specifications of attrition. The empirical results are consistent with the presence of Bayesian learning that leads to a considerable upward bias in the estimated effect of incentives and tenure on performance when the fixed effects approach is used. They also show that switching from regime 1 to regime 2 leads to a decline in effort but also to an improvement in average ability as tenure increases.

5.1 Results for the Attrition Model

Table 2 reports the results from estimating the attrition model by MLE. Model 1 is estimated under the restriction of known ability of match quality, or $\lambda = 0$. Model 3 is estimated under the restriction of learning about ability or $\lambda = 1$. Model 2 nests both Models 1 and 3 as special cases and estimates λ. The performance equations under all three models are the same: the explanatory variables include third degree orthogonal polynomials of tenure and calendar time, dummies for regimes of operation and regimes of hiring, and controls. Regime 2 enters additively as implied by the theoretical model. Since regime 3 has the same pay schedule as regime 2 but conditions pay on the quality of service, the performance equation incorporates interaction terms between the tenure polynomials and regime 3. The attrition equations include third degree orthogonal polynomials interacted with regimes and, depending on the specification, θ_i or μ_{it}, controls, calendar time, and regime of hiring.

The estimated λ under Model 2 is 0.74 and is significantly different from 0, but not significantly different from 1, which is consistent with the hypothesis of learning. A likelihood ratio test fails to reject the restriction $\lambda = 1$, while rejecting the restriction $\lambda = 0$. I take these results to imply that Bayesian learning is present and that Model 3 is the correct model. Accordingly, in what follows I discuss the estimated coefficients under the restriction of Bayesian learning. The variance of initial ability is 0.48 calls per hour, significantly different from 0, and accounts for the greater part of the variance in performance in the first months of employment. Moreover, it has an important effect on attrition. Figure 1 presents the random truncations of the distribution of ability at different tenure horizons under regime 1.[15] The value of θ is on the horizontal axis, while the vertical axis represents the proportion

[15] Here truncation of $f(x)$ with lower limit a is defined as $\int_a f(x)\,dx$. That is, one has removed the part of

of agents of a certain match quality who are present in the firm at a given tenure horizon. The figure shows that the conditional distribution of ability in a cohort of employees shifts to the right as tenure increases: by 0.62 calls per hour within the first six months on the job. Most workers in the bottom quartile of the distribution quit within the first 2 months of entry in the firm, and most workers in the bottom half of the distribution within 6 months. The figure suggests that only workers with very high match quality face a low probability of quitting. As expected, the switch from regime 1 to the less generous regimes 2 and 3 leads to an increase in turnover at any tenure horizon, as shown on Figure 2. Noticeably, the truncations under regime 2 and 3 are very similar, suggesting that the quality standard under regime 3 did not have a significant effect on effort choice and quitting.

The variance of the disturbance term in the performance equation is estimated at 0.17 which implies that the ratio of the variance of match quality over the variance of noise is approximately 2.6 initially. After 6 months the variance of the posterior beliefs declines to approximately 0.05 and the weight that the worker puts on observed signals when forming her beliefs, $k(t)$, approaches 1, as evident from Figure 3. The figure indicates that both the shape of the trajectory is consistent with what theory predicts and the coefficient is significantly different from zero at all tenure horizons.

Recall that relative to regime 1, regime 2 offered less incentives to exert effort and lower base pay. The estimated parameters for the performance equation are broadly consistent with the theoretical predictions; the effect of regimes 2 and 3 on effort is negative and significant. However, the effect of regime 2 does not differ significantly from the effect of regime 3. This finding is confirmed by plotting on Figure 4 the tenure–performance profile for an entering employee, conditional on staying, across regimes. Regime 2 is restricted only to a downward shift in performance across tenure horizons, while regime 3 is also interacted with tenure. The restrictions on the way regime 2 enters in the performance equation follow directly from the restriction imposed on the stochastic technology. The fact that the estimated trajectories of performance under regime 2 and 3 is comforting. Switching from regime 1 to regime 2 leads to

the distribution less than a but not scaled up the distribution to integrate to one over its domain.

a decline in effort that translates into 0.19 fewer calls per hour, which is approximately 9% of the average initial performance under regime 1. In economic terms, the change in incentives leads to a decline in worker's hourly pay by approximately \$2, which is 20% of the average hourly pay in the first month of employment under regime 1.

All models estimate a significant improvement in performance over time due to accumulation of experience: in the first 6 months of employment performance increases by approximately 0.67 calls per hour, or 23% growth in the first six months under regime 1. Under regime 1, this growth translates in an increase in hourly pay by approximately \$2.2. Finally, the dummies for regimes of hiring are not significant, which indicates that there is no self-selection into the firm on the basis of the pay regime at the time of hiring.[16] Among the rest of the covariates, the percentage of outbound calls has a negative effect on performance and the average of past percentages of outbound calls has a negative and significant effect on performance. Women have on average lower performance than men, and marriage has a positive effect on performance but a negative effect on the probability of staying.

Figures 1 and 2 made clear that pay incentives encourage workers of high ability to stay and workers of low ability to quit. Thus, selective attrition is another channel through which pay incentives affect average performance and profits. Figure 5 compares the magnitudes of the effects of attrition, experience and effort choice. By inducing workers of low ability to quit, regime 1 leads to the successive improvement of the quality mix after each month on the job which translates into an increase in average performance among staying workers by 0.62 calls per hour in the first six months of employment. This effect is approximately two-thirds of the growth in performance induced by the accumulation of experience. Since regime 2 offers both less incentives to exert effort and lower base pay, the probability of quitting increases across different types of ability. Yet, this increase is not uniform: workers of low and average ability are more affected than workers of high ability. As a result, the average performance among the staying employees increases by 0.73 calls per hour in the first six months of employment, which is slightly more than one standard deviation of the distribution

[16] The test for selection at entry is the same asthe one used in Lazear (2000). At least in principle, the effect of selection at entry may affect not only the mean of the distribution of θ, but also other moments.

of ability in the population. Thus, the quality mix after 6 months under regime 2 is actually better than the quality mix under the more generous regime 1.

In more detail, Figure 7 investigates how regimes 1 and 2 affect the probability of staying for employees of different abilities from the first to the sixth month of employment. Under both regimes 1 and 2, workers of low ability leave the firm within the first two months of employment. The greatest difference is in their effect on the workers of average and high ability. Regime 2 practically forces workers of average ability to leave within the first six months of employment. It also reduces the number of high ability workers who stay, partly due to the effect of bad signals in the early stages of the employment relation. Figure 8 plots the probability of quitting at tenure $t = 6$ as a function of the posterior mean. By the sixth month of employment, workers have a much more precise beliefs about their ability than at the time of entry and the accumulation of experience has already plateaued. Thus, one may regard Figure 8 as an approximation to the state at which quitting behavior depends on incentives and ability only. The figure shows that the impact of regimes 1 and 2 is similar for workers in the tails of the distribution of ability and differs greatly for those in between. Since workers of low ability would have already left, as indicated on Figure 7, the main impact of regime 2 relative to regime 1 is that it "weeds out" the workers of average and slightly better than average ability.

Combining the effects of incentives and selective attrition gives the net impact of changing pay incentives on average performance. Note that the decline in performance due to low effort is partially offset by the increase of the quality of the remaining employees: by the sixth month it is only 0.1 calls per hour. Figure 6 plots the effect of selective attrition on average performance under different regimes. It confirms the finding that a decrease in the base pay and in the slope of incentives actually induces an improvement in the quality mix of the workforce. This seemingly counterintuitive result is consistent with the firm-specific nature of ability: since workers cannot export their high ability to other jobs, the employer is in a position to gain much of the surplus generated by the employment relation.

5.2 The Fixed Effects Approach

This subsection presents a set of regression results that are obtained by applying the fixed effects approach to estimate (1) for the first six months of employment on the subsample of workers who stay at least six months in the firm. The specification of the performance equation is identical to the one in the attrition model.

Table 4 summarizes the results under the fixed effects approach; it omits the estimates of parameters that are not of direct interest to the discussion. Model 1 is estimated using fixed effects. Nijman and Verbeek (1992) propose a simple test for nonrandom attrition which in the current setting involves the inclusion of a dummy for the separation decision of the agent at the end of the sixth month. Under the null hypothesis of no attrition bias, the coefficient of this dummy variable is 0. If the null hypothesis is rejected, then the estimates are not valid and attrition needs to be modelled explicitly. The results from performing this test are reported under Model 2. The estimates under Model 1 are in line with what Lazear (2000) and Lazear and Shaw (2009) find in a similar environment. Namely, regimes 2 and 3 have a highly significant negative effect on performance relative to regime 1. However, there does not appear to be a significant difference between the levels of effort under regime 2 and 3. Finally, the coefficients of the tenure terms imply significant accumulation of experience during the first 6 months of the employment relation. The percentage of outbound calls has a positive effect on performance: this result is counterintuitive, since an employee is more likely to collect payment during an inbound than during an outbound call. Most importantly, the estimated effects of incentives on effort and the effect of accumulated experience are considerable larger than their counterparts under the attrition model.

The dummy for quitting at the end of the sixth month has a highly significant negative effect on performance, implying that the null hypothesis of no attrition bias is soundly rejected. Intuitively, the negative coefficient of the dummy variable implies that individuals who do not stay for an extra period have lower performance than those who stay. The rejection of the hypothesis of random attrition implies that the estimated effects of incentives and tenure on performance are biased for the reasons discussed in section 2. The discrepancy between the estimates under the fixed effects approach and the ones reported in the preceding subsection suggests that neglecting the effect of Bayesian learning on turnover may lead to considerable bias in the estimated effects of incentives. This issue is investigated formally in the following subsection.

