Habermalz, Steffen

Working Paper

Rational inattention and employer learning

IZA Discussion Papers, No. 5311

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Habermalz, Steffen (2010) : Rational inattention and employer learning, IZA Discussion Papers, No. 5311, Institute for the Study of Labor (IZA), Bonn

This Version is available at:
https://hdl.handle.net/10419/51921

DISCUSSION PAPER SERIES

IZA DP No. 5311

Rational Inattention and Employer Learning

Steffen Habermalz

November 2010

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Rational Inattention and Employer Learning

Steffen Habermalz
Northwestern University
and IZA

Discussion Paper No. 5311
November 2010

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 5311
November 2010

ABSTRACT

Rational Inattention and Employer Learning*

Research on employer learning has provided important insights into the dynamic process that determines individual wages, especially during the early part of a worker's career. However, the recent evidence on the absence of employer learning for college graduates by Arcidiacono et al. (2008) and results that economic conditions at labor market entry have persistent effects on wages (for example Oreopoulos et al. (2008)) cast doubt on the model's validity. This paper extends the employer learning model with the theory of rational inattention introduced by Sims (2006). In the model firms optimally allocate attention (=information processing capacity) to learning about the productivity of different worker groups. I find that firms allocate more attention to learning about the productivities of workers who have a higher impact on profits. Furthermore, firms learn about workers' productivities as quickly as possible. Taken together these results resolve the discrepancy between the data and the employer learning model.

NON-TECHNICAL SUMMARY

The theory of rational inattention basically states that a firm cannot attend to everything that affects profits perfectly. However, the firm will allocate more attention to those decisions that affect profits the most. This results in an intriguing situation in which the firm makes mistakes yet behaves rationally! This paper analyzes how a firm with rational inattention learns about the productivity of its workers. The theory predicts that the firm will devote more resources to more "important" workers and will try to learn as fast as possible. These two results are able to explain two very different observations. First, that there seems to be no employer learning for college graduates when the standard employer learning model is estimated for college graduates and high school graduates separately. Second, that there is a wage penalty for college graduates that graduate during a recession.

JEL Classification: D21, D22, J21, J24

Keywords: employer learning, rational inattention, endogenous information

Corresponding author:

Steffen Habermalz
Department of Economics
Northwestern University
2001 Sheridan Road
Evanston, Illinois 60208
USA
E-mail: s-habermalz@northwestern.edu

* I would like to express gratitude to Mirko Wiederholt who (irrationally?) devoted large amounts of attention to an infinite stream of my questions. I also benefited from discussions with Fabian Lange, Giorgio Primiceri and Matthias Fahn. The remaining errors are mine.

1 Introduction

Understanding the dynamic process in which an individual's wages are determined after entry into the labor force is one of the most important research questions in labor economics. Part of this research agenda has been the inquiry into the process by which employers learn about the productivities of their workers. The contributions of Farber and Gibbons (1996), Altonji and Pierret (2001) and Lange (2007) have made enormous strides in this direction and furthered our understanding significantly.

The employer learning model assumes that wages are determined competitively and therefore the following two predictions are warranted. First, if employer learning correctly models the behavior of firms one should expect the empirical pattern of employer learning to hold for different educational groups and certainly for college graduates. Second, in a competitive labor market economic conditions at the time of labor market entry should not have an effect on wages years later. Yet, recent research shows that the data do not seem to be consistent with these two predictions. Specifically, Oreopoulos et al. (2008) find persistent wage effects of graduating during a recession and Arcidiacono et al. (2008) find that the empirical pattern predicted by the employer learning model holds only for workers with a high school diploma and is absent for college educated workers. Together those results imply a failure of the employer learning model.

This paper develops a simple employer learning model that is consistent with both findings. The innovation of this paper is to demonstrate that the addition of the theory of rational inattention developed by Sims (2006) to the employer learning model resolves the inconsistencies discussed above. In the rational inattention (RI) framework it is assumed that all information is publicly available but that the amount of information that economic agents can process is limited. The agents are otherwise identical to the standard neoclassical agents and would, given infinite information processing capacity, make optimal decisions. As a consequence of this set-up the agent has to divide attention (=information processing capacity) between different sources of uncertainty that affect her optimal decision. The striking feature of this framework is that the agent is bound to make mistakes (due to imperfect information) while behaving rationally!

So far rational inattention has primarily been applied to macroeconomic topics. For example, in a recent paper Machkowiak and Wiederholt (2009) use the theory of rational inattention to explain

the empirical puzzle that the aggregate price level responds slowly to monetary shocks while firm-specific prices change frequently. In their model firms optimally devote more of their attention to firm-specific shocks because firm-specific shocks have a greater impact on the firm's profits than aggregate shocks. Therefore, prices respond quickly to firm-specific shocks but slowly to aggregate shocks. Although RI also seems to provide an intuitive way to model the behavior of economic agents on the micro level to date no study has done so.

I find that firms allocate more attention to learning about the productivity of those workers who have a higher impact on profits. Furthermore, firms attempt to learn about workers' productivity as quickly as possible. Taken together, I show that these results resolve the inconsistencies mentioned above. Intuitively, firms want to learn about "more important" workers as fast as possible and are content to learn less about workers with a lower impact on profits. The contributions of this paper are twofold. First, the paper shows that an employer learning model with rational inattention can explain two *unrelated* empirical results. Second, by resolving the aforementioned empirical inconsistencies the paper shows the potential of applying rational inattention to learning processes at the micro level. Clearly, employer learning is only one of many situations where a learning agent is faced with competing sources of uncertainty and the paper suggests that applying the theory of rational attention to other situations should be fruitful.

The rest of the paper is organized as follows. Section 2 introduces the theory of rational inattention in a general model of employer learning. Section 3 adds specificity to the model by extending it to 2 periods and connecting it to the earlier employer learning literature, specifically to Lange (2007). Section 4 provides empirical evidence that is consistent with the predictions of the model. Section 5 concludes.

2 A Model of Employer Learning under Rational Inattention

This section describes the behavior of a firm under rational inattention in a partial-equilibrium setting[1] Firms are assumed to follow a 3-step process when maximizing profits.

1. *Derivation of the perfect information solution:* Firms calculate the profit-maximizing decision

[1]This section follows section 2 in Machkowiak and Wiederholt (2009).

rule under perfect information. Consider a firm in the short-run that has the profit function $\pi(x, a_1, a_2)$. Here x is the choice variable (i.e. price or input ratios) and a_1 and a_2 are shocks. If the firm had perfect information (it would observe the realizations of the shocks) it would choose $x^* = x^*(a_1, a_2)$.

2. *Derivation of the expected value of the perfect information solution conditional on the signals the firm chooses to receive:* In this step uncertainty is introduced. The firm receives information about the two shocks in the form of two independent signals s_1 and s_2. It uses this information to calculate the expectation of the profit-maximizing value of the choice variable conditional on the signals it received $x^{\odot} = E(x^*|s_1, s_2)$.

3. *Minimizing the expected loss from deviating from the profit-maximizing value of x:* As will become clear below, under rational inattention the firm is able to choose the signals it wants to receive and thus the precision of the signals s_1 and s_2. If the firm had infinite information processing capacity it would be able to reduce uncertainty to zero and achieve $x^{\odot} - x^* = 0$. Given the information processing constraint formalized below the firm has to decide how much of its information processing capacity it will allocate to improving the precision of each signal. Thus the firm faces a constraint maximization problem with the constraint given by the amount of information that can be processed.

$$\min_{s_1,s_2} E[\pi(x^{\odot}, a_1, a_2) - \pi(x^*, a_1, a_2)]$$

subject to information constraint

I will now illustrate each of those three steps with a concrete example of a firm hiring labor under rational inattention.

