Incentives to Work or Incentives to Quit?

Raicho Bojilov*

First draft: November 21, 2010
This draft: September 26, 2011

Abstract

Pay incentives affect not only effort choice but also turnover and thereby the quality mix of the workforce. This paper investigates how considerations about the quality mix shape pay policy and profits within a structural model of effort choice, learning about match quality, and turnover. Using unique data from a call center in North Carolina, I estimate the model in two steps adapting estimation methods for dynamic structural models to the analysis of employment dynamics. The results provide the basis for counterfactual policy analysis. The optimal policy, in the class of linear contracts in output, not only induces employees to exert effort but also acts as a selection mechanism that helps the firm build a workforce of high match quality over time. Simulation results show that turnover is the major channel through which pay incentives affect profits.

KEYWORDS: Piece Rates, Learning, Turnover
JEL CODE: J22, J33, D82, D83, M51, M52

*Raicho Bojilov, Department of Economics, Ecole Polytechnique, 91128 Palaiseau Cedex, France. E-mail: raicho.bojilov@gmail.com. I would like to thank Professors V. J. Hotz, B. MacLeod, L. Munasinghe, and B. O’Flaherty for their helpful comments and suggestions. I am greatly indebted to Professors P.A. Chiappori and B. Salanié for their advice and support in my work on this project.
1 Introduction

Pay incentives affect profits not only through their impact on effort choice but also through their effect on the quality mix of the workforce. In this paper, I investigate the relative importance of these two channels to maximizing profits in a structural model of employment dynamics that includes effort choice, learning about match quality, and labor turnover. My empirical analysis focuses on contracts that are linear in output: compensation is equal to the sum of a base pay and a bonus proportional to hourly output (performance from now on). I limit my attention to this class of linear contracts for two reasons. First, the firm whose personnel records I use itself implemented such linear contracts and one of the objectives of this paper is to characterize the profitability of the firm’s compensation policies. Second, firms often apply simple compensation policies based on such linear contracts and the problem of finding and characterizing the optimal linear contract is of interest on its own. ¹

The firm’s data are ideally suited for the empirical analysis of pay incentives; they come from a call center in North Carolina and contain an objective measure of individual performance (defined as output per hour), a known compensation policy based on linear contracts in performance, and a variation in the pay policies that does not depend on what the firm learns about its employees. My estimates show that steeper incentives are associated with higher performance, that persistent differences in individual performance are driven by differences in the quality of the employer-employee match, and that employees learn about the quality of the match on the job. 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 pay incentives on profits. Thus, I conclude that finding the optimal pay policy requires the explicit modelling of all three: effort choice, learning about match quality, and separation decisions.

¹Within a dynamic setting as the one in this paper, the firm’s flexibility in designing the contract is reduced. This general idea is first explored by Holmstrom and Milgrom (1987) who show that in some settings the optimal compensation is to provide workers with incentives that are linear in output. Besides, the cost of implementing a complicated nonlinear contract is high: see, for example, Holmstrom and Milgrom (1990) as well as Ferrall and Shearer (1999).
Unobserved effort, labor turnover and learning about match quality are the subject of intensive study in the structural literature but usually separately from one another. For example, Shearer and Paarsch (2009) analyze the effect of incentives on effort and conduct a related policy analysis, but the experimental design of their study does not allow them to analyze the effects of incentives on the pool of entering employees and turnover. However, Lazear (2000) points out that in the context of his study about one-third of the improvement in performance after the introduction of pay incentives can be traced back to the improvement in the quality mix of entering employees. When the quality of the match between a potential employee and the firm becomes known to the worker in the hiring process, the firm may use its pay policy not only to induce effort but also to shape the quality mix of the newly hired employees, as discussed in Lazear (1998); the firm also faces a trade-off between the extra revenue generated by workers who stay and the associated increase in pay that is necessary to make them stay. The interaction of all these considerations determine the firm’s pay policy and the associated turnover is not necessarily low or nonexistent but depends on the characteristics of the technology, the workforce and the alternative jobs. When the firm and workers learn about match quality over time, an additional consideration arises. Depending on how much starting employees know about their match, turnover becomes the primary channel through which pay incentives affect the quality mix. In the special case of a common prior, there is no selection at entry, and turnover is the only source of changes in the quality mix at the workplace.

Similarly, the models in the structural literature on learning about match quality and turnover do not incorporate effort choice. The dynamic programs in such models are complicated enough even as they are, since they involve heterogeneity across employees and a sequence of posterior beliefs. To my knowledge, Miller (1984) is the first paper to estimate a model with learning about match quality. Pastorino (2009) considers a variation of this model that incorporates correlation between ability at one job and ability at others. Furthermore, Nagypal (2007) applies the method of indirect inference to distinguish between learning about match quality and learning-by-doing within a structural setting. More re-
Recently, Camargo and Pastorino (2010) estimate a structural model of career concerns with learning-by-doing. While these models consider selection and job mobility, they assume away potential problems of moral hazard. Furthermore, the computational complexity associated with estimating the models usually requires some strong assumptions about the production technology and the heterogeneity among workers.

Here, I propose a simple two-step procedure to estimate a structural model of learning about match quality, effort choice, and turnover. In the first step, I estimate a semi-structural attrition model and recover the stochastic technology up to a constant, as well as a scaled version of the value of continued employment. Bojilov (2010) shows that the first-step provides consistent estimates of the effect of changes in pay incentives on effort, while popular alternatives may overestimate the same effect by a factor of two. I use these estimates in the second step to recover the remaining parameters using the method of moments. That is, I use an indirect approach to estimate the value of continued employment without directly solving for the value function. The same principle underpins the literature starting with Hotz and Miller (1993) and including the recent works by Aguirregabiria and Mira (2007) and by Pesendorfer and Schmidt-Dengler (2003). Thus, from a methodological point of view, this paper adapts recent structural estimation methods to the analysis of employment relations in the presence of Bayesian learning.

The estimates of the model are then used as a basis for counterfactual policy analysis. My results suggest that firms choose their pay policy for reasons that go beyond effort choice. Most of the increase in profits from switching to the optimal policy from hourly wage can be traced back to the effect of incentives on the quality mix. The optimal linear contract induces low quality employees to quit and in this way it helps the firm build a workforce of high match quality over time. This effect more than offsets the loss associated with replacing an experienced worker with a newly hired one of no experience and unknown ability. Furthermore, the employer exploits the firm-specific nature of match quality to capture most of the surplus generated by the employment relation. To achieve that, the firm offers pay incentives that induces little effort, so high level of effort and low turnover are not necessarily attributes of
the profit-maximizing pay policy. Finally, the optimal policy that I find generates only 4.8% higher profits than one of the actually implemented pay regimes. An exercise in comparative statics shows that as turnover costs grow, the firm increases compensation to induce lower turnover by offering much steeper incentives. The result is a decline in the quality mix of the workforce and an increase in the importance of effort to the firm’s profits. Given the strong evidence of high turnover costs in some industries, this finding cautions that models of job mobility, such as Keane and Wolpin (1997) and Hoffmann (2010), should incorporate turnover costs. Finally, another counterfactual experiment shows that the firm’s profits would have been 27% higher if match quality was known to the employees at the time of hiring. The primary reason is that workers of high match quality self-select into the firm which leads to low turnover and high level of experience.

The results indicate that incentives matter and in this way they are consistent with the literature devoted to investigating incentive effects represented by Paarsch and Shearer (2000), Lazear (2000) and Shearer (2004). More specifically, they show that workers are very responsive to changes in the slope of incentives. This is consistent with previous results obtained in Paarsch and Shearer (1999, 2009), as well as Haley (2003). The novelty, relative to Shearer and Paarsch (2009), is that optimal pay incentives are allowed to affect not only effort choice but also the composition of the workforce at different tenure horizons. Thus, the paper extends the work in Lazear (1998, 2000) on the effect of incentives on the quality mix by studying how turnover shapes the properties of the optimal pay policy. In particular, the results show that turnover may be the primary channel through which pay incentives affect profits when workers learn about match quality on the job. The results also indicate that the considerable contribution of improved match quality to worker’s compensation that is estimated in some structural papers of job mobility, such as Hoffmann (2010), may depend on the strong assumptions of low or non-existent turnover costs.

The rest of this paper is organized as follows. Section 2 presents the model and section 3 the data. Section 4 introduces the estimation of the model. Section 5 discusses the estimates of the structural parameters and presents the policy analysis. Section 6 presents some counteractual
experiments and Section 7 concludes with an overview of future research.

2 Model

2.1 Worker’s Problem

The model is a variation of the classical model in Jovanovic (1979) which in addition to match quality also includes effort choice. The crucial parameter in the model is match quality \( \theta_i, \theta_i \in R \). Match quality is time-invariant, independent and identically distributed across workers \( i \), and normally distributed, \( \theta_i \sim N \left( \mu_\theta, \sigma^2_\theta \right) \), with a probability density function denoted as \( f_\theta \). At the beginning of each period the worker decides whether to stay or quit by comparing the value of continued employment and the realization of an outside offer \( \xi^*_i \).

The outside offer \( \xi^*_i, \xi^*_i \in R \), is normally distributed, \( \xi^*_i \sim N \left( \mu_{\xi^*}, \sigma^2_{\xi^*} \right) \), independent and identically distributed across tenure horizons \( t \) and \( i \), where \( t = 1, 2, \ldots \), independent from \( \theta_i \), and has a probability density function denoted \( f_{\xi^*} \). If the worker stays, she observes a noisy performance signal \( y_{it} \). The noise in the performance signal \( \epsilon_{it}, \epsilon_{it} \in R \), is continuous, independent and identically distributed across tenure horizons \( t \) and workers \( i \), independent from \( \theta_i \), normally distributed, \( \epsilon_{it} \sim N \left( 0, \sigma^2_{\epsilon} \right) \), and has a probability density function denoted as \( f_{\epsilon} \). The performance signal \( y_{it} \) is generated by the following technology\(^2\):

\[
y_{it} = \theta_i + g(t) + l_{it} + \epsilon_{it}
\]

where \( g, g(t) \in R_+ \), represents the accumulation of firm-specific knowledge or experience and \( l_{it} \) is effort, \( l_{it} \in L \subset R_+ \), where \( L \) is compact. \( g(t) \) is increasing and continuous. Bojilov (2010) provides evidence in support of the choice of this additive functional form for the technology. The worker is paid \( w_{it} = \alpha_{it} + \beta_{it} y_{it} \), according to a linear compensation regime \( R_{it} = (\alpha_{it}, \beta_{it})' \), where \( \alpha_{it} > 0 \) and \( \beta_{it} > 0 \). Regime \( R_{it} \) is said to be more generous than regime \( R'_{it} = (\alpha'_{it}, \beta'_{it})' \), if both \( \alpha_{it} > \alpha'_{it} \) and \( \beta_{it} > \beta'_{it} \). The worker does not expect the

\(^2\)Since both the actual and estimated performance are always greater than 0, the restriction \( y_{it} \geq 0 \) never binds.
compensation regime to change in the future. The VNM utility of worker \( i \) is:

\[
u(\tilde{R}_{it}, \tilde{l}_{it}, y_{it}) = \alpha_{it} + \beta_{it} y_{it} - \frac{\gamma}{1 + \frac{1}{\psi}} \tilde{l}_{it}^{1 + \frac{1}{\psi}}.3
\]

This specification for the disutility of labor is popular in the related literature; for example it is used in Shearer (2004) and Paarsch and Shearer (2009). Here \( \psi \) is the elasticity of effort to its return. Since \( \tilde{\theta}_i \) and \( \tilde{l}_t \) enter additively in the utility function, posterior beliefs do not depend on effort and optimal effort choice does not depend on beliefs, so it is a function only of \( \tilde{R}_{it}, \tilde{l} (\tilde{R}_{it}) \). Optimal effort is then

\[
\tilde{l}_{it} = \left( \frac{\beta_{it}}{\gamma} \right)^{\psi}
\]

Intuitively, this assumption about the functional form of the utility implies that conditional on one’s ability, output is proportionate to \( \beta^{\psi} \).