5.3 Simulations

Section 2 establishes that if the effect of Bayesian learning on nonrandom attrition is ignored, the estimated effect of incentives on performance and the tenure-performance profile are biased. The crucial question is how large this bias is. One way to approach the question is to estimate the bias using data simulated from the estimated model. In what follows, I use this approach to evaluate the magnitude of the "learning" bias in the context of the used data.

I simulate 1,000 data sets from the explanatory variables in the original data and the estimated parameters of the model, Model 3 in Table 1. In each simulated data set, individuals enter at the calendar time of their actual entry in the firm, but their quitting decision is endogenously determined by the simulated performance signals and the piece rate of operation. The characteristics, and duration of piece rates in the simulated data sets is exactly the same as in the original data set. For each individual I draw 1,000 error paths and a θ from the corresponding distributions, and with the help of the estimated model parameters I generate the performance and separation series. If the econometrician estimates (1) subject to staying for at least 6 periods using the fixed effects approach, she overestimates considerably the effect of incentives and of tenure on performance. Table 5 reports the mean and the variance of the estimated parameters. The results indicate that on average the fixed effects approach overestimates the effect of switching from regime 1 to regime 2 on effort by a factor of two. The fixed effects approach also overestimates the effect of tenure on performance by a similar magnitude. This last point is illustrated on Figure 10 which plots the true and the estimated tenure-performance profiles.

The channels through which Bayesian learning leads to the bias are complex. Under regime 1, the true match quality of those who survive for the first six months varies widely.

Many of the survivors under regime 1 have just been lucky, since in the following months they quit. In contrast, under the less generous regime 2 and 3, only workers with θ in the top quartile of the distribution of match quality survive. In what follows, I will discuss how the average performance error varies across regimes and at different tenure horizons, conditional on staying for at least 6 months at the firm. While the average of the noise in the performance equations, conditional on staying, is positive and increases in the first three months under both regimes, its magnitude is considerably larger under regime 1 due to the survival of individuals with low θ under that regime. Furthermore, there is a great heterogeneity in the observed performance-tenure profiles under regime 1 with low θ being associated with steep profiles. Given the additive separability of equation (1), the estimated effect of tenure on performance becomes the demeaned weighted average of the performance-tenure profiles across different θ and across different regimes. Since many more workers stay under regime 1 than under regime 2 and 3, the estimated effect of tenure is heavily influenced by the average of the conditional performance errors under regime 1. Thus, the differences in the tenure-performance profiles under regimes 1, 2, and 3 find their way into the estimated effect of differences in pay incentives.

5.4 Robustness Checks

The model was estimated under a number of strong assumptions. In this subsection, I perform some nonparametric test for consistency of the performance data with the imposed restrictions on the stochastic technology. Then, I move to discuss some postestimation tests of the normality and independence assumptions that underlie the MLE results. Finally, I consider several alternative specifications for the attrition model.

5.4.1 Technology Tests

The model of section 2 is based on strong distributional and technology assumptions which can be tested nonparametrically. The following observation presents one such nonparametric test.[17]

If workers start with a common prior and learn the quality of their match with the employer over time, the distribution of match quality does not vary across different pay regimes: each employee knows that the turnover after the first period will be higher under a less generous regime than under a more generous one, but at the time of hiring everyone faces the same odds of staying more than one period. Since match quality does not interact with effort, only the mean of performance in the first period varies across regimes. That is, the distributions of performance across regimes are the same up to a location parameter. Observation 1 states this argument formally.

Observation 1. *Consider the model defined by (1) and (2) and suppose that the workers share a common prior at the time of hiring. Then the demeaned distribution of performance at* $t = 1$ *is the same across piece rates:*

$$F\left(y_1^0|R\right) = F\left(y_1^0|R'\right),$$

for any R and R', *where* $y_1^0 = y_1 - E\left(y_1|R\right)$.

Proof: Since effort enters additively in the stochastic technology and optimal effort does not vary with θ_i or θ_{it} across i, the pay regime affects only the first moment of the conditional distribution of performance. Furthermore, under the assumption of a common prior belief at the time of hiring, the entry decision is not affected by the pay regime in place, so $F\left(\theta_i|R\right) = F\left(\theta\right)$. Therefore, $F\left(y_1^0|R\right) = F\left(y_1^0|R'\right)$.■

The proof relies crucially on the assumption of common priors: if some workers had a more accurate belief about the quality of the match than others, the probability of staying more than one period will differ with beliefs leading to differences in the distribution of newly hired employees. For example, heterogeneity in priors arises when θ_i stands for industry-specific rather than firm-specific match quality parameter. Moreover, known ability at the time of hiring is a special case of heterogeneity in priors. Thus, for single-peaked distributions

[17] In what follows, the subscript i is omitted where no confusion arises.

with non-zero probability for every possible match quality this property is also enough to distinguish between Bayesian learning with a common prior and known ability.

Observation 1 imposes necessary restrictions on the observed performance series that are strong. Table 6 presents the results from testing for the technology restrictions implied by Observation 1. A casual look at the standard deviations of performance in period 1 under the different regimes verifies the plausibility of the hypothesis of equal variance: the standard deviations vary between 0.45 and 0.47. This observation is confirmed by the results of the Mann-Whitney tests for equality of the demeaned distributions of performance under regimes 1, 2, and 3 in the first month: the tests fail to reject the hypothesis of equality of the demeaned distributions across regimes.

5.4.2 Postestimation Tests and Alternative Specifications

The estimation also relies on a number of additional assumptions; some of the more important ones are the assumptions of normality and independence of the error terms in the performance and attrition equations across individuals and tenure horizons. These considerations provide the basis for a simple Kolmogorov-Smirnov normality test for the sum of match quality and noise in the first month of employment, y_1^0, where the subindex indicates time, fails to reject the hypothesis of normality at the 5% significance level.

There are also a number of alternative specifications for the attrition process. I have explored these in the standard way and arrived at the attrition specification for the observables reported here; it includes all controls, the interaction terms between beliefs and tenure, as well as tenure and pay regime, but exclude second-order interactions, as well as quadratic terms. With respect to the tenure-varying independent variables, I find that the average percentage outbound calls in the past has the greatest effect on attrition among all functions of the lags of the percentage outbound calls. Furthermore, I consider a number of alternatives to the proposed model of Bayesian learning.[18] These include a model under which Bayesian learning ends within 6 or 12 months of the start of the employment relation; a more general form

[18] For example, workers may engage in adaptive learning or base beliefs on only the most recent signals.

of dependence of the attrition equation on past signals about θ, possibly adaptive learning; individual-specific heterogeneity in outside offers; and heterogeneity in prior beliefs at the beginning of the employment relation.[19] The results from estimating each of these alternative models are reported in Table 3.

Model 4 in Table 3 incorporates an additional heterogeneity term that enters additively in the performance equation. It may represent a time-invariant individual-specific heterogeneity in the outside offers. Due to the presence of Bayesian learning, it is assumed that this additional heterogeneity is independent of θ.[20] The boundary χ^2 test indicates that the variance of this term is not significantly different from zero at the 1% significance level, implying that the hypothesis of heterogeneity in outside offers is rejected. Similarly, Model 5 rejects the hypothesis that attrition depends on past signals about θ through a more general functional form than the simple average of past signals. In particular, the model implies that agents do not assign disproportionately large weight on recent signals when they decide to stay or quit. Due to computational considerations, Model 5 incorporates only the last six signals starting from $t-2$ and their coefficients are estimated freely. The sign of the estimated coefficients varies but is never significantly different from zero. Furthermore, the likelihood ratio test rejects the hypothesis that Model 5 is significantly different from the basic model. Finally, Model 6 allows for heterogeneity in the means of prior beliefs. This model is a special case of a more general test for heterogeneity of priors that allows for both individual-specific means and variance. The specification of Model 3 indicates that, while agents may not share the same common prior, the precision of their initial beliefs remains the same due to the firm-specific nature of the productivity parameter. In addition, I impose the restriction that this individual-specific effect is independent of θ. The estimated variance of this effect is not significantly different from zero, which suggests that the assumption of a common prior is reasonable in the context of the study.

[19] Note that the possibility for heterogeneity in learning rates is ruled out by the nonparametric test of Observation 2 that establishes that for the sub-sample of top performers performance does not vary with tenure.

[20] Such an assumption may be justified when the outside option can be decomposed into two terms: one that depends linearly on theta and one that is orthogonal to theta.

6 Conclusion and Related Research

This paper demonstrates that neglecting the interaction between learning about ability and separation decisions leads to biased estimates of the effect of incentives and returns to monthly tenure. With the help of testable restrictions on the stochastic technology, I estimate the effect of incentives within a model for the employment dynamics at a call center in North Carolina that controls for the interaction between learning and attrition. The results indicate that incentives induce workers of low and average ability to quit which, depending on the pay regime, leads in the first 6 months of employment to 24% and 31% increase in average performance for regimes 1 and 2 respectively. Furthermore, they show that growth in performance is primarily due to quitting decisions and the accumulation of experience; the estimated effect of incentives is significant but small. Simulating the estimated model, I find that the fixed effects approach, popular in the existing literature, overestimates the effect of incentives on effort by a factor of two and the effect of tenure on performance by a similar magnitude. To the extent that Bayesian learning and nonrandom attrition are likely in many environments[21], the issues discussed in this paper relate to a large body of empirical work in labor economics, applied microeconomics, and industrial organization.