2.1 Derivation of the perfect information solution

In this section I am investigating the effect of random shocks on the hiring decisions of a firm. Consider a firm that uses two types of labor (L_1 and L_2) in production. I am assuming a constant returns to scale production technology which implies that, given a level of output, the firm chooses the labor ratio $lr = \frac{L_1}{L_2}$. Further assume that there are two shocks (a_1 and a_2) that alter the

effective amount of each type of labor used. The shocks are assumed to be normally distributed and mutually independent:

$$a_1 \sim N(0, \sigma_{a_1}^2)$$
$$a_2 \sim N(0, \sigma_{a_2}^2)$$

I will model this as $e^{a_1} \cdot L_1$ and $e^{a_2} \cdot L_2$ entering the production function. In general then the profit function can be written as $\pi(lr, e^{a_1}, e^{a_2})$. Under perfect information the firm chooses the profit-maximizing labor ratio $lr^* = lr^*(a_1, a_2)$. To analyze the labor ratio that maximizes expected profits under general conditions I will use a second-order Taylor approximation of the profit function in log deviations. The approximation will be taken around the non-stochastic solution (i.e. the solution in which $a_1 = a_2 = 0$ and $lr = lr^*(0, 0)$). We first rewrite the variables in the profit function in terms of log deviations from their non-stochastic equilibrium values [2]

$$\pi(\frac{L_1}{L_2}, e^{a_1}, e^{a_2}) = \pi(\frac{\overline{L}_1 \cdot e^{l_1}}{\overline{L}_2 \cdot e^{l_2}}, e^{a_1}, e^{a_2}) = \pi(\frac{\overline{L}_1}{\overline{L}_2} \cdot e^{\Delta l}, e^{a_1}, e^{a_2})$$

$$\text{where } x = \ln X - \ln \overline{X},\ \Delta l = l_1 - l_2 \text{ and } X = \overline{X} \cdot e^x$$

Non-stochastic equilibrium values are indicated by a bar [3]. This leads to a "new" profit function in terms of variables that are log-deviations around the non-stochastic equilibrium.

$$\widetilde{\pi}(\Delta l, a_1, a_2)$$

The second-order Taylor expansion of this function yields

[2] I am using a log-quadratic approximation (as opposed to a quadratic approximation) so that that the resulting objective function is quadratic in normally distributed variables.

[3] Note that $e^{a_i} = \overline{e^{a_i}} \cdot e^{\ln e^{a_i} - \ln \overline{e^{a_i}}}$

$$\begin{aligned}\widehat{\pi}(\Delta l, a_1, a_2) &= \widetilde{\pi}_1 \cdot \Delta l + \widetilde{\pi}_2 \cdot a_1 + \widetilde{\pi}_3 \cdot a_2 \\ &+ \frac{1}{2} \cdot \widetilde{\pi}_{11} \cdot \Delta l^2 + \frac{1}{2} \cdot \widetilde{\pi}_{22} \cdot a_1^2 + \widetilde{\pi}_{33} \cdot a_2^2 \\ &+ \widetilde{\pi}_{12} \cdot \Delta l \cdot a_1 + \widetilde{\pi}_{13} \cdot \Delta l \cdot a_2 \\ &+ \widetilde{\pi}_{23} \cdot a_1 \cdot a_2\end{aligned}$$

where $\widetilde{\pi}_i$ is the derivative of the profit function with respect to its i^{th} argument and $\widetilde{\pi}_{ij}$ are second derivatives. To find the perfect information solution we maximize this function with respect to Δl under the assumption that both shocks are known. Note that $\widetilde{\pi}_1 = 0$ because the firm optimizes at the non-stochastic solution. This yields

$$\Delta l^* = \frac{\widetilde{\pi}_{12}}{|\widetilde{\pi}_{11}|} \cdot a_1 + \frac{\widetilde{\pi}_{13}}{|\widetilde{\pi}_{11}|} \cdot a_2 \tag{1}$$

Equation (1) represents the profit-maximizing labor ratio under perfect information.

2.2 Derivation of the conditional expected value of the profit-maximizing labor ratio

Firms receive the signals about the shocks a_1 and a_2

$$s_1 = a_1 + \epsilon_1$$
$$s_2 = a_2 + \epsilon_2$$

where $\epsilon_1 \sim N(0, \sigma_{\epsilon_1}^2)$ and $\epsilon_2 \sim N(0, \sigma_{\epsilon_2}^2)$ are mutually independent and independent of a_1 and a_2. Given signals s_1 and s_2 and the quadratic profit function the firm chooses a labor ratio equal to the

conditional expected value of the profit-maximizing labor ratio given s_1 and s_2.

$$\begin{aligned}
\Delta l^{\odot} &= E(\Delta l^{*}|s_1, s_2) \\
&= \frac{\widetilde{\pi}_{12}}{|\widetilde{\pi}_{11}|} \cdot E(a_1|s_1) + \frac{\widetilde{\pi}_{13}}{|\widetilde{\pi}_{11}|} \cdot E(a_2|s_2) \qquad (2) \\
&= \frac{\widetilde{\pi}_{12}}{|\widetilde{\pi}_{11}|} \cdot \frac{\sigma^2_{a_1}}{\sigma^2_{a_1} + \sigma^2_{\epsilon_1}} \cdot (a_1 + \epsilon_1) + \frac{\widetilde{\pi}_{13}}{|\widetilde{\pi}_{11}|} \cdot \frac{\sigma^2_{a_2}}{\sigma^2_{a_2} + \sigma^2_{\epsilon_2}} \cdot (a_2 + \epsilon_2)
\end{aligned}$$

The third line of Equation (2) follows from the fact that the signal and the shocks are jointly normally distributed [4]. The expected loss due to imperfect information equals

$$\begin{aligned}
\textit{Expected Loss} &= E[\widetilde{\pi}(\Delta l^{*}, a_1, a_2) - \widetilde{\pi}(\Delta^{\odot}, a_1, a_2)] \\
&= \frac{1}{2} \cdot |\widetilde{\pi}_{11}| \cdot E(\Delta l^{\odot} - \Delta l^{*})^2 \qquad (3)
\end{aligned}$$

See Appendix 2 for a derivation. Equation (3) shows that the loss due to imperfect information depends on the expected squared deviation of the labor ratio from its profit-maximizing value. Equation (3) can be written as

$$\begin{aligned}
E(\Delta l^{\odot} - \Delta l^{*})^2 &= \left(\frac{\widetilde{\pi}_{12}}{|\widetilde{\pi}_{11}|}\right)^2 \cdot \left[\sigma^2_{\epsilon_1} \cdot \left(\frac{\sigma^2_{a_1}}{\sigma^2_{a_1} + \sigma^2_{\epsilon_1}}\right)^2 + \sigma^2_{a_1} \cdot \left(\frac{\sigma^2_{\epsilon_1}}{\sigma^2_{a_1} + \sigma^2_{\epsilon_1}}\right)^2\right] \\
&+ \left(\frac{\widetilde{\pi}_{13}}{|\widetilde{\pi}_{11}|}\right)^2 \cdot \left[\sigma^2_{\epsilon_2} \cdot \left(\frac{\sigma^2_{a_2}}{\sigma^2_{a_2} + \sigma^2_{\epsilon_2}}\right)^2 + \sigma^2_{a_2} \cdot \left(\frac{\sigma^2_{\epsilon_2}}{\sigma^2_{a_2} + \sigma^2_{\epsilon_2}}\right)^2\right] \qquad (4)
\end{aligned}$$

Note that due to the independence assumptions all cross-products have zero expectations and drop out. The firm then minimizes (4) subject to an information processing constraint that will be introduced in the next section

[4] If two random variables are jointly normally distributed then $E(Y|X) = \mu_Y + \frac{\sigma_{XY}}{\sigma^2_X} \cdot (X - \mu_X)$

2.3 The information Processing Constraint: Mutual Information

The extent to which the firm faces uncertainty is different before and after the firm receives the signals. Mutual Information measures this reduction in uncertainty. It is limited by the firm's information processing capacity. Formally Mutual Information is defined as

$$I(X;Y) = H(Y) - H(Y|X)$$

where $H(Y)$ denotes the entropy of random variable X and $H(Y|X)$ denotes the conditional entropy of random variable Y given X. Entropy is a measure of uncertainty. Conditional entropy is a measure of conditional uncertainty. Thus mutual information is a measure of how much information one random variable contains about another. Or in other words, how much does knowing the outcome of one random variable (the signal) reduce uncertainty about the outcomes another (the state variable)? For a discrete random variable Y the entropy $H(Y)$ is defined as

$$H(Y) = -\sum_{y} p(y) \cdot \log_2 p(y)$$

where $p(y)$ is the probability distribution of random variable Y. For example, consider the entropy of a Bernoulli random variable for success probabilities between 0 and 1. Entropy equals zero when the success probability is either zero or one which corresponds to the case of no uncertainty. Intuitively entropy is maximized at at success probability p=0.5. Conditional entropy is the uncertainty of a random variable Y after observing random variable X. Formally,

$$H(Y|X) = -\sum_{x} p(x) \sum_{y} p(y|x) \cdot \log_2 p(y|x)$$

Therefore, mutual information represents the reduction in uncertainty in random variable Y due to observing random variable X. In order to be able to work with mutual information I have to specify a distribution for the random variables X and Y. So far, most of the literature is working with the multivariate normal distribution [5]. I do the same.

