IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15807

Approaches to Learn about Employer Learning

Mahmut Ablay
Fabian Lange

DECEMBER 2022

IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15807

Approaches to Learn about Employer Learning

Mahmut Ablay
McGill University

Fabian Lange
McGill University, NBER, IZA, CIREQ and CIRANO

DECEMBER 2022

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Approaches to Learn about Employer Learning*

The empirical literature on employer learning assumes that employers learn about unobserved ability differences across workers as they spend time in the labor market. This article describes testable implications that arise from this basic hypothesis and how they have been used to quantify the contribution of Job Market Signaling and human capital in measured returns to education. While the empirical basis is still thin, the results suggest that Signaling contributes at most about 25% to the observed returns to education.

JEL Classification: E24, J31
Keywords: Job Market Signaling, human capital, returns to education, employer learning

Corresponding author:
Fabian Lange
McGill University
845 Rue Sherbrooke O
Montréal
QC H3A 0G4
Canada
E-mail: fabian.lange@mcgill.ca

* This research was undertaken, in part, thanks to funding from the Canada Research Chairs Program. We thank Georg Graetz for helpful feedback on this paper. While the research discussed in Section 4 was completed Prof Lange was also affiliated with the University of Oslo, Norway.

1 Introduction

What employers believe about the skills of workers is bound to determine their labor market outcomes. How schooling affects these beliefs is thus crucial for determining the returns to an additional year of schooling. Thus, schooling might act as a signal to the job market and this function might distort decisions to acquire schooling from the social optimum (Spence, 1973).

Job Market Signaling is difficult to test since it is about objects that are inherently hard to observe: the latent skills of workers and how costs of schooling vary with these skills. As such, it is difficult to measure how much Job Market Signaling contributes to the private returns to education and how much the social and the private returns diverge from each other due to Job Market Signaling.

This viewpoint describes one approach to quantify how large this gap might be, an approach that builds on the specification proposed of the common employer learning model proposed by Farber and Gibbons (1996) and further developed by Altonji and Pierret (2001). Central to this approach are the assumptions that (i) individual productivity has a persistent, time-invariant component, that (ii) individual compensation equals expected productivity, and (iii) that employers learn about this persistent component as workers spend time in the labor market. To empirically operationalize these assumptions, the literature on common (or public) employer learning that grew out of Farber and Gibbons (1996) and Altonji and Pierret (2001) explicitly formulates both how potential employers learn about unobserved skills of workers and what information about underlying skills researchers have access to. These assumptions are leveraged to derive testable predictions of the employer learning model for how earnings vary with schooling and proxies of productivity across the life-cycle.

In this paper, we use a simplified version of the employer learning model to illustrate this approach to quantify how important job market signaling is. Two papers form the basis of this discussion. Lange (2007) follows Altonji and Pierret (2001) and assumes that the researcher has access to a correlate of productivity that employers do not use in wage setting. That paper shows (i) how functional form assumptions can be used to summarize the process of employer learning in a single parameter, the "Speed of Employer Learning," (ii) derives conditions under which we can identify this parameter, and (iii) proceeds to estimate it. Lange (2007) goes on to impose additional structure on the schooling decision to bound the contribution of signaling to the economic returns from years of schooling.

The second paper, Aryal et al. (2022), takes a different but related approach. Again, the analysis starts with the specification proposed by Farber and Gibbons (1996) and shows under what circumstances one can interpret instrumental variable estimates of the causal effect of schooling on earnings as either representing private or social returns to education.[1] The main insight is that how one interprets the IV estimates depends on whether or not potential employers know that the instrument induces the variation in schooling in the population. To the extend that employers do know about this variation, instrumental variable estimates of the returns to education represent social returns to education. To the extend that they do not, they represent private returns.

As we discuss these approaches to bound or estimate the returns to signaling, we also reference empirical work that implement these approaches. In two appendices,

[1] A natural interpretation of the gap between social and private returns is the returns to job market signaling.

we both provide additional evidence (appendix A) and review the existing empirical literature (appendix B). Our view is that the evidence in this field suggests that on balance the reported evidence suggests that the returns to signaling make up at most 25% of the returns to education. However, we also find that the empirical basis for this conclusion is narrow and more work, using either of the approaches highlighted above, is necessary to firm up our confidence in this assessment. Ideally, new work will use the basic model following Lange (2007) to provide estimates of the learning process and the returns to signaling that can be compared with the existing estimates.

We begin by first introducing in Section 2 the basic model of wage setting and learning that we will rely on throughout the remainder of the article. It is based on but less general than the model analyzed by Farber and Gibbons (1996), Altonji and Pierret (2001), and Lange (2007). We then show in section 3 how access to a hidden correlate, that is a correlate of productivity not observed by employers, enables the researcher to test this model and estimate its parameters using the partial correlation of earnings with schooling and the hidden correlate across the life-cycle. The section also briefly reviews the empirical work based on a hidden correlate and discusses how Lange (2007) proposes to bound the returns to signaling. The following section, section 4, explores how to interpret instrumental variable regressions in a model of employer learning. Again, we need to make assumptions on what employers observe, notably how much they know about the variation in schooling induced by the instrument. Instrumental variables have the advantage that rather than bound the returns to schooling, they allow, under suitable assumptions, to point-identify the private and/or the social returns to schooling. Section 5 concludes.

2 A Model of Productivity, Wages, and Information

Assume that log productivity y_{it} is additively separable into an individual heterogeneity A_i, an experience-component $H(t)$ common to all, and an iid shock ε_{it}.

$$y_{it} = A_i + H(t) + \varepsilon_{it} \tag{1}$$

Next, use $\delta^{A|S}$ to define the causal effect of schooling on productivity and write

$$A_i = \delta^{A|S} S_i + \tilde{A}_i \tag{2}$$

$\tilde{A}_i$ than stands for unobserved components of ability that are not caused by schooling. Crucially, we do not restrict $cov(\tilde{A}_i, S_i) = 0$ so that schooling can be correlated with productivity not caused by schooling. The coefficient $\delta^{A|S}$ represents the social return to education as it captures the increase in productivity independent of whether it accrues to the worker.

Combining, we obtain that log output y_{it} depends on $(S_i, \tilde{A}_i, t)$ as follows:

$$y_{it} = \delta^{A|S} S_i + \tilde{A}_i + H(t) + \varepsilon_{it} \tag{3}$$

Further, assume that $(\tilde{A}_i, S_i, \varepsilon_{it})$ are jointly normally distributed and $\{\varepsilon_{it}\}$ is iid.

Throughout, we maintain that compensation W_{it} of individual i with experience t equals expected productivity conditional on the information employers at time t have access to. This information at experience level t is $\mathcal{E}_{it} = \{S_i, \{y_{i,\tau}\}_{\tau<t}, t\}$. This

information set grows with experience t which implies that employers learn about unobserved productivity $\tilde{A}_i$. The literature on employer learning hypothesizes that with time in the labor market more information is revealed and $\lim_{t\to\infty}(W_{it} - E[exp(y_{it})|\mathcal{E}_{it}]) = 0$. This becomes useful for differentiating between Human Capital and Signaling as we make additional functional form and substantive assumptions.

Two brief remarks are in order. First, this is a model of common employer learning in the sense that any information is shared among a sufficient number of competing employers to ensure that compensation equals expected productivity. We also rule out long-term contracts linking compensation to expected productivity over longer time-periods. Common employer learning is a strong assumption, but models of private or asymmetric employer learning entail complex strategic interactions on the part of firms and workers that make them hard to use to model the life-cycle of earnings.[2]

Second, it is straightforward to incorporate additional observable controls in the analysis. Importantly, Farber and Gibbons (1996) and all authors following in their footsteps also allow for employers to observe productivity correlates Q_i that researchers do not have access to. This implies that standard Mincer earnings equations will not identify the parameters of the learning model nor the private or the social returns to education as estimated returns to education will always be biased by the correlation between Q_i and S_i. This paper abstracts from both Q_i and observable controls as this simplifies the expressions and since we believe we can still communicate the basic arguments without these.