The model estimated in this paper is semi-structural in the sense that it incorporates restrictions on the stochastic technology but does not specify a utility function. The estimation of a fully structural model that specifies the worker's utility explicitly is the topic of a companion paper, Bojilov (2010). The main benefit from performing such an exercise is that the estimation of a fully structural model provides the basis for counterfactual policy analysis. I study the problem of optimal pay within a model very similar to the one considered in this paper; it incorporates effort choice, labor turnover, and learning about worker's ability. The novelty, relative to Shearer and Paarsch (2009), is that incentives affect not only effort choice, but also the composition of the workforce. The structural model is estimated using a two-step procedure. In the first step, I estimate the attrition model as done in the

[21] A number of studies, including Pakes and Ericson (1999), Chiappori, Salanié, and Valentin (1999), and Gibbons, Katz, Lemieux, and Parent (2005), have found evidence that that observed data patterns are consistent with Bayesian learning.

present paper and recover the stochastic technology up to a constant, as well as a scaled version of the expected utility of continued employment. I use these estimates in the second step to estimate the remaining structural parameters using the method of minimum distance estimation. The estimates are used to find and characterize the optimal piece rate. The main result is that switching from hourly wage to the optimal piece rate has much greater impact on profits through the effect of incentives on turnover than through the effect of incentives on effort choice. Thus, the companion paper shows that turnover is a major channel through which pay incentives affect profits.

7 References

Abraham K.,; Farber, H.; 1987. "Job Duration, Seniority, and Earnings." *American Economic Review*, vol. 77 iss. 3, pp. 278 - 97.

Aghion, P.; Bolton, P.; Harris, C.; Jullien B.; 1991. "Optimal Learning by Experimentation." *Review of Economic Studies*, vol 58, pp. 621 - 654.

Altonji, J. G.; Williams, N.; 2005. "Do Wages Rise with Job Seniority? A Reassessment." I*ndustrial and Labor Relations Review*, vol. 58, iss. 3, pp. 370-97

Bandiera, O.; Barankay, I.; Rasul, I.; 2007. "Incentives for Managers and Inequality among Workers: Evidence from a Firm-Level Experiment." *Quarterly Journal of Economics*, v. 122, iss. 2, pp. 729-73

Bandiera, O.; Barankay, I.; Rasul, I. 2005; "Social Preferences and the Response to Incentives: Evidence from Personnel Data." *Quarterly Journal of Economics*, v. 120, iss. 3, pp. 917-62.

Bandiera, O.; Barankay, I.; Rasul, I.; 2009. "Team Incentives: Evidence from a Field Experiment." Working paper.

Bojilov, R.; 2010. "Incentives to Work or Incentives to Quit?" Working paper.

Chiappori, P. A.; Salanie, B.; Valentin, J.; 1999. *Journal of Political Economy*, vol. 107 no. 4, pp 731 - 762.

Easley, D.; Kiefer, N.; 1988. "Controlling a Stochastic Process with Unknown Parameters." *Econometrica*, vol. 56, no. 5, pp 1045-1064.

Fernie, S. and Metcalf, D.; 1996. " It's not what you pay." *CEPDP 295. Centre for Economic Performance.*

Florens, J. P.; Heckman, J..; Meghir, C.; Vytlacil, E.; 2008. "Identification of Treatment Effects Using Control Functions in Models with Continuous, Endogenous Treatment and Heterogeneous Effects." *NBER Working Papers: 14002*

Gibbons, R.; Katz, L. F.; Lemieux, T.; Parent, D.l; 2005. "Comparative Advantage, Learning, and Sectoral Wage Determination." *Journal of Labor Economics*, vol. 23, iss. 4, pp. 681-723

Jovanovic, B.; 1979. "Job Matching and the Theory of Turnover." *Journal of Political Economy*, vol. 87, no. 5, pp 972 - 990.

Jovanovic, B.; Nyarko, Y.; 1994. " Learning by Doing and the Choice of Technology." *Econometrica*, vol. 64., no. 6 pp 1299 - 1310.

Harris, M.; Holmstrom, B.; 1982. "A Theory of Wage Dynamics." *Review of Economic Studies*, vol. 49, no. 3, pp 315 - 333.

Hausman, J.; Wise, D.; 1979. "Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment." *Econometrica* v. 47, iss. 2, pp. 455-73.

Kanemoto, Y.; MacLeod, W. B.; 1992. "The Ratchet Effect and the Market for Second-hand Workers." *Journal of Labor Economics*, vol. 10, iss. 1, pp. 85-98

Kahn L., Sherer, P.; 1990. "Contingent pay and managerial performance." *Industrial and Labor Relations Review*, vol. 43, iss. 3, pp. 107-120

Lemieux, T.; MacLeod, W. B.; Parent, D.; 2009. "Performance Pay and Wage Inequality." *Quarterly Journal of Economics*, vol. 124, iss. 1, pp. 1-49

Lazear, E.; 2000. "Performance Pay and Productivity." *American Economic Review*, vol. 90, no. 5, pp 1346 - 1361.

Lazear, E.; Shaw, K.; 2008. "Tenure and Output" *Journal of Labor Economics*, forthcoming.

MacLeod, W. B.; Malcomson, J.; 1989. "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment." *Econometrica* vol. 57, iss. 2, pp. 447-80

McMillan, J., Whalley, J., Zhu, L.; 1989. "The Impact of China's Economic Reforms on Agricultural Productivity Growth," J*ournal of Political Economy*, vol. 97, iss 4, pp 781-807

Mortensen, D.; 1988. "Wages, Separations, and Job Tenure: On-the-job Specific Training or Matching?" *Journal of Labor Economics*, vol. 6, no 4, pp 445 - 471.

Nagypal, E.; 2007. "Learning by Doing and Learning about Match Quality: Can We tell Them Apart?" *Review of Economic Studies*, vol 74, pp. 537 - 566.

Neumann, M.; von Sachs, R.; 2000. "A wavelet–based test for stationarity." *Journal of Time Series Analysis*, vol. 21, pp. 597–613.

Paarsch, H.; Shearer, B. S.; 2009. "The response to incentives and contractual efficiency: Evidence from a field experiment ." *European Economic Review.* vol. 53, iss. 5, pp. 481-494.

Pakes, A.; Ericson, R.; 1999. "Empirical Implications of Alternative Models of Firm Dynamics." *Journal of Economic Theory*, vol. 79, pp 1-45.

Prendergast, C.; 1999. "The Provision of Incentives in Firms." *Journal of Economic Literature*, vol. 37, iss. 1, pp. 7-63.

Prendergast, C.; 2002. "The Tenuous Trade-Off between Risk and Incentives." *Journal of Political Economy*, vol. 110, iss. 5, pp. 1071-1102

Priestley, M. B.; Rao, T. S.; 1969. "A test for non-stationarity of time-series." *Journal of the Royal Statistical Society*, Series B, vol. 31, no. 1, pp. 140-149

Shearer, B.; 2004. "Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment." Review of E*conomic Studies* v. 71, iss. 2, pp. 513-34

Taylor, F. W.; 1911. *The Principles of Scientific Management.* New York, NY: Harper & Brothers.

Wooldridge, J.; 2002. *Econometric Analysis of Cross Section and Panel Data.* The MIT Press: Cambridge, MA

8 Appendix A[22]

The main ingredients of the model are: a continuous individual time-invariant productivity parameter θ with a CDF denoted by F_θ; an outside offer ξ ,with a CDF denoted by F_ξ, equal to the utility that the worker will receive if she quits; if she stays, then she must decide on effort l_t, $l_t \in L \subset R_+$, on the basis of the piece rate $R_{it} = (\alpha_{it}, \beta_{it})'$ and the other relevant parameters of the utility function defined as

$$u(R_{it}, l_t, y_t) = \alpha_t + \beta_t y_t - \psi(l_t),$$

where y_t is the performance signal at t and ψ represents the disutility of labor; upon observing the performance signal y_t the agent updates her belief about θ denoted θ_t, whose CDF is denoted F_θ^t. The piece rate is taken to be exogenous by the employee and is not expected to change. The noise ε_t in the performance signal is continuous and iid over time with CDF denoted by F_ε.

The following assumptions on individual behavior are maintained throughout.

AA1. *The stochastic technology governing* y_t is *defined by (1) and is strictly increasing in its arguments, bounded, and jointly continuous.*

AA2. $u(R_{it}, l_t, y_t)$ *is strictly increasing in its first argument and strictly decreasing and strictly concave in its second argument, bounded and jointly continuous. L is compact.*

AA3. ε_t, θ, *and* ξ *are continuous. Furthermore,* F_ε *and* F_θ *are log-concave and have full support.*

AA4. *The sequence of signals* $\{y_{ik} - g(k) + l_{ik}\}_{k=1}^{t}$ *is ordered in the sense of the likelihood ratio property.*

The continuity assumptions on the production function and the distributions are necessary for the proof of existence. The monotonicity assumptions ensure monotonicity of the value function and the optimal policy. The last assumption establishes a link between signals

[22] In this section, I drop the individual subscript "i"

and beliefs. In this role, it has a crucial role in the characterization of the solution and the identification of learning.