[5]For an exception see Sims (2003). Machkowiak and Wiederholt (2009) also investigate the non-normal case.

2.4 Entropy and Mutual Information under Normality

The mutual information for two random variables X and Y that are jointly normally distributed equals

$$I(Y;X) = H(Y) - H(Y|X) = \frac{1}{2} \cdot \log_2 \left(\frac{\sigma_Y^2}{\sigma_{Y|X}^2} \right) \tag{5}$$

See Appendix 1. Here $\sigma_{Y|X}^2$ is the conditional variance of Y given X. For our current application mutual information is written as

$$\begin{aligned} I(a_1, a_2; s_1, s_2) &= H(a_1, a_2) - H(a_1, a_2|s_1, s_2) \\ &= H(a_1) - H(a_1|s_1) + H(a_2) - H(a_2|s_2) \\ &= I(a_1; s_1) + I(a_2; s_2) \\ &= \frac{1}{2} \cdot \log_2 \left(\frac{\sigma_{a_1}^2}{\sigma_{a_1|s_1}^2} \right) + \frac{1}{2} \cdot \log_2 \left(\frac{\sigma_{a_2}^2}{\sigma_{a_2|s_2}^2} \right) \leq \kappa \end{aligned} \tag{6}$$

The second line follows because of the independence assumption across signals and shocks, see Cover and Thomas (2006). The fourth line makes use of the fact that signals and shocks for each worker group are jointly normally distributed. Equation (6) tells us that the overall reduction in uncertainty is (less or) equal to the sum of the reductions in uncertainty about the two shocks.

2.5 The Firm Problem

Equations (4) and (6) jointly imply that the firm's problem can be written as

$$\begin{aligned} \min_{s_1, s_2} E(\Delta l^{\odot} - \Delta l^*)^2 &= \left(\frac{\widetilde{\pi}_{12}}{|\widetilde{\pi}_{11}|} \right)^2 \cdot \left[\sigma_{\epsilon_1}^2 \cdot \left(\frac{\sigma_{a_1}^2}{\sigma_{a_1}^2 + \sigma_{\epsilon_1}^2} \right)^2 + \sigma_{a_1}^2 \cdot \left(\frac{\sigma_{\epsilon_1}^2}{\sigma_{a_1}^2 + \sigma_{\epsilon_1}^2} \right)^2 \right] \\ &+ \left(\frac{\widetilde{\pi}_{13}}{|\widetilde{\pi}_{11}|} \right)^2 \cdot \left[\sigma_{\epsilon_2}^2 \cdot \left(\frac{\sigma_{a_2}^2}{\sigma_{a_2}^2 + \sigma_{\epsilon_2}^2} \right)^2 + \sigma_{a_2}^2 \cdot \left(\frac{\sigma_{\epsilon_2}^2}{\sigma_{a_2}^2 + \sigma_{\epsilon_2}^2} \right)^2 \right] \end{aligned}$$

$$subject\ to\ \ I(a_1, a_2; s_1, s_2) = \underbrace{\frac{1}{2} \cdot \log_2 \left(\frac{\sigma^2_{a_1}}{\sigma^2_{a_1|s_1}} \right)}_{\kappa_1} + \underbrace{\frac{1}{2} \cdot \log_2 \left(\frac{\sigma^2_{a_2}}{\sigma^2_{a_2|s_2}} \right)}_{\kappa_2} \leq \kappa$$

The constraint shows that we can partition the total amount of uncertainty reduction into uncertainty reduction with respect to the first shock (κ_1) and the second shock (κ_2). In addition, the equation shows that the firm faces the constraint $\kappa_1 + \kappa_2 \leq \kappa$.

Note that we can rewrite the first part of the constraint as

$$\begin{aligned} \kappa_1 &= \frac{1}{2} \log_2 \left(\frac{\sigma^2_{a_1}}{\sigma^2_{a_1|s_1}} \right) \\ &= \frac{1}{2} \log_2 \left(\frac{\sigma^2_{a_1} + \sigma^2_{\epsilon_1}}{\sigma^2_{\epsilon_1}} \right) \qquad (7) \\ \Rightarrow 2^{-2 \cdot \kappa_1} &= \frac{\sigma^2_{\epsilon_1}}{\sigma^2_{a_1} + \sigma^2_{\epsilon_1}} \end{aligned}$$

Equation (7) is fundamental as it shows that there is an inverse relationship between the amount of attention the firm devotes to the first shock (κ_1) and the variance of the signal (through lowering $\sigma^2_{\epsilon_1}$). If the firm had infinite processing capacity $\sigma^2_{\epsilon_1}$ would be driven to zero and the firm would receive a signal that matches the true realization of the shock.

$$\lim_{\kappa_1 \to \infty} \left(\frac{\sigma^2_{\epsilon_1}}{\sigma^2_{a_1} + \sigma^2_{\epsilon_1}} \right) = 0$$

Equation (7) also implies that we can substitute the information processing constraint into the objective function to get an unconstrained minimization problem in κ_1. This yields

$$\min_{\kappa_1} \ E(\Delta l^{\odot} - \Delta l^*)^2 = \left(\frac{\widetilde{\pi}_{12}}{|\widetilde{\pi}_{11}|} \right)^2 \cdot 2^{-2 \cdot \kappa_1} \cdot \sigma^2_{a_1} + \left(\frac{\widetilde{\pi}_{13}}{|\widetilde{\pi}_{11}|} \right)^2 \cdot 2^{-2 \cdot (\kappa - \kappa_1)} \cdot \sigma^2_{a_2} \qquad (8)$$

Equation (8) shows that, given infinite information processing capacity ($\kappa = \infty$), the firm could make maximum profits as it would eliminate the difference between the optimal and the actual decision. It also shows that the overall variance has two sources and that the firm faces a trade-off when allocating attention. Minimizing this loss function with respect to κ_1 yields

$$\kappa_1 = \frac{1}{2} + \frac{1}{4} \cdot \ln(x) \ \ where\ x = \left(\frac{\widetilde{\pi}_{12}}{\widetilde{\pi}_{13}} \right)^2 \cdot \frac{\sigma^2_{a_1}}{\sigma^2_{a_2}} \qquad (9)$$

$$\kappa_1^* = \begin{cases} \kappa & \text{if } \ln(x) \geq 2^{\kappa} \\ \frac{1}{2} + \frac{1}{4}\ln(x) & \text{if } 2^{\kappa} < ln(x) < 2^{-\kappa} \\ 0 & \text{if } \ln(x) \leq 2^{-\kappa} \end{cases}$$

Equation (9) is the main result of the general model. It shows that the allocation of attention to the first shock depends on the relative variability of the two shocks. The firm will allocate attention to the shock that affects the profit-maximizing labor ratio the most. Equation (9) also shows that the relative share of the variability that is attributable to each shock will vary with the relative variances of the shocks $\left(\frac{\sigma_{a_1}^2}{\sigma_{a_2}^2}\right)$ and with the specific way in which the shocks affects the labor ratio $\left(\frac{\widetilde{\pi}_{12}}{\widetilde{\pi}_{13}}\right)$. I now investigate the effect the latter derivative ratio on the allocation of attention. At this point a comment is warranted because one might ask the question Wouldn't the whole model be simpler if one would just attach a cost to reducing the variance of the signals and dispense with the rest of the complex machinery? Yes, the model would be simpler but it would also miss one of the main points of rational inattention which is that the agent *chooses* the signal. The signal is endogenous. The particular signal that we work with is optimal in the above model but it's simplicity relies partially on the constant state variables that the firm wants to learn about[6]. Rational inattention is best thought of as a state-space representation of a process in which the agent chooses the observation equation given the informational constraint. In addition, the above model assumes that a exogenously given amount of attention is available to the firm. Again, this simplification is made because the paper focuses on the main aspect of rational inattention, the allocation of attention between activities. Machkowiak and Wiederholt (2009) extend the model to make the total amount of attention a choice variable and show that the main results are not affected by this.