$Y_{i,t}$ is log normally distributed and we can write:

[2]Those interested in empirical models of asymmetric employer learning might want to consult, among others, Waldman (1984), Schönberg (2007), and Kahn (2013).

$$log(W_{it}) = log(\mathbb{E}[Y_{i,t}|\mathcal{E}_{it}]) = \delta^{A|S} S_i + H(t) + \frac{var(\tilde{A}_i + \varepsilon_{i,t}|\mathcal{E}_{it})}{2} + \mathbb{E}[\tilde{A}_i|\mathcal{E}_{it}] \tag{4}$$

One of the features of updating expectations about normal variables using normal signals is that the posterior variance does not depend on the realization of the signal itself. This property of normal-normal learning applies here and we therefore have that $var(\tilde{A}_i + \varepsilon_{i,t}|\mathcal{E}_{it}) = var(\tilde{A}_i + \varepsilon_{i,t}|t)$. Thus, we can control for $\tilde{H}(t) = H(t) + \frac{var(\tilde{A}_i+\varepsilon_{i,t}|t)}{2}$ in equation (4) with a set of experience controls. The interesting object is $\mathbb{E}[\tilde{A}_i|\mathcal{E}_{it}]$ and we can again draw on the properties of normal random variables to characterize this:

$$\mathbb{E}\left[\tilde{A}_i \mid \mathcal{E}_{\text{it}}\right] = \theta(t,\kappa) \underbrace{\left(\phi_{A|S} S_i\right)}_{\mathbb{E}\left[\tilde{A}_i|S\right]} + \left(1 - \theta(t,\kappa)\right) \bar{\xi}_i^t \tag{5}$$

where $\bar{\xi}_i^t = \frac{1}{t}\Sigma_{\tau \leq t}\{y_{i\tau} - \delta^{A|S} S_i - H(\tau)\}$. That is, at t, the observed output y_{it} net of the predictable component $\delta^{W|S} S_i + H(t)$ serves as a signal ξ_{it} on unobserved ability $\tilde{A}_i$ with noise ε_{it}. The average of these signals received up to period t is $\bar{\xi}_i^t$.

Equation (4) and (5) together produce the experience-t private return to an additional year of education. At $t = 0$, an additional year of education raises expected earnings because of the direct productivity effect $\delta^{A|S}$ and because it raises the inferred $\tilde{A}_i$ by $\phi_{A|S}$. At experience t, the private return $\delta_t^{W|S}$ an individual can expect from an additional year of schooling is thus

$$\delta_t^{W|S} = \delta^{A|S} + \theta(t,\kappa)\phi_{A|S} \tag{6}$$

Equation (5) shows that the conditional expectation of ability $\tilde{A}_i$ at experience t is a weighted average of the conditional expectation of ability prior to entry to the labor market (at $t = 0$) and the average signal received up to t. The relative weights $\theta(t, \kappa)$ depend only on the number of signals t received and one single parameter, κ.

$$\theta(t, \kappa) = \frac{1 - \kappa}{1 + (t - 1)\,\kappa} \text{ with } \kappa = \frac{\sigma_0}{\sigma_0 + \sigma_\varepsilon} \tag{7}$$

Here $var(\tilde{A}_i) = \sigma_0^2$ and $var(\varepsilon_{it})$ and the parameter κ is the signal-to-noise ratio in the learning process, ie the ratio of the variance of the object to learn about (ie $\tilde{A}_i$) and the variance of the signal $\xi_{it} = \tilde{A}_i + \varepsilon_{it}$.

This model of wages developed so far forms the basis of the discussion that follows. We invoke stronger assumptions than strictly required, sometimes significantly so, but they allow uncovering in a fairly transparent way the reasoning behind the empirical employer learning literature.

3 The Hidden Correlate Approach

Predictions

The two foundational papers in this literature, Farber and Gibbons (1996) and Altonji and Pierret (2001) combine a wage-setting model similar to the one developed in Section 2 with the assumption that the researcher has access to a variable Z that employers do not have direct access to. We will refer to this variable as a *hidden* correlate. Farber and Gibbons (1996), Altonji and Pierret (2001), and Lange (2007) assume access to such a hidden correlate and derive predictions for the partial regression coefficients on S and Z in log earnings regressions and how these evolve with experience.[3] A main reason for the popularity of this literature is that these predictions conform to simple intuitions about learning models.

So, what do we predict for a regression of log wages on a hidden correlate Z as well as schooling S at experience t? Substituting (5) into (4) we obtain the linear projection (denoted by $\mathbb{E}^*$) of log wages at experience t on (S_i, Z_i).

$$\begin{aligned} &\mathbb{E}^*[log(W_{it})|S, Z, t] = \alpha_t + \delta^{A|S} S_i + \mathbb{E}^*[\mathbb{E}[\tilde{A}_i|\mathcal{E}_{\text{it}}]] \\ &= \alpha_t + (\delta^{A|S} + \theta(t,\kappa)\phi_{A|S})S_i + (1-\theta(t,\kappa))\mathbb{E}^*[\bar{\xi}_i^t|S, Z] \end{aligned} \tag{8}$$

We note that $\mathbb{E}^*[\bar{\xi}_i^t|S, Z] = \mathbb{E}^*[\tilde{A}_i|S_i, Z_i]$ and use $(\hat{\beta}_{AS}, \hat{\beta}_{AZ})$ for the coefficients that project A_i on (S_i, Z_i). From this follow the coefficients projecting wages on (S_i, Z_i):

$$\hat{\beta}_{wS,t} = \delta^{A|S} + \theta(t,\kappa)\phi_{A|S} + (1-\theta(t,\kappa))\hat{\beta}_{AS} \tag{9}$$

$$\hat{\beta}_{wZ,t} = (1-\theta(t,\kappa))\hat{\beta}_{AZ} \tag{10}$$

These coefficients and their evolution over the life-cycle are the primary empirical objects of interest in the literature following Farber and Gibbons (1996), especially in Altonji and Pierret (1997), Altonji and Pierret (2001), and Lange (2007). Each of these papers can be associated with a different set of predictions on these coefficients.

[3] Farber and Gibbons (1996) actually consider regressions in levels.

First, Farber and Gibbons (1996) (in a level specification analogous to this log specification) show that $\{\hat{\beta}_{wZ,t}\}$ (eq. (10)) increase in experience if Z_i correlates positively with unobserved ability $\tilde{A}_i$. Second, Altonji and Pierret (2001) observe that when projecting log wages on S_i but not Z_i the projection coefficient on S_i exceeds $\hat{\beta}_{wS,t}$ by $(1-\theta(t,\kappa))\hat{\beta}_{AZ}cov(Z,S)/var(S)$.[4] Thus, across experience, one can use the emerging differences between the projection coefficients on schooling in a projection of log wages on schooling alone and on schooling and Z jointly to test whether employers screen for unobserved ability using schooling. Third, Lange (2007) exploits the functional form restrictions implied by normal-normal learning on $\theta(t,\kappa)$ (see eq. (7)) to summarize the learning process with the single parameter κ. This parameter can be estimated using the partial regression coefficients $\{\hat{\beta}_{wS,t}, \hat{\beta}_{wZ,t}\}$.

The contrast between equation (9) and the private return to education defined in equation (6) also shows that the projection of log earnings on education will not directly deliver an estimate of the social returns to education even when controlling for the hidden correlate Z_i as long as schooling correlates with the component of ability not caused by education.

Evidence

Empirically, the literature on employer learning has relied heavily on the NLSY 1979, the data used by Farber and Gibbons (1996), Altonji and Pierret (2001), and Lange (2007). The NLSY 1979 is a long panel of individuals during the early part of their careers with good earnings data as well as a cognitive skill measure that can serve as a hidden correlate.[5] This measure, the Armed Forces Qualification Test (AFQT)

[4] To derive the latter use the omitted variable formula.

[5] Of course, the fact that the NLSY 1979 is US data is another major advantage in securing a prominent place for this data-set in the literature.

score is based on a battery of tests administered to respondents to the NLSY 1979.