The belief at the beginning of \( t \) is denoted as \( \tilde{\theta}_{it} \) and is formed in a Bayesian way. Let the initial prior be \( \theta_{i1} \) and suppose that employees share a common prior at the time of hiring - the distribution of match quality in the population of potential employees, \( \theta_{i1} \sim N (\mu_{\theta}, \sigma_{\theta}^2) \). Given the normality assumptions from above, the posterior belief \( \tilde{\theta}_{it} \) is normally distributed for all \( t > 1 \), \( \tilde{\theta}_{it} \sim N (\mu_{it}, \sigma_{it}^2) \), where

\[
\mu_{it} = (1 - K_t) \mu_{it-1} + K_t (y_{it-1} - \tilde{l} (\tilde{R}_{it-1}) - g (t - 1))
\]

\[
\sigma_{it}^2 = \frac{\sigma_{\tilde{\theta}}^2 \sigma_{\tilde{\theta}}^2}{\sigma_{\tilde{\theta}}^2 (t - 1) + \sigma_{\tilde{\theta}}^2}
\]

\[
K_t = \frac{\sigma_{\tilde{\theta}}^2}{\sigma_{\tilde{\theta}}^2 (t - 1) + \sigma_{\tilde{\theta}}^2}
\]

Note that precision of beliefs depends only on \( t \), so the average of the demeaned past signals is a sufficient statistic to characterize posterior beliefs. The posterior mean can be rewritten as

\[
\mu_{it} = k (t) \cdot \left( \frac{1}{t - 1} \sum_{k}^{t-1} (y_{ik} - l (R_{ik}) - g (k)) \right) + (1 - k (t)) \mu_{\theta}
\]
where \( k(t) = \frac{\sigma_i^2(t-1)}{\sigma_i^2(t-1)+\sigma^2} \). Then, the expected utility from working in period \( t \) is

\[
U(\mu_{it}, R_{it}, t, \gamma, \psi)) = \alpha_{it} + \beta_{it} \left( \mu_{it} + \left( \frac{\beta_{it}}{\gamma} \right)^\psi + g(t) \right) - \frac{\gamma}{1 + \frac{1}{\psi}} \left( \frac{\beta_{it}}{\gamma} \right)^{\psi+1}
\]

By these observations, the independence of \( \theta_i \) from \( \varepsilon_{it} \) and \( \xi_{it}^* \), and by the additivity of \( \varepsilon_{it} \) in the stochastic technology, the optimal problem of the worker can be formulated as functional equation (P)

\[
v(\mu_{it}, R_{it}, t) = \int [\max(\xi_{it}^*, U(\mu_{it}, R_{it}, t, \gamma, \psi)) + \delta \int v(\mu_{it+1}, R_{it}, t+1) f(\mu_{it+1}|\mu_{it}, t) d\mu_{it+1})] f_{\xi^*}(\xi_{it}^*) d\xi_{it}^*\]

where \( f(\mu_{it+1}|\mu_{it}, t) \) is the conditional density of \( \mu_{it+1} \), given \( \mu_{it} \) and \( t \).

**Proposition 1.** Given the specification of the model above

i. The functional equation (P) has a unique continuous solution \( V(\mu_{it}, R_{it}, t) \) and the optimal policy

\[
A(\mu_{it}, R_{it}, t) = \{ l_t \in L \mid (P) \text{ holds.} \}
\]

is a continuous function.

ii. Optimal effort \( l(R_{it}) > l(R_{it}') \) if \( R_{it} > R_{it}' \).

iii. \( V(\mu_{it}, R_{it}, t) > V(\mu_{it}, R_{it}, t) \) if \( R_{it} > R_{it}' \), and \( V(\mu_{it}, R_{it}, t) \) increases \( \mu_{it} \).

The proof of Proposition 1 is presented in Appendix A. Let

\[
H(\mu_{it}, R_{it}, t) = U(\mu_{it}, R_{it}, t, \gamma, \psi))
\]

\[
+ \delta \int \int \max\{\xi_{it+1}^*, H(\mu_{it+1}, R_{it}, t+1)\} \varphi(\mu_{it+1}|\mu_{it}, t) f_{\xi}(\xi_{it+1}^*) d\mu_{it+1} d\xi_{it+1}^*
\]

After the realization of the outside offer, \( i \) decides to stay if the value of continued employment
is higher than the value of the outside offer

\[ H(\mu_{it}, R_{it}, t) - \xi_{it}^* > 0. \]

### 2.2 Firm’s Problem

Based on the firm’s records, I take the revenue from a successfully processed call to be \( r = $8.5 \). This approximation is based on the firm records for average outbound and inbound calls, the reward that the firm receives from processing each type of calls, and the relation between the number of processed calls and accounts serviced by the company.\(^4\) I consider only contracts \( R(\alpha, \beta) \) that are linear in the performance signal: compensation \( w_{it} \) is equal to a base pay \( \alpha \) plus a bonus proportionate to performance: \( w_{it} = \alpha + \beta y_{it} \). I assume that inbound and outbound calls, as well as the number of processed calls per account is independent from the implemented contract. Furthermore, I assume that the firm’s monthly discount factor is \( \delta = 0.99 \), implying an annual discount factor of just below 0.9. Quitting disrupts the production process and necessitates spending money to advertise the available job position, and train the replacement. In what follows, I incorporate turnover costs, which according to some estimates of the firm itself amount to approximately $750. Furthermore, I also allow the firm to hire a replacement immediately after a worker quits. The firm is assumed to face constant returns to scale. Finally, the profit function that I consider below abstracts away from fixed costs.

Given these assumptions, the expected profits per employee in period \( t \), conditional on staying, match quality, tenure \( t \), and the pay policy, are defined as

\[
\pi_{it}(\theta_i, R, t) = (r - \beta) \cdot (\theta_i + l(R) + g(t)) - \alpha,
\]

\(^4\)The actual contract between the call center and the cable TV company was stated in more complicated terms. The cable TV company transfered accounts to the call center after the latter had successfully processed previously transfered calls. Thus, the call center made its profits from successfully processing accounts; the cable TV company expected more than 95% rate of collection. Yet, the contract recognized that not all attempts to contact a cable TV subscriber are successful, so it conditioned pay per account on the inbound and outbound calls that operators make to the client. Nevertheless, the underlying factor that drives profits is the successful collection of debt because that leads to the transfer of more accounts. The approximation establishes a relation between the processed calls and serviced accounts.
where $r$ is the revenue per call. Let the probability of staying at least until period $t$ be $p_{it}(R, \theta_i, t, \{\xi_{ik}, \xi_{ik}^{*}\}_{k=1}^{t-1})$. The expected profits of the firm are:

$$
\pi (R) = E_\theta \left\{ \sum_{t=1}^{\infty} \delta^{t-1} [p_{it}(R, \theta_i, t, \{\xi_{ik}, \xi_{ik}^{*}\}_{k=1}^{t-1}) \pi_{it}(\theta_i, R, t) + (1 - p_{it}(R, \theta_i, t, \{\xi_{ik}, \xi_{ik}^{*}\}_{k=1}^{t-1})) (\pi (R) - c)] \right\}.
$$

The expectation operator $E_\theta$ indicates that the expectation is taken with respect to the initial prior belief, which coincides with the distribution of match quality in the population. Each period, the employee either stays with probability $p_{it}(R, \theta_i, t, \{\xi_{ik}, \xi_{ik}^{*}\}_{k=1}^{t-1})$ and generates profits $\pi_{it}(\theta_i, R, t)$ or quits and the firm hires a new employee who at entry is expected to generate exactly the same profits as the original employee, $\pi (R)$. This equation can be solved for $\pi (R)$

$$
\pi (R) = E_\theta \left\{ \sum_{t=1}^{\infty} \delta^{t-1} \frac{p_{it}(R, \theta_i, t, \{\xi_{ik}, \xi_{ik}^{*}\}_{k=1}^{t-1}) \pi_{it}(\theta_i, R, t) - (1 - p_{it}(R, \theta_i, t, \{\xi_{ik}, \xi_{ik}^{*}\}_{k=1}^{t-1})) c}{1 - \sum_{t=1}^{\infty} \delta^{t-1} (1 - p_{it}(R, \theta_i, t, \{\xi_{ik}, \xi_{ik}^{*}\}_{k=1}^{t-1}))} \right\}
$$

The firm chooses $(\alpha, \beta)$ to maximize $\pi (R)$ subject to

$$
l (R) = \left( \frac{\beta}{\gamma} \right)^\psi.
$$

The probability of staying $p_{it}(R, \theta_i, t, \{\xi_{ik}, \xi_{ik}^{*}\}_{k=1}^{t-1})$ and the optimal effort conditions connect the firm’s problem to that of the worker presented above.

3 Data

The data set contains a clean performance measure and three known compensation regimes that were implemented in a way that allows to identify each one’s effect on performance.
The data come from 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.

As part of a reorganization plan, the central management implemented a linear contract at the beginning of January 2005: a linear function of the performance metric, $x$, the number of calls per hour that end with collection of the outstanding debt. Before January 2005, compensation was based on an hourly wage of $9.5. The central management was concerned that the company was paying "too much," so it implemented a new regime 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, and increased the slope of the piece rate for those with performance greater than 3.8 (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 1 shows a time line for the implementation of the three regimes.