$F_{\theta_{t+1}}$ provides the update of Bayesian beliefs given an output realization y_t. Since the stochastic technology is additively separable in θ and effort, effort choice does not affect the precision of posterior beliefs, so the transitional map $Q : P(\Theta) \times L \rightarrow P(P(\Theta))$ does not depend on effort choice in t and is the beliefs in the following period $t+1$, conditional on the available information at t. The expected utility at time t becomes

$$U(R_t, l_t, t) = \int_{\Theta} \int_{Y} u(R_t, l_t, y_t) f(y_t|\theta_t, t, R_t, l_t) \gamma(dy_t) F_{\theta_t}(d\theta_t).$$

Then, the worker's dynamic program (P) is :

$$v(\theta_t, R_t, t) = \int_{\Theta} \left[\max[\xi_t, \max_{l_t \in L} [U(R_t, l_t, t) + \beta \int_{\Theta} v(\theta_{t+1}, R_t, t+1) F_{\theta_{t+1}}(d\theta|\theta_t, t)]] \right] F_{\xi}(d\xi)$$

Under AA1-AA4, a unique continuous solution to this problem exists and the value function is convex in the appropriate sense, and the optimal policy is unique.These results are summarized in the following two propositions.

Proposition 1. *Under Assumptions AA1-AA4:*

i. *The functional equation (P) has a unique continuous solution $V(\theta_t, R_t, t)$ and the optimal policy*

$$A(\theta_t, R_t, t) = \{l_t \in L \mid (P) \text{ holds.}\}$$

is a continuous function.

ii. Optimal effort, $l(R_t) > l(R_t')$ if $R_t > R_t'$.

iii. $V(\theta_t, R_t, t) > V(\theta_t, R_t, t)$ if $R_t > R_t'$, and $V(\theta_t, R_t, t)$ increases θ_t in the sense of the likelihood ratio property.

Proof of Proposition 1:

Part (i). B(.) and Q(.) are continuous by Lemma 1 and 2 in Easley and Kiefer (1988), so the proof of existence is reduced to a problem which can be solved using Blackwell (1965). Define the operator T by

$$(Tw)(\theta_t, R_t, t) = \int_{\Theta} \left[\max[\xi_t, \max_{l_t \in L}[U(R_t, l_t, t) + \beta \int_{\Theta} w(\theta_{t+1}, R_t, t+1) F_{\theta_{t+1}}(d\theta|\theta_t, t)]] \right] F_{\xi}(d\xi)$$

Let C denote the set of bounded functions on $P(\Theta)$. Under the supnorm metric, $\|.\|$, C is a Banach space. By the contraction mapping theorem, a contraction operator $T : C \to C$ has a unique fixed point and by Blackwell's contraction mapping lemma, T is a contraction if

(1). (Monotonicity) $w_1 \geq w_2$ implies $Tw_1 \geq Tw_2$ and

(2). (Discounting) there exists $\beta \in (0, 1)$, such that $T(w + c) \leq Tw + \beta c$, for any constant $c \geq 0$.

Consequently, to prove existence it is sufficient to show that **(i)** the operator T is a contraction and that **(ii)** T maps continuous bounded functions into the space of continuous bounded functions, C.

(i). This result follows by establishing that conditions (1) and (2) of the Blackwell's contraction mapping lemma are satisfied. It is obvious that if $w_1 \geq w_2$ uniformly, then $Tw_1 \geq Tw_2$. Furthermore, for discount factor β

$$\begin{aligned} T(w + c) &= \int \left[\xi, \max_{l \in L} (U + \beta w + \beta c) \right] dF_{\xi} \\ &< \int \left[\xi, \max_{l \in L} (U + \beta w) \right] dF_{\xi} + \beta c \\ &= Tw + \beta c \end{aligned}$$

(ii). Suppose that $w(\theta_{t+1}, R_t, t+1)$ is continuous. Since L is compact-valued,

$$\widetilde{w}(\theta_t, R_t, t) = \max_{l_t \in L} \left[U(R_t, l_t, t) + \beta \int_{\Theta} w(\theta_{t+1}, R_t, t+1) F_{\theta_{t+1}}(d\theta|\theta_t, t) \right]$$

has a solution. Observe that L is a constant correspondence, so it is a continuous correspondence and the theorem of the maximum applies, so $\widetilde{w}(F_{\theta_t}, h_t)$ is continuous. The function

$max(a, b)$ is continuous if a and b are continuous, and the integral over ξ is also continuous if ξ is continuous. Thus, T is a contraction that maps bounded continuous functions into bounded continuous functions. The proofs of (i) and (ii) imply that a unique solution $V(\theta_t, R_t, t)$. By the theorem of the maximum, the optimal policy correspondence $A(\theta_t, R_t, t)$ is upper-hemicontinuous, and since u(.) is concave in l_t, the optimal policy is a continuous function.[23]

Part (ii). The absence of interaction between effort and beliefs makes the problem of choosing optimal effort static. Since the utility function obeys increasing differences in (β, l_t), optimal effort $l(R_t) > l(R_t')$ if $R_t > R_t'$.

Part (iii). Suppose that $R_t > R_t'$ and $V(\theta_{t+1}, R_t, t+1) > V(\theta_{t+1}, R_t', t+1)$. Since ξ_t is constant in R_t, and $l(R_t) > l(R_t')$, the first part of the statement follows. Finally, suppose that $V(\theta_t, R_t, t)$ increases in θ_{t+1} in the sense of the likelihood ratio property; then, the integral of $V(\theta_t, R_t, t)$ over the distribution of θ_{t+1} conditional on θ_t and t is also increasing in θ_t in the sense of the likelihood ratio property. Similarly, $u(\theta_t, R_t, l_t, t)$ increases in θ_t in the sense of the likelihood ratio property. Thus, $V(\theta_t, R_t, t)$ increases in θ_t in the sense of the likelihood ratio property. ■

[23] Note that nothing in the proof of this section requires that θ must be unidimensional. All proofs hold for an arbitrary finite number of unknown parameters. The assumption that θ is unidimensional is important in the following section, since vectors are only partially ordered.

9 Appendix B

Let the joint distribution of $Y_{it} = (y_{i1}, ..., y_{it})$ and $S_{it} = (s_{i1}, ..., s_{it})$, conditional on $M_i = (M_{i1}, ..., M_{it})$, $M_{it} = (R_{it}, k, X_{ik}, \lambda\mu_{ik} + (1-\lambda)\theta_i)$ and parameters Θ_1, be given by $F(Y_{it}, S_{it}|M_i, \Theta_1)$. By the definition of conditional distribution:

$$
\begin{aligned}
& f(Y_{it}, S_{it}|M_i, \Theta_1) \\
& = f_t(s_{it}|y_{it}, Y_{it-1}, S_{it-1}, M_i, \Theta_1) . f_t(y_{it}|Y_{it-1}, S_{it-1}, M_i, \Theta_1) . f(Y_{it-1}, S_{it-1}, M_i, \Theta_1)
\end{aligned}
$$

Assumption 4 below plays a crucial role in deriving the likelihood and is justified by the model presented above.

AA5. Dynamic Completeness *Suppose that*

$$f_t(y_{it}|Y_{it-1}, S_{it-1}, M_i, \Theta_1) = f_t(y_{it}|M_{it}, \Theta_1)$$

$$f_t(s_{it}|Y_{it}, S_{it-1}, M_i, \Theta_1) = f_t(s_{it}|M_{it}, \Theta_1).$$

By AA5, the conditional density becomes:

$$
\begin{aligned}
& f(Y_{it}, S_{it}|M_i, \Theta_1) \\
& = f_t(s_{it}|Y_{it,}M_{it}, \Theta_1) . f_t(y_{it}|M_{it}, \Theta_1) . f(Y_{it-1}, S_{it-1}, M_i, \Theta_1) \\
& = f_t(s_{it}|Y_{it}, M_{it}, \Theta_1) . f_t(y_{it}|M_{it}, \Theta_1) . f_{t-1}(s_{it}|Y_{it-1}, M_{it-1}, \Theta_1) . \\
& . f_{t-1}(y_{it-1}|M_{it-1}, \Theta_1) . f(Y_{it-2}, S_{it-2}, M_i, \Theta_1) \\
& = ... \\
& = \prod_{k=1}^{t} [f(s_{ik}|Y_{ik}, M_{ik}, \Theta_1) f(y_{ik}|M_{ik}, \Theta_1)]
\end{aligned}
$$

By assumptions (i) and (ii), for all k

$$\begin{pmatrix} \varepsilon_{ik} \\ \xi_{ik} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

To save on notation, the MLE is developed for the case when $\rho = 0$. If $s_{it} = 1$, then

$$f\left(Y_{it}, S_{it} = (1, ..., 1)' | M_i, , \Theta_1\right) = \prod_{k=1}^{t} \left[\Pr\left(s_{ik} = 1 | Y_{ik}, M_{ik}, \Theta_1\right) f\left(y_{ik} | M_{ik}, \Theta_1\right)\right],$$

where:

$$\Pr\left(s_{ik} = 1 | Y_{ik}, M_{ik}, \Theta_1\right) = \Phi\left(G\left(R_{ik}, k, X_{ik}, \lambda\mu_{ik} + (1-\lambda)\theta_i\right)\right)$$

$$f\left(y_{ik} | M_{ik}, \Theta_1\right) = \frac{1}{\sigma}\varphi\left(\frac{y_{ik} - g\left(W_{ik}, k\right) - \theta_i}{\sigma}\right)$$