In 1979, the NLSY 1979 started collecting data on a panel of US residents born between 1957 and 1964. From 1997 on, collection of data on a new cohort born between 1980 and 1984 commenced. This panel is known as the NLSY 1997. Figure 1 shows the partial regression coefficients on schooling and the AFQT in regressions by year of experience in both data. Each scatter point represents the coefficient estimate on either schooling or the AFQT (measured in standard deviations) up to experience 17.[6]

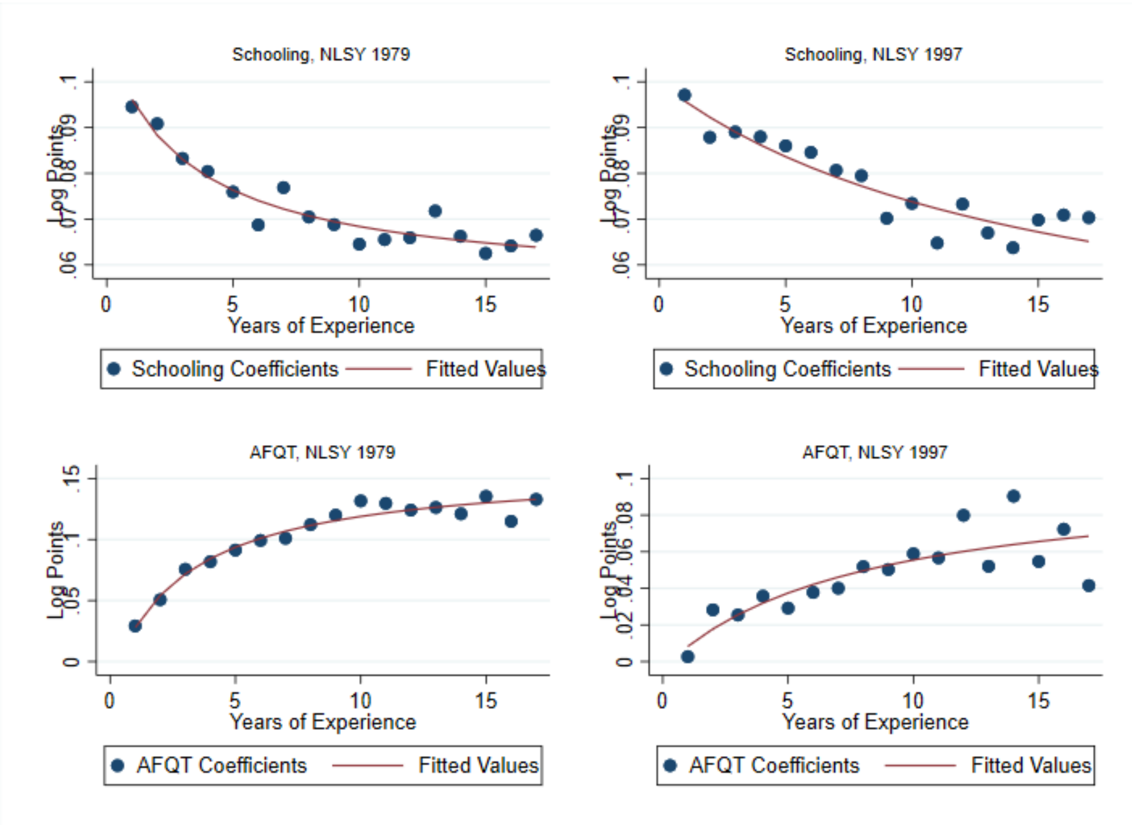
For both data-sets, we obtain that the regression coefficients on Z increase and those on S decrease with experience, as we expect in a learning model. Visual inspection suggests that the functional form implied by the normal-normal learning model developed above fits the data quite well. The line fitted to the experience profiles is obtained by estimating the parameter κ using the regression coefficients on S or Z and the restrictions of the normal learning model. As described by Lange (2007), these regression coefficients at experience t are weighted averages of the regression coefficients at experience t=0 and a limit value to which the coefficient converges as $t \to \infty$. Thus, the projection coefficients at any experience level are a function of 3 parameters only (the initial and limit values of the projection coefficients and the parameter κ) and we can estimate these by fitting the estimated coefficients using non-linear least squares.[7]

Data generated by other learning processes would generate the basic features

[6]The online appendix A contains information on the precise specification and data selection criteria.

[7]Our results differ from Castex and Kogan Dechter (2014) in that we find increasing returns to the AFQT and decreasing returns with experience for both males and females. The differences in the findings can be largely accounted for by the fact that we have access to a longer panel and estimate a non-linear specification. Additional details are available from the authors on request.

Figure 1: Returns to schooling and AFQT over the life cycle



Notes: The scatters display the estimated coefficients on schooling and the standardized AFQT score for each experience level. The line shows the predicted returns to schooling and AFQT score over the life cycle implied by the estimates in table 1. The estimation of these parameters is described in Online Appendix Section A2.

shown in figure 1 as well. The usefulness of the normal learning model is not in that it delivers a particularly good fit (even though it does), but that it allows summarizing the learning process in a simple manner whose main features can be summarized in a single parameter κ.

Table 1 shows the estimated speed of learning parameter κ for both data-sets and both variables (S_i, Z_i). Columns 1 and 2 reproduce the main empirical results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	NLSY 1979		NLSY 1997		Aryal, Bhuller, Lange (2022)		
	Years of Schooling	AFQT Score	Years of Schooling	AFQT Score	Years of Schooling	IQ Test Score	Years of Schooling (IV)
Speed of Learning κ	0.236	0.253	0.061	0.114	0.476	0.126	0.550
	(0.002)	(0.001)	(0.001)	(0.002)	(0.069)	(0.034)	(0.126)
Initial Errors Decline by %50 (Years)	3.238	2.953	15.455	7.808	1.101	6.936	0.818
Initial value b_0	0.109	-0.017	0.100	-0.004	0.100	0.011	0.195
	(0.007)	(0.001)	(0.000)	(0.012)	(0.007)	(0.005)	(0.012)
Limit value b_∞	0.055	0.159	0.034	0.102	0.023	0.093	0.055
	(0.001)	(0.000)	(0.027)	(0.042)	(0.002)	(0.006)	(0.008)

Table 1: The Speed of Employer Learning

Note: The reported parameters are estimated by nonlinear least squares using the coefficient estimates on schooling and Armed Forces Qualification Test (AFQT) score at different experience levels. The reported initial errors decline is obtained using eq. (7). The standard errors in columns 1-4 are obtained by bootstrapping with 5,000 repetitions. Columns 5-7 are from Aryal et al. (2022).

from Lange (2007) using data stemming from a longer time period and document identical findings as in Lange (2007). The estimated coefficient on κ for AFQT score (column 2) implies that initial expectation errors by employers on average decline by 25% after workers spend around one year in the labor market. After three years, the initial error will have declined by around 50%. After workers spend around 9 years in the labor market, initial expectation errors will on average have fallen by 75%, and after around 27 years they will have declined by 90%. The point estimate of κ, therefore, suggests that a worker's productivity is mostly revealed to the employer within the first few years of her labor market career.

The point estimates in columns 3 and 4 on the other hand suggest that employer learning is relatively slow for the cohort entering the labor market in the early 2000s. Specifically, the estimate of κ from column 4 implies that initial expectation errors by employers on average decrease by around 11% during the first year a worker spends in the labor market. It takes about 8 years for initial expectation errors to on average decline by 50%, 23 years to decline by 75%, and a full 70 years to decline by 90%.

This finding suggests that the speed of employer learning has slowed down between the 1980s and 2000s.

Columns 5 and 6 report analogous estimates based on a sample of Norwegian males and their IQ test scores collected as part of military conscription procedures.[8] The estimates from the Norwegian data bracket the estimates from the two NLSY cohorts. Those using schooling coefficients imply very rapid learning, while those using the IQ score imply learning similar to those from the NLSY 1979. Nevertheless, even the lower point estimate indicate that substantial learning about individual skills does take place in the first 10 years of an individual's career.[9]

Unfortunately, the evidence from other countries and settings is not as abundant as desirable. And, unfortunately, few of these papers explicitly use the learning structure proposed by Lange (2007) to provide parameter estimates of the Speed of Learning that would allow comparing and synthesizing across sometimes quite different economic environments. We review this literature in appendix B.

