Rational inattention and employer learning

By Steffen Habermalz

Department of Economics, Northwestern University, 2001 Sheridan Road, Evanston IL, 60208, United States; email: s-habermalz@northwestern.edu, and IZA

Abstract

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. (2010) and results that economic conditions at labor market entry have persistent effects on wages (for example Oreopoulos et al. (2012)) cast doubt on the model’s validity. This paper extends the employer learning model with the theory of rational inattention introduced by Sims (2003). In the model firms optimally allocate attention (i.e. information processing capacity) to learning about the productivities of different workers. 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.

JEL classifications:

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. (2012) find persistent wage effects of graduating during a recession and Arcidiacono et al. (2010) find that the empirical pattern predicted by the employer learning model holds only for workers with a high school diploma and is mostly absent for college educated workers which questions the validity of the employer learning model.¹

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 (2003) 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 (i.e. information processing capacity) between different sources of uncertainty that affect her optimal decision.

So far rational inattention has primarily been applied to macroeconomic topics. For example, in a recent paper Maćkowiak 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

---

¹This is not to imply that these authors do not forward explanations for the observed effects. Arcidiacono et al. (2010) provide evidence that college educated workers are able to reveal their true productivity to employers directly and Oreopoulos et al. (2012) forward an explanation based on a search model. The current paper’s contribution is that the employer learning model with rational inattention serves as a single explanation for both phenomena.
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 hiring labor under rational inattention in a partial-equilibrium setting. I will first introduce the objective (profit) function and then introduce the information processing constraint that the firm faces. Finally, the last sub-

---

2 This section follows section 2 in Maçkowskiak and Wiederholt (2009).
section will combine the objective function with the information processing constraint and derive the solution of the firm problem under rational inattention.

2.1 The Profit Function

Consider a firm that employs two workers and has to choose how many units of labor ($L_1$ and $L_2$) of each worker it uses 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 $\left(\frac{L_1}{L_2}\right)$. Further assume that the productivities of the two workers ($A_1$ and $A_2$) are stochastic and alter the workers’ effective amount of labor used. The productivities are assumed to be normally distributed and mutually independent:

$$A_1 \sim N(0, \sigma^2_{A_1}) \quad A_2 \sim N(0, \sigma^2_{A_2}) \quad A_1 \perp A_2$$

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

$$\pi \left(\frac{L_1}{L_2}, e^{A_1}, e^{A_2}\right).$$

To analyze the model 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 optimal solution when $A_1 = A_2 = 0$). We first rewrite the variables in the profit function in terms of log deviations from their non-stochastic equilibrium values$^3$ Indicating non-stochastic equilibrium values by a bar over the variables. This implies for example that $L_1 = \bar{L}_1 e^{l_1}$ where $l_1 = \ln L_1 - \ln \bar{L}_1$.$^4$ Therefore the profit function in

$^3$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.

$^4$Note that $e^{A_1} = e^{\bar{A}_1} \cdot e^{\ln e^{A_1} - \ln \bar{A}_1}$. 

4
log-deviations is equal to
\[
\pi\left(\frac{L_1}{L_2}, e^{A_1}, e^{A_2}\right) = \pi\left(\frac{L_1}{L_2} \cdot \frac{e^{\Delta l}}{e^{\Delta l}}, e^{A_1}, e^{A_2}\right)
= \pi\left(\frac{L_1}{L_2} \cdot e^{\Delta l}, e^{A_1}, e^{A_2}\right)
= \tilde{\pi}(\Delta l, A_1, A_2)
\]
where \(\Delta l = l_1 - l_2\). The 2nd order Taylor expansion of \(\tilde{\pi}(\Delta l, A_1, A_2)\) yields
\[
\hat{\pi}(\Delta l, A_1, A_2) = \tilde{\pi}(0, 0, 0) + \tilde{\pi}_1 \cdot \Delta l + \tilde{\pi}_2 \cdot A_1 + \tilde{\pi}_3 \cdot A_2
+ \frac{1}{2} \cdot \tilde{\pi}_{11} \cdot \Delta l^2 + \frac{1}{2} \cdot \tilde{\pi}_{22} \cdot A_1^2 + \tilde{\pi}_{33} \cdot A_2^2
+ \tilde{\pi}_{12} \cdot \Delta l \cdot A_1 + \tilde{\pi}_{13} \cdot \Delta l \cdot A_2
+ \tilde{\pi}_{23} \cdot A_1 \cdot A_2
\]
where \(\tilde{\pi}_i\) is the derivative of the profit function with respect to its \(i^{th}\) argument and \(\tilde{\pi}_{ij}\) are second derivatives. To find the perfect information solution we maximize this function with respect to \(\Delta l\) under the assumption that both productivities are known. Note that \(\tilde{\pi}_1 = 0\) because the firm optimizes at the non-stochastic solution. This yields
\[
\Delta l^* = \frac{\tilde{\pi}_{12}}{|\tilde{\pi}_{11}|} \cdot A_1 + \frac{\tilde{\pi}_{13}}{|\tilde{\pi}_{11}|} \cdot A_2 \quad (1)
\]
Equation (1) represents the profit-maximizing labor ratio under perfect information. Naturally the firm will not observe the realizations of the workers’ productivities. However, I am assuming that the firm receives two independent signals \((S_1, S_2)\) that contain information about the productivities \((A_1, A_2)\). The firm will then form conditional expectations of equation (1) based on these signals. This yields
\[
\Delta l^\circ = E(\Delta l^*|S_1 = s_1, S_2 = s_2)
= \frac{\tilde{\pi}_{12}}{|\tilde{\pi}_{11}|} \cdot E(A_1|S_1 = s_1) + \frac{\tilde{\pi}_{13}}{|\tilde{\pi}_{11}|} \cdot E(A_2|S_2 = s_2) \quad (2)
\]
Finally, the expected loss from deviating from the optimal solution equals