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 stayers is higher than the simple average, suggesting that poor performers quit. This evidence suggests that steep pay incentives lead to high performance; that attrition appears to be non-random, since workers with higher performance are more likely to stay; that individual-specific effects are present; and finally that workers accumulate experience or knowledge in the course of their first six months of employment.

4 Estimation

Moral hazard, learning about match quality and labor mobility have been studied intensively but separately in the structural literature. Still, moral hazard and labor turnover are defining features of the analytical environment at most workplaces; their interaction shapes employment outcomes and through them profits and individual welfare. The estimation of a structural model including all these components is, however, a complicated exercise. The dynamic programs in models with Bayesian learning are quite complicated, since they involve posterior beliefs about an unobserved individual-specific parameter. As a result, estimation methods that rely on solving for the value function at each step of the optimization algorithm are computationally intensive.\(^5\)

Here, I propose a simple two-step procedure to estimate the structural model incorporating effort choice, learning about match quality, and separation decisions. In principle, the structural parameters can be recovered by estimating the following model

\[
y_{it} = \theta_i + l(R_{it}) + g(t) + \varepsilon_{it}
\]

\[
s_{ik} = 1 \left[ H(\mu_{ik}, R_{it}, k) - \xi_{ik}^* > 0 \right],
\]

\(^5\)See Nagypal (2007) for an application of indirect inference to the estimation of a model of learning about match quality. Smith (2003) provides a summary of the econometric challenges associated with the use of indirect inference to discrete choice problems and outlines a smoothing approach that addresses them.
where $y_{it}, t > 1$, is observed if $s_{ik} = 1$ for all $k = 1, ..., t$. Doing so, however, involves solving for the value function of each individual for each belief at each step of the optimization algorithm, which is computationally intensive. Therefore, in practice I estimate the model in two steps. In the first step, I estimate a semi-structural attrition model and recover the stochastic technology up to a constant, as well as a scaled version of the value of continued employment. I use these estimates in the second step to recover the remaining structural parameters using the method of moments. The main advantage of this two-step estimator is its computational simplicity, but it has also one important limitation. The initial flexible approximation of the value of continued employment can be imprecise in small samples and this can generate a finite sample bias. One way to investigate the magnitude of the potential problem and limit its effect is to apply a K-step procedure as presented in Aguirregabiria and Mira (2007). This issue is left for future research. Once the structural parameters are recovered, I use simulation methods to evaluate the profitability of the implemented regimes and to find the optimal linear contract. The simulation method and the optimization algorithm are discussed in the last subsection.

4.1 Worker’s Problem: Step 1

The first step is based on the same estimation method used in Bojilov (2010). Recall that I estimate the following semi-structural model:

$$ y_{it} = \theta_i + l(R_{it}) + g(t) + \varepsilon_{it} $$

$$ s_{ik} = 1 \left[ G(\mu_{ik}, R_{it}, k) - \xi_{ik} > 0 \right], $$

where $y_{it}, t > 1$, is observed if $s_{ik} = 1$ for all $k = 1, ..., t$, $\xi_{it} \sim N(0,1)$, and

$$ G(\theta_{ik}, R_{ik}, k) = \frac{1}{\sigma_{\xi^*}} \left( H(\theta_{ik}, R_{it}, k) - \mu_{\xi^*} \right) $$

\footnote{This is the approach taken in the literature on labor mobility, starting with Keane and Wolpin (1997).}
for all \( t \). Moreover, \( G(\theta_{ik}, R_{it}, k) \) is approximated using a linear combination of orthogonal polynomials of the explanatory variables; \(^7\) I assume that the following condition holds for the approximation \( \hat{G}(\mu_{it}, R_{it}, t) \):

\[
E \left[ \hat{G}(\mu_{it}, R_{it}, t) - \frac{1}{\sigma_{\xi}} H(\mu_{it}, R_{it}, t) - \frac{\mu_{it}}{\sigma_{\xi}} \right] = 0
\]

This model incorporates the restrictions on the stochastic technology, but imposes no structure on the utility function. As a result, it does not impose a link between effort in the performance equation and disutility of effort in the attrition equation. By estimating the semi-structural model, I recover the stochastic technology up to a constant and \( G(\mu_{it}, R_{it}, t) \) up to a scaling parameter and an additive constant. I estimate the model using maximum likelihood as discussed in Appendix B. Here I provide a short summary of the estimation method.

Let the observable information about individual \( i \) be \( W_i \) and \( \Theta_1 \) be the vector of parameters to be estimated conditional on \( \theta_i \). The likelihood for individual \( i \) conditional on the data and \( \theta_i \) can be written in a standard way as follows:

\[
l_i(\Theta_1|\theta_i, W_i) = \prod_{t=1}^{T_i} \left( \varphi \left( \frac{y_{it} - g(t) - l(R_{it}) - \theta_i}{\sigma} \right) \Phi \left( G(R_{it}, t, \mu_{it}) \right) \right)^{S_{it}} \\
(1 - \Phi \left( G(R_{iT_i}, T_i, \mu_{iT_i}) \right) \right)^{1-S_{iT_i}}
\]

where \( S_{it} = \prod_{k=1}^{t} s_{ik} \) and \( T_i \) is the last period in which \( i \) is observed. Since \( \theta_i \) is not observed, it is integrated out to obtain the individual contribution to the likelihood:

\[
l_i(\Theta|W_i) = \int l_i(\Theta_1|\theta_i, W_i) \cdot \varphi(\theta_i|W_i, \Theta_2) d\theta_i
\]

\(^7\)Bellman, Kaleba, and Kotkin (1963) first propose the use of such a linear approximation to the value function. The approximation method remains popular in both economics and machine learning, where it is still the workhorse for approximating dynamic programs as discussed in Kveton and Hauskrecht (2004).
where $\Theta_2$ is a vector of parameters that define the distribution of $\theta_i$ and $\Theta$ is a vector that contains all parameters in $\Theta_1$ and $\Theta_2$. Finally, the log-likelihood is obtained by taking logs and summing over $i$:

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

These estimates are used in the second step to obtain the remaining structural parameters: the marginal disutility of one unit of effort $\gamma$, the curvature of the disutility of effort $\psi$, and the mean and variance of the outside offer, $\mu_\xi$ and $\sigma_\xi^2$, as well as the discount factor $\delta$.

4.2 Workers Problem: Step 2

Let the difference in exerted effort under regimes 1 and 2 be $\Delta l$. Then from the performance equation,

$$\Delta l = \left( \frac{1}{\gamma} \right)^\psi \left( \beta_1^{\psi} - \beta_2^{\psi} \right)$$

The first step of the estimation provides the empirical counterpart of $\Delta l$, $\hat{\Delta} l$. Thus, one can solve

$$\Delta \hat{l} = \left( \frac{1}{\gamma} \right)^\psi \left( \beta_1^{\psi} - \beta_2^{\psi} \right)$$

for $\gamma$ in terms of $\psi$ and $\Delta \hat{l}$; let the solution be $\gamma \left( \psi, \Delta \hat{l} \right)$ and substitute the solution in the expression for the disutility of labor to obtain $U \left( \mu_{it}, R_{it}, t, \Delta \hat{l}, \psi \right)$.

To save on notation, define

$$\lambda(G(\mu_{it}, R_{it}, t))$$

$$= E_\xi \max \{\xi_{it}, G(\mu_{it}, R, t)\}$$

$$= G(\mu_{it}, R_{it}, t). \Phi (G(\mu_{it}, R_{it}, t)) + \varphi (G(\mu_{it}, R_{it}, t))$$
Note that $V(\mu_{it}, R, t)$ can be expressed in terms of $G(\mu_{it}, R_{it}, t)$ as follows

$$V(\mu_{it}, R_{it}, t) = E_{\xi^*} \max \{\xi_{it}^*, H(\mu_{it}, R_{it}, t)\}$$

$$= \mu_{\xi^*} + \sigma_{\xi^*} E_{\xi} \max \{\xi_{it}, G(\mu_{it}, R, t)\}$$

$$= \mu_{\xi^*} + \sigma_{\xi} [\lambda(G(\mu_{it}, R_{it}, t))]$$

From the definition of $H(\mu_{it}, R_{it}, t)$ and the above representation of $V(\mu_{it}, R_{it}, t)$,

$$H(\mu_{it}, R_{it}, t) = \mu_{\xi^*} + \sigma_{\xi^*} G(\mu_{it}, R_{it}, t)$$

$$H(\mu_{it}, R_{it}, t) = U(\mu_{it}, R_{it}, t, \gamma, \psi)) + \delta [\mu_{\xi^*} + \sigma_{\xi^*} E_{\mu_{it+1}}(\lambda(G(\mu_{it+1}, R_{it}, t + 1)))]$$

where $E_{\mu_{it+1}}$ indicates that expectation is taken with respect to the distribution of the mean of the posterior beliefs in $t + 1$ given the information available at $t$, $\sigma_{\xi}^2$ and $\mu_{it}$. Consequently,

$$\mu_{\xi^*} + \sigma_{\xi^*} G(\mu_{it}, R_{it}, t) = U(\mu_{it}, R_{it}, t, \gamma, \psi)) + \delta [\mu_{\xi^*} + \sigma_{\xi^*} E_{\mu_{it+1}}(\lambda(G(\mu_{it+1}, R_{it}, t + 1)))]$$

By this identity and the assumptions on the approximation of $G(\mu_{it}, R_{it}, t)$, the delta method implies the following conditions for $t = 1, ..., T$

$$E\left(\tilde{G}(\mu_{it}, R_{it}, t) - M_{it}(\mu_{it}, R_{it}, t, \Theta_2)\right) = 0,$$

where

$$M_{it}(\mu_{it}, R_{it}, t, \Theta_2)$$

$$= \frac{1}{\sigma_{\xi^*}} \left\{ U(\mu_{it}, R_{it}, t, \Delta\hat{\psi}, \psi) + \delta \left[\mu_{\xi^*} + \sigma_{\xi^*} E_{\mu_{it+1}}(\lambda(\tilde{G}(\mu_{it+1}, R_{it}, t + 1)))) - \mu_{\xi^*}\right] \right\},$$

and $\Theta_2 = (\psi, \delta, \sigma_{\xi^*}, \mu_{\xi^*})$ is the vector of structural parameters recovered at the second stage. $\Theta_2$ does not contain $\gamma$ as it can be recovered from $\gamma(\psi, \Delta\hat{\psi})$. 