Bounds on the Returns to Signaling

Lange (2007) takes the analysis one step further and uses the estimates of the speed of learning to quantify the contribution of the private returns to signaling in the total gains from schooling.[10]

[8] The same data is discussed in section 4. The data consists of administrative data on the population of Norwegian males born between 1950 and 1980 with earnings data between 1967 and 2014.

[9] The NLSY cohorts have also been a rich source of data to explore heterogeneity and channels of learning. Arcidiacono et al. (2010), Mansour (2012), Light and McGee (2015) explore the role of education, occupations, and skill types in the employer learning framework. We review this literature in online appendix B.

[10] See Altonji and Pierret (1997) for a similar attempt to use estimates of how rapidly information is revealed to inform the question of how large signaling returns are.

Two aspects of this approach bear particular mentioning. First, since the information available to employers Q_i (which we abstracted from above) is not available to researchers, the estimate Lange (2007) arrives at is an upper bound for the contribution of signaling to the total return to education.

Second, he relies on additional information in the form of an estimate of the cost of schooling. Lange (2007) here follows the tradition in the literature that views schooling primarily as an investment and assumes that the cost of schooling is largely determined by the opportunity costs and thus the discount rate on future earnings. His estimate of the bound on the returns to signaling thus depends on the assumption that schooling costs are largely due to discounting as well as the particular rate of discounting used. Both are strong assumptions are central to the bound he provides. For discount rates between 3 and 7% and using his preferred estimate of the Speed of Learning parameter of $\kappa = 0.26$, he finds that the upper bound on the contribution of signaling to the overall private returns to schooling varies between 3 and 26%. The lower bound of the 95% confidence interval on the speed of learning parameter is $\kappa = 0.14$. If learning is this slow, then the upper bound varies between 6% and 47%.

4 Instrumental Variables and Employer Learning

The hidden correlate approach discussed above delivers testable predictions for Employer Learning and allows summarizing in a single, estimable parameter how rapidly firms learn about unobserved ability. And, with suitable assumptions, we can bound the contribution of signaling to the private returns to education. However, this approach does not produce a point estimate of the private or the social returns to education. And, the additional assumptions required to arrive at the bound on the Signaling returns are strong.

Aryal et al. (2022) show that instrumental variable estimates can provide useful information on the social and the private returns to education without having to invoke strong assumptions on the costs of schooling as required by Lange (2007). Rather, if the researcher is willing to take a stand on whether or not an instrument can be observed by employers, then the instrumental variable estimates can inform discussions about the productivity and signaling returns to education.

We follow Aryal et al. (2022) and discuss a binary instrument for schooling.[11] How to interpret the IV estimates depends on whether or not employers understand and price the variation induced by the instrument correctly. The usual reason offered to justify instrumenting is ability bias. Using our formulation in equation (2) the concern is that $\tilde{A}_i$ correlates with schooling S_i. Various instruments have been proposed claiming to increase schooling but to also be orthogonal to $\tilde{A}_i$. We are not interested in re-opening debates about these claims but posit that there indeed is an instrument $L_i \in \{0, 1\}$ that satisfies both the rank and orthogonality condition. That is, we assume that

$$\mathbb{E}[S_i|L_i = 1] \neq \mathbb{E}[S_i|L_i = 0] \tag{11}$$

$$\mathbb{E}[\tilde{A}_i|L_i] \equiv 0 \tag{12}$$

[11]Importantly, their argument extends beyond instrumental variables (binary or not) to all quasi-experimental approaches.

The IV coefficient for our binary instrument is (Wald, 1940):

$$b_{IV,t} = \frac{\mathbb{E}[log(W_{it})|L_i = 1] - \mathbb{E}[log(W_{it})|L_i = 0]}{\mathbb{E}[S_i|L_i = 1] - \mathbb{E}[S_i|L_i = 0]}.$$

And, using eq. (4) we have that $\mathbb{E}[log(W_{it})|L_i = 1] - \mathbb{E}[log(W_{it})|L_i = 0] = \delta^{A|S} * (\mathbb{E}[S_i|L_i = 1] - \mathbb{E}[S_i|L_i = 0]) + \mathbb{E}[\mathbb{E}[\tilde{A}_i|\mathcal{E}_{it}]|L_i = 1] - \mathbb{E}[\mathbb{E}[\tilde{A}_i|\mathcal{E}_{it}]|L_i = 0]$. Thus, we have

$$b_{IV,t} = \delta^{A|S} + \frac{\mathbb{E}[\mathbb{E}[\tilde{A}_i|\mathcal{E}_{it}]|L_i = 1] - \mathbb{E}[\mathbb{E}[\tilde{A}_i|\mathcal{E}_{it}]|L_i = 0]}{\mathbb{E}[S_i|L_i = 1] - \mathbb{E}[S_i|L_i = 0]} \tag{13}$$

Equation (13) shows that interpreting $b_{IV,t}$ comes down to making assumptions about how the instrument L_i enters into the information $\mathcal{E}_{it}$ held by employers. Aryal et al. (2022) focus on two polar cases. First, they consider "transparent instruments" that form part of the information held by firms so that $\mathcal{E}_{it} = \{S_i, \xi_i^t, L_i\}$. And, they constrast this with "hidden instruments" that are not observed by employers and thus do not form part of the firms information set. For these, $L_i \notin \mathcal{E}_{it}$.

For transparent instruments, applying the LIE to (13) and using (12) implies that $b_{IV,t} = \delta^{A|S}$. Intuitively, when potential employers are aware of the instrument, they will not use the induced variation in schooling to infer $\tilde{A}_i$. By assumption, we imposed that the instrument is orthogonal to the endogenous unobserved ability $\tilde{A}_i$. Thus, transparent IVs identify the effect of schooling on productivity itself: $\delta^{A|S}$.

For hidden instruments, we have from equation (5) that $\mathbb{E}[\tilde{A}_i|\mathcal{E}_{it}] = \theta(t,\kappa)\underbrace{(\phi_{A|S}S_i)}_{\mathbb{E}[\tilde{A}_i|S]} + (1-\theta(t,\kappa))\,\bar{\xi}_i^t$ and thus $\mathbb{E}[\mathbb{E}[\tilde{A}_i|\mathcal{E}_{it}]|L_i] = \theta(t,\kappa)\phi_{A|S}\mathbb{E}[S_i|L_i]$ since $\mathbb{E}[\bar{\xi}_i^t|L_i] = 0$. Substituting in (13), we find that for a hidden instrument, the instrumental coefficient is $b_{IV,t} = \delta^{A|S} + \phi_{A|S}\theta(t,\kappa)$. Hidden instruments thus identify the private returns to additional schooling which consist of both the productivity return and the signaling return. Employers are unaware of the instrument and thus variation in schooling induced by the instrument induces them to update their conjectures of $\tilde{A}_i$. Thus, hidden instruments identify the private returns to schooling.

Causal estimation using instruments already requires a priori assumptions on the validity of the instrument that cannot be tested. As such, when researchers employ instruments they need to carefully argue based on contextual evidence that an instrument will in fact satisfy the exclusion restrictions. In such cases, researchers might also be able to determine whether or not an instrument is in fact plausibly hidden or transparent. As Aryal et al. (2022) show, being willing to make such a determination reaps rewards in being able to identify deep policy-relevant parameters when the analysis is embedded in the employer learning framework.

For instance, as argued above, with a transparent instrument one can recover immediately the policy-relevant parameters $\delta^{A|S}$ that describe the productivity effect of schooling. A hidden instrument delivers the private returns $b_{IV,t} = \delta^{A|S} + \phi_{A|S}\theta(t, \kappa)$. But, using data from repeated cross-sections allows recovering $\delta^{A|S}$ as the limit to which $b_{IV,t}$ converges as $t \to \infty$. And, it also allows estimating κ using an approach analogous to that using a hidden correlate.

Of course, all this does not come for free. The assumptions on whether employers are aware of the instrument will be debatable. And, the strong identification results described above derive partially from the functional form assumptions, in particular the assumption that $\delta^{A|S}$ does not vary with experience. Aryal et al. (2022) and their companion working paper Aryal et al. (2019) however do show ways to relax some of these assumptions, especially if access to multiple instruments with different information assumptions is available.