\[ \text{Expected Loss} = E[\tilde{\pi}(\Delta l^\ast, A_1, A_2) - \tilde{\pi}(\Delta l^\odot, A_1, A_2)] \]

\[ = \frac{1}{2} \cdot |\tilde{\pi}_{11}| \cdot E(\Delta l^\odot - \Delta l^\ast)^2 \]  
\[ = \frac{1}{2} \cdot |\tilde{\pi}_{11}| \cdot \left[ \left( \frac{\tilde{\pi}_{12}}{|\tilde{\pi}_{11}|} \right)^2 \cdot E[V(A_1|S_1 = s_1)] + \left( \frac{\tilde{\pi}_{13}}{|\tilde{\pi}_{11}|} \right)^2 \cdot E[V(A_2|S_2 = s_2)] \right] \]  

(3)

See Appendix 1 for a derivation. Note that the value of the objective function depends on the signals (\(S_1\) and \(S_2\)) the firm chooses to receive. Specifically, the firm attempts to minimize the expected value of the conditional variances of the workers' productivities. This is the main feature of rational inattention: The firm is free to choose any signals and is only constrained by the amount of information the signal can provide about the productivities of the workers. The next section will introduce the information processing constraint that limits the choice of signals for firms.

### 2.2 The Information Processing Constraint: Mutual Information

A widely used measure of the uncertainty of a random variable is its entropy defined as

\[ h(Y) = - \int_N f(Y) \log_2 f(Y)dy \]  

(4)

where \(N\) is the support of random variable \(Y\). For more intuition consider the entropy of a Bernoulli random 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 \(Y\). 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 success probability \(p=0.5\). The extent to which the firm faces uncertainty is different before and after the firm receives the signals. The uncertainty of a random variable \(Y\) after another random variable \(X\) is observed is called conditional entropy and is given by

\[ h(Y|X) = - \int_N f(X,Y) \log_2 f(Y|X)dx\,dy \]
We can now compare the uncertainty that the firm faces before and after observing \( X \). Mutual Information measures this reduction in uncertainty. It is limited by the firm’s information processing capacity. Formally Mutual Information is defined as

\[
I(X; Y) = h(Y) - h(Y|X) \leq \kappa
\]

where \( h(Y) \) denotes the entropy of random variable \( Y \) 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 another random variable (worker productivities)? The variable \( \kappa \) indicates that the reduction of uncertainty is limited.

In the current model the workers’ productivities \((A_1, A_2)\) are independent and the firm receives two independent signals \((S_1, S_2)\) about them. Mutual Information is easily adapted to the multivariate case as follows.

\[
I(A_1, A_2; S_1, S_2) = h(A_1, A_2) - h(A_1, A_2|S_1, S_2) \\
= h(A_1) - h(A_1|S_1) + h(A_2) - h(A_2|S_2) \\
= I(A_1; S_1) + I(A_2; S_2) \leq \kappa
\] (5)

The second line follows because of the independence assumption across signals and productivities, see Cover and Thomas (2006). Equation (5) tells us that the overall reduction in uncertainty is (less or) equal to the sum of the reductions in uncertainty about the two productivities. It also provides the general constraint for the firm’s minimization problem. If productivities and signals are jointly normally distributed\(^5\) mutual information equals

\[
I(A_1; S_1) + I(A_2; S_2) = \frac{1}{2} \cdot \log_2 \left( \frac{\sigma_{A_1}^2}{\sigma_{A_1|S_1}^2} \right) + \frac{1}{2} \cdot \log_2 \left( \frac{\sigma_{A_2}^2}{\sigma_{A_2|S_2}^2} \right) \leq \kappa
\] (6)

\(^5\) Most of the literature to date is working with the multivariate normal distribution. For an exception see Sims (2006). Mańkowski and Wiederholt (2009) also investigate the non-normal case.
See Appendix 2 for a derivation.