16
Define
\[ M_t(\Theta_2) = \sum_i M_i(\mu_{it}, R, t, \Theta_2) \] and \[ \hat{G}_t = \sum_i \hat{G}(\mu_{it}, R, t) \]
where the summation is over the individuals who make the decision to stay or quit in period \( t \). Furthermore, let
\[
M(\Theta_2) = \begin{pmatrix}
\cdot & \\
\cdot & \cdot \\
\cdot & \cdot \\
\cdot & \cdot
\end{pmatrix}
\quad \text{and} \quad
\hat{G} = \begin{pmatrix}
\cdot \\
\cdot \\
\cdot \\
\cdot
\end{pmatrix}
\]
The remaining parameters \( \Theta_2 \) are the solution to
\[
\min_{\Theta_2} \left( \hat{G} - M(\Theta_2) \right)' \Omega^{-1} \left( \hat{G} - M(\Theta_2) \right),
\]
where \( \Omega \) is the optimal weighing matrix. Under the specified model \( \delta, \psi, \gamma, \mu_{\xi^*}, \sigma_{\xi^*} \) are identified; identification is discussed in Appendix C.

Let the covariance matrix of the structural parameters \( \Theta_1 \) estimated in the first step be \( \Sigma \). By the Delta method
\[
\Omega = W'\Sigma W
\]
where \( W = \frac{\partial \hat{G}}{\partial \Theta_1} \). Finally, the covariance matrix \( \Sigma \) is obtained from the first-step MLE estimates. Then, following Hansen (1982), the asymptotic covariance matrix for \( \Theta_2 \) is \( (J'\Omega^{-1}J)^{-1} \), where \( J = \frac{\partial M}{\partial \Theta_2} \).

Note that the estimation of the structural parameters \( \Theta_2 \) is in its essence a consistency check for the estimates of the first-step attrition model: the second step can be interpreted as a search for structural parameters that generate a data process that is consistent with the findings in the first step. The criterion function evaluated at the optimum has \( \chi^2 \) distribution.
with 12 degrees of freedom under the null hypothesis that the theoretical model is valid.

4.3 Firm's Problem

For any regime, the probability of staying at tenure \( t, p_{it} \left( R, \theta_i, t, \{ \xi_{ik}, \xi_{ik} \}_{k=1}^{t-1} \right) \), cannot be estimated analytically, so I resort to simulations to evaluate the profitability of pay regimes. For the set of employees who enter the firm, I draw paths of \( \varepsilon_t \) and \( \xi_t^* \) and \( \theta \) to generate 1000 data sets. Using the point estimates from steps one and two, I generate the sequence of noisy performance signals and posterior beliefs. The generation of the separation indicators, \( s_{ik} \), requires some care. Given a regime \( R \), I solve for the value function of each individual for each posterior belief. I assume that conditional on staying for 2 years employees know the true value of their match quality and the accumulation of experience has stopped. Then, the worker’s problem becomes

\[
V(\theta_i, R) = \int \max \left[ \xi_{it}^*, U(\theta_i, R) + \delta V(\theta_i, R) \right] dF_{\xi^*},
\]

where \( U(\theta_i, R) \) stands for the expected utility after the individual knows her match quality \( \theta_i \), there is no more experience to be gained, and \( V(\theta_i, R) \) is the value of continued employment. This problem can be solved as a standard fixed-point problem using value function iteration. The starting value for the iterations is the discounted sum of expected utility, i.e.

\[
V^0(\theta_i, R) = \frac{1}{1 - \delta} U(\theta_i, R)
\]

Then, I solve backwards for the utility of continued employment \( V(\mu_{it}, R, t) \), using the appropriate posterior. I use the Gauss-Hermite method with 8 nodes of integration. This approach to solving for the value function is similar to the one employed in Nagypal (2007). Comparing the drawn outside offers and the values of continued employment from above generates the sequence of separation indicators. It should be noted that all workers eventually quit. For each of the simulated data sets, I find the expected profits per entering employee by averaging the discounted some of individual profits for the duration of stay. To find the expected profits per
workstation, I take into account that all quits are replaced by new workers who have exactly the same expected profits at entry as the original cohort. This simulation method is used to estimate the profits of the firm under the actually implemented regimes and to evaluate the candidates for the optimal linear contract at each step of the optimization. Given the low dimension of the optimization problem, I use a version of the simplex algorithm to find the optimal linear contract.

5 Results

This section presents the results from estimating the structural model and then shows how they can be used to find the profit-maximizing pay policy under various assumptions about the employment environment. The policy analysis indicates that turnover is a major channel through which pay incentives affect both performance and profits.

5.1 Estimates of Structural Parameters

In this subsection, I present the results from estimating the structural model and characterize the employment environment. The attrition model of the first step is estimated using MLE. The results, their econometric implications, and possible alternative specifications are discussed in greater detail in Bojilov (2010). The explanatory variables for the performance equations include second degree orthogonal polynomials of tenure and calendar time, dummies for regimes of operation and regimes of hiring, unobserved match quality, and controls. Specifically, 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 orthogonal polynomials interacted with regimes and, depending on the specification, \( \theta_i \) or \( \mu_{it} \), controls, calendar time, and regime of hiring. As a preliminary step, I conduct a specification search for the degrees of the orthogonal polynomials in the performance and attrition equations. I find that orthogonal polynomials of degree 2 for the
performance equation and orthogonal polynomials of degree 3 for the attrition equations fit the data best. The estimates can be found in Tables 1-3. They are very similar to the ones reported in Bojilov (2010) for the basic attrition model. The main difference is that some not significant variables have been omitted, along with the dummies for regime of hiring. In what follows, I make a brief summary of those results that are directly related to the second step of estimation and profits. Furthermore, I measure the contribution of effort, match quality, and experience to performance in terms of successful calls per hour (just calls per hour for short from now on).
Table 3: Estimates of parameters related to ability and learning in the attrition model

<table>
<thead>
<tr>
<th>Parameter or explanatory variable</th>
<th>Attrition Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
</tr>
<tr>
<td>$\sigma_\epsilon^2$</td>
<td>0.17</td>
</tr>
<tr>
<td>$\sigma_\xi^2$</td>
<td>1</td>
</tr>
<tr>
<td>$\rho(\epsilon, \xi)$</td>
<td>0.01</td>
</tr>
<tr>
<td>$\sigma_\beta^2$</td>
<td>0.48</td>
</tr>
<tr>
<td>$\mu_{it}$</td>
<td>0.18</td>
</tr>
<tr>
<td>$t_1 = \mu_{it}$, orthog. pol. 1</td>
<td>0.06</td>
</tr>
<tr>
<td>$t_2 = \mu_{it}$, orthog. pol. 2</td>
<td>0.02</td>
</tr>
<tr>
<td>$t_3 = \mu_{it}$, orthog. pol. 3</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Log-likelihood: -3745.73

Notes: The specification also includes calendar time orthogonal polynomials of degree 2 and individual controls: gender, age, marriage status, distance from home, and race.

In the first step, I estimate the distribution of match quality at entry and characterize the associated dynamics of learning and turnover. The variance of match quality is 0.48 and accounts for the greater part of the variance in performance at entry under regimes 1 and 2. Moreover, it has an important effect on attrition. Figure 1 presents the distribution of match quality at entry and how it changes by the sixth months of employment. The value of $\theta$ is on the horizontal axis, while the vertical axis represents the proportion 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 $\theta$ shifts to the right as tenure increases under both regimes 1 and 2 and only workers with very high match quality remain employed after six months of work. As expected, the switch from regime 1 to regime 2 generates an increase in turnover at any tenure horizon. The variance of the disturbance term in the performance equation is estimated at 0.17 which implies that the signal-to-noise ratio, defined as the ratio of the variance of match quality over the variance of noise, is approximately 2.6. Consequently, within 6 months the variance of the posterior beliefs declines to approximately 0.05 and the weight on the initial belief declines to almost zero.

Furthermore, I recover the technology up to an additive constant. The estimated parameters for the performance equation are broadly consistent with the theoretical predictions.
In economic terms, the switch from regime 1 to regime 2 leads to a decline in worker’s effort and in turn performance by about 0.2 calls per hour which translates into a decline in hourly pay by approximately $2. The estimates imply a significant improvement in performance over time due to the accumulation of experience: in the first 6 months of employment performance increases by approximately one successful call per hour, or 35% growth in the first six months under regime 1. Under regime 1, this growth translates in an increase in hourly pay by approximately $3.3. Finally, I also estimate flexibly the normalized and scaled value of continued employment \( G(\mu_{it}, R, t, X_{it}) \) which provides the basis for the second step estimation.

Table 4 presents the estimates from the second step. Before discussing the results from the second step, I first check the validity of the overidentifying restrictions. The \( \chi \)-square test with 12 degrees of freedom for the overidentifying restrictions fails to reject the null hypothesis that the restrictions are valid, since the test statistic is 5.16. Thus, I conclude that the data is consistent with the restrictions imposed by the model. 8 The second step results allow for the characterization of optimal effort choice under regimes 1 and 2. The elasticity of effort to pay incentives \( \psi \) is estimated at 3.27 with a standard error of 0.28, implying that workers’ supply of effort is highly sensitive to changes in pay incentives. The relative benefit of effort to its subjective cost represented by \( \gamma \) is estimated at 3.9 with a standard error of 0.3, so that effort amounting to one call per hour costs to the individual $3.9. Given the estimates of \( \gamma \) and \( \psi \), the level of effort under regime 1 translates into an increase in performance by 0.59 calls per hour and under regime 2 by 0.39 calls per hour. Compared to the variation in match quality, the contribution of effort to performance is relatively small: less than one standard deviation under regime 1 and even less under regime 2. In contrast, the mean of match quality in the population of entering workers is approximately 2, i.e. independent of pay incentives an employee of average quality successfully completes two calls per hour when starting work. Furthermore, the results indicate that in the absence of selection at entry and

---

8Note that the estimation of the structural parameters \( \Theta_2 \) is a test for consistency of the first-step estimates with the specified utility in the model.
Table 4: Estimates of the structural parameters of the model.