Aryal et al. (2022) illustrate how to use this approach using a compulsory schooling law lengthening the mandatory school length. The data draws on administrative records for Norwegian males that also includes IQ scores collected as part of the military draft.[12]

The effect of the law was to extend minimum required schooling from 7 to 9 years across Norway, but the law was locally implemented at different times during the 1960-1975 period. The effect was to induce variation in the law across different jurisdictions which were often geographically very close. Frequently, small communities with a few thousand inhabitants would extend the compulsory schooling period at earlier or later dates than much larger jurisdictions within the same broad labor market. Empirically, this provides variation across locations and time that can be used to identify the causal impact of the law on wages. More important in the present context is that it allows dividing the sample in distinct populations that are separated by how plausible it is that the instrument is either hidden or transparent. As such, this setting provides an opportunity to illustrate and apply the ideas on transparent and hidden instruments.

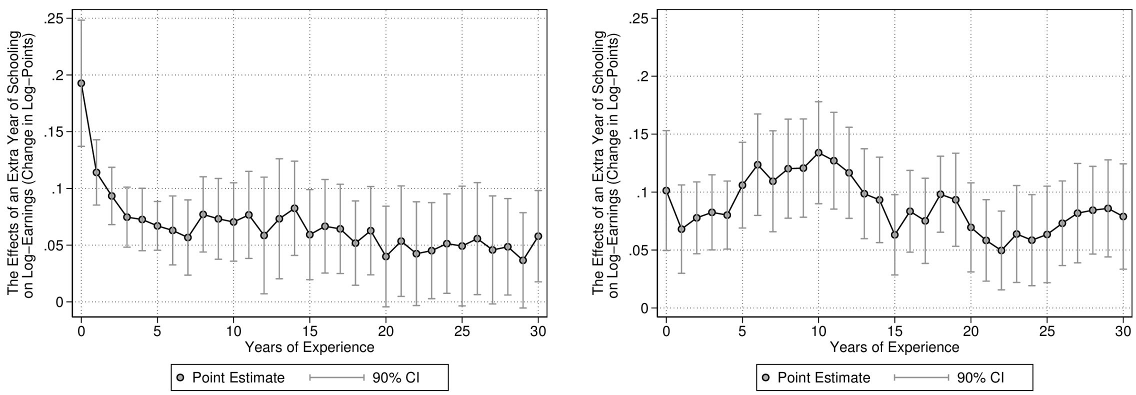
In particular, Aryal et al. (2022) argue that for the central communities in commuting zones in which the majority of the population of a given labor market lives, the variation induced by the instrument is plausibly understood by market participants. They interpret the estimates obtained from this sample as generated by a transparent instrument and thus they identify the social (or productivity) returns to additional schooling. By contrast, employers are much less likely to be informed of variation in the time when the law was implemented across small localities forming part of a large labor market. The instrument can thus plausibly be argued to be hidden for the sample of individuals growing up in smaller communities surrounding

[12]Columns 5 and 6 in Table 1 report estimates of the Speed of Learning measure based on these IQ scores and schooling measures using the hidden correlate approach discussed in Section 3.

the larger central community forming the core of a labor market. Thus, Aryal et al. (2022) construct two samples, one for which the instrument is plausibly hidden and one for which it is observed.

The coefficients on years of schooling over the life-cycle estimated using this instrument are shown in figure 2.

Figure 2: Returns to Schooling estimated using a Hidden and a Transparent Instrument



Notes: This figure reproduces figure 3 in Aryal et al. (2022). It shows the returns to education estimated using the compulsory schooling reform discussed in the main text. The right panel shows the estimates on the sample that permits interpreting the instrument as transparent. The left panel shows the estimates on the sample for which the instrument is hidden.

The right panel shows the estimates using the transparent instruments. These estimates do not display the typical decline in the returns to education associated with learning model, suggesting that the assumption that this instrument is transparent is indeed valid. We see that the returns to education are high and roughly stable across experience levels. By contrast, the estimates from the hidden sample (left panel) display the typical shape expected from a learning model when the instrument is not observed. The rapid decline in the estimated returns to education during the first few years in the labor market gives rise to the high estimated value for κ reported in column 7 of table 1.

What does all this imply for the productivity and signaling returns to education? Aryal et al. (2022) report a total return to education of 7.9% per year of schooling. 70% of is accounted for by a causal effect of schooling on productivity with the remainder due to signaling effects. These estimates are local average treatment effects applying to an extension of required minimum schooling 50 years in the past and might therefore be of limited utility for current use. We hope that more current estimates will emerge from other settings that explicitly try to implement this approach to inform on the social and private returns to education.

Aryal et al. (2022) are able to distinguish signaling and productivity effects of schooling because they are willing to take a stand on whether or not employers price the instrument correctly. This idea is related to a literature examining variation in education credentials induced by variation in test scores across a cut-off required for graduation (see the review of the related literature in online appendix B).[13] Clark and Martorell (2014) employ this approach to estimate the causal effect of a diploma by comparing individual who just passed or failed high school exit exams. This approach is clearly related to the instrumental variable approach as discussed by Aryal et al. (2022) in that employes don't observe the variation in test scores around the discontinuity. Thus, this regression discontinuity approach sets up a contrast between the treated group (those with scores just above the discontinuity) and the control group (just below) that is analogous to the hidden instrument case. Graetz

[13] An early related literature pointed to higher than usual returns to years of schooling required to complete a degree as evidence of information rents (Hungerford and Solon (1987),Layard and Psacharopoulos (1974)). However, those not completing degree years are likely to be negatively selected, casting doubts on these conclusions (Lange and Topel, 2006). Hungerford and Solon (1987) themselves note that diploma effects can't be understood without relying on screening.

(2021) provides a useful discussion of how RD estimates of returns to degrees or diplomas should be interpreted in light of the employer learning model.

5 Conclusion

This paper discussed how one approach to quantify the relative contributions of signaling versus human capital to the private return to education based on the assumption that firms learn over time about unobserved productive skills of individuals. In the process, we showed how to estimate the speed of employer learning. Our view of the literature is that the evidence on balance suggests that learning is rapid but not instantaneous. There is evidence from a variety of countries that supports the basic linear tests of employer learning following Altonji and Pierret (2001), but few papers estimate the Speed of Learning parameter (Lange, 2007) or use this parameter to bound the returns to signaling. Those that do tend to find that signaling accounts at most for one quarter of the returns to education, but we argue that the empirical basis for these conclusions is narrow.

The model of Job Market Signaling emphasizes that incomplete information distorts the returns to invest into observable signals of productivity, such as schooling. Incomplete information however also lowers the returns to other investments into one's productivity as long as these investments are not directly observable. How large these distortions are will generally depend on the Speed of Employer Learning. Kahn and Lange (2014) for instance show evidence of employer learning throughout individuals' careers. This evidence suggests a wedge between the social and private returns to on-the-job investments. Graetz (2022) explores how gradual employer learning affects the returns for unobserved investments through studying and effort while in school. Again, as in Kahn and Lange (2014), individuals will underinvest in their skills if their investments can not be directly observed. And, the degree to which they invest less than the first best depends on the Speed of Employer Learning. Obtaining better estimates of the Speed of Employer Learning and more generally of learning processes in the labor market should thus be high on the priority list of empirical labor economists.