If $s_{it} = 0$, while $s_{it-1} = 1$, then the unobserved y_{it} must be integrated out, leading to:

$$\begin{aligned} & f\left(Y_{it}, S_{it} = (1, ...1, 0)' | M_i, \Theta_1\right) \\ & = \Pr\left(s_{it} = 0 | Y_{it-1}, M_{it}, \Theta_1\right) \prod_{k=1}^{t-1} \left[\Pr\left(s_{ik} = 1 | Y_{it-1}, M_{ik}, \Theta_1\right) f\left(y_{ik} | M_{ik}, \Theta_1\right)\right], \ t \geq 2 \end{aligned}$$

where

$$\Pr\left(s_{it} = 0 | Y_{it-1}, M_{it}, \Theta_1\right) = 1 - \Phi\left(G\left(R_{ik}, k, X_{ik}, \lambda\mu_{ik} + (1-\lambda)\,\theta_i\right)\right)$$

$$\Pr\left(s_{ik} = 1 | Y_{ik-1}, M_{ik}, \Theta_1\right) = \Phi\left(G\left(R_{ik}, k, X_{ik}, \lambda\mu_{ik} + (1-\lambda)\,\theta_i\right)\right)$$

$$f\left(y_{ik} | M_{ik}, \Theta_1\right) = \frac{1}{\sigma}\varphi\left(\frac{y_{ik} - g\left(k\right) - m\left(X_{ik}\right) - l\left(R_{ik}\right) - \theta_i}{\sigma}\right)$$

AA5 allows for the use of conditional MLE, so the likelihood for individual i is:

$$\begin{aligned} & l_i(\theta_i, \Theta_1|W_i) \\ & = \Psi_{i2}.\frac{1}{\sigma}\varphi\left(\frac{y_{i1} - g(1) - m(X_{i1}) - l(R_{i1}) - \theta_i}{\sigma}\right). \end{aligned}$$

where

$$\Psi_{ik} = \left[\Psi_{ik+1}.\Phi\left(G\left(R_{ik}, k, X_{ik}, (1-\lambda)\theta_i + \lambda.\mu_{ik}\right)\right).\frac{1}{\sigma}\varphi\left(\frac{y_{ik} - g(k) - m(X_{ik}) - l(R_{ik}) - \theta_i}{\sigma}\right)\right]^{S_{ik}}.$$
$$.\left[1 - \Phi\left(G\left(R_{ik}, k, X_{ik}, (1-\lambda)\theta_i + \lambda.\mu_{ik}\right)\right)\right]^{1-S_{ik}}$$

for $k = 2, ..., \tau_i$, where τ_i is the last period in which i is observed. Note that the likelihood for individual i is separable in θ_i. The likelihood can be rewritten in a standard way as follows:

$$\begin{aligned} & l_i(\Theta_1|\theta_i, W_i) \\ & = \prod_{t=1}^{T}\left(\varphi\left(\frac{y_{it} - g(t) - m(X_{it}) - l(R_{it}) - \theta_i}{\sigma}\right)\Phi\left(G\left(R_{it}, t, X_{it}, (1-\lambda)\theta_i + \lambda.\mu_{it}\right)\right)\right)^{\prod_{k=1}^{t} s_{ik}} \\ & \left(1 - \Phi\left(G\left(R_{it}, t, X_{it}, (1-\lambda)\theta_i + \lambda.\mu_{it}\right)\right)\right)^{1-\prod_{k=1}^{t} s_{ik}}, \end{aligned}$$

where T is the last period for which the econometrician observed performance.

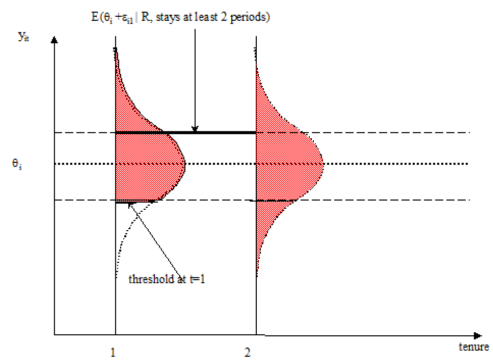
Next, we integrate out θ_i

$$l_i(\Theta|W_i) = \int_{\Theta} l_i(\Theta_1|W_i, \theta_i).\varphi(\theta_i|W_i, \Theta_2)d\theta_i,$$

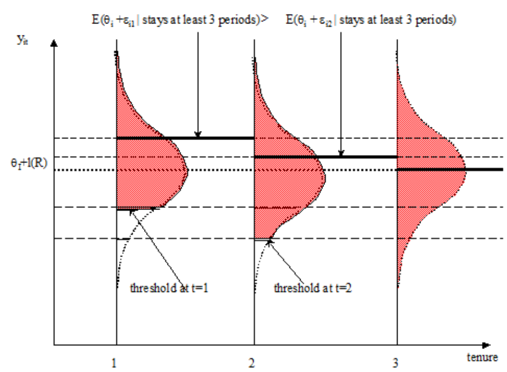
where Θ_2 is a vector of parameters that govern the distribution of θ_i and Θ is a vector of all parameters in Θ_1 and Θ_2. Then the log-likelihood becomes

$$l\left(\Theta|\{W_i\}_{i=1}^{N}\right) = \sum_{i=1}^{N}\log l_i(\Theta|W_i).$$

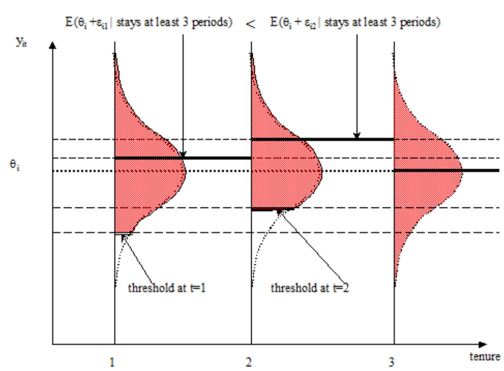
10 Appendix C



Panel A



Panel B



Panel C

Diagram 1: Implications of learning about ability for observed signals

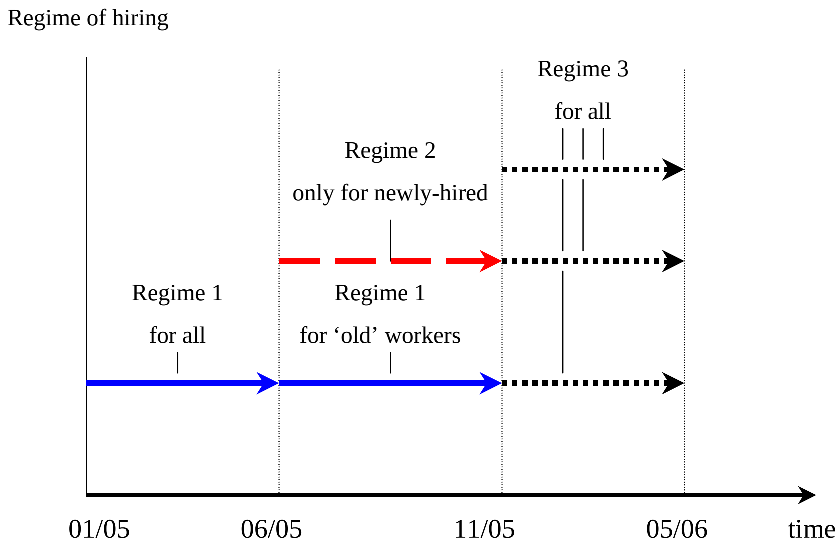


Diagram 2. Timeline of pay regimes.

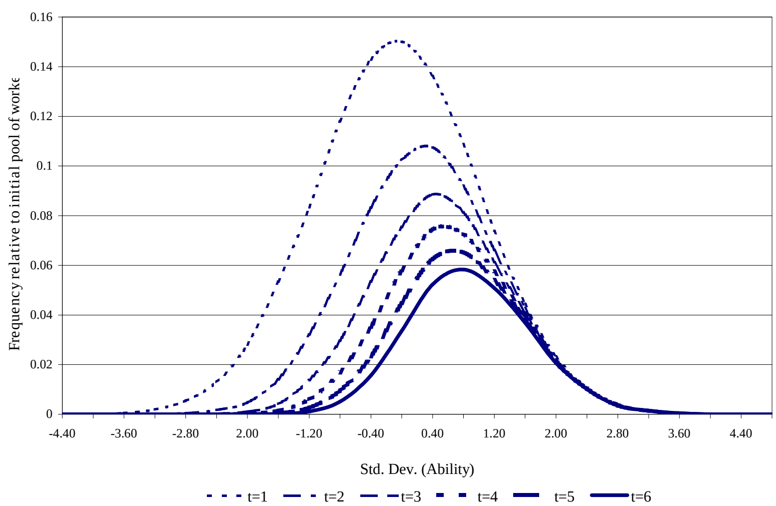


Figure 1. Distribution of θ at t, conditional on staying at least t months, under regime 1(estimates are based on the learning specification of the attrition model)

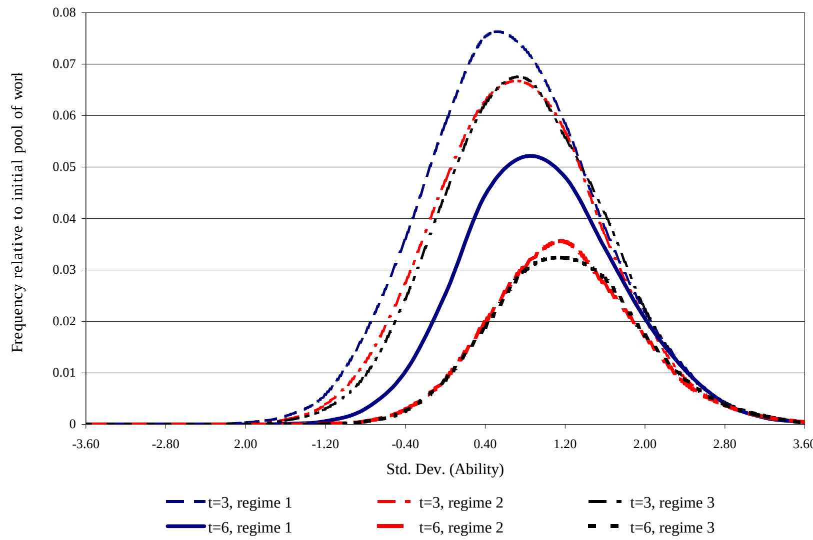


Figure 2. Distribution of θ at t, conditional on staying at least t months under regime 1-3 (estimates are based on the learning specification of the attrition model).