<table>
<thead>
<tr>
<th>Parameter:</th>
<th>Coefficient</th>
<th>Std. Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\psi$</td>
<td>3.26</td>
<td>0.21</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>3.89</td>
<td>0.24</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.74</td>
<td>0.08</td>
</tr>
<tr>
<td>$\sigma_{\xi}^2$</td>
<td>32.75</td>
<td>0.31</td>
</tr>
<tr>
<td>$\mu_{\xi}$</td>
<td>45.6</td>
<td>2.36</td>
</tr>
<tr>
<td>$\Delta l$</td>
<td>-0.19</td>
<td>0.07</td>
</tr>
<tr>
<td>$\Delta$disutility</td>
<td>-0.64</td>
<td>0.07</td>
</tr>
<tr>
<td>experience by $t = 6$</td>
<td>0.99</td>
<td>0.09</td>
</tr>
<tr>
<td>$\sigma_{\theta}^2$</td>
<td>0.48</td>
<td>0.02</td>
</tr>
<tr>
<td>$\mu_{\theta}$</td>
<td>2.02</td>
<td>0.11</td>
</tr>
<tr>
<td>$\chi^2_{12}$ test stat.</td>
<td>5.06</td>
<td></td>
</tr>
</tbody>
</table>

exit the optimal piece rate involves $\beta = 6.55^9$. The fact that the implemented pay regimes have $\beta$ much lower suggests that turnover has a nontrivial effect on profits.

The monthly discount factor is estimated at 0.75 with standard error of 0.08, indicating a strong preference for present to future consumption: when making decisions workers assign a weight 0.001 to consumption after two years. The mean of the distribution of outside options is $45.6 with a standard error of 2.46, where an outside option stands for the present value of hourly compensation at an alternative job. The variance of outside offers is estimated at $32.5. To give some perspective, a job with an hourly wage of $11.25 has a present value of $45, assuming that the worker has no right to leave after entry. From these characteristics of the distribution of outside offers one may guess that the workers may find alternative employment at low-skill service jobs or low-skill manufacturing jobs whose hourly wage varies between $8 and $14.

\[ \beta = \frac{p\psi}{\psi + 1} \]

where $p$ is the revenue generated from the employment relation and in the present context is $8.5 per call. Substituting the estimated $\psi$ gives $\beta = 6.55$

---

9Solving the profit-maximization problem of the firm subject to a participation constraint yield
5.2 Profits and Policy Analysis

The estimates of the structural model provide the basis for counterfactual policy analysis. In this subsection, I start by discussing the profitability of the implemented regimes 1 and 2, as well as the contribution of effort, experience, and match quality to profits under these regimes. Then, I consider the problem of maximizing the profits of the firm. I compare the results to those previously presented for regimes 1 and 2. Finally, I consider some counterfactual changes in the firm environment and their effect on profits. In particular, I consider a higher level of turnover costs than the one reported by the firm. I also evaluate the implications for profits and the optimal compensation policy when workers learn about their match quality before deciding whether to enter the firm.

The decomposition of profits is difficult because profits depend on both performance and the probability of staying, while the latter is a highly nonlinear function of effort, beliefs about match quality and experience. I approach the problem in the following way. The additive structure of the technology allows for isolating the contribution of effort, ability, and experience to profits. I define

$$\pi_\theta (R) = E_\theta \left\{ \sum_{t=1}^{\infty} \delta^{t-1} P_{it} \left( R, \theta_i, t, \{ \varepsilon_{ik}, \xi_{ik}^* \}_{k=1}^{t-1} \right) (r - \beta) \theta_i \right\}$$

where

$$P_{it} \left( R, \theta_i, t, \{ \varepsilon_{ik}, \xi_{ik}^* \}_{k=1}^{t-1} \right) = \frac{p_{it} \left( R, \theta_i, t, \{ \varepsilon_{ik}, \xi_{ik}^* \}_{k=1}^{t-1} \right)}{1 - \sum_{t=1}^{\infty} \delta^{t-1} \left( 1 - p_{it} \left( R, \theta_i, t, \{ \varepsilon_{ik}, \xi_{ik}^* \}_{k=1}^{t-1} \right) \right)}$$

to be the profits associated with match quality. In a similar way,

$$\pi_1 (R) = E_\theta \left\{ \sum_{t=1}^{\infty} \delta^{t-1} P_{it} \left( R, \theta_i, t, \{ \varepsilon_{ik}, \xi_{ik}^* \}_{k=1}^{t-1} \right) (r - \beta) l (R) \right\}$$

$$\pi_t (R) = E_\theta \left\{ \sum_{t=1}^{\infty} \delta^{t-1} P_{it} \left( R, \theta_i, t, \{ \varepsilon_{ik}, \xi_{ik}^* \}_{k=1}^{t-1} \right) (r - \beta) g (t) \right\}$$
\(\pi_l(R)\) and \(\pi_t(R)\) stand for the contribution to profits by effort and with experience, respectively. I take hourly wage as a benchmark regime with respect to which I evaluate how total profits and the contributions of effort, experience and match quality defined above change as the pay regime changes. An alternative approach that I also apply to the study of the effect of effort on profits is to compare profits under the same pay regime when effort affects performance and stay and when it is restricted to have no effect on them: the difference in the profits provides a conservative estimate for the contribution of effort to profits.

5.2.1 Implemented Regimes

Table 5 presents the pay policies that I analyze, along with profits,\(^{10}\) effort, average match quality and average tenure per workstation under each of them. Table 6 considers the channels through which pay incentives affect profits. It considers the effects on the contributions of effort, match quality, and tenure to profits when switching from the initial hourly wage to some alternative pay regimes. Under an hourly wage, employees do not exert effort. Moreover, equal hourly pay implies that workers of different match quality are equally likely to quit at any tenure horizon. I fix the hourly wage to $9.5 which was actually implemented by the firm prior to January 2005. This hourly wage is clearly quite low relative to the mean of the outside offer and leads to a very high turnover: more than 93% of the employees last at most six months in the firm. This high turnover leads to a low level of experience in the workforce as indicated by an average tenure of 3.23. Furthermore, the failure of the hourly wage to distinguish between workers of high and low match quality leads to an average match quality of 1.99 calls per hour. Taken together, these effects of the hourly wage lead to total profits of $19.4.\(^{11}\)

Switching from the hourly wage to regime 1 induces all workers to exert effort of 0.58 calls per hour but also rewards workers of high match quality more than workers of low match quality \(\pi_B\) are $28.89 and the profits associated with experience \(\pi_t\) are $10.4.

\(^{10}\)Recall that from the definition of profits, total profits stands for the discounted infinite sum of hourly profits starting from the month of hiring the worker.

\(^{11}\)Under the given specification of the stochastic technology and the utility, the profits associated with match quality \(\pi_B\) are $28.89 and the profits associated with experience \(\pi_t\) are $10.4.
quality. As discussed in Chapter 3, the result is that workers of high match quality stay longer in the firm than workers of low match quality. These differences lead to an increase in average match quality to 2.88 calls per hour. The net effect of the change in the compensation policy on the separation decisions is a decline in turnover illustrated with an increase in average tenure to 11.2 months. The retention of employees of high match quality, along with the decline in their probability of quitting at any tenure horizon leads to an impressive increase in \( \pi_{t} \) by $90. The lower turnover also leads to an increase in the profits associated with experience by approximately $49. Finally, the introduction of the bonus rate of $3.3 per successful call induces effort that generates profits associated with effort in the amount of $31.42.\(^\text{12}\) Total profits jump to $167. These numbers indicate that the increase in \( \pi_{t} \), followed by the increase in \( \pi_{t} \), rather than the increase in \( \pi_{t} \) makes the greatest contribution to the increase in profits when switching from hourly wage to regime 1. The results suggest that the firm benefits considerably from the accumulation of workers of high match quality through turnover.

Next, I consider the effect of regime 2 on profits. Recall that this regime stipulates both lower base pay and lower piece rate. This less generous compensation policy leads to a sharp increase in the probability of quitting during the first six months which approaches the levels under the hourly wage. The result is average tenure of 6.2 months, a decrease by more than 40% relative to regime 1, which implies also lower levels of accumulated experience. While the probability of quitting increases at each tenure horizon, the firm still retains workers of very high match quality, as discussed in Bojilov (2010). However, average match quality under regime 2 is not higher but slightly lower than average match quality under regime 1: 2.82 calls per hour. This result indicates that the negative effect of high turnover more than offsets the effect of retaining only the workers of highest match quality. At the same time, effort declines to 0.34 calls per hour. The combined effect of these factors implies that regime 2 yields much lower profits than regime 1. Despite the fact that the piece rate declines by $0.8 calls per hour, the profits associated with match quality are still $70 higher compared to their level under the hourly wage. However, as a result of the high rate of destruction of

\(^{12}\text{Recall that the firm incurs a flat hourly pay of } \alpha \text{ and turnover costs which must be subtracted to obtain the total profits.}\)
accumulated experience, $\pi_t$ is quite close to its level under the hourly wage: it is only $16$ higher. The profits associated with effort $\pi_t$ are approximately $10$, and total profits amount to about $110$. Thus, under regime 2 match quality continues to be a crucial determinant of profits and the decline relative to regime 1 is smallest in the case of $\pi_\theta$.

5.2.2 Optimal Regime

The solution of the profit-maximization problem is the optimal pay regime $R^w$ defined by $\alpha_w = 3.64$ and $\beta_w = 3.26$. Several factors affect its properties. While steep incentives induce more effort and increase the probability of staying, they also surrender a larger proportion of the revenues to the employees. Since quitting of an employee comes with the possibility of hiring a better one in the future, the firm chooses a pay schedule that among other things, balances the benefit from continued employment of a worker and the benefit from finding one of higher quality. The results here depend crucially on the firm-specific nature of the match quality parameter: in particular, the ability of the firm to extract much of the surplus
Figure 2: Comparison between profits under regime 1, the optimal regime when turnover cost is $750 and when it is $8800. Turnover cost is $750.

from the employment relation will be limited if workers can export their match quality \( \theta \) to alternative jobs. The findings also depend to some extend on the simple nature of the compensation policy: for example, the properties of the optimal pay regime will change if the firm can condition base pay \( \alpha \) on posterior beliefs about match quality.

Table 5: Profits, effort, ability and tenure per workstation under different regimes.