References

Altonji, Joseph G., and C. Pierret (1997) 'Employer learning and the signaling value of education.' In *Industrial Relations, Incentives, and Employment,* ed. Ohashi and Tachibanaki (Macmillan Press Ltd., London) pp. 159–195

Altonji, Joseph G., and Charles R. Pierret (2001) 'Employer learning and statistical discrimination.' *The Quarterly Journal of Economics* 116(1), 313–350

Altonji, Joseph G, Prashant Bharadwaj, and Fabian Lange (2012) 'Changes in the characteristics of american youth: Implications for adult outcomes.' *Journal of Labor Economics* 30(4), 783–828

Araki, Shota, Daiji Kawaguchi, and Yuki Onozuka (2016) 'University prestige, performance evaluation, and promotion: Estimating the employer learning model using personnel datasets.' *Labour Economics* 41, 135–148

Arcidiacono, Peter, Patrick Bayer, and Aurel Hizmo (2010) 'Beyond signaling and human capital: Education and the revelation of ability.' *American Economic Journal: Applied Economics* 2(4), 76–104

Aryal, Gaurab, Manudeep Bhuller, and Fabian Lange (2019) 'Signaling and employer learning with instruments.' *NBER Working Paper*

_ (2022) 'Signaling and employer learning with instruments.' *American Economic Review* 112(5), 1669–1702

Barrera-Osorio, Felipe, and Hernando Bayona-Rodríguez (2019) 'Signaling or better human capital: Evidence from colombia.' *Economics of Education Review* 70, 20–34

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

Bordón, Paola, and Breno Braga (2020) 'Employer learning, statistical discrimination and university prestige.' *Economics of Education Review* 77, 101995

Broecke, Stijn (2015) 'Experience and the returns to education and skill in oecd countries: Evidence of employer learning?' *OECD Journal: Economic Studies* 2015(1), 123–147

Busso, Matías, Sebastián Montano, and Juan S Munoz-Morales (2022) 'Signaling specific skills and the labor market of college graduates'

Carroll, David, Chris Heaton, and Massimiliano Tani (2019) 'Does it pay to graduate from an'elite'university in australia?' *Economic Record* 95(310), 343–357

Castex, Gonzalo, and Evgenia Kogan Dechter (2014) 'The changing roles of education and ability in wage determination.' *Journal of Labor Economics* 32(4), 685–710

Chatterji, Monojit, Paul T Seaman, and Larry D Singell Jr (2003) 'A test of the signalling hypothesis.' *Oxford Economic Papers* 55(2), 191–215

Cheung, Stephen L (2010) 'A test of employer learning in the labour market for young australians.' *Applied Economics Letters* 17(1), 93–98

Clark, Damon, and Paco Martorell (2014) 'The signaling value of a high school diploma.' *Journal of Political Economy* 122(2), 282–318

Comi, Simona Lorena, and Mara Grasseni (2021) 'Employer learning during apprenticeship.' *Applied Economics Letters* 28(18), 1548–1551

Di Pietro, Giorgio (2017) 'Degree classification and recent graduates' ability: Is there any signalling effect?' *Journal of Education and Work* 30(5), 501–514

Ehrmantraut, Laura, Pia Pinger, and Renske Stans (2020) 'The expected (signaling) value of higher education.' *CESifo Working Paper Series 8589, CESifo.*

Falter, Jean-Marc (2006) 'Employer learning and literacy scores in switzerland.' *Available at SSRN 952519*

Farber, Henry S., and Robert Gibbons (1996) 'Learning and wage dynamics.' *Quarterly Journal of Economics* 111(4), 1007–1047

Feng, Andy, and Georg Graetz (2017) 'A question of degree: the effects of degree class on labor market outcomes.' *Economics of Education Review* 61, 140–161

Freier, Ronny, Mathias Schumann, and Thomas Siedler (2015) 'The earnings returns to graduating with honors—evidence from law graduates.' *Labour Economics* 34, 39–50

Galindo-Rueda, Fernando (2003) 'Employer learning and schooling-related statistical discrimination in britain.' *Available at SSRN 412483*

Ge, Suqin, Andrea Moro, and Beibei Zhu (2021) 'Testing for asymmetric employer learning and statistical discrimination.' *Applied Economics* 53(12), 1361–1377

Graetz, Georg (2021) 'On the interpretation of diploma wage effects estimated by regression discontinuity designs.' *Canadian Journal of Economics/Revue canadienne d'économique* 54(1), 228–258

_ (2022) 'Imperfect signals.' *Working Paper*

Hansen, Anne Toft, Ulrik Hvidman, and Hans Henrik Sievertsen (2021) 'Grades and employer learning.' Technical Report, IZA Discussion Papers

Hensvik, Lena, and Oskar Nordström Skans (2016) 'Social networks, employee selection, and labor market outcomes.' *Journal of Labor Economics* 34(4), 825–867

Hungerford, Thomas, and Gary Solon (1987) 'Sheepskin effects in the returns to education.' *The review of economics and statistics* pp. 175–177

Jaeger, David A, and Marianne E Page (1996) 'Degrees matter: New evidence on sheepskin effects in the returns to education.' *The review of economics and statistics* pp. 733–740

Kahn, L. B. (2013) 'Asymmetric information between employers.' *American Economic Journal: Applied Economics* 5(4), 165–205

Kahn, Lisa B., and Fabian Lange (2014) 'Employer learning, productivity, and the earnings distribution: Evidence from performance measures.' *Review of Economic Studies* 81(4), 1575–1613

Kane, Thomas J, and Cecilia E Rouse (1993) 'Labor market returns to two-and four-year colleges: is a credit a credit and do degrees matter?'

Khoo, Pauline, and Ben Ost (2018) 'The effect of graduating with honors on earnings.' *Labour Economics* 55, 149–162

Kim, Seik, and Emiko Usui (2021) 'Employer learning, job changes, and wage dynamics.' *Economic Inquiry* 59(3), 1286–1307

Lang, Kevin, and Erez Siniver (2011) 'Why is an elite undergraduate education valuable? evidence from israel.' *Labour Economics* 18(6), 767–777

Lange, Fabian (2007) 'The speed of employer learning.' *Journal of Labor Economics* 25(1), 1–35

Lange, Fabian, and Robert Topel (2006) 'The social value of education and human capital.' *Handbook of the Economics of Education* 1, 459–509

Layard, Richard, and George Psacharopoulos (1974) 'The screening hypothesis and the returns to education.' *Journal of political economy* 82(5), 985–998

Lesner, Rune V (2018) 'Testing for statistical discrimination based on gender.' *Labour* 32(2), 141–181

Light, Audrey, and Andrew McGee (2015) 'Employer learning and the "importance" of skills.' *Journal of Human Resources* 50(1), 72–107

MacLeod, W Bentley, Evan Riehl, Juan E Saavedra, and Miguel Urquiola (2017) 'The big sort: College reputation and labor market outcomes.' *American Economic Journal: Applied Economics* 9(3), 223–61

Mansour, Hani (2012) 'Does employer learning vary by occupation?' *Journal of Labor Economics* 30(2), 415–444

Mazrekaj, Deni, and Sofie Cabus (2019) 'Does a high school diploma matter? evidence using regression discontinuity design.' Technical Report, KU Leuven working paper

Pan, Shih-Yi (2005) 'Employer learning and statistical discrimination in the canadian labour market.' PhD dissertation

Petre, Melinda (2018) ‘Are employers omniscient? employer learning about cognitive and noncognitive skills.’ *Industrial Relations: A Journal of Economy and Society* 57(3), 323–360

Pinkston, Joshua C (2009) ‘A model of asymmetric employer learning with testable implications.’ *The Review of Economic Studies* 76(1), 367–394

Schönberg, Uta (2007) ‘Testing for asymmetric employer learning.’ *Journal of Labor Economics* 25(4), 651–691

Schwerdt, Guido, and Ludger Woessmann (2017) ‘The information value of central school exams.’ *Economics of Education Review* 56, 65–79

Spence, Michael (1973) ‘Job market signaling.’ *The quarterly journal of economics* pp. 355–374

Strobl, Eric (2004) ‘Do employers use education as a signal for ability in developing countries? evidence from ghana.’ *Applied Economics Letters* 11(4), 259–261

Wald, Abraham (1940) ‘The fitting of straight lines if both variables are subject to error.’ *The annals of mathematical statistics* 11(3), 284–300

Waldman, Michael (1984) ‘Job assignments, signalling, and efficiency.’ *The RAND Journal of Economics* 15(2), 255–267

Wang, Jun, and Bo Li (2020) ‘Does employer learning with statistical discrimination exist in china? evidence from chinese micro survey data.’ *International Review of Economics & Finance* 69, 319–333

Zhangaliyeva, Aigerim, Masaki Nakabayashi et al. (2019) ‘Legacy of czar: The russian dual system of schooling and signaling.’ Technical Report, Institute of Social Science, The University of Tokyo

Online Appendix

A Data, Estimation, and Sensitivity Analysis

A.1 Data and Sample Selection

A.1.1 NLSY 1979

The data used in this study stem from the 1979 and 1997 waves of the National Longitudinal Survey of Youth (NLSY). The NLSY 1979 is a nationally representative sample of 12,686 men and women. The survey was conducted annually from 1979 to 1993 and since then participants were interviewed biannually. The NLSY 1979 consists of three samples. The cross-sectional main sample consists of 6,111 young non-institutionalized men and women aged 14 to 21 at the first interview in 1979. The supplemental sample, which consists of 5,295 young men and women, oversamples the Hispanic, black, or disadvantaged white individuals. The military sample consists of 1,280 youths aged 17-21 at the time of first interview.