### 2.3 The Firm Problem and its Solution

Now we can combine the objective function (4) and the information constraint together to formulate the general firm problem under rational inattention.

\[
\min_{\tilde{s}_1, \tilde{s}_2} \left[ \left( \frac{\tilde{\pi}_{12}}{\tilde{\pi}_{11}} \right)^2 \cdot \mathbb{E} \left[ V(A_1|S_1 = s_1) \right] + \left( \frac{\tilde{\pi}_{13}}{\tilde{\pi}_{11}} \right)^2 \cdot \mathbb{E} \left[ V(A_2|S_2 = s_2) \right] \right]
\]

subject to \( I(A_1; S_1) + I(A_2; S_2) \leq \kappa \)

(7)

However, in the case in which the objective function is quadratic and the shocks are normally distributed it can be shown (see Mańkowski and Wiederholt (2009)) that the following signals are optimal

\[
S_1 = A_1 + E_1
\]

\[
S_2 = A_2 + E_2
\]

where \( E_1 \sim N(0, \sigma_{E_1}^2) \) and \( E_2 \sim N(0, \sigma_{E_2}^2) \) are mutually independent and independent of \( A_1 \) and \( A_2 \). This choice of signals by the firm implies that the productivities and the signals are jointly normally distributed. Given joint normality the conditional variances \( V(A_i|S_i = s_i) \) are constant and we can rewrite the optimization problem as

\[
\min_{\kappa_1, \kappa_2} \left[ \left( \frac{\tilde{\pi}_{12}}{\tilde{\pi}_{11}} \right)^2 \cdot \sigma_{A_1|S_1}^2 + \left( \frac{\tilde{\pi}_{13}}{\tilde{\pi}_{11}} \right)^2 \cdot \sigma_{A_2|S_2}^2 \right]
\]

subject to \[
\frac{1}{2} \cdot \log_2 \left( \frac{\sigma_{A_1}^2}{\sigma_{A_1|S_1}^2} \right) + \frac{1}{2} \cdot \log_2 \left( \frac{\sigma_{A_2}^2}{\sigma_{A_2|S_2}^2} \right) \leq \kappa
\]

\( \kappa_1 \geq 0 \) \( \kappa_2 \geq 0 \)

(8)

Given the choice of signals the constraint shows that we can partition the total amount of uncertainty reduction (attention) into uncertainty reduction with respect to the first worker's
productivity ($\kappa_1$) and the second worker’s productivity ($\kappa_2$). Noting that the firm cannot reduce both uncertainties to zero it allocates its attention optimally between the two workers. It is rationally inattentive. Equation (8) also implies that we can substitute the information processing constraint into the objective function to get an unconstrained minimization problem in $\kappa_1$. This yields
\[
\min_{\kappa_1 \in [0, \kappa]} \left[ \left( \frac{\tilde{\pi}_{12}}{\tilde{\pi}_{11}} \right)^2 \cdot 2^{-2\kappa_1} \cdot \sigma_{A_1}^2 + \left( \frac{\tilde{\pi}_{13}}{\tilde{\pi}_{11}} \right)^2 \cdot 2^{-2(\kappa - \kappa_1)} \cdot \sigma_{A_2}^2 \right] \tag{9}
\]
Equation (9) 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 $\kappa_1$ yields
\[
\kappa_1^* = \begin{cases} 
\kappa & \text{if } x \geq 2^{2\kappa} \\
\frac{1}{2}\kappa + \frac{1}{4}\log_2(x) & \text{if } 2^{-2\kappa} < x < 2^{2\kappa} \\
0 & \text{if } x \leq 2^{-2\kappa}
\end{cases} \tag{10}
\]
where $x = \left[ \left( \frac{\tilde{\pi}_{12}}{\tilde{\pi}_{11}} \right)^2 \cdot \frac{\sigma_{A_1}^2}{\sigma_{A_2}^2} \right]$. Equation (10) is the main result of the general model. It shows that the allocation of attention devoted to learning about the productivity of the first worker depends on the relative effect of the two workers on profits. The firm will allocate more attention to the worker that has a bigger effect on the profit-maximizing labor ratio. Equation (10) also shows that the impact of the two workers on the profit-maximizing labor ratio has two components: the relative variability of their productivities $\left( \frac{\sigma_{A_1}^2}{\sigma_{A_2}^2} \right)$ and the specific way in which their productivities affect the labor ratio $\left( \frac{\tilde{\pi}_{12}}{\tilde{\pi}_{11}} \right)$. The next section investigates the effect the latter derivative ratio on the allocation of attention.

At this point a few comments are warranted. First, note that we can rewrite the first part
of the constraint as

\[
\kappa_1 = \frac{1}{2} \log_2 \left( \frac{\sigma^2_{A_1}}{\sigma^2_{A_1|S_1}} \right)
\]

\[
\Rightarrow 2^{-2\kappa_1} = \frac{\sigma^2_{A_1|S_1}}{\sigma^2_{A_1}} = \left( \frac{\sigma^2_{E_1}}{\sigma^2_{A_1} + \sigma^2_{E_1}} \right)
\]

Equation (11) shows that there is an inverse relationship between the amount of attention \((\kappa_1)\) the firm devotes to the first worker’s productivity and the conditional variance of the productivity given the signal \((\sigma^2_{A_1|S_1})\). This is achieved through lowering \(\sigma^2_{E_1}\). If the firm had infinite processing capacity \(\sigma^2_{E_1}\) would be driven to zero and the firm would receive a signal that matches the true realization of the worker’s productivity.