<table>
<thead>
<tr>
<th>Pay policy:</th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( t )</th>
<th>( E(\alpha) )</th>
<th>( E(\theta) )</th>
<th>( \pi )</th>
</tr>
</thead>
<tbody>
<tr>
<td>hourly wage, ( \bar{w} )</td>
<td>9.5</td>
<td>0</td>
<td>0</td>
<td>3.23</td>
<td>1.99</td>
<td>19.41</td>
</tr>
<tr>
<td>regime 1</td>
<td>3.8</td>
<td>3.3</td>
<td>0.58</td>
<td>11.22</td>
<td>2.86</td>
<td>167.72</td>
</tr>
<tr>
<td>regime 2</td>
<td>3.5</td>
<td>2.8</td>
<td>0.34</td>
<td>6.23</td>
<td>2.82</td>
<td>109.81</td>
</tr>
<tr>
<td>regime ( R_w )</td>
<td>3.65</td>
<td>3.24</td>
<td>0.55</td>
<td>9.85</td>
<td>2.91</td>
<td>174.24</td>
</tr>
<tr>
<td>regime ( R_w ), known ability</td>
<td>3.65</td>
<td>3.24</td>
<td>0.55</td>
<td>12.30</td>
<td>3.12</td>
<td>215.92</td>
</tr>
<tr>
<td>regime ( R_n ), known ability</td>
<td>3.74</td>
<td>3.09</td>
<td>0.47</td>
<td>11.61</td>
<td>3.17</td>
<td>221.74</td>
</tr>
<tr>
<td>regime ( R_h ), high costs</td>
<td>1.82</td>
<td>5.44</td>
<td>2.91</td>
<td>19.12</td>
<td>2.17</td>
<td>162.30</td>
</tr>
</tbody>
</table>
Table 6: Effects of different pay regimes relative to hourly wage.

<table>
<thead>
<tr>
<th>Pay policy:</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\Delta \pi_l$</th>
<th>$\Delta \pi_t$</th>
<th>$\Delta \pi_\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>regime 1</td>
<td>3.8</td>
<td>3.3</td>
<td>35.06</td>
<td>49.14</td>
<td>89.85</td>
</tr>
<tr>
<td>regime 2</td>
<td>3.5</td>
<td>2.8</td>
<td>9.74</td>
<td>15.68</td>
<td>70.33</td>
</tr>
<tr>
<td>regime $R^w$, known ability</td>
<td>3.65</td>
<td>3.24</td>
<td>31.42</td>
<td>99.32</td>
<td>99.32</td>
</tr>
<tr>
<td>regime $R^m$, known ability</td>
<td>3.65</td>
<td>3.24</td>
<td>33.31</td>
<td>49.93</td>
<td>131.18</td>
</tr>
<tr>
<td>regime $R^h$, high costs</td>
<td>3.74</td>
<td>3.09</td>
<td>29.18</td>
<td>48.37</td>
<td>140.57</td>
</tr>
<tr>
<td></td>
<td>1.82</td>
<td>5.44</td>
<td>191.04</td>
<td>56.20</td>
<td>23.42</td>
</tr>
</tbody>
</table>

Note: $\Delta \pi_x = \pi_x (R) - \pi_x (\overline{w})$, $x = l, t, \theta$.

The slope and base pay of the optimal pay regime $R^w$ are very close to those implemented under regime 1. The optimal pay regime $R^w$ induces considerable turnover: only about 55% of the employees stay more than six months in the firm. Furthermore, it not only induces a similar rate of turnover but also leads to a similar quality mix at different tenure horizons as regime 1. Figure 1 shows that the conditional distributions of match quality after six months under regime 1 and regime $R^w$ are almost identical. It also indicates that only workers of high match quality (match quality greater than $\theta + \sigma_\theta$) experience little or no turnover. The small slope of incentives induces effort that translates into only 0.55 calls per hour, less than one standard deviation of match quality among starting employees. The distributions of expected profits under the optimal pay regime $R^w$ and regime 1 are again very similar but some differences are also present, as evident from Figure 2. Regime 1 generates more income on employees of average quality while regime $R^w$ generates more profits on the top performers. This pattern is explained by the fact that under regime $R^w$ the firm captures more of the surplus from the top performers while the slightly higher slope of incentives and base pay under regime 1 induce more workers of average ability to stay and work. The net effect is that optimal pay regime $R^w$ generates approximately 4.8% higher profits than regime 1.

Relative to regime 1, the optimal regime is less generous which leads to a small increase in the probability of quitting across posterior beliefs and tenure horizons. The result is average tenure of about 10 months, slightly lower than the 11.2 months under regime 1. This finding implies that the level of accumulated experience under the optimal regime $R^w$ is just below that for regime 1. The probability of quitting increases at each tenure horizon relative to regime 1, but the firm still retains workers of very high match quality, as discussed in the
previous paragraph in the context of Figure 2. The net effect is that the average match quality under $R^w$ is 2.91 calls per hour, slightly higher than its counterpart for regime 1 which leads also to higher profits. Furthermore, shaving off 6 cents from the bonus rate reduces effort only to 0.55 calls per hour relative to the 0.58 call per hour under regime 1. Due to the improved quality mix, the small negative effect on turnover, and the small reduction in the variable costs, the contribution of match quality to profits, $\pi_\theta$, increases by $99 relative to its level under hourly wage. The contribution of tenure to profits, $\pi_t$, increases by $44 when switching from hourly wage to regime $R^w$. The contribution of effort, $\pi_l$, declines slightly relative to regime 1 to $31. Combining all of these with the costs of turnover and base pay yields total profits of $174. The results show that switching from the benchmark hourly wage to the optimal regime $R^w$ leads to an increase in profits by more than eight times, or in absolute terms by $154. The results show that much of this change is due to the effect of incentives on the quality mix of the workforce rather than the effect of incentives on effort.
An alternative approach to evaluate the effect of incentives on profits through effort is to compare profits under the optimal pay regime when effort affects performance and stay and when it is restricted to have no effect on them. My analysis starts with the quality mix, assuming that effort is not a channel through which incentives affect performance or separation decisions. Figure 3 shows that under the hourly wage of $9.5, implemented until January 2005, the firm makes losses on some workers of below average ability and its expected profits amount only to approximately $20, due to a high quitting rate and the associated costs. The figure also indicates that the introduction of the optimal regime induces high quality employees to stay while low quality employees to quit. The firm captures 75 % of the additional surplus and profits increase by more than a factor of three.

Next, I relax the restriction that incentives do not affect effort, but still maintain that effort choice has not effect on separation decisions. Figure 3 shows that the exerted effort leads to an additional increase in profits by 114%. Finally, I also allow effort choice to affect separation decisions, but Figure 3 indicates that only but a few separation decisions remain unchanged: the combined effect of effort choice and match quality for those who switch from quitting to staying accounts for a 19% increase in profits. Thus, the total effect of switching from hourly wage to the benchmark rate results in a dramatic increase in profits, but two-thirds of the increase would have materialized even if pay incentives did not affect effort choice or the separation decisions.

To summarize, these results show that most of the increase in profits from switching to the optimal pay regime can be traced back to the effect of incentives on the quality mix. Pay incentives not only induce high quality employees to stay but also act as a selection mechanism that helps the firm build a workforce of high match quality over time. In the present context, the firm exploits the firm-specific nature of the relation to capture most of the surplus generated by the employment relation.
Figure 4: Comparison between profits under regime 1, the optimal regime when turnover cost is $750 and the optimal regime when it is $8800. Turnover cost of $8800.

6 Counterfactual Experiments

The turnover costs of $750 reported by the firm appear very low relative to industry averages published in Superb Staff Services (2011) which vary between $4,100 and $25,000. In this subsection, I explore the effect of high turnover costs on profits under the regime $R^w$, optimal under a turnover cost of $750, and search for the optimal regime under turnover costs equal to the industry average of $8,800. Furthermore, I study the effect on profits when the worker knows her match quality before deciding to start working but the employer does not.

6.1 Turnover Costs

The optimal pay regime $R^h$ when turnover costs are $8,800 is defined by $\alpha_h = 1.55$ and $\beta_h = 5.42$. These slope and base pay are much different from the one’s implemented by the firm. The high-powered incentives induce little turnover, mainly in the first two months of employment, and a high level of effort resulting in 2.91 calls per hour. Figure 4 presents profits under regimes 1, $R^w$ and $R^h$ when turnover costs are $8,800. A comparison of profits under
pay regime \( R^h \) and regime \( R^w \) reveals that the two have a very similar expected profits from the top performers, while regime \( R^h \) accumulates much higher profits on the employees of low and average quality. Thus, the top performers capture much of the revenue under regime \( R^h \), while the firm increases its profits from the higher effort exerted by employees who would not have stayed under regimes 1 or \( R^w \). Table 5 shows that the low turnover rate leads not only to a high average tenure of about 20 months but also to a low average match quality of 2.17 calls per hour. In contrast to the case of the optimal contract when turnover cost is only $750, the profits under optimal regime \( R^h \) come mainly from high levels of effort: \( \pi_l \) for this regime is $191. Still, total profits are only $162 because of the high turnover costs.

Next, I analyze the composition of profits. I start with the quality mix, assuming that implementing regime \( R^h \) does not affect performance or separation decisions through effort choice. Figure 5 shows that the introduction of regime \( R^h \), even in the absence of any effect through effort choice, allows many employees to remain in the firm and in turn generate revenue of which more than 67% go to the workers. Then, I relax the restriction that incentives do not affect effort but still maintain that effort is not a channel through which incentives affect
separation decisions. Figure 5 shows that, under these new restrictions, pay incentives induce effort that increases revenues considerably in contrast to the case of the optimal regime when turnover costs are $750. Finally, I allow effort choice to affect separation decisions. Given the assumptions of the model about the utility function, the regime induces higher effort and higher utility. Thus, the introduction of effort choice changes some but not all separation decisions: some workers who would have otherwise left now decide to stay. The employees who now stay contribute to profits with their match quality and effort: as evident from Figure 5, the effect is not negligible.

Summing up, the total effect of switching from hourly wage to regime $R^e$ results in an impressive increase in profits, but only 27% of this growth would have materialized if pay incentives did not affect performance and separation decisions through effort choice. Thus, this counterfactual experiment indicates the sensitivity of the solution to the profit maximization problem to turnover costs.

6.2 Workers Know the Match Quality

Table 5 and 6 report average match quality, average tenure, profits, and their decomposition when workers know their match quality before deciding to join the firm. When regime $R^w$, optimal when workers learn their match quality, is implemented in this environment, profits increase to $216$. Much of this increase can be traced back to self-selection at entry: some workers know that their match with the firm is of low quality and decide to opt out for an alternative. Figure 6 shows the mean of the distribution of match quality at $t = 1$ under $R^w$ when workers know their match quality is 2.35 calls per hour compared to 2.01 calls per hour when they learn about it. As a result, the firm accumulates workers of high match quality faster than when workers learn about match quality and the average match quality increases to 3.12 calls per hour, while the average tenure increases to 12.3 months. These effects lead to a considerable increase of $141$ in $\pi_\theta$ relative to its level under hourly wages, which is largely responsible for the increase of total profits to $216$.

The next step is to find the optimal pay regime when workers know their match quality
Figure 6: Ability at entry under the optimal regimes when workers know match quality, when they learn about it, and when the latter is applied to an environment in which workers know their match quality. Turnover cost is $750.