[6]The signal in the general form of (s=true state + noise) is optimal as long as the optimal decision (labor ratio) is normally distributed and the objective function is quadratic, see Machkowiak and Wiederholt (2009) for more details.

3 Model Extensions

3.1 Derivation of the Ratio of Second Derivatives

In order to evaluate the impact of the derivative ratio $\left(\frac{\widetilde{\pi}_{12}}{\widetilde{\pi}_{13}}\right)$ on the amount of attention that a firm devotes to each shock I assume that the firm produces output using a CES production function. I also assume that the firm sells its output at a given price P and hires labor a constant wages. Then the profit-function can be written as

$$\pi = P \cdot \left[\delta \cdot (e^{b \cdot a_1} \cdot L_1)^{-\rho} + (1-\delta) \cdot (e^{a_2} \cdot L_2)^{-\rho}\right]^{-\frac{1}{\rho}} - w_1 \cdot L_1 - w_2 \cdot L_2$$

where a_1 and a_2 are shocks to the effective labor input of the two types of labor as defined above, L_1 and L_2 are the quantities of labor the firm hires and are w_1 and w_2 the respective wages. The (constant) elasticity of substitution between the two types of labor is equal to $\frac{1}{1+\rho}$. For simplicity think of the firm hiring two workers drawn at random from two groups with known productivity distributions. In this case the firm maximizes profits by adjusting the hours of work of each worker. The parameter $b > 1$ represents the sensitivity of effective labor input to shocks and is restricted to equal one for the second type of labor. This is meant to imply that the first type of labor is of higher quality or more skilled and thus has a higher influence on profits. Imagine a firm employing janitors and CEOs. It is evident that the same percentage deviation in productivity of the CEO would have a greater effect on profit. The production function exhibits constant returns to scale which implies that, given output, the firm cannot choose L_1 and L_2 independently. It therefore chooses the labor ratio ($lr = \frac{L_1}{L_2}$). Therefore I use profit-per-unit by dividing the profit function by output.

$$\pi_Y = \frac{\pi}{Y} = P - \left[\delta \cdot \left(\frac{w_1}{e^{b \cdot a_1}}\right)^{\rho} + (1-\delta) \cdot \left(\frac{w_1}{e^{a_2}}\right)^{\rho} \cdot lr^{\,\rho}\right]^{\frac{1}{\rho}} - \left[\delta \cdot \left(\frac{w_2}{e^{b \cdot a_1}}\right)^{\rho} \cdot lr^{-\rho} + (1-\delta) \cdot \left(\frac{w_2}{e^{a_2}}\right)^{\rho}\right]^{\frac{1}{\rho}}$$

The second derivatives in the solution in the rational inattention model are taken from the 2^{nd} order Taylor approximation of the profit-function in which the variables are in log-deviations from the non-stochastic equilibrium. Therefore I write the profit function as

$$\widetilde{\pi}_Y = P \quad - \quad \left[\delta \cdot \left(\frac{w_1}{e^{b \cdot a_1}}\right)^{\rho} + (1-\delta) \cdot \left(\frac{w_1}{e^{a_2}} \cdot \frac{\overline{L}_1}{\overline{L}_2}\right)^{\rho} \cdot e^{\rho \cdot \Delta l}\right]^{\frac{1}{\rho}}$$
$$- \quad \left[\delta \cdot \left(\frac{w_2}{e^{b \cdot a_1}} \cdot \frac{\overline{L}_2}{\overline{L}_1}\right)^{\rho} \cdot e^{-\rho \cdot \Delta l} + (1-\delta) \cdot \left(\frac{w_2}{e^{a_2}}\right)^{\rho}\right]^{\frac{1}{\rho}}$$

where $\Delta l = l_1 - l_2$. Taking the derivative with respect to Δl , solving for $\left(\frac{\overline{L}_2}{\overline{L}_1}\right)$ and taking logs yields

$$\ln\left(\frac{\overline{L}_1}{\overline{L}_2}\right)^{*} = \frac{1}{1+\rho} \cdot \ln\left(\frac{\delta}{1-\delta}\right) + \frac{1}{1+\rho} \cdot \ln\left(\frac{w_2}{w_2}\right) - \frac{1}{1+\rho} \cdot (b \cdot a_1 - a_2) \tag{10}$$

This derivative is equal to the first derivative of the 2^{nd} order Taylor approximation of the profit function with respect to Δl (See Appendix 3). This enables us to easily find the desired ratio by differentiating Equation (10) with respect to a_1 and a_2 and taking the ratio. Thus the desired derivative ratio equals

$$\frac{\widetilde{\pi}_{12}}{\widetilde{\pi}_{13}} = \frac{-\frac{1}{1+\rho} \cdot b}{\frac{1}{1+\rho}} = -b$$

This means we can rewrite equation (9) as

$$\kappa_1^{t=1} = \frac{1}{2} + \frac{1}{4} \cdot \ln\left(b^2 \cdot \frac{\sigma_{a_1}^2}{\sigma_{a_2}^2}\right) \tag{11}$$

This implies that the firm will allocate more attention to the workers that influence profits more.

3.2 Dynamic Solution in a 2-period model

The model outlined above is easily interpreted as a model in which firms allocate attention to find out about the (constant) productivity of their workers. Since employer learning is a dynamic process we also need to know how firms will allocate attention over time. This section develops a 2-period model of employer learning under rational inattention and relates it to the model put forward by Altonji and Pierret (2001) and Lange (2007). The model will be analyzed under two different

scenarios. First, to investigate the allocation of attention between worker groups *within* a period I assume that the amount of attention in each time period is constant. Second, to investigate the allocation of attention *between* periods I assume that the amount of attention allocated to a specific worker group is constant. Taken together the two scenarios characterize the dynamic allocation of attention.

Result 1: The firm will allocate more attention to workers with a higher impact on profits Let the total amount of attention available for allocation be equal to $\kappa = \kappa_1^{t=1} + \kappa_1^{t=2} + \kappa_2^{t=1} + \kappa_2^{t=2}$. This section assumes that $\kappa_1^{t=1} + \kappa_2^{t=1} = \kappa^{t=1}$. In other words the firm cannot "transfer" attention between periods and faces the same static problem each period. The results here are similar to the static problem described above. The optimal amount of attention allocated to workers group one is equal to

$$\kappa_1 = \frac{1}{2}\kappa^{t=1} + \frac{1}{4}\ln\left[\left(\frac{\widetilde{\pi}_{12}}{\widetilde{\pi}_{13}}\right)^2 \frac{\sigma_{a_1}^2}{\sigma_{a_2}^2}\frac{(1+2^{-\kappa_1^{t=2}})}{(1+2^{-\kappa_2^{t=2}})}\right]. \tag{12}$$

See Appendix 4. The firm allocates more attention to the type of worker that impacts profits more. The solution is almost identical to the static model the only difference being that the amount of attention in the second period is a determinant of the amount allocated in the first.

Result 2: The firm will try to transfer all attention to the first period In this scenario the firm is restricted to use a constant total amount of attention per worker group over time: $\kappa_1^{t=1} + \kappa_1^{t=2} = \kappa_1$ and can allocate attention across time periods but not between workers. The optimal amount of attention the firm allocates to worker group *1* in the first period is

$$\kappa_1^{t=1} = \kappa_1 \tag{13}$$

See appendix 4. The firm uses all available attention in the first period. The intuition for this result is based on the fact that the productivity of a worker (the state variable) is constant over time. Hence, when the firm has the option to find out about the true productivity it wants to do so sooner than later.

3.3 Summary of 2-period model

The two-period employer learning model shows that the firm allocates more attention to the workers that impact profits more and that the firm will try to learn the workers' true productivity as fast as it can by allocating as much attention to the first period as possible. Yet the model presented above is very general. The next section relates the findings to the employers learning model put forward by Farber and Gibbons (1996) and fully developed by Altonji and Pierret (2001) and specifically Lange (2007).