At this point one might ask the question if it is possible to minimize the objective function with respect to the variances \((\sigma^2_{E_1})\) and \((\sigma^2_{E_2})\) instead. The answer is a definitive yes. One can set up a Lagrangian using the information constraint. However, the problem isn’t easier to solve and the solutions are harder to interpret. In essence minimizing with respect to \(\kappa_1\) is merely a change-in-variable which leads to a much simpler problem. In addition, minimizing with respect \(\kappa_1\) is general and doesn’t depend on the specific choice of signals.\(^6\) The above model also 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 the two workers. 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 \((s=\text{true state} + \text{noise})\) is optimal as long as the the shocks are 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 \( \frac{\pi_{12}}{\pi_{13}} \) on the amount of attention that a firm devotes to each worker 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 as before \( A_1 \) and \( A_2 \) represent the stochastic productivities of the two workers. \( L_1 \) and \( L_2 \) are the quantities of labor the firm hires and \( w_1 \) and \( w_2 \) are the respective wages. The (constant) elasticity of substitution between the two workers is equal to \( \frac{1}{1 + \rho} \). In essence this amounts to 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 of the first worker to productivity shocks and is restricted to equal one for the second worker. This is meant to imply that the first worker is of higher quality or more skilled and thus has a higher influence on profits. Imagine a firm employing a janitor and a 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
\[ \tilde{\pi}_Y = P - \left[ \delta \cdot \left( \frac{w_1}{e^{b - A_1}} \right)^\rho + (1 - \delta) \cdot \left( \frac{w_1}{e^{A_2}} \cdot \frac{L_1}{L_2} \right)^\rho \cdot e^{\rho \Delta l} \right]^{\frac{1}{\rho}} \]

\[ - \left[ \delta \cdot \left( \frac{w_2}{e^{b - A_1}} \cdot \frac{L_2}{L_1} \right)^\rho \cdot e^{-\rho \Delta l} + (1 - \delta) \cdot \left( \frac{w_2}{e^{A_2}} \right)^\rho \cdot e^{\rho \Delta l} \right]^{\frac{1}{\rho}} \]

where \( \Delta l = l_1 - l_2 \). Taking the derivative with respect to \( \Delta l \), solving for \( \left( \frac{L_2}{L_1} \right) \) and taking logs yields

\[
\ln \left( \frac{L_1}{L_2} \right)^* = \frac{1}{1 + \rho} \cdot \ln \left( \frac{\delta}{1 - \delta} \right) + \frac{1}{1 + \rho} \cdot \ln \left( \frac{w_2}{w_2} \right) - \frac{1}{1 + \rho} \cdot \left( b \cdot A_1 - A_2 \right)
\]

(12)

This derivative is equal to the first derivative of the 2nd order Taylor approximation of the profit function with respect to \( \Delta l \) (See Appendix 3). This enables us to easily find the desired ratio by differentiating Equation (12) with respect to \( A_1 \) and \( A_2 \) and taking the ratio. Thus the desired derivative ratio equals

\[
\frac{\tilde{\pi}_{12}}{\tilde{\pi}_{13}} = \frac{-\frac{1}{1 + \rho} \cdot b}{\frac{1}{1 + \rho}} = -b
\]

This means we can rewrite equation (10) as

\[
\kappa_1 = \frac{1}{2} + \frac{1}{4} \cdot \ln \left( b^2 \cdot \frac{\sigma_{A_1}^2}{\sigma_{A_2}^2} \right)
\]

(13)

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 shows how 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 workers 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 is constant. Taken together the two scenarios characterize the dynamic allocation of attention.

**Result 1: The firm will allocate more attention to the worker with a higher impact on profits**

Let the total amount of attention available for allocation be equal to $\kappa = \kappa_{t=1}^1 + \kappa_{t=2}^1 + \kappa_{t=1}^2 + \kappa_{t=2}^2$. This section assumes that $\kappa_{t=1}^1 + \kappa_{t=1}^2 = \kappa_{t=1}^1$. 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 the first worker is equal to

$$\kappa_{t=1}^1 = \frac{1}{2} \kappa_{t=1}^1 + \frac{1}{4} \ln \left( \left( \frac{\bar{\pi}_{12}}{\bar{\pi}_{13}} \right)^2 \frac{\sigma_{A_1}^2}{\sigma_{A_2}^2} \left( 1 + 2^{-\kappa_{t=2}^1} \right) \left( 1 + 2^{-\kappa_{t=2}^2} \right) \right).$$

(14)

See Appendix 4. The firm allocates more attention to the 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 over time: $\kappa_{t=1}^1 + \kappa_{t=2}^1 = \kappa_1$ and can allocate attention across time periods but not between workers. The optimal amount of attention the firm allocates to the first worker in the first period is

$$\kappa_{t=1}^1 = \kappa_1$$

(15)

See appendix 4 for the derivation of this result. 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.

To summarize, the two-period employer learning model shows that the firm allocates more attention to the worker that impacts 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).