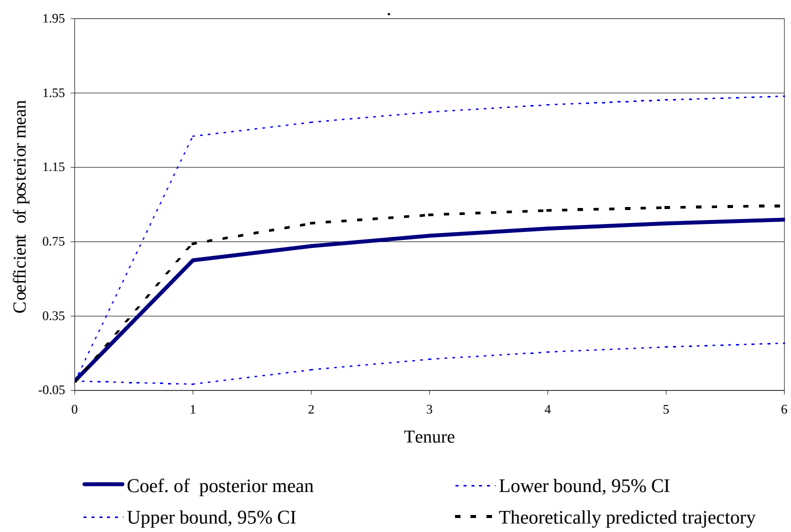


Figure 3. Increase in the importance of observed signals relative to the initial prior when workers decide to stay or quit (estimates based on the learning specification of theattrition model).

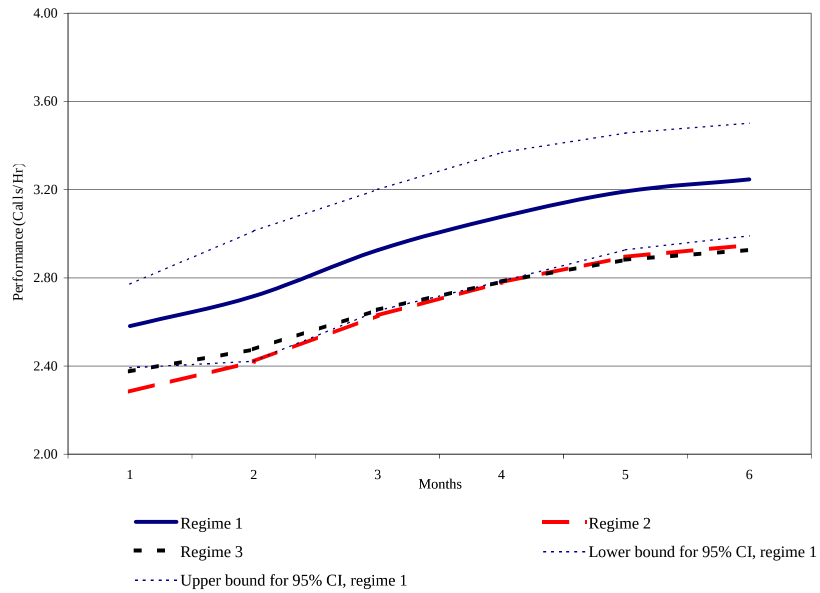


Figure 4. Predicted perf.-tenure profile at entry for $\theta_i = 0$ under regime 1, 2, and 3 (estimates are based on the learning specification of the attrition model).

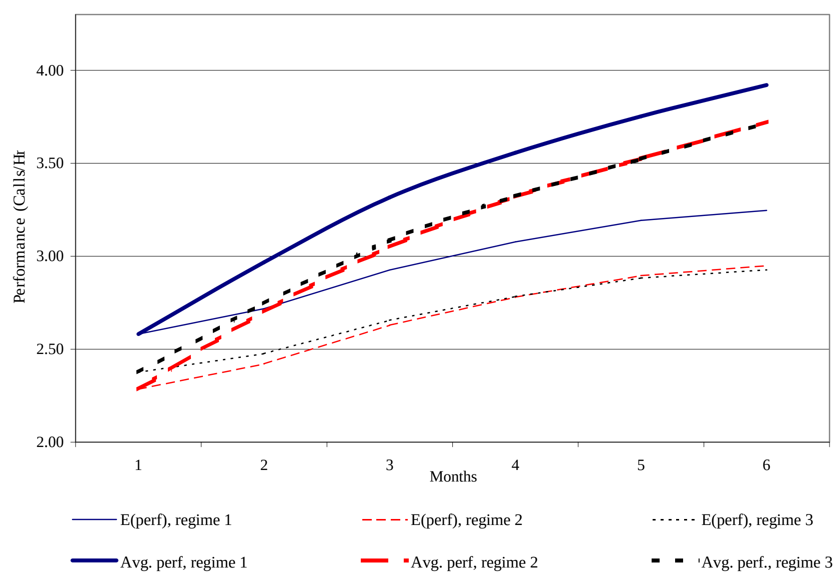


Figure 5. Comparison between expected performance at entry for $\theta_i = 0$ and avg. performance (estimates are based on learning specification of the attrition model).

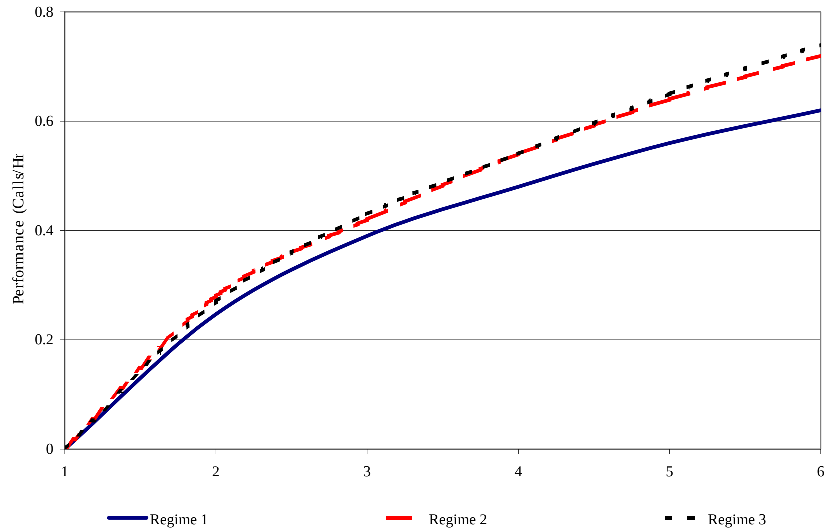


Figure 6. Improvement in average performance due to selective turnover under regime 1- 3 (estimates are based on learning specification of the attrition model).

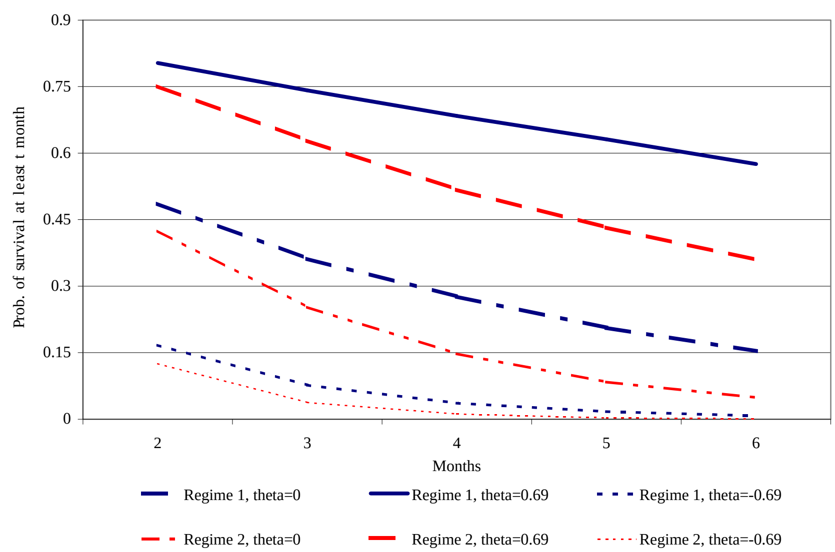


Figure 7. Probability of staying for $\theta_i = 0$, and $\theta_i = \pm\sigma_\theta$ under regime 1-2 (estimates are based on the learning specification of the attrition model).