3.4 Relationship between Lange (2007) and Employer Learning with Rational Inattention

Lange (2007) develops the model of Altonji and Pierret (2001) further. In his model employers use all observable information to form an initial guess about the productivity of a worker, $E(\widetilde{X}|Edu, q)$, where $\widetilde{X}$ represents the stationary part of the worker's productivity [7], Edu represents her level of education and q represents characteristics observable to the firm (but not the econometrician). In subsequent periods the employer receives signals about the worker's productivity in the following form, $y_t = \widetilde{X} + \epsilon_t$ where $\epsilon_t \sim N(0, \sigma_\epsilon^2)$ and independent of $\widetilde{X}$. Lange (2007) shows that the employer's conditional expectation of the worker's productivity (which are equal to wages) and its conditional variance evolve according to the following formulas.

$$E_t(\widetilde{X}|Edu, q, y_1, y_2, .., y_t) = (1 - \theta_t) \cdot E(\widetilde{X}|Edu, q) + \theta_t \cdot \left(\frac{1}{t} \cdot \sum_{x=0}^{t} y_x \right) \tag{14}$$

$$V_t(\widetilde{X}|Edu, q, y_1, y_2, .., y_t) = \frac{1}{\frac{1}{\sigma_0^2} + \frac{t}{\sigma_\epsilon^2}} \tag{15}$$

Here σ_0^2 represent the variance of the initial guess ($E(\widetilde{X}|Edu, q)$) and σ_ϵ^2 represents the variance of the noise part of the signal. The formula shows that the conditional expectation of the worker's productivity given received signals in each period is a weighted average of the initial guess and new information that the firm receives via the signals. Over time the coefficient θ_t converges to 1

[7] For simplicity, let the non-stationary part of the worker's productivity, that is usually modeled as deterministic and independent of the stationary part, be zero.

implying that the worker's true productivity is revealed. Lange (2007) also shows that the coefficient θ_t is equal to

$$\theta_t = \frac{t \cdot K_1}{1 + (t-1) \cdot K_1} \text{ where } K_1 = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_\epsilon^2}.$$

Lange (2007) calls K_1 the speed of employer learning. Introducing rational inattention into the model enables me to rewrite the change in the weights (θ_t) in the conditional expectations formula in terms of the attention the firm devotes to the worker. This yields

$$\begin{aligned} \Delta\theta_t = \theta_t - \theta_{t-1} &= -\%\Delta V_t \cdot (1 - \theta_{t-1}) \\ &= (1 - 2^{-2\kappa 1}) \cdot (1 - \theta_{t-1}). \end{aligned} \tag{16}$$

See appendix 5. Here $\%\Delta V_t$ represents the percentage change of the conditional variance in equation (15) due to the arrival of a signal in period t. Since the two conditional variances depend on the amount of attention allocated equation (16) identifies the direct link between employer learning with rational inattention and the employer learning model in Lange (2007)[8]. A second subtle but important impact of rational inattention in the employer learning model concerns the firm's initial guess about the productivity of a worker, $E(\widetilde{X}|Edu, q)$ which is based on a fixed information set in the earlier employer learning models. However, under rational inattention there is no clear division between variables observable and unobservable to the employer. The only distinction we can make is between variables that the employer *chooses to observe* and variables that the employer does not observe[9]. To make this distinction clear let's rewrite the initial guess as

$$E_t[\widetilde{X}|I(\kappa_t)] \tag{17}$$

where $I(\kappa_t)$ represents the information set of the employer at the time when the conditional expectation about the true productivity of a worker is formed given a certain level of attention. The point I want to make here is that the size of $I(\kappa_t)$ is not exogenous but depends on the amount of attention allocated to learning about the workers. I will show in the empirical section that this distinction is very helpful in aligning the model with recent empirical facts. The following two

[8]This implies that the speed of employer learning might not be constant if the firm allocates varying amounts of attention to learning about the true productivity of a worker over time

[9]Either by choice or because of the information processing constraint is binding

sections present empirical evidence that is consistent with the predictions of the employer learning model with rational inattention.

4 Empirical Evidence

This section presents empirical results of other studies to show that the employer learning model with rational inattention is consistent with results from related *and* seemingly unrelated studies. I will first show that the model is predicts the results in Arcidiacono et al. (2008) that were hard to reconcile with the employer learning framework in general. I will then show that the model is also consistent with the results in the literature on the wage-effects of graduating during a recession.

4.1 Empirical Evidence I: Arcidiacono et al. (2008)

Farber and Gibbons (1996) and Altonji and Pierret (2001) develop a model of employer learning which makes predictions about two distinct kinds of variables. The first category includes all variables that the firm and the econometrician observe (i.e. education). Their model predicts that the importance of these variables is constant (Farber and Gibbons (1996)) or non-increasing (Altonji and Pierret (2001)) over time as new information about the actual productivity of workers becomes available. The second category includes variables that correlate with the true productivity of a worker and are observed by the econometrician only (unobserved by the employer). Those variables should become more important in a wage equation as the firm acquires more information about the worker. Altonji and Pierret (2001) are able to confirm the models prediction using the following wage equation estimated with NLSY79 data.

$$\begin{aligned} \ln wage_{it} &= \beta_0 + \beta_{Edu} \cdot Edu_i + \beta_{EduX} \cdot Edu_i \cdot X_{it} \\ &+ \beta_{AFQT} \cdot AFQT_i + \beta_{AFQTX} \cdot AFQT_i \cdot X_{it} + other + \epsilon_{it} \end{aligned} \tag{18}$$

where AFQT stands for standardized scores on the Armed Forces Qualification Test administered to all individuals in the survey, Edu is years of education and X actual labor market experience. Their model of employer learning predicts that $\beta_{Edu \cdot X} < 0$ and $\beta_{AFQT \cdot X} > 0$. They estimate the above equation and are able to confirm these predictions. Arcidiacono et al. (2008) re-estimate

(18) separately for high school and college graduates. Their results in table 2 (reproduced from their paper) show that employer learning as defined by the earlier literature takes place among high school graduates but does not matter much for college graduates[10]. In the high school regression the coefficient on the individual's AFQT score (β_{AFQT}) is low and its increase over time (β_{AFQTX}) is large. The reverse is true for college graduates where coefficient on the AFQT score high and does not grow much over time. Arcidiacono et al. (2008) take this as evidence against the employer learning model and suggest that workers instead somehow reveal their productivity.

The patterns estimated in Arcidiacono et al. (2008) are remarkably consistent with the rational inattention model developed above. Results 1 and 2 imply that the firm will allocate more attention to the workers that have a higher impact on profits and will also try to transfer all attention to the first period. It is reasonable to assume that college educated workers have a higher impact on the profit-maximizing decision of the firm than workers with a high school degree. Therefore the firm accumulates a lot of information on college educated workers and little on high school educated workers. This implies that there is little left to learn about the true productivities of college educated workers and a lot to learn about the true productivity of high school educated workers. The estimating in Arcidiacono et al. (2008) shows exactly this pattern and the employer learning model with rational inattention is consistent with their empirical results. Of course, one could ask why after transferring all attention to the first period, the firm still learns about the productivity of workers at all in the subsequent periods. There are two possible scenarios.

1. *Limit to transferring attention over time.* Simply put it might not be possible to transfer all attention to the first period.

2. *Non-constant productivity of workers.*The model above is a special case as it assumes that worker productivity is constant. It is therefore perfectly correlated over time which gives the firm an incentive to learn fast. In other scenarios in which this is not true the firm intuitively has less of an incentive to transfer all attention to the first period. In those cases the firm will keep on learning over time and the differential rates at which firms learn about college educated and high school educated workers stem from the fact that there is little left to learn about the former but lots learn about the latter.

[10]Similar results are obtained in Bauer and Haisken-DeNew (2001). These authors find evidence of employer learning for blue collar workers but not for white collar workers.

3. *Free signals.* Even if the firm is able to allocate all attention to the first period it is conceivable and very likely that there are some signals of productivity that the firm can obtain for free. After all, the workers work for the firm every day and the firm would have a hard time not to learn anything about their employees. Again the empirical pattern in Arcidiacono et al. (2008) would result from the differences in residual information that is available about the two groups.