### 3.3 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, $Edu$ 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 $\epsilon_t \sim N(0, \sigma^2_\epsilon)$ 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.

$$E_t(\widetilde{X}|Edu, q, y_1, y_2, ..., y_t) = (1 - \theta_t) \cdot E(\widetilde{X}|Edu, q) + \theta_t \cdot \left( \frac{1}{t} \sum_{x=0}^{t} y_x \right)$$  \hspace{1cm} (16)

$$V_t(\widetilde{X}|Edu, q, y_1, y_2, ..., y_t) = \frac{1}{t} \frac{1}{\sigma^2_0 + \sigma^2_\epsilon}$$  \hspace{1cm} (17)

---

In this section I deviate from the convention to denote random variables by capital letters and realizations of random variables by lower case letters to be consistent with the cited literature.

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.
Here $\sigma_0^2$ represent the variance of the initial guess $(E(\tilde{X}|Edu, q))$ and $\sigma_2^2$ 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 $\theta_t$ converges to 1 implying that the worker’s true productivity is revealed. Lange (2007) also shows that the coefficient $\theta_t$ is equal to

$$\theta_t = \frac{t \cdot K_1}{1 + (t - 1) \cdot K_1} \text{ where } K_1 = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_2^2}.$$ 

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 $(\theta_t)$ in the conditional expectations formula in terms of the attention the firm devotes to the worker. This yields

$$\Delta \theta_t = \theta_t - \theta_{t-1} = \% \Delta V_t \cdot (1 - \theta_{t-1})$$

$$= (1 - 2^{-\kappa_1}) \cdot (1 - \theta_{t-1}).$$

(18)

See appendix 5. Here $\% \Delta V_t$ represents the percentage change of the conditional variance in equation (17) due to the arrival of a signal in period $t$. Since the two conditional variances depend on the amount of attention allocated equation (18) identifies the direct link between employer learning with rational inattention and the employer learning model in Lange (2007)$^9$. 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, $E(\tilde{X}|Edu, q)$ 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$^{10}$. To make

\[\]

$^9$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.

$^{10}$Either by choice or because of the information processing constraint is binding.
this distinction clear let’s rewrite the initial guess as

\[ E_t[\bar{X}|I(\kappa_t)] \]

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 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 predicts the results in Arcidiacono et al. (2010) 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. (2010)

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.

\[
\ln \text{wage}_{it} = \beta_0 + \beta_{Edu} \cdot Edu_i + \beta_{EduX} \cdot Edu_i \cdot X_{it} + \beta_{AFQT} \cdot AFQT_i + \beta_{AFQT,X} \cdot AFQT_i \cdot X_{it} + \text{other} + \epsilon_{it} \tag{20}
\]

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 \( \beta_{Edu,X} < 0 \) and \( \beta_{AFQT,X} > 0 \). They estimate the above equation and are able to confirm these predictions. Arcidiacono et al. (2010) re-estimate (20) 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\(^{11}\). In the high school regression the coefficient on the individual’s AFQT score (\( \beta_{AFQT} \)) is low and its increase over time (\( \beta_{AFQT,X} \)) is large. The reverse is true for college graduates where coefficient on the AFQT score is high and does not grow much over time. Arcidiacono et al. (2010) 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. (2010) are remarkably consistent with the results derived from the rational inattention model developed above. Result 1 from the 2-period model implies that the firm will allocate more attention to the learning about the group that has a higher impact on profits. 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 less on high school educated workers. Result 2 from the 2-period model states that the firm will try to transfer all attention into the first period. These two results together imply that

\(^{11}\text{Similar results are obtained in Bauer and Haisken-DeNew (2001). These authors find evidence of employer learning for blue collar workers but not for white collar workers.}\)
there is little left to learn about the true productivities of college educated workers after the first period and a lot to learn about the true productivity of high school educated workers. The estimation in Arcidiacono et al. (2010) shows exactly this pattern and hence 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 three 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 to learn about the latter.

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. (2010) would result from the differences in residual information that is available about the two groups.

### 4.2 Empirical Evidence II: Graduating during a Recession

Every year many individuals enter the labor market. This section relates the employer learning model with rational inattention to how graduating during a recession affects the wages of those individuals years later. There exists a growing literature on the effects of initial labor market experiences on subsequent wages. In the latest installment Oreopoulos
et al. (2012) 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 (2010) 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 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. 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. (2012) 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 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 $E_t[\tilde{X}|I(\kappa_t)]$ 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 wages.

---

*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. For evidence of persistent effects of economic downturns on labor market outcomes using U.K. data see Bukodi and Goldthorpe (2011). They find that men of the 1958 birth cohort who entered the labor market during a deep recession seemed to have experienced lasting effects on their occupational histories.*
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

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

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

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

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 difference 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.