We closely follow Lange (2007) to construct our sample. We employ the cross-sectional main sample of the NLSY 1979. Our analysis use individual-year observations from all genders and races. We removed 808 individuals with no valid information on AFQT score. Subsequently, we drop individuals for which the only reported observations occur prior to graduation and restrict the sample to individuals with at least 6 years of completed education. We restrict the analysis to years in which individuals are not enrolled in school, work for pay and earn between \$1 and \$100 hourly wages. We follow Lange (2007) and limit the sample to observations with 17 years of experience or less. The analysis sample consists of 5,253 individuals with 47,261 observations stemming from the 1979–2014 waves of the NLSY 1979.

Panel A of table A1 summarizes the statistics of the key variables in the NLSY 1979 sample. The wage is the natural logarithm of the real hourly rate of pay measured in cents for the current or most recent job. Hourly wages are indexed to 2019 using the Consumption Price Index (CPI). We follow Lange (2007) and calculate experience as years since left school for the first time. The year of graduation is defined as the last year respondents attended formal education before leaving school for the first time. The AFQT scores are measured by the respondent's score on the Armed Forces Qualifying Test (AFQT). Altonji et al. (2012) construct a mapping of the AFQT score across NLSY waves to account for differences in age-at-test, test format, and other idiosyncrasies. We take comparable AFQT scores from Altonji et al. (2012) and normalize them separately for each sample to have a mean of 0 and a standard deviation of 1.

A.1.2 NLSY 1997

The NLSY 1997 is a nationally representative sample of 8,984 young men and women aged 12–17 in 1997. The survey was conducted annually from 1997 to 2011 and biennially starting from 2011. The NLSY 1997 follows a nearly identical structure to the NLSY 1979 and provides information on respondents' background characteristics, education, and labor market outcomes. We follow the same sample selection criteria as above to construct the NLSY 1997 sample. It is not possible to construct the AFQT score for 1,982 individuals. We drop individuals with less than 6 years of completed education. Subsequently, we restrict the sample to years in which individuals are not enrolled in school, work for pay, and earn between \$1 and \$100 hourly wages. We drop observations with less than 1 and more than 17 years of experience. This leaves us with 6,462 individuals with 45,391 observations stemming from the 1997–2019 waves of the NLSY 1997.

Panel B of table A1 summarizes the statistics of the key variables in the NLSY 1997 sample. The variables are defined in the same way as in the NLSY 1979 sample. Hourly wages are indexed to 2019 using the Consumption Price Index (CPI). We measure the first graduation year using actual graduation dates along with monthly schooling histories.

	(1) Observation	(2) Mean	(3) Std. dev.	(4) Min	(5) Max
	Panel A: NLSY 1979				
ln (wage)	47,261	7.418	0.524	4.629	9.205
Highest grade completed	47,261	13.155	2.299	6	20
Standardized AFQT	47,261	0	1	-2.769	1.729
Experience	47,261	8.428	4.623	1	17
Female	47,261	0.487	0.500	0	1
Black	47,261	0.109	0.311	0	1
Hispanic	47,261	0.064	0.245	0	1
	Panel B: NLSY 1997				
ln (wage)	45,391	7.367	0.545	4.640	9.210
Highest grade completed	45,391	13.257	2.751	6	20
Standardized AFQT	45,391	0	1	-2.591	1.820
Experience	45,391	7.321	4.541	1	17
Female	45,391	0.481	0.50	0	1
Black	45,391	0.250	0.433	0	1
Hispanic	45,391	0.201	0.400	0	1

Table A1: Summary statistics

Note: All statistics are unweighted. Hourly wages are indexed to 2019 using the CPI. The data source is the 1979 and 1997 waves of the National Longitudinal Survey of Youth. The data stems from the 1979–2014 waves of the NLSY 1979 and the 1997–2019 waves of the NLSY 1997. The AFQT scores are measured by the respondent's score on the Armed Forces Qualifying Test (AFQT) and are normalized to have a mean of 0 and a standard deviation of 1 for each sample. Test format and age-at-test adjusted comparable AFQT scores are taken from Altonji et al. (2012). Education is the completed years of education in a given year.

A.2 Estimating the Speed of Employer Learning

We follow Lange (2007) and perform our analysis in two steps. In the first step, we estimate a wage regression that relates log wages to schooling and ability over the life cycle. Specifically, the following estimation equation regresses log wages on schooling s and ability z interacted with a complete set of experience dummies.

$$\log\left(W_{t,x}\right) = \Sigma_x \beta_{ws,x}\left(sD_x\right) + \Sigma_x \beta_{wz,x}\left(zD_x\right) + \beta_\Phi' \Phi_{i,t} + \varepsilon_x \tag{A1}$$

$W_{t,x}$ is the real hourly rate of pay at time t at experience level x indexed to 2019 dollars. D_x is an indicator function taking the value one if the experience is x and zero otherwise. $\Phi_{i,t}$ is a vector of year-fixed effects and demographic dummies including gender and race. We estimate this specification separately for both NLSY samples and obtain the coefficient estimates $\left\{\hat{\beta}_{ws,x}, \hat{\beta}_{wz,x}\right\}_{x=0}^{x=T}$. As described in great detail in Lange (2007), these are known functions of the structural parameters, $\{b_{s,0}, b_{s,\infty}, b_{z,0}, b_{z,\infty}, \kappa\}$ where $b_{s,0}$ and $b_{z,0}$ are initial values of schooling and ability coefficients and $b_{s,\infty}$ and $b_{z,\infty}$ are limit values. The following equation characterizes the structural relationship between five parameters of interest and coefficients.

$$\{\beta_{ws,x}, \beta_{wz,x}\}_{x=0}^{T} = \{(1-\theta_x)\, b_{s,0} + \theta_x b_{s,\infty}\, (1-\theta_x)\, b_{z,0} + \theta_x b_{z,\infty}\}_{x=0}^{T}. \tag{A2}$$