4.2 Empirical Evidence II: Graduating during a Recession

There exists a growing literature on the effects of initial labor market experiences on subsequent wages. In the latest installment Oreopoulos et al. (2008) use Canadian data to estimate the effect of graduating from college during a recession on wages later in the worker's career. They show that these effects are substantial and also non-homogeneous across the population of college graduates. They conjecture that the underlying driving force of this effect is the lower quality of jobs that are available during a recession. Kahn (2009) provides further evidence that graduation during a recession has lasting effects of an individual's wages later in life. In addition she finds that labor market conditions at the time of graduation also affect the type of job taken by graduates. Baker et al. (1994) analyze data from a single firm and find that persuasive evidence of cohort effects indicating that those cohorts that had a higher starting wage are able to maintain that advantage over time. Finally, Beaudry and DiNardo (1991) find that the past unemployment rates are statistically related to current wages.[11] Most of the authors of these studies attempt to provide evidence for/against a particular theory or try to distinguish between theories of the labor market. For example, Oreopoulos et al. (2008) forward a search explanation in which the intensity of job search is dependent on the worker's age and the worker accumulates firm-specific human capital while Beaudry and DiNardo (1991) forward an explanation based on the theory of implicit contracts (see Rosen (1985) for a survey). I do not attempt to distinguish between those explanations. My goal in this section to convince the reader that an employer-learning model with rational inattention

[11] Other studies are Oyer (2008) which investigates the effect of stock market performance at graduation on the career choices of investment bankers, Oyer (2006) which looks at the relationship between economic conditions during graduation and the career success of economics PhDs, and Devereux (2002) who investigates the effect of initial job quality on wages and job quality later in life. Another related paper is Gardecki and Neumark (1998). However, the authors investigate the relationship between early career experiences and subsequent labor market outcomes by using variables that proxy for an instable early career (mobility, tenure, training, etc) in their analysis and don't focus on early initial economic conditions.

is (also)consistent with the observed persistence in the effect of initial labor market conditions on wages.

Above I showed that the initial guess that a firm makes about a worker's true productivity under rational inattention is written as $E_t[\widetilde{X}|I(\kappa_t)]$ where $I(\kappa_t)$ is the information set that is available to the employer at time t. The theoretical part also showed that the firm will allocate more attention to workers that have a higher impact on profits. Now assume two jobs (J_1 and J_2, not necessarily in the same firm) and assume that Job 2 has less of an impact on profits. Also assume that, under normal conditions, a given worker would usually obtain the higher impact job (J_1). In this scenario the worker would earn

$$w_{J_1} = E[\widetilde{X}|I(\kappa^{J_1})].$$

Now consider the counterfactual outcome in which the worker, due to a recession, has to settle for the lower-impact job (J_2). In this case the worker earns

$$w_{J_2} = E[\widetilde{X}|I(\kappa^{J_2})].$$

The salient feature here is that the same worker will earn different starting salaries depending on prevailing economic conditions. Assuming that the starting salary is less in job 2 the diference in (counterfactual) wages is result of rational inattention because firms will allocate less attention to the second job such that

$$I(\kappa^{J_1}) > I(\kappa^{J_2}).$$

This means that firms will accumulate more information on workers in Job 1 than worker in Job 2. Given lower starting salaries for lower impact jobs the same worker earns less in the lower-impact job because less attention is devoted to her. Does this difference in wages for identical workers persist over time? Results 1 and 2 show that firms will allocate more attention to workers with a bigger impact on profits and will try to transfer all attention to the current period. This implies that firms will, after allocating most attention to the initial guess, take longer to learn about the true productivity if the worker is in Job 2 and the difference in wages, while not permanent, could be quiet persistent. So far we looked at two identical workers but the results in Oreopoulos et al. (2008) also suggest that, within the group of college graduates, less able workers will be hurt more. This is also consistent with the results from the employer learning model with rational inattention.

Consider two workers whose productivity at the time of hiring is equal to $\widetilde{X}^H$ and $\widetilde{X}^L$ respectively. Given a recession the results above imply that each worker will face a wage penalty. The penalties for the two workers are

$$w_{J_1}^H - w_{J_2}^H = E[\widetilde{X}^H | I(\kappa_{J_1})] - E[\widetilde{X}^H | I(\kappa_{J_2})]$$
$$w_{J_1}^L - w_{J_2}^L = E[\widetilde{X}^L | I(\kappa_{J_2})] - E[\widetilde{X}^L | I(\kappa_{J_3})]$$

Each of the wage penalties is dependent on the change in the amount of attention that the firm devotes to the worker. Given that the firm devotes less attention to lower-impact jobs ($\kappa_{J_1} > \kappa_{J_2} > \kappa_{J_3}$) we would expect the penalties to be negatively correlated with the profit-impact of a job. However, this conjecture is probably only correct for for the group that have a relatively high impact on profits (like college workers) since the amount of attention devoted low-impact workers is small to start with.

5 Conclusion

This paper develop a simple model of employer learning under rational inattention. The predictions of the model are such that the firm will allocate more attention to the group of employees that have a relatively high impact on profits. The firm will also try to learn as quickly as possible by transferring attention to the first period. In addition to the intuitive behavioral appeal of the theory the paper contributes by showing that the model is consistent with a variety of empirical phenomena. Another contribution of the paper is that it suggests the applicability of the theory of rational inattention to other microeconomic topics. For example, investigating the effect of rational inattention on the firm's wage setting could affect the employment response to changes in the minimum wage. In addition, extensions should be developed that investigate the symmetry of employer learning since the model suggests that firms (want to) learn differently about different worker groups. Other examples are the role of educational signaling, educational investments and/or occupational choice under rational inattention. Future work should also develop a more detailed model that would yield structural, testable predictions.

A Appendix

A.1 Mutual Information under Normality

In general, the entropy of a continuously distributed random variable can be written as

$$h(x) = -\int_S f(x)\log_2 f(x)dx \tag{19}$$

If the random variable is normally distributed we can write

$$\begin{aligned} h(x) &= -\int_{-\infty}^{\infty} f(x)\ln\left[\frac{1}{\sqrt{2\pi}\sigma_X}e^{-\frac{1}{2}\left(\frac{X-\mu_X}{\sigma_X}\right)^2}\right]dx = -\int_{-\infty}^{\infty} f(x)\left[\ln\left(\frac{1}{\sqrt{2\pi}\sigma_X}\right) - \frac{1}{2}\left(\frac{X-\mu_X}{\sigma_X}\right)^2\right]dx \\ &= -\ln\left(\frac{1}{\sqrt{2\pi}\sigma_X}\right)\underbrace{\int_{-\infty}^{\infty} f(x)}_{=1} + \frac{1}{2\sigma_X^2}\underbrace{\int_{-\infty}^{\infty}(X-\mu_X)^2 f(x)dx}_{\sigma_X^2} \\ &= \ln(\sqrt{2\pi}\sigma_X) + \frac{1}{2} = \ln(2\pi\sigma_X^2)^{\frac{1}{2}} + \frac{1}{2}\ln e \\ &= \frac{1}{2}\ln(2\pi e\sigma_X^2) \end{aligned} \tag{20}$$

The conditional entropy is given by $h(x|y) = -\int_{-\infty}^{\infty} f(x,y)\ln f(x|y)dxdy$

Going through the same derivation as above yields $h(x|y) = \frac{1}{2}\ln(2\pi e\sigma_{X|Y}^2)$ Therefore we can write mutual information as

$$\begin{aligned} I(X;Y) &= h(x) - h(x|y) = \frac{1}{2}\ln(2\pi e\sigma_X^2) - \frac{1}{2}\ln(2\pi e\sigma_{X|Y}^2) \\ &= \frac{1}{2}\ln\left(2\pi e\frac{\sigma_X^2}{\sigma_{X|Y}^2}\right) \end{aligned} \tag{21}$$

A.2 Derivation of Loss Function

$$\begin{aligned} \hat{\pi}(\Delta l^*, a_1, a_2) - \hat{\pi}(\Delta l^{\odot}, a_1, a_2) &= \frac{1}{2}\widetilde{\pi}_{11}\Delta l^{*2} - \frac{1}{2}\widetilde{\pi}_{11}\Delta l^{\odot 2} + \widetilde{\pi}_{12}a_1(\Delta l^* - \Delta l^{\odot}) + \widetilde{\pi}_{13}a_2(\Delta l^* - \Delta l^{\odot}) \\ &= \frac{1}{2}\widetilde{\pi}_{11}\Delta l^{*2} - \frac{1}{2}\widetilde{\pi}_{11}\Delta l^{\odot 2} + |\widetilde{\pi}_{11}|\underbrace{\left(\frac{\widetilde{\pi}_{12}}{|\widetilde{\pi}_{11}|}\cdot a_1 + \frac{\widetilde{\pi}_{13}}{|\widetilde{\pi}_{11}|}\cdot a_2\right)}_{\Delta l^*}(\Delta l^* - \Delta l^{\odot}) \\ &= -\frac{1}{2}|\widetilde{\pi}_{11}|\Delta l^{*2} + \frac{1}{2}|\widetilde{\pi}_{11}|\Delta l^{\odot 2} + |\widetilde{\pi}_{11}|\Delta l^{*2} - |\widetilde{\pi}_{11}|\Delta l^*\Delta l^{\odot} \\ &= \frac{1}{2}|\widetilde{\pi}_{11}(\Delta l^* - \Delta l^{\odot})^2 \end{aligned} \tag{22}$$