In summary, the employer learning model with rational inattention explains the persistent wage effects of graduating during a recession because a) Under the assumption\(^{13}\) that given

\(^{13}\)The reason why a worker will accept a lower impact job stems from the fact that, during a recession, employers reduce the relative number of high impact jobs to low impact jobs available to new hires as they cut back on their investments (including investments in workers). This labor demand driven change increases the likelihood of a mismatch between the worker and the job either through employers raising job requirements, workers taking a job that they would not have taken in better economic times or both. There is ample empirical evidence on this. For example, Hall(2005) finds that employment adjustments are primarily achieved through hiring and reports lower job finding rates during recessions. Bowlus (1995) reports that the match quality between employers and workers is lower
adverse economic conditions the worker ends up in lower impact job b) Employer’s devote less attention to lower impact jobs (Result 1) resulting in a lower initial wage for the worker (compared to a non-recession economic environment) c) Due to the ”front-loading” of attention (Result 2) this wage gap is persistent since employer learning after the initial period is slow and it takes time until the worker will receive his non-recession wage.

Naturally, a worker is free to change jobs. Assume that potential employers have the same information as the current employer. This implies that any new wage offer depends on how much information the current firm gathered about the worker. According to the results derived above this means that workers in lower impact jobs will generate wage offers that are lower than offers by workers generated by workers in high impact jobs. This implies that the counterfactual wage difference will persist until the workers productivity is fully revealed. According to result 2 this can take a long time.\(^\text{14}\)

So far we looked at two identical workers but the results in Oreopoulos et al. (2012) 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 \(\tilde{X}_H\) and \(\tilde{X}_L\) respectively. Given a recession the results above imply that each worker will face a wage penalty. The penalties for the two workers are

\[
\begin{align*}
    w^H_{J_1} - w^H_{J_2} &= E[\tilde{X}^H | I(\kappa_{J_1})] - E[\tilde{X}^H | I(\kappa_{J_2})] \\
    w^L_{J_1} - w^L_{J_2} &= E[\tilde{X}^L | I(\kappa_{J_2})] - E[\tilde{X}^L | I(\kappa_{J_3})]
\end{align*}
\]

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 positively correlated with the profit-impact of a job. However, this conjecture is probably only correct for for the group that during recessions. Devereux (2004) and Oreopoulos et al. (2012) report that new hires get lower quality jobs in adverse economic times. For international evidence see van Ours and Ridder (1995) and Teulings (1993) who report similar findings using Dutch data. Most of the works cited find that mismatches occur primarily for higher skilled workers, a fact the matches up nicely with the current model.

\(^{14}\)The wage difference would be less persistent if the worker could somehow credibly signal his true productivity to an outside firm after new high impact jobs become available with improving economics conditions.
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 develops 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. Although outside the scope of this paper, the model is also applicable to a number of 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.

Acknowledgement: 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, Matthias Fahn and two anonymous referees. The remaining errors are mine.
References


Appendix

1 Derivation of Loss Function

\[ \hat{\pi}(\Delta^*, A_1, A_2) - \hat{\pi}(\Delta^\circ, A_1, A_2) = \frac{1}{2} \pi_{11}(\Delta^r - \Delta^\circ) + \pi_{12} A_1(\Delta^* - \Delta^\circ) + \pi_{13} A_2(\Delta^* - \Delta^\circ) \]

\[ = \frac{1}{2} \pi_{11}(\Delta^r - \Delta^\circ) + |\pi_{11}| \left( \frac{\pi_{12}}{|\pi_{11}|} \cdot A_1 + \frac{\pi_{13}}{|\pi_{11}|} \cdot A_2 \right)(\Delta^* - \Delta^\circ) \]

\[ = \frac{1}{2} |\pi_{11}|(\Delta^* - \Delta^\circ)^2 \]

Substituting the expressions for \( \Delta^r \) and \( \Delta^\circ \), squaring and taking expectations yields equation (4)

2 Mutual Information under Normality

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

\[ h(x) = - \int_{\mathbb{S}} f(x) \log_2 f(x) dx \]

If the random variable is normally distributed we can write

\[ h(X) = - \int_{-\infty}^{\infty} f(x) \ln \left( \frac{1}{\sqrt{2\pi\sigma_X}} e^{-\frac{1}{2} \left( \frac{x-\mu_X}{\sigma_X} \right)^2} \right) dx = - \int_{-\infty}^{\infty} f(x) \left[ \ln \left( \frac{1}{\sqrt{2\pi\sigma_X}} \right) - \frac{1}{2} \left( \frac{x-\mu_X}{\sigma_X} \right)^2 \right] dx \]

\[ = - \ln \left( \frac{1}{\sqrt{2\pi\sigma_X}} \right) \int_{-\infty}^{\infty} f(x) + \frac{1}{2\sigma_X^2} \int_{-\infty}^{\infty} (X - \mu_X)^2 f(x) dx \]

\[ = \ln(\sqrt{2\pi\sigma_X}) + \frac{1}{2} = \ln(2\pi\sigma_X^2)^{\frac{1}{2}} + \frac{1}{2} \ln e \]

\[ = \frac{1}{2} \ln(2\pi e\sigma_X^2) \]

The conditional entropy is given by \( h(X|Y) = - \int_{-\infty}^{\infty} f(x, y) \ln f(x|y) dx dy \)

Going through the same derivation as above yields \( h(X|Y) = \frac{1}{2} \ln(2\pi e\sigma_X^2|Y) \). Therefore we can write
mutual information as
\[ I(X;Y) = h(X) - h(X|Y) = \frac{1}{2} \ln(2\pi e \sigma_X^2) - \frac{1}{2} \ln(2\pi e \sigma_{X|Y}^2) \]
\[ = \frac{1}{2} \ln \left( \frac{2\pi e \sigma_X^2}{\sigma_{X|Y}^2} \right) \]