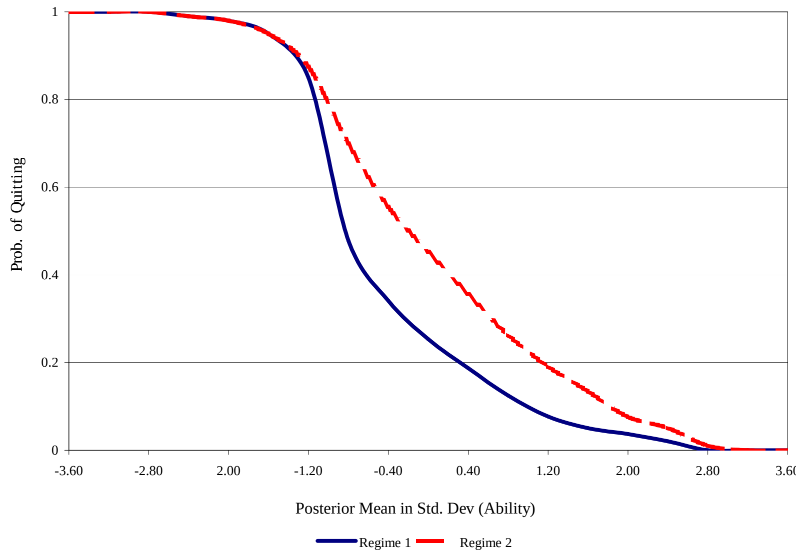


Figure 8. Probability of quitting in $t = 6$ by posterior mean μ_{i6} and pay regime, $\sigma_\theta = 0.68$ and μ is normalized to 0 (estimates are based on the learning specification of the attrition model).

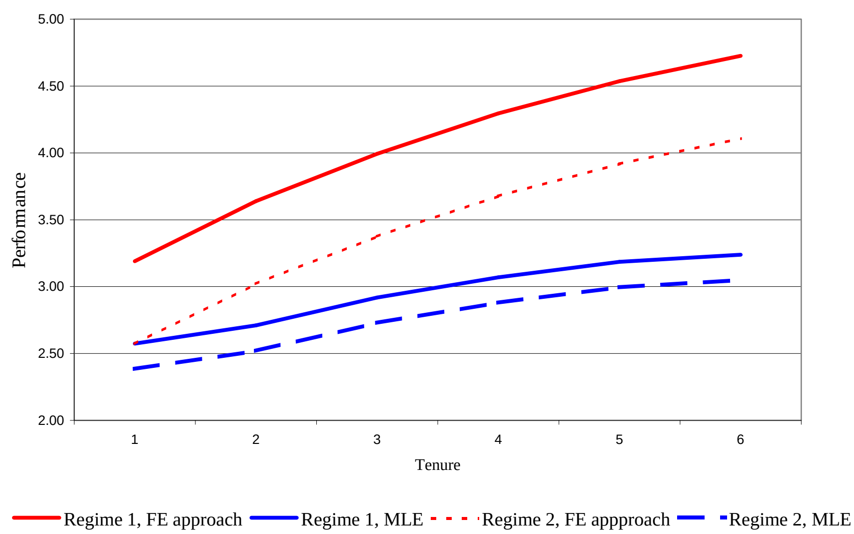


Figure 9. Performance-tenure profile at entry for $\theta_i = 0$ under regime 1 and 2: comparison between the estimates of FE approach and learning specification of the attrition model.

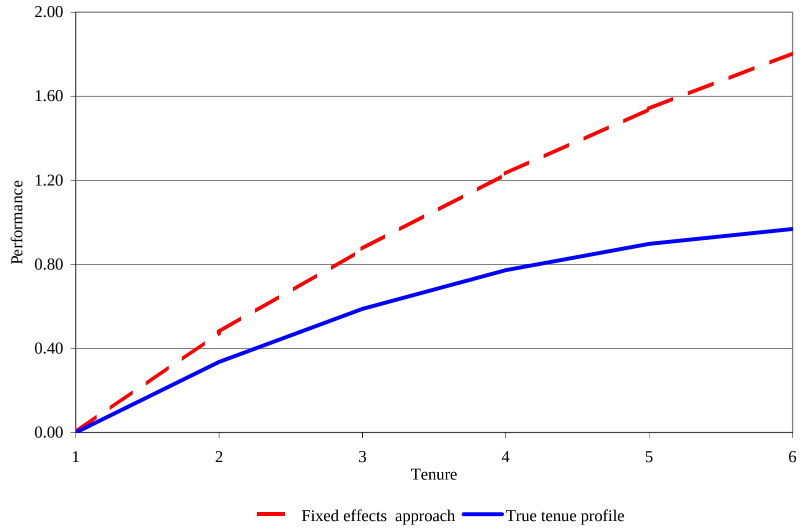


Figure 10. Attrition bias in the tenure effect estimated by the FE approach on the simulated data.

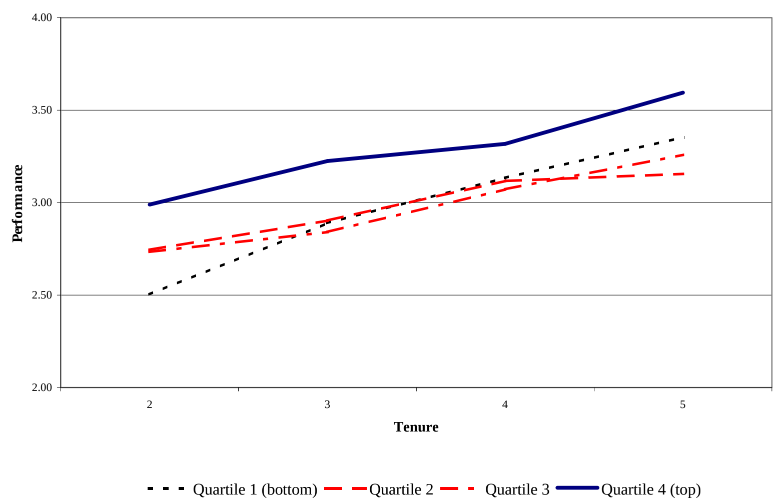


Figure 11. Descriptive statistics: avg. perf. for months 2 to 5 of workers who stay at least 5 months, conditional on initial performance quartile.

11 Appendix D

Table 1. Summary statistics for the first 6 months (includes only workers who start and work under the same regime).

Variable:	Regime 1	Regime 2	Regime 3
β (in \$)	3.3	2.5	2.5
α (in \$)	3.8	3.5	3.5
Avg. Perf. (call/hr.)	2.74	2.6	2.66
Std. Dev. (Perf.)	0.68	0.77	0.7
Avg. Perf., stay$\geq$6	2.91	2.76	2.71
Std. Dev. (Perf., stay$\geq$6)	0.65	0.67	0.6
Turnover	0.52	0.68	0.67
Obs., stay>6	113	59	9

Table 2a. Estimates for the performance equation in the attrition model. In model 1 ability is known ($\lambda = 0$), in model 3 workers learn about it $(\lambda = 1)$, and model 2 estimates λ.

Dependent Variable: Performance	Model 1 $\lambda = 0$		Model 2 λ estimated		Model 3 $\lambda = 1$	
Explanatory Variable	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
t, orthog. pol. 1	-0.41	0.16	-0.56	0.15	-0.57	0.15
t, orthog. pol. 2	-0.46	0.09	-0.62	0.09	-0.62	0.09
regime 2	-0.24	0.08	-0.19	0.07	-0.19	0.07
regime 3	0.23	0.12	0.13	0.08	0.13	0.08
$t.\,(\text{regime } 3)$, orthog. pol. 1	0.88	0.13	0.77	0.13	0.78	0.13
$t.\,(\text{regime } 3)$, orthog. pol. 2	0.53	0.09	0.43	0.08	0.43	0.08
% outbound calls	-0.1	0.05	-0.06	0.05	-0.06	0.05
Hired under regime 2	-0.11	0.08	-0.11	0.08	-0.10	0.08
Hired under regime 3	-0.36	0.29	-0.34	0.23	-0.34	0.23
Constant	3.59	0.11	3.43	0.11	3.43	0.11
Log-likelihood	-3756.22		-3745.07		-3745.71	

Notes: The specification includes calendar time, orthogonal polynomials of degree 3 and inidividual controls: gender, age, marriage status, distance from home, and race. Obs.=3,675.

Table 2b. Estimates for the separation decisions in the attrition model . In model 1 ability is known ($\lambda = 0$), in model 3 workers learn about it ($\lambda = 1$) ,and model 2 estimates λ.

Dependent Variable: Decision to stay	Model 1 $\lambda = 0$		Model 2 λ estimated		Model 3 $\lambda = 1$	
Explanatory Variable:	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
t, orthog. pol. 1	-0.44	0.18	-0.50	0.16	-0.50	0.17
t, orthog. pol. 2	-0.19	0.08	-0.21	0.08	-0.21	0.08
t, orthog. pol. 3	0.23	0.14	0.24	0.13	0.23	0.13
$t.$(regime 2) , orthog. pol. 1	0.52	0.16	0.41	0.14	0.41	0.14
$t.$(regime 2) , orthog. pol. 2	0.45	0.11	0.40	0.11	0.41	0.11
$t.$(regime 2) , orthog. pol. 3	-0.68	1.68	-0.78	1.57	-0.78	1.57
$t.$(regime 3) , orthog. pol. 1	0.41	0.16	0.40	0.16	0.40	0.16
$t.$(regime 3) , orthog. pol. 2	0.23	0.08	0.22	0.08	0.23	0.08
$t.$(regime 3) , orthog. pol. 3	-0.03	0.13	-0.03	0.12	-0.03	0.13
avg. % outbound calls in past	-0.12	0.09	-0.29	0.09	-0.31	0.09
Hired under regime 2	-0.02	0.09	-0.07	0.09	-0.07	0.09
Hired under regime 3	-0.25	0.14	-0.25	0.13	-0.25	0.13
Constant	0.46	0.10	0.56	0.10	0.58	0.10
Log-likelihood	-3756.22		-3745.07		-3745.71	

Notes: The specification includes calendar time, orthogonal polynomials of degree 3 and individual controls: gender, age, marriage status, distance from home, and race. Obs.=3,675.