A.3 Equality of FOC of 2^{nd}-order approximation and 1^{st}-order approximation of first derivative

Consider the function $Y = f(x, a, b)$. A second-order Taylor approximation yields

$$\widetilde{Y} = f(0,0,0) + f_1 x + f_2 a + f_3 b + \frac{1}{2} f_{11} x^2 + \frac{1}{2} f_{22} a^2 + \frac{1}{2} f_{33} b^2 + f_{12} xa + f_{13} xb + f_{23} ab \tag{23}$$

The first-order condition (FOC) with respect to x is $\widetilde{Y}' = f_1(0,0,0) + f_{11}x + f_{12}a + f_{13}b$ The first-order condition of the original function Y is f_1 to which a second-order Taylor approximation yields

$$\hat{Y}' = f_1(0,0,0) + f_{11}x + f_{12}a + f_{13}b$$

A.4 Derivation of Result 1 and Result 2

In the 2-period case the profit function (in log-deviations from the non-stochastic solution and written as a 2^{nd}-order Taylor approximation) is equal to

$$\pi = \pi_{t=1}(\Delta l_{t=1}, a_1, a_2) + \pi_{t=2}(\Delta l_{t=1}, a_1, a_2) \tag{24}$$

Maximizing π with respect to $lr_{t=1}$ and $lr_{t=2}$ yields

$$\Delta l^*_{t=1} = \Delta l^*_{t=2} = \frac{\widetilde{\pi}_{12}}{|\widetilde{\pi}_{11}|} \cdot a_1 + \frac{\widetilde{\pi}_{13}}{|\widetilde{\pi}_{11}|} \cdot a_2 \tag{25}$$

The equality of the two solutions arises because the shocks a_1 and a_2 are drawn in the first period and are constant from then on. The firm forms conditional expectations of the optimal labor ratio on each period. These are given by

$$\begin{aligned}
\Delta l^{\odot}_{t=1} &= E(\Delta l^*_{t=1} | s_1^{t=1}, s_2^{t=1}) \\
&= \frac{\widetilde{\pi}_{12}}{|\widetilde{\pi}_{11}|} \cdot E(a_1 | s_1^{t=1}) + \frac{\widetilde{\pi}_{13}}{|\widetilde{\pi}_{11}|} \cdot E(a_2 | s_2^{t=1}) \\
\Delta l^{\odot}_{t=2} &= E(\Delta l^*_{t=1} | s_1^{t=1}, s_2^{t=1}, s_1^{t=2}, s_2^{t=2}) \\
&= \frac{\widetilde{\pi}_{12}}{|\widetilde{\pi}_{11}|} \cdot E(a_1 | s_1^{t=1}, s_1^{t=2}) + \frac{\widetilde{\pi}_{13}}{|\widetilde{\pi}_{11}|} \cdot E(a_2 | s_2^{t=1}, s_2^{t=2})
\end{aligned}$$

As in the text the profit-maximization is equal to minimizing the expected squared deviations of the optimal and actual labor ratios.

$$\begin{aligned}
\min_{s_i^{t=1}, s_i^{t=2}} \quad & \left[E(\Delta l^{\odot}_{t=1} - \Delta l^{*}_{t=1})^2 + E(\Delta l^{\odot}_{t=2} - \Delta l^{*}_{t=2})^2\right] \\
= \quad & \left(\frac{\widetilde{\pi}_{12}}{|\widetilde{\pi}_{11}|}\right)^2 \left(2^{-2\kappa_1^{t=1}}\sigma^2_{a_1} + 2^{-2\kappa_1^{t=2}}\sigma^2_{a_1|s_1^{t=1}}\right) + \left(\frac{\widetilde{\pi}_{13}}{|\widetilde{\pi}_{11}|}\right)^2 \left(2^{-2\kappa_2^{t=1}}\sigma^2_{a_2} + 2^{-2\kappa_2^{t=2}}\sigma^2_{a_2|s_2^{t=1}}\right) \\
= \quad & \left(\frac{\widetilde{\pi}_{12}}{|\widetilde{\pi}_{11}|}\right)^2 \left(2^{-2\kappa_1^{t=1}}\sigma^2_{a_1}(1 + 2^{-2\kappa_1^{t=2}})\right) + \left(\frac{\widetilde{\pi}_{13}}{|\widetilde{\pi}_{11}|}\right)^2 \left(2^{-2\kappa_2^{t=1}}\sigma^2_{a_2}(1 + 2^{-2\kappa_2^{t=2}})\right)
\end{aligned}$$

The last equality is due to the fact that $\sigma^2_{a_i|s_1} = 2^{-2\kappa_i^{t=1}}$ and reflects the fact that attention devoted to learning about the shock in the first period lowers the conditional variance of the shock that serves as a baseline in the second period.

Derivation of Result 1
Assume that the total amount of attention *per time* period is fixed. $\kappa_1^{t=1} + \kappa_1^{t=2} + \kappa_2^{t=1} + \kappa_2^{t=2} = \kappa$ reduces to $\underbrace{\kappa^{t=1}}_{\kappa_1^{t=1}+\kappa_2^{t=1}} + \underbrace{\kappa^{t=2}}_{\kappa_1^{t=2}+\kappa_2^{t=2}} = \kappa$. Substituting the period one constraint in the objective function and maximizing with respect to $\kappa_1^{t=1}$ yields equation (12) in the text.

Derivation of Result 2
Assume that the total amount of attention *per worker group* is fixed. This reduces $\kappa_1^{t=1} + \kappa_1^{t=2} + \kappa_2^{t=1} + \kappa_2^{t=2} = \kappa$ to $\underbrace{\kappa_1}_{\kappa_1^{t=1}+\kappa_1^{t=2}} + \underbrace{\kappa_2}_{\kappa_2^{t=1}+\kappa_2^{t=2}} = \kappa$. . Substituting the constraint for worker group one in the equation and maximizing with respect to $\kappa_1^{t=1}$ yields equation (13) in the text.

A.5 Derivation of Equation 16

Equation (14) in the text is derived by Bayesian updating. The following formula applies to the first period

$$E_{t=1}(\widetilde{X}|Edu, q, y_1) = \frac{\frac{1}{\sigma_0^2}E_{t=0}(\widetilde{X}|Edu, q) + \frac{1}{\sigma^2_{\epsilon_{t=1}}}y_1}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2_{\epsilon^{t=1}}}} \tag{26}$$

For the second period this expectation equals

$$\begin{aligned}
E_{t=2}(\widetilde{X}|Edu, q, y_1, y_2) \quad = \quad & \frac{\left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2_{\epsilon_{t=1}}}\right)E_{t=1}(\widetilde{X}|Edu, q, y_1) + \left(\frac{1}{\sigma^2_{\epsilon_{t=2}}}\right)y_2}{\left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2_{\epsilon_{t=1}}} + \frac{1}{\sigma^2_{\epsilon_{t=2}}}\right)} \\
= \quad & \frac{\left(\frac{1}{\sigma_0^2}\right)E_{t=0}(\widetilde{X}|Edu, q) + \left(\frac{1}{\sigma^2_{\epsilon_{t=2}}}\right)y_2 + \left(\frac{1}{\sigma^2_{\epsilon_{t=1}}}\right)y_1}{\left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2_{\epsilon_{t=1}}} + \frac{1}{\sigma^2_{\epsilon_{t=2}}}\right)}
\end{aligned} \tag{27}$$