3 Equality of FOC of 2nd-order approximation and 1st-order approximation of first derivative

Consider the function \( Y = f(x, a, b) \). A second-order Taylor approximation yields
\[ \tilde{Y} = f(0, 0, 0) + f_1 x + f_2 a + f_3 b + \frac{1}{2} f_{11} x^2 + \frac{1}{2} f_{22} a^2 + \frac{1}{2} f_{33} b^2 + f_{12} xa + f_{13} xb + f_{23} ab \]

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

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 2nd-order Taylor approximation) is equal to
\[ \pi = \pi_{t=1}(\Delta l_{t-1}, A_1, A_2) + \pi_{t=2}(\Delta l_{t-1}, A_1, A_2) \]

Maximizing \( \pi \) with respect to \( lr_{t=1} \) and \( lr_{t=2} \) yields
\[ \Delta l^*_t = \Delta l^*_{t=2} = \frac{\pi_{12}}{\pi_{11}} \cdot A_1 + \frac{\pi_{13}}{\pi_{11}} \cdot A_2 \]

The equality of the two solutions arises because the productivities \( A_1 \) 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
\[
\Delta l^*_{t=1} = E(\Delta l_{t=1}^* | S_{t=1}^1 = s_{t=1}^1, S_{t=1}^2 = s_{t=1}^2) = \frac{\pi_{12}}{\pi_{11}} \cdot E(A_1 | S_{t=1}^1 = s_{t=1}^1) + \frac{\pi_{13}}{\pi_{11}} \cdot E(A_2 | S_{t=1}^2 = s_{t=1}^2)
\]
\[
\Delta l^*_{t=2} = E(\Delta l_{t=2}^* | S_{t=1}^1 = s_{t=1}^1, S_{t=1}^2 = s_{t=1}^2, S_{t=2}^2 = s_{t=2}^2) = \frac{\pi_{12}}{\pi_{11}} \cdot E(A_1 | S_{t=1}^1 = s_{t=1}^1, S_{t=1}^2 = s_{t=1}^2) + \frac{\pi_{13}}{\pi_{11}} \cdot E(A_2 | S_{t=2}^2 = s_{t=2}^2)
\]
As in the text the profit-maximization is equal to minimizing the expected squared deviations of the optimal and actual labor ratios.

\[
\begin{align*}
\min_{s_1^{t=1}, s_1^{t=2}} & \quad [E(\Delta l^{\circ}_{t=1} - \Delta l^*_t)^2 + E(\Delta l^{\circ}_{t=2} - \Delta l^*_t)^2] \\
& = \left( \frac{\tilde{\pi}_{12}}{\tilde{\pi}_{11}} \right)^2 (2^{-2\kappa^{t=1}_1} \sigma^2_{A_1} + 2^{-2\kappa^{t=2}_1} \sigma^2_{A_1|s_1^{t=1}}) + \left( \frac{\tilde{\pi}_{13}}{\tilde{\pi}_{11}} \right)^2 (2^{-2\kappa^{t=1}_2} \sigma^2_{A_2} + 2^{-2\kappa^{t=2}_2} \sigma^2_{A_2|s_1^{t=1}})
\end{align*}
\]

The last equality is due to the fact that \(\sigma^2_{A_1|s_1^{t=1}} = 2^{-2\kappa^{t=1}_1}\) and reflects the fact that attention devoted to learning about the productivity in the first period lowers the conditional variance of the productivity 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. \(\kappa^{t=1}_1 + \kappa^{t=2}_1 + \kappa^{t=1}_2 + \kappa^{t=2}_2 = \kappa\). Substituting the period one constraint in the objective function and maximizing with respect to \(\kappa^{t=1}_1\) yields equation (14) in the text.

**Derivation of Result 2**

Assume that the total amount of attention per worker group is fixed. This reduces \(\kappa^{t=1}_1 + \kappa^{t=2}_1 + \kappa^{t=1}_2 + \kappa^{t=2}_2 = \kappa\) to \(\kappa^{t=1}_1 + \kappa^{t=2}_1 + \kappa^{t=1}_2 + \kappa^{t=2}_2 = \kappa\). Substituting the constraint for worker group one in the equation and maximizing with respect to \(\kappa^{t=1}_1\) yields equation (15) in the text.