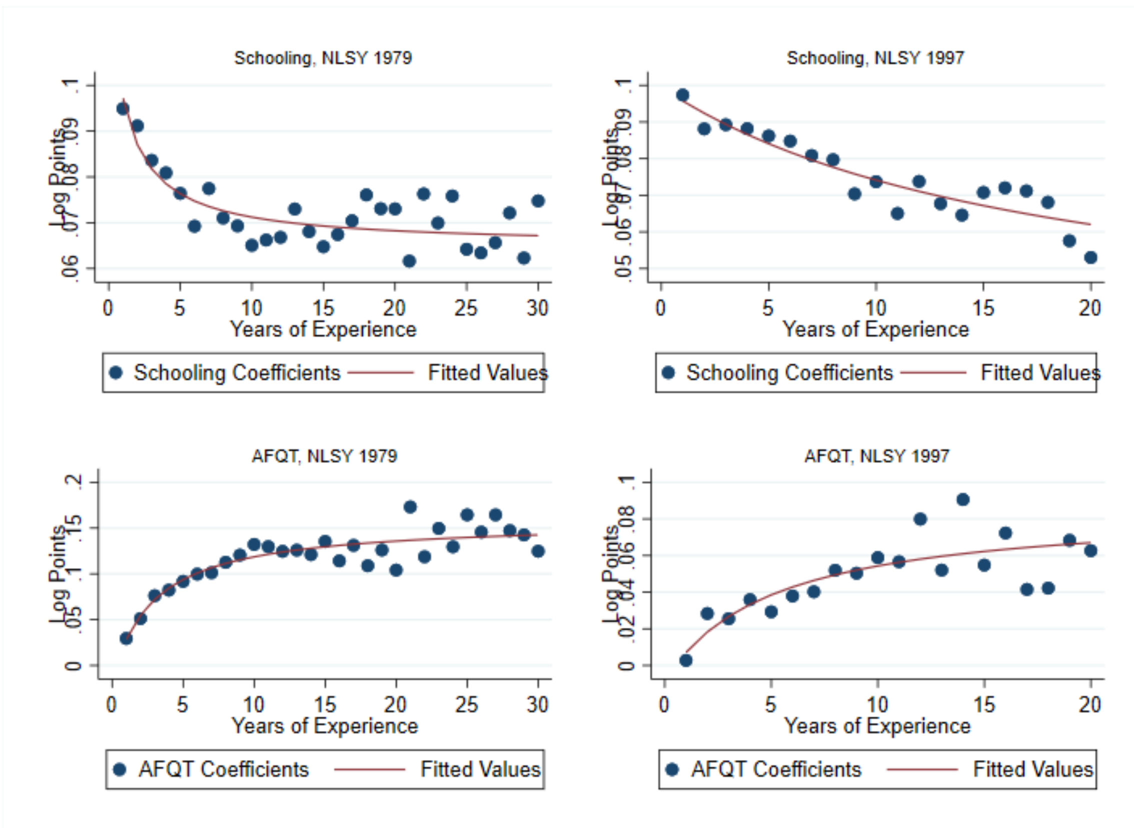
where θ_x is defined as $\theta_x = x\kappa / \left[1 + (x-1)\kappa\right])$ and represents the weights on initial and limit values of regression coefficients. This equation implies that the regression coefficients at experience level x are weighted averages of regression coefficients at experience level $x = 0$ (initial values $b_{s,0}$ and $b_{z,0}$) and at experience level $x = \infty$ (limit values $b_{s,\infty}$ and $b_{z,\infty}$). This suggests that the estimated coefficients, $\left\{\hat{\beta}_{ws,x}, \hat{\beta}_{wz,x}\right\}_{x=0}^{x=T}$, at any experience level are characterized by three parameters: the speed of learning parameter κ, initial values, and limit values.

We treat each of the estimated coefficients $\left\{\hat{\beta}_{ws,x}, \hat{\beta}_{wz,x}\right\}_{x=0}^{x=T}$ as an observation to estimate five structural parameters $\{b_{s,0}, b_{s,\infty}, b_{z,0}, b_{z,\infty}, \kappa\}$ and thus obtain the speed of employer learning. We estimate structural parameters by fitting two non-linear functions to the estimated coefficients using the method of nonlinear least squares. Specifically, we fit the nonlinear function of schooling, $b_s(x) = (1 - \theta_x)\, b_{s,0} + \theta_x b_{s,\infty}$, to $\left\{\hat{\beta}_{ws,x}\right\}_{x=0}^{x=T}$ by the choice of $\{b_{s,0}, b_{s,\infty}, \kappa\}$. Similary, we fit the non-linear function of ability, $b_z(x) = (1 - \theta_x)\, b_{z,0} + \theta_x b_{z,\infty}$, to $\left\{\hat{\beta}_{wz,x}\right\}_{x=0}^{x=T}$ by the choice of $\{b_{z,0}, b_{z,\infty}, \kappa\}$. We follow Lange (2007) and set T to 17 in the main analysis. Also, we allow T to go up to 30 for the NLSY 1979 and 20 for the NLSY 1997 and find similar results (see figure A1 and table A2).

A.3 Sensitivity Analysis

A.3.1 Restrict sample to a new maximum experience

Figure A1: Returns to schooling and AFQT over the life cycle



Note: The scatters display the estimated coefficients on schooling and the standardized AFQT score for each experience level. The line shows the predicted returns to schooling and AFQT score over the life cycle implied by the estimates in table A2.

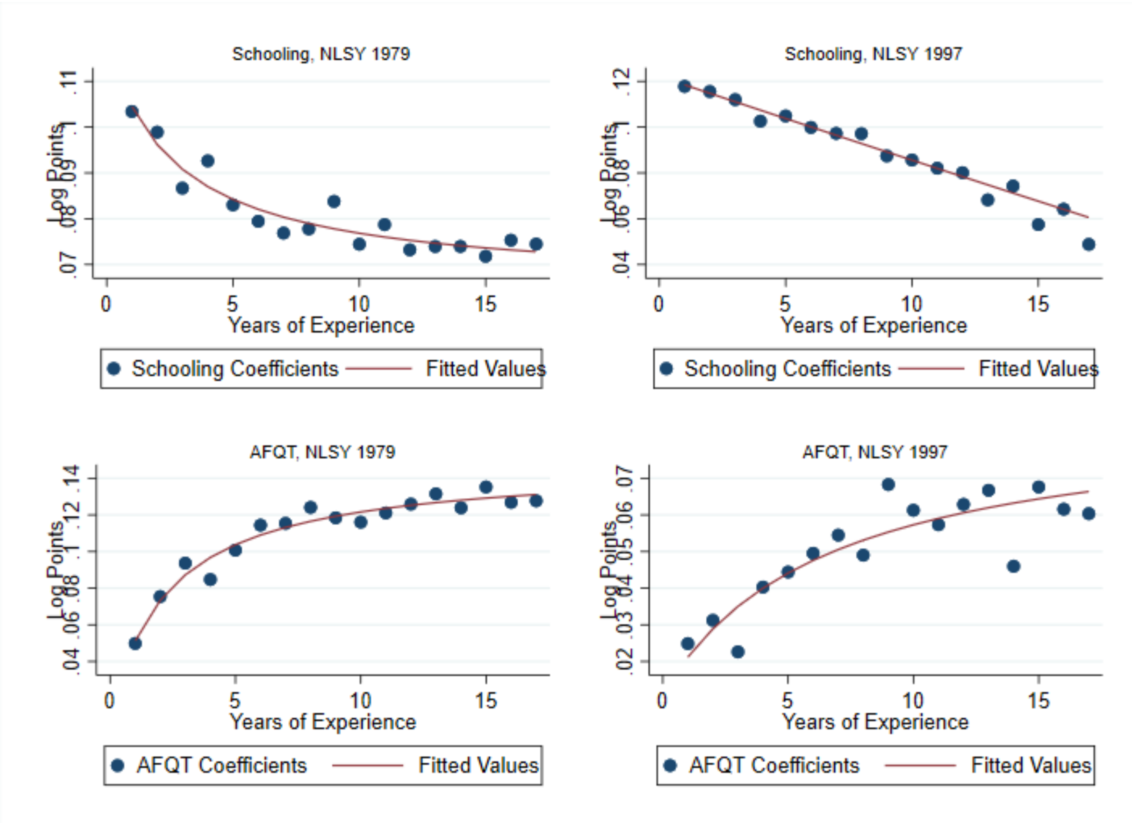
	(1)	(2)	(3)	(4)
	NLSY 1979		NLSY 1997	
	Schooling	AFQT	Schooling	AFQT
Speed of Learning κ	0.459	0.255	0.052	0.165
	(0.003)	(0.001)	(0.001)	(0.052)
Initial value b_0	0.125	-0.016	0.099	-0.009
	(0.022)	(0.000)	(0.000)	(0.019)
Limit value b_∞	0.065	0.158	0.028	0.086
	(0.000)	(0.000)	(0.037)	(0.010)

Table A2: The Speed of Employer Learning

Note: The reported parameters are estimated by nonlinear least squares using the coefficient estimates on schooling and Armed Forces Qualification Test (AFQT) score at different experience levels. The NLSY 1979 sample consists of individuals who accumulate less than or equal to 30 years of labor market experience and the NLSY 1997 sample includes individuals who accumulate less than or equal to 20 years of labor market experience. The standard errors are obtained by bootstrapping with 5,000 repetitions.

A.3.2 Classic potential experience measure

Figure A2: Returns to schooling and AFQT over the life cycle



Note: The scatters display the estimated coefficients on schooling and the standardized AFQT score for each experience level. The line shows the predicted returns to schooling and AFQT score over the life cycle implied by the estimates in table A3.

	(1)	(2)	(3)	(4)
	NLSY 1979		NLSY 1997	
	Schooling	AFQT	Schooling	AFQT