For the t^{th} period this expectation equals

$$
\begin{aligned}
E_{t=n}(\widetilde{X}|Edu,q,y_1) &= \frac{\left(\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2_{\epsilon_{t=1}}}+\frac{1}{\sigma^2_{\epsilon_{t=2}}}+...+\frac{1}{\sigma^2_{\epsilon_{t=n-1}}}\right)E_{t=n-1}(\widetilde{X}|Edu,q,y_1,y_2,...,y_n)+\left(\frac{1}{\sigma^2_{\epsilon_{t=n}}}\right)y_2}{\left(\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2_{\epsilon_{t=1}}}+\frac{1}{\sigma^2_{\epsilon_{t=2}}}+...+\frac{1}{\sigma^2_{\epsilon_{t=n}}}\right)} \\
&= \frac{\left(\frac{1}{\sigma_0^2}\right)E_{t=0}(\widetilde{X}|Edu,q)+\left(\frac{1}{\sigma^2_{\epsilon_{t=n}}}\right)y_n+\left(\frac{1}{\sigma^2_{\epsilon_{t=n-1}}}\right)y_{n-1}+..+\left(\frac{1}{\sigma^2_{\epsilon_{t=1}}}\right)y_1}{\left(\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2_{\epsilon_{t=1}}}+\frac{1}{\sigma^2_{\epsilon_{t=2}}}+...+\frac{1}{\sigma^2_{\epsilon_{t=n}}}\right)} \\
&= \frac{\left(\frac{1}{\sigma_0^2}\right)}{\left(\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2_{\epsilon_{t=1}}}+\frac{1}{\sigma^2_{\epsilon_{t=2}}}+...+\frac{1}{\sigma^2_{\epsilon_{t=n}}}\right)}E_{t=0}(\widetilde{X}|Edu,q)+\sum_{i=1}^{n}(signals) \\
&= \left(1-\underbrace{\left[\frac{\frac{1}{\sigma^2_{\epsilon_{t=1}}}+\frac{1}{\sigma^2_{\epsilon_{t=2}}}+...+\frac{1}{\sigma^2_{\epsilon_{t=n}}}}{\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2_{\epsilon_{t=1}}}+\frac{1}{\sigma^2_{\epsilon_{t=2}}}+...+\frac{1}{\sigma^2_{\epsilon_{t=n}}}}\right]}_{\theta_t}\right)E_{t=0}(\widetilde{X}|Edu,q)+\sum_{i=1}^{n}(signals)
\end{aligned}
$$

The change in θ_t is equal to

$$
\begin{aligned}
\Delta\theta_t &= \theta_t-\theta_{t-1} \\
&= \left[1-\frac{(1-\theta_t)}{(1-\theta_{t-1})}\right]\cdot(1-\theta_{t-1}) \\
&= \left[1-\frac{\left(1-\frac{\frac{1}{\sigma^2_{\epsilon_{t=1}}}+\frac{1}{\sigma^2_{\epsilon_{t=2}}}+...+\frac{1}{\sigma^2_{\epsilon_{t=n}}}}{\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2_{\epsilon_{t=1}}}+\frac{1}{\sigma^2_{\epsilon_{t=2}}}+...+\frac{1}{\sigma^2_{\epsilon_{t=n}}}}\right)}{\left(1-\frac{\frac{1}{\sigma^2_{\epsilon_{t=1}}}+\frac{1}{\sigma^2_{\epsilon_{t=2}}}+...+\frac{1}{\sigma^2_{\epsilon_{t=n-1}}}}{\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2_{\epsilon_{t=1}}}+\frac{1}{\sigma^2_{\epsilon_{t=2}}}+...+\frac{1}{\sigma^2_{\epsilon_{t=n-1}}}}\right)}\right]\cdot(1-\theta_{t-1}) \\
&= \left[1-\frac{\left(\frac{\frac{1}{\sigma_0^2}}{\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2_{\epsilon_{t=1}}}+\frac{1}{\sigma^2_{\epsilon_{t=2}}}+...+\frac{1}{\sigma^2_{\epsilon_{t=n}}}}\right)}{\left(\frac{\frac{1}{\sigma_0^2}}{\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2_{\epsilon_{t=1}}}+\frac{1}{\sigma^2_{\epsilon_{t=2}}}+...+\frac{1}{\sigma^2_{\epsilon_{t=n-1}}}}\right)}\right]\cdot(1-\theta_{t-1}) \\
&= \left[1-\frac{V_t(\widetilde{X}|Edu,q,y_1,y_2,..,y_n)}{V_{t-1}(\widetilde{X}|Edu,q,y_1,y_2,..,y_{n-1})}\right]\cdot(1-\theta_{t-1}) \\
&= (1-2^{-2\kappa_1})\cdot(1-\theta_{t-1}) \\
&= \%\Delta V_t\cdot(1-\theta_{t-1})
\end{aligned}
\tag{28}
$$

Table 2: The Effects of AFQT on Log Wages for High School and College Graduates

	High School		College	
	(1)	(2)	(3)	(4)
Model:				
Standardized AFQT	.0060	.0078	.1467**	.1403**
	(.0130)	(.0129)	(.0364)	(.0369)
AFQT x experience/10	.1265**	.1187**	.0179	.0281
	(.0176)	(.0173)	(.0611)	(.0608)
Black	-.0628*	-.0483	.0922	.0867
	(.0267)	(.0259)	(.0575)	(.0563)
Black x experience/10	-.0369	-.0398	-.0823	-.0681
	(.0350)	(.0345)	(.0959)	(.0961)
R^2	0.1628	0.1871	0.1553	0.1688
Additional variables	No	Yes	No	Yes
No. Observations	11798	11775	3373	3373

Experience measure: Years since left school for the first time

Note - Specifications (1) and (3) control for urban residence, a cubic in experience and year effects. Specifications (2) and (4) also control for region of residence and for part time vs full time jobs. There are fewer observations in specification (2) versus specification (1) because some observations are missing region of residence and/or full/part time status. Potential experience is restricted to less than thirteen and ten years for the high school and the college market, respectively. The White/Huber standard errors in parenthesis control for correlation at the individual level.

* statistical significance at the 95% level

** statistical significance at the 99% level

References

Altonji, Joseph G. and Charles R. Pierret, "Employer Learning and Statistical Discrimination," *The Quarterly Journal of Economics*, 2001, *116* (1), 313–50.

Arcidiacono, Peter, Patrick Bayer, and Hizmo Aurel, "Beyond signaling and human capital: Education and the revelation of ability," *NBER Working Paper*, 2008, (13951).

Baker, George, Michael Gibbs, and Bengt Holmstrom, "The wage policy of a firm," *The Quarterly Journal of Economics*, 1994, *109* (4), 921–55.

Bauer, Thomas K. and John P. Haisken-DeNew, "Employer learning and the returns to schooling," *Labour Economics*, 2001, *8* (2), 161 – 180.

Beaudry, Paul and John DiNardo, "The effect of implicit contracts on the movement of wages over the business cycle: Evidence from micro data," *The Journal of Political Economy*, 1991, *99* (4), 665–88.

Cover, Thomas M. and Joy A. Thomas, *Elements of information theory*, 2. ed., Wiley, 2006.

Devereux, Paul, "The importance of obtaining a high-paying job," Working Paper, University of California - Los Angeles 2002.

Farber, Henry S. and Robert Gibbons, "Learning and wage dynamics," *The Quarterly Journal of Economics*, 1996, *111* (4), 1007–47.

Gardecki, Rosella and David Neumark, "Order from chaos? The effects of early labor market experiences on adult labor market outcomes," *Industrial and Labor Relations Review*, 1998, *51* (2), 299–322.

Kahn, Lisa, "The long-term labor market consequences of graduating from college in a bad economy," Working Paper, Yale School of Management 2009.

Lange, Fabian, "The Speed of Employer Learning," *Journal of Labor Economics*, 2007, *25* (1), 1–35.

Machkowiak, Bartosz and Mirko Wiederholt, "Optimal Sticky Prices under Rational Inattention," *American Economic Review*, 2009, *99* (3), 769–803.

Oreopoulos, Philip, Till von Wachter, and Andrew Heisz, "The short- and long-term career effects of graduating in a recession: Hysteresis and heterogeneity in the market for college graduates," Discussion Paper 3578, IZA 2008.

Oyer, Paul, "Initial labor market conditions and long-term outcomes for economists," *Journal of Economic Perspectives*, 2006, *20* (3), 143–60.

_ , "The making of an investment banker: Stock market shocks, career choice, and lifetime income," *The Journal of Finance*, 2008, *63* (6), 2601–28.

Rosen, Sherwin, "Implicit contracts: A survey," *Journal of Economic Literature*, 1985, *23* (3), 1144–75.

Sims, Christopher A., "Implications of Rational Inattention," *Journal of Monetary Economics*, 2003, *50* (3), 665–90.

_ , "Rational Inattention: Beyond the Linear-Quadratic Case," *American Economic Review*, 2006, *96* (2), 158–63.