Figure 7: Comparison between profits under the optimal regimes when workers know match quality at entry, when they learn about it, and when the latter is applied to an environment in which workers know it. Turnover cost of $8800.
before deciding to enter the firm. This problem is a special case of the more general model presented above: the prior belief is a degenerate distribution centered at the true value of match quality. The solution of the profit-maximization problem is the optimal pay regime $R^n$ defined by $\alpha_n = 3.74$ and $\beta_n = 3.09$. The results for this regime are reported in tables 5 and 6. The firm offers lower incentives to exert effort in order to capture a greater share of the profits associated with match quality which is partially offset by a modest increase in the base pay. Thus, the growth in income and the variance of the distribution of income in this environment are smaller than their counterparts when workers learn about match quality. Still, one cannot generalize too much from this result because the properties of regime $R^n$ depend considerably on the restriction to search for the optimal regime within the family of linear contracts only. The average match quality under $R^n$ is 3.17 calls per hour and the distribution of match quality among the entering employees is not much different from that under $R^w$, as shown on Figure 6. Average tenure is 11.6 months, compared to 12.3 months under $R^w$, while effort amounts to only 0.47 calls per hour. Total profits are $221. Figure 7 presents profits under regime $R^n$ when workers know their match quality, under regime $R^w$ when workers learn about their match quality, and when they know it. It shows that the profits under $R^n$ and $R^w$ when workers know their match quality are similar. This is driven by the fact that more or less the same type of people enter the firm under both regimes and all differences arise from the fact that $R^n$ shaves off more of the revenue from top performers by decreasing the bonus rate at the expense of a slightly higher turnover. Consequently, the results indicate that employers can benefit considerably if they can introduce a technology that helps workers find out their match quality before they decide to enter the firm.

7 Conclusion

This paper considers a structural model of effort choice, learning about match quality, and turnover. It shows how such a model can be estimated with a two-step procedure that borrows ideas from the literature on estimation of dynamic structural models. The results indicate that
employees are very responsive to pay incentives, impatient to postpone future consumption, and face a large variety of outside options primarily selected from low-skill service and manufacturing industries. Workers accumulate experience during the first six months on the job which improves performance. Still, variability in the quality of the employer-employee match accounts for most variation in performance across individuals under a given pay regime. The paper examines a variety of regimes to find that the firm maximizes profits by selecting and keeping the high quality employees, even at the expense of inducing low effort. It also shows that most gains from switching to the optimal pay regime from an hourly wage can be traced to the improvement in the match quality of the workforce.

The question arises whether the profit-maximizing regime is also socially optimal. The social surplus per workstation in a given period is the collected revenue minus disutility denominated in dollars and turnover costs. At first sight, the criterion functions for maximizing profits and social surplus are different, so there is no reason to expect that the solutions of the two optimization problems are going to be the same. The analysis of this issue is left for future research. Furthermore, in this paper, I limited my attention to linear contracts in the performance signal in order to characterize the optimality of the firm’s policy and because such linear contracts are common place. Another avenue for future work would be to allow for nonlinear contract in the performance signal. Given the presence of learning about match quality and the accumulation of experience, it is likely that the firm would also optimally choose to have different nonlinear compensation schedules for each tenure horizon. Moreover, the optimal compensation schedule may depend on all past performance signals, possibly through a sufficient statistic such as their average. Finally, the analysis in this and the preceding chapters has abstracted away from the impact of gender, race, and age differences on individual response to pay incentives. This is a topic also left for future research.
8 Appendix A

Assumption 1. (i) Suppose that ability $\theta^{13}$, $\theta \in R$, is time-invariant, continuous and has a probability density function $f_\theta$. (ii) The outside offer $\xi^+_t$, $\xi^+_t \in R$, is continuous, independent and identically distributed across tenure horizons $t$, where $t = 1, 2, \ldots$, independent from $\theta$, and has a probability density function $f_{\xi^+}$. (iii). The noise in the performance signal $\varepsilon_t$, $\varepsilon_t \in R$, is continuous, independent and identically distributed across tenure horizons $t$, independent from $\theta$, and has a probability density function $f_{\varepsilon}$.

Assumption 2. The performance signal $y_t$ is generated by the following technology:

$$y_t = \theta + g(t) + l_t + \varepsilon_t$$

where $g(t)$, $g(t) \in R_+$, represents the accumulation of firm-specific knowledge or experience and $l_t$ is effort, $l_t \in L \subset R_+$, where $L$ is compact. $g(t)$ is increasing and continuous.

Assumption 3. The belief at the beginning of $t$ is denoted as $\theta_t$ and is formed in a Bayesian way. Let the initial prior be $\theta_1$ and suppose that it has the same distribution as $\theta$.

Assumption 4. The worker is paid $w_t = \alpha_t + \beta_t y_t$, according to a linear compensation regime $R_t = (\alpha_t, \beta_t)'$, where $\alpha_t > 0$ and $\beta_t > 0$. Regime $R_t$ is said to be more generous than regime $R'_t$, $R_t > R'_t$, if both $\alpha_t > \alpha'_t$ and $\beta_t > \beta'_t$. The worker does not expect the compensation regime to change in the future.

Assumption 5. Workers are assumed to be risk-neutral with a utility function

$$u(R_t, l_t, t, \theta, \varepsilon_t) = \alpha_t + \beta_t (\theta + g(t) + l_t + \varepsilon_t) - \psi(l_t),$$

where $\psi(l_t)$ is strictly convex, increasing in effort, and $\psi(0) = 0$.

Assumption 6. $\varepsilon_t \sim N(0, \sigma^2)$, $\xi^+_t \sim N \left( \mu_{\xi^+}, \sigma^2_{\xi^+} \right)$ and $\theta_t \sim N \left( \mu_{\theta}, \sigma^2_{\theta} \right)$.

I maintain assumption 6 because normality of $\varepsilon_t$, $\xi^+_t$, and $\theta_t$ is imposed on the models taken to the data in Chapters 2 and 3. However, the statement of the optimal problem of

\[13\] In this section, I drop the individual subscript "i"
the worker, the proof of existence of a solution to the problem and its characterization can be stated more generally. Without normality, in addition to assumptions 1 to 5, one needs to impose that

**Assumption 6’** (i) \( \varepsilon_t \) and \( \theta \) are log-concave, and (ii) the sequence of noisy signals about ability \( \{y_{ik} - g(k) + l_{ik}\}_{k=1}^t \) is ordered in the sense of the likelihood ratio property.

Since \( \theta \) and \( l_t \) enter additively in the utility function, posterior beliefs do not depend on effort and optimal effort choice does not depend on beliefs and is function only of \( R_t, l(R_t) \). The assumptions on the disutility of effort imply the existence of a unique interior solution to the problem of choosing optimal effort. Furthermore, they imply that as the bonus rate \( \beta_t \) increases, optimal effort increases, too. By assumption 6 the posterior belief \( \theta_{it} \) is normally distributed for all \( t > 1 \) : \( \theta_t \sim N(\mu_t, \sigma_t^2) \), where

\[
\mu_t = (1 - K_t)\mu_{t-1} + K_t(y_{t-1} - l(R_{t-1}) - g(t - 1) - m(X_{t-1}))
\]

\[
\sigma_t^2 = \frac{\sigma_\varepsilon^2 \sigma_\theta^2}{\sigma_\theta^2 (t - 1) + \sigma_\varepsilon^2}
\]

\[
K_t = \frac{\sigma_\theta^2}{\sigma_\theta^2 (t - 1) + \sigma_\varepsilon^2}
\]

Note that precision of beliefs depends only on \( t \), so the average of the demeaned past signals is a sufficient statistic to characterize posterior beliefs. The posterior mean can be rewritten as

\[
\mu = k(t) \cdot \left( \frac{1}{t - 1} \sum_{k=1}^{t-1} (y_k - l(R_k) - g(k) - m(X_k)) \right) + (1 - k(t)) \mu_0
\]

where \( k(t) = (t - 1)K_t \). By these observations, the independence of \( \theta \) from \( \varepsilon_t \) and \( \xi_t^* \), and by the additivity of \( \varepsilon_t \) in the stochastic technology, the optimal problem of the worker can be formulated as (P)

\[
v(\mu_t, R_t, t) = \int \max(\xi_t^*, \alpha_t + \beta_t (\mu_t + g(t) + l(R_t)) - \psi(l(R_t)))
\]

\[
+ \delta \int v(\mu_{t+1}, R_t, t + 1) f(\mu_{t+1}|\mu_t, t) d\mu_{t+1} \right] f_{\xi_t^*}(\xi_t^*) d\xi_t^*,
\]

39
where \( f(\mu_{t+1}|\mu_t, t) \) is the conditional density of \( \mu_{t+1} \), given \( \mu_t \) and \( t \).

**Proposition 1.** Under Assumptions 1 - 6:

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

\[
A(\mu_t, R_t, t) = \{ l_t \in L \mid (P) \text{ holds.} \}
\]

is a continuous function.

ii. Optimal effort \( l(R_t) > l(R_t') \) if \( R_t > R_t' \).

iii. \( V(\mu_t, R_t, t) > V(\mu_t, R_t', t) \) if \( R_t > R_t' \), and \( V(\mu_t, R_t, t) \) increases \( \mu_t \).

**Proof of Proposition 1:**

**Part (i).** The sequence of posterior beliefs is continuous which also implies that \( f(\mu_{t+1}|\mu_t, t) \) is continuous\(^{14}\), so the proof of existence is reduced to a problem which can be solved using Blackwell (1965). Define the operator \( T \) by

\[
(Tw)(\mu_t, R_t, t) = \int \max\left[ \xi^*_t, (\alpha_t + \beta_t (\mu_t + g(t) + l(R_t)) - \psi(l(R_t)) \right] + \delta \int w(\mu_{t+1}, R_t, t + 1) f(\mu_{t+1}|\mu_t, t) d\mu_{t+1} | f(\xi^*_t, \xi_t^*) \, d\xi_t^*]
\]

Let \( C \) denote the set of bounded functions on \( P(\Theta) \). Under the supnorm metric, \( \|\cdot\| \), \( C \) 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 \( \delta \in (0, 1) \), such that \( T(w + c) \leq Tw + \delta c \), 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, \( C \).