Table 2c. Estimates for the unobserved heterogeneity and Bayesian l earning in the attrition model . In model 1 ability is known ($\lambda = 0$), in model 3 workers learn about it ($\lambda = 1$) ,and model 2 estimates λ.

Parameter or explanatory variable	Model 1 $\lambda = 0$		Model 2 λ estimated		Model 3 $\lambda = 1$	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
σ_ε^2	0.17	0.01	0.17	0.01	0.17	0.01
σ_ξ^2	1		1		1	
$\rho(\varepsilon, \xi)$	0.16	0.04	0.05	0.04	0.01	0.04
σ_θ^2	0.49	0.02	0.48	0.02	0.48	0.02
θ	0.14	0.06	0.06	0.06		
$t.\theta$, orthog. pol. 1	0.00	0.00	0.05	0.05		
$t.\theta$, orthog. pol. 2	-0.01	0.05	0.00	0.04		
$t.\theta$, orthog. pol. 3	0.04	0.04	0.00	0.03		
μ_{it}			0.17	0.06	0.18	0.05
$t.\mu_{it}$, orthog. pol. 1			0.06	0.03	0.06	0.03
$t.\mu_{it}$, orthog. pol. 2			0.02	0.04	0.02	0.04
$t.\mu_{it}$, orthog. pol. 3			0.03	0.03	0.03	0.03
λ			0.74	0.29		
Log-likelihood	-3756.22		-3745.07		-3745.71	

Notes: The specification includes calendar time orthogonal polynomials of degree 3 and individual controls: gender, age, marriage status, distance from home, and race. Obs.=3,675.

Table 3a. Estimates for the performance equation under alternative specifications of the attrition model. Model 4 allows for heterogeneity in outside offers, model 5 for a more general dependence of turnover on past performance and model 6 tests for heterogeneity in the means of prior beliefs at entry.

Dependent Variable: Performance	Model 4 Heterogeneity in offers		Model 5 Gen. form of learning		Model 6 No common prior	
Explanatory Variable	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
t, orthog. pol. 1	-0.55	0.15	-0.55	0.15	-0.59	0.15
t, orthog. pol. 2	-0.61	0.09	-0.63	0.09	-0.63	0.09
regime 2	-0.21	0.07	-0.19	0.08	-0.24	0.07
regime 3	0.14	0.08	0.11	0.08	0.13	0.08
$t.$(regime 3) , orthog. pol. 1	0.76	0.30	0.69	0.30	0.75	0.31
$t.$(regime 3) , orthog. pol. 2	0.39	0.20	0.40	0.20	0.44	0.21
% outbound calls	-0.08	0.05	-0.06	0.05	-0.04	0.04
Constant	3.73	0.20	3.85	0.18	3.54	0.19
Log-likelihood	-3745.34		-3744.29		-3744.52	

Notes: The specification also includes calendar time, orthogonal polynomials of degree 3 and individual controls: gender, age, marriage status, distance from home, and race. Obs.=3,675.

Table 3b. Estimates for the separation decisions under alternative specifications of the attrition model. Model 4 allows for heterogeneity in outside offers, model 5 for a more general dependence of turnover on past performance and model 6 tests for heterogeneity in the means of prior beliefs at entry.

Dependent Variable: Decision to stay	Model 4 Heterogeneity in offers		Model 5 Gen. form of learning		Model 6 No common prior	
Explanatory Variable:	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
t, orthog. pol. 1	-0.48	0.17	-0.51	0.17	-0.52	0.17
t, orthog. pol. 2	-0.19	0.08	-0.23	0.08	-0.21	0.08
t, orthog. pol. 3	0.24	0.13	0.22	0.14	0.23	0.13
$t.$(regime 2), orthog. pol. 1	0.41	0.14	0.42	0.14	0.41	0.14
$t.$(regime 2), orthog. pol. 2	0.41	0.11	0.39	0.11	0.44	0.11
$t.$(regime 2), orthog. pol. 3	-0.79	1.57	-0.99	1.57	-0.78	1.57
$t.$(regime 3), orthog. pol. 1	0.40	0.16	0.43	0.16	0.38	0.16
$t.$(regime 3), orthog. pol. 2	0.21	0.08	0.25	0.08	0.23	0.08
$t.$(regime 3), orthog. pol. 3	-0.03	0.13	-0.03	0.14	-0.03	0.13
avg. % outbound calls in past	-0.31	0.09	-0.27	0.10	-0.32	0.12
Hired under regime 2	-0.07	0.09	-0.07	0.09	-0.07	0.09
Hired under regime 3	-0.25	0.13	-0.25	0.13	-0.25	0.13
Constant	1.07	0.33	1.08	0.34	2.15	0.53
Log-likelihood	-3745.34		-3744.29		-3744.52	

Notes: The specification also includes calendar time, orthogonal polynomials of degree 3 and individual controls: gender, age, marriage status, distance from home, and race. Obs. = 3,675.

Table 3c. Estimates for ability and learning under alternative attrition model.

Parameter or explanatory variable	Model 4 Heterogeneity in offers		Model 5 Gen. form of learning		Model 6 No common prior	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
σ_{ε}^{2}	0.17	0.01	0.16	0.01	0.17	0.01
$\rho(\varepsilon,\xi)$	0.15	0.03	-0.02	0.03	0.07	0.04
σ_{θ}^{2}	0.48	0.02	0.47	0.02	0.48	0.02
μ_{it}	0.17	0.04	0.22	0.05	0.16	0.05
$t.\mu_{it}$, orthog. pol. 1	0.06	0.05	0.05	0.06	0.08	0.07
$t.\mu_{it}$, orthog. pol. 2	0.02	0.05	0.04	0.05	0.05	0.06
$t.\mu_{it}$, orthog. pol. 3	0.03	0.03	0.00	0.03	0.04	0.04
σ_{μ}^{2}	0.00	0.01				
$y_{it-1}-\mu_{it}-g(t-1)-l(R_{it-1})$			0.01	0.03		
$y_{it-2}-\mu_{it}-g(t-2)-l(R_{it-2})$			-0.06	0.03		
$y_{it-3}-\mu_{it}-g(t-3)-l(R_{it-3})$			0.02	0.03		
$y_{it-4}-\mu_{it}-g(t-4)-l(R_{it-4})$			-0.05	0.04		
$y_{it-5}-\mu_{it}-g(t-5)-l(R_{it-5})$			0.02	0.04		
$y_{it-6}-\mu_{it}-g(t-6)-l(R_{it-6})$			0.00	0.04		
$\sigma_{\theta_1}^{2}$					0.00	0.01
$t.\mu_{i1}$, orthog. pol. 1					109.43	4316.80
$t.\mu_{i1}$, orthog. pol. 2					2609.56	$1.065.10^{5}$
Log-likelihood	-3745.34		-3744.29		-3744.52	

Notes: The specification also includes calendar time, orthogonal polynomials of degree 3 and individual controls: gender, age, marriage status, distance from home, and race. Obs. = 3,675.

Table 4. Estimates based on the fixed effects approach. Model 1 is estimated on the set of workers who stay for at least six months. Model 2 is the same as model 1 but also includes a dummy for the decision to stay or quit at $t = 7$.

Dependent Variable:	FE Approach		FE + attrition test	
Performance	Coef.	S.E.	Coef.	S.E.
t, orthog. pol. 1	-14.78	0.32	-15.25	0.32
t, orthog. pol. 2	-0.61	0.18	-0.55	0.18
regime 2	-0.56	0.21	-0.58	0.21
regime 3	0.09	0.31	0.08	0.22
$t.\,(\text{regime }3)$, orthog. pol. 1	2.81	0.33	2.57	0.36
$t.\,(\text{regime }3)$, orthog. pol. 2	0.85	0.05	0.83	0.05
% outbound calls	0.15	0.07	0.13	0.07
Quit in $t = 7$			-0.28	0.09
Constant	1.71	0.35	1.62	0.35
Obs.	1131		1131	
R^2	0.45		0.46	

Notes: The specification also includes calendar time, orthogonal polynomials of degree 3 and individual controls: gender, age, marriage status, distance from home, and race. Obs. = 3,675.

Table 5. Simulation results showing the presence of attrition bias. The true parameters are equal to the estimated parameters under the attrition model with learning, Model 3. The table reports the average of the estimated coefficients by the fixed effects approach, along with the standard deviation of the estimates.

Variable	True value	Coef. Avg.	Coef., Std. Dev.
regime 2	-0.19	-0.57	0.18
regime 3	0.13	0.06	0.06
t, orthog. pol. 1	-0.57	-14.89	0.29
t, orthog. pol. 2	-0.62	-0.55	0.16
(t, orthog. pol. 1)*(regime 3)	0.78	2.78	0.31
(t, orthog. pol. 2)*(regime 3)	0.43	0.85	0.05

Note: The FE approach is applied to estimate the performance equation on the "stayers" for each simulated data set. The specification also includes calendar time orthogonal polynomials of degree 3 and individual controls: gender, age, marriage status, distance from home, and race.

Table 6. Mann-Whitney tests for equal distributions of perf. under different regimes in the first month: reported prob. of identical distributions.

	Regime 1	Regime 2	Regime 3
Std. Dev.	0.465	0.454	0.457
Regime 1		Pr=0.907	Pr=0.564
Regime 2			Pr=0.408