### 5 Derivation of Equation 18

Equation (16) in the text is derived by Bayesian updating. The following formula applies to the first period

\[
E_{t=1}(\tilde{X}|Edu, q, y_1) = \frac{1}{\sigma_0} E_{t=0}(\tilde{X}|Edu, q) + \frac{1}{\sigma^2_{t=1}} y_1
\]

For the second period this expectation equals

\[
E_{t=2}(\tilde{X}|Edu, q, y_1, y_2) = \frac{\left( \frac{1}{\sigma_0} + \frac{1}{\sigma^2_{t=1}} \right) E_{t=1}(\tilde{X}|Edu, q, y_1) + \left( \frac{1}{\sigma^2_{t=2}} \right) y_2}{\left( \frac{1}{\sigma_0} + \frac{1}{\sigma^2_{t=2}} + \frac{1}{\sigma^2_{t=2}} \right)}
\]

\[
= \left( \frac{1}{\sigma_0} \right) E_{t=0}(\tilde{X}|Edu, q) + \left( \frac{1}{\sigma^2_{t=2}} \right) y_2 + \left( \frac{1}{\sigma^2_{t=1}} \right) y_1
\]

\[
= \frac{\left( \frac{1}{\sigma_0} + \frac{1}{\sigma^2_{t=1}} \right) E_{t=1}(\tilde{X}|Edu, q, y_1) + \left( \frac{1}{\sigma^2_{t=2}} \right) y_2}{\left( \frac{1}{\sigma_0} + \frac{1}{\sigma^2_{t=2}} + \frac{1}{\sigma^2_{t=2}} \right)}
\]

\[27\]
For the  \( t^{th} \) period this expectation equals

\[
E_{t=n}(\tilde{X}|Edu,q) = \frac{\left( \frac{1}{\sigma_0^2} + \frac{1}{\sigma_{t=1}^2} + \frac{1}{\sigma_{t=2}^2} + \ldots + \frac{1}{\sigma_{t=n}^2} \right) E_{t=n-1}(\tilde{X}|Edu,q, y_1, y_2, \ldots, y_n) + \left( \frac{1}{\sigma_{t=n}^2} \right) y_2}{\left( \frac{1}{\sigma_0^2} + \frac{1}{\sigma_{t=1}^2} + \frac{1}{\sigma_{t=2}^2} + \ldots + \frac{1}{\sigma_{t=n}^2} \right)}
\]

\[
= \left( \frac{1}{\sigma_0^2} \right) E_{t=0}(\tilde{X}|Edu,q) + \left( \frac{1}{\sigma_{t=1}^2} + \frac{1}{\sigma_{t=2}^2} + \ldots + \frac{1}{\sigma_{t=n}^2} \right) E_{t=0}(\tilde{X}|Edu,q) + \sum_{i=1}^{n} \text{(signals)}
\]

\[
= \left( 1 - \left( \frac{\frac{1}{\sigma_0^2} + \frac{1}{\sigma_{t=1}^2} + \frac{1}{\sigma_{t=2}^2} + \ldots + \frac{1}{\sigma_{t=n}^2}}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma_{t=1}^2} + \frac{1}{\sigma_{t=2}^2} + \ldots + \frac{1}{\sigma_{t=n}^2}} \right) \theta_t \right) E_{t=0}(\tilde{X}|Edu,q) + \sum_{i=1}^{n} \text{(signals)}
\]

The change in \( \theta_t \) is equal to

\[
\Delta \theta_t = \theta_t - \theta_{t-1}
\]

\[
= \left[ 1 - \frac{(1 - \theta_t)}{(1 - \theta_{t-1})} \right] \cdot (1 - \theta_{t-1})
\]

\[
= \left[ 1 - \left( \frac{\frac{1}{\sigma_0^2} + \frac{1}{\sigma_{t=1}^2} + \frac{1}{\sigma_{t=2}^2} + \ldots + \frac{1}{\sigma_{t=n}^2}}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma_{t=1}^2} + \frac{1}{\sigma_{t=2}^2} + \ldots + \frac{1}{\sigma_{t=n}^2}} \right) \right] \cdot (1 - \theta_{t-1})
\]

\[
= \left[ 1 - \left( \frac{\frac{1}{\sigma_0^2} + \frac{1}{\sigma_{t=1}^2} + \frac{1}{\sigma_{t=2}^2} + \ldots + \frac{1}{\sigma_{t=n}^2}}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma_{t=1}^2} + \frac{1}{\sigma_{t=2}^2} + \ldots + \frac{1}{\sigma_{t=n}^2}} \right) \right] \cdot (1 - \theta_{t-1})
\]

\[
= \left[ 1 - \frac{V_t(\tilde{X}|Edu,q)}{V_{t-1}(\tilde{X}|Edu,q, y_1, y_2, \ldots, y_n)} \right] \cdot (1 - \theta_{t-1})
\]

\[
= (1 - 2^{2\kappa_1}) \cdot (1 - \theta_{t-1})
\]

\[
= %\Delta V_t \cdot (1 - \theta_{t-1})
\]
6 Derivation of equation 10

The FOC for the problem is:

$$FOC : -2 \ln 2 \cdot \left( \frac{\tilde{\pi}_{12}}{|\tilde{\pi}_{11}|} \right)^2 \cdot \sigma_{A_1}^2 \cdot 2^{-2\kappa_1} + 2 \ln 2 \cdot \left( \frac{\tilde{\pi}_{13}}{|\tilde{\pi}_{11}|} \right)^2 \cdot \sigma_{A_2}^2 \cdot 2^{-(\kappa - \kappa_1)}$$

Setting the FOC equal to zero and solving for $\kappa_1$ yields equation (10)