\(^{14}\)See Lemma 1 and 2 in Easley and Kiefer (1988) to establish proof of continuity of the transitional kernel under Assumption 6' and the more general formulation of the worker’s problem in Chapter 2, Section 2.
(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 \( \delta \)

\[
T (w + c) = \int \max[\xi^*, \alpha_t + \beta_t(\mu_{it} + g(t) + l(R_t)) - \psi l(R_t)]
\]

\[
+ \delta w(\mu_{it+1}, R_t, t + 1) + \delta c f_{\xi^*}(\xi_t^*) d\xi_t^*
\]

\[
< \int \max[\xi^*, \alpha_t + \beta_t(\mu_{it} + g(t) + l(R_t)) - \psi l(R_t)]
\]

\[
+ \delta \int w(\mu_{it+1}, R_t, t + 1) f(\mu_{it+1}|\mu_t, t) d\mu_{it+1}] f_{\xi^*}(\xi_t^*) d\xi_t^* + \delta c
\]

\[
= Tw + \delta c
\]

(ii). As discussed above, the assumptions on the utility function and the production technology imply that there is a unique interior solution to the problem of choosing optimal effort. Again, by the assumptions of the model expected utility in the current period is continuous. Suppose that \( w(\mu_{it+1}, R_t, t + 1) \) is continuous, then

\[
(\alpha_t + \beta_t(\mu_{it} + g(t) + l(R_t)) - \psi l(R_t)) + \delta \int w(\mu_{it+1}, R_t, t + 1) f(\mu_{it+1}|\mu_t, t) d\mu_{it+1}
\]

is also continuous. The function \( \max(a, b) \) is continuous if \( a \) and \( b \) are continuous, and the integral over \( \xi_t^* \) is also continuous if \( \xi_t^* \) 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 \( V(\theta_t, R_t, t) \) exists. By the theorem of the maximum, the optimal policy correspondence \( A(\theta_t, R_t, t) \) is upper-hemicontinuous, and since the utility is concave in \( l_t \), the optimal policy is a continuous function.

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 \( (\beta, l_t) \), optimal effort \( l(R_t) > l(R'_t) \) if \( R_t > R'_t \).

Part (iii). Suppose that \( R_t > R'_t \) and \( V(\mu_{it+1}, R_t, t + 1) > V(\mu_{it+1}, R'_t, t + 1) \). Since \( \xi_t^* \) does not depend on \( R_t \), and \( l(R_t) > l(R'_t) \), the first part of the statement follows. Finally,
suppose that \( V(\mu_{t+1}, R_t, t+1) \) increases in \( \mu_{t+1} \); then, the integral of \( V(\mu_{t+1}, R_t, t+1) \) over the distribution of \( \mu_{t+1} \) conditional on \( \mu_t \) and \( t \) is also increasing in \( \mu_t \) because the conditional distributions of \( \mu_{t+1} \) are ordered in the sense of the likelihood ratio property with respect to \( \mu_t \). Similarly, expected utility for the current period increases in \( \mu_t \). Thus, \( V(\mu_t, R_t, t) \) increases in \( \mu_t \) in the sense of the likelihood ratio property. ■

9 Appendix B

Let the joint distribution of \( Y_{it} = (y_{i1}, ..., y_{it}) \) and \( S_{it} = (s_{i1}, ..., s_{it}) \), conditional on \( M_i = (M_{i1}, ..., M_{it}) \), \( M_{it} = (R_{it}, t, X_{it}, \mu_{it}) \), \( \theta_i \), and parameters \( \Theta_1 \), be given by \( F (Y_{it}, S_{it} | M_i, \theta_i, \Theta_1) \).

By the definition of conditional distribution:

\[
\begin{align*}
&f (Y_{it}, S_{it} | M_i, \theta_i, \Theta_1) \\
= &f_t (s_{it} | y_{it}, Y_{it-1}, S_{it-1}, M_i, \theta_i, \Theta_1) \cdot f_t (y_{it} | Y_{it-1}, S_{it-1}, M_i, \theta_i, \Theta_1) \cdot f_t (Y_{it-1}, S_{it-1}, M_i, \theta_i, \Theta_1)
\end{align*}
\]

Assumption 7 below plays a crucial role in deriving the likelihood and is justified by the model presented above.

**Assumption 7. Dynamic Completeness** Suppose that

\[
\begin{align*}
f_t (y_{it} | Y_{it-1}, S_{it-1}, M_i, \theta_i, \Theta_1) &= f_t (y_{it} | M_{it}, \theta_i, \Theta_1) \\
f_t (s_{it} | Y_{it}, S_{it-1}, M_i, \theta_i, \Theta_1) &= f_t (s_{it} | M_{it}, \theta_i, \Theta_1)
\end{align*}
\]

42
By Assumption 7, the conditional density becomes:

\[
\begin{align*}
    f(Y_{it}, S_{it}|M_i, \theta_i, \Theta_1) & = f_t(s_{it}|Y_{it}, M_{it}, \theta_i, \Theta_1) \cdot f_t(y_{it}|M_{it}, \theta_i, \Theta_1) \cdot f_{t-1}(s_{it-1}|Y_{it-1}, M_{it-1}, \theta_i, \Theta_1) \cdot f_{t-1}(y_{it-1}|M_{it-1}, \Theta_1) \cdot f(Y_{it-2}, S_{it-2}, M_i, \Theta_1) \\
    & = \ldots \\
    & = \prod_{k=1}^{t} \left[ f(s_{ik}|Y_{ik}, M_{ik}, \theta_i, \Theta_1) f(y_{ik}|M_{ik}, \theta_i, \Theta_1) \right]
\end{align*}
\]

By assumptions (i) and (ii) in the text of Chapter 3, 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

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

where

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

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

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

\[
f(Y_{it}, S_{it} = (1, \ldots, 0')|M_i, \theta_i, \Theta_1) \]

\[
= \Pr(s_{it} = 0|Y_{it-1}, M_{it}, \theta_i, \Theta_1) \prod_{k=1}^{t-1} \left[ \Pr(s_{ik} = 1|Y_{it-1}, M_{ik}, \theta_i, \Theta_1) f(y_{ik}|M_{ik}, \theta_i, \Theta_1) \right], \ t \geq 2
\]

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

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

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

Assumption 7 allows for the use of conditional MLE, so the likelihood for individual \(i\) is:

\[
l_i(\Theta_1|\theta_i, W_i)
\]

\[
= \prod_{t=1}^{T} \left( \varphi \left( \frac{y_{it} - g(t) - m(X_{it}) - l(R_{it}) - \theta_i}{\sigma} \right) \Phi \left( G \left( R_{it}, t, X_{it}, (1 - \lambda) \theta_i + \lambda \mu_{it} \right) \right) \right) \prod_{k=1}^{t} s_{ik}^{1 - \prod_{k=1}^{t} s_{ik}}
\]

where \(T\) is the last period for which the econometrician observes performance. Next, I integrate out \(\theta_i\):

\[
l_i(\Theta|W_i) = \int_{\Theta} l_i(\Theta_1|W_i, \theta_i) \cdot \varphi(\theta_i|W_i, \Theta_2) d\theta_i,
\]

where \(\Theta_2\) is a vector of parameters that govern the distribution of \(\theta_i\) and \(\Theta\) is a vector of all parameters in \(\Theta_1\) and \(\Theta_2\). 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

There are 18 moment conditions and 5 parameters to be estimated. The conditions are nonlinear which complicates identification. By the assumption on the approximation of \(G(\mu_{it}, R_{it}, t)\), the minimum distance problem has a solution. The following discussion addresses the uniqueness of that solution. The discount factor is identified from variation in \(G(\mu_{it}, R_{it}, t + 1)\) with
that is associated with changes in the precision of beliefs. The results from step 1 show that by $t = 12$ the accumulation of experience has come to an end. Therefore, conditional on the information available at $t$, $t > 12$,

$$U(\mu_{it+1}, R_{it}, t, \gamma, \psi) - U(\mu_{it}, R_{it}, t, \gamma, \psi) = 0.$$ 

Consequently,

$$G(\mu_{it+1}, R_{it}, t + 1) - G(\mu_{it}, R_{it}, t) = \delta \left[ E_{\mu_{it+1}}(\lambda(G(\mu_{it+1}, R_{it}, t + 1))) - E_{\mu_{it+2}}(\lambda(G(\mu_{it+2}, R_{it}, t + 2))) \right]$$

Conditional on the information available at $t$ and $R_{it} = R_{it+1}$, variation in $G(\mu_{it}, R_{it}, t)$ across periods originates from changes in the precision of posterior beliefs. Therefore, variation in the first differences on the left-hand and the right-hand side of the condition above identifies the discount factor.

Given $\delta$, $\sigma_{\xi^*}$ is identified from variation in $G(\mu_{it}, R_{it}, t + 1)$ that is associated with accumulated experience and variation in the means of the posterior beliefs. If $R_{it} = R_{it+1},$

$$U(\mu_{it+1}, R_{it}, t, \gamma, \psi) - U(\mu_{it}, R_{it}, t, \gamma, \psi) = \beta_{it} (\mu_{it+1} + g(t + 1) - \mu_{it} - g(t)).$$

Then,

$$G(\mu_{it+1}, R_{it}, t + 1) - G(\mu_{it}, R_{it}, t) = \frac{1}{\sigma_{\xi^*}} (U(\mu_{it+1}, R_{it}, t, \gamma, \psi) - U(\mu_{it}, R_{it}, t, \gamma, \psi))$$

$$+ \delta \left[ E_{\mu_{it+1}}(\lambda(G(\mu_{it+1}, R_{it}, t + 1))) - E_{\mu_{it+2}}(\lambda(G(\mu_{it+2}, R_{it}, t + 2))) \right]$$

Therefore, variation in the first differences of expected utility and in the first-difference in the left-hand side in the above condition identifies $\sigma_{\xi^*}$, with changes in beliefs and experience.
when the pay regime does not change variation in $H(\theta_{it}, R, t)$ originates from the accumulated experience and variation in beliefs, so $\sigma_{\xi}$ is identified from the ratio of the first-difference in

$$H(\theta_{it}, R, t) - \delta E_{\theta_{it+1}|\theta_{it}}(\lambda(H(\theta_{it}, R, t + 1))$$

and the first-difference in $U(\theta_{it}, R_{it}, t)$. Given $\delta$ and $\sigma_{\xi^*}$, the structural parameters $\gamma$ and $\psi$ are identified from $\gamma = \gamma(\psi, \Delta l)$ and from variation in $G(\mu_{it}, R_{it}, t)$ with $t$ that is associated with changes in the pay regime. Given $\delta, \sigma_{\xi^*}, \gamma, \psi$, the mean of the outside offer is identified from variation in $G(\mu_{it}, R_{it}, t)$ and the definition of the VNM utility.
References


gramming for factored continuous state Markov decision processes." In ICAPS-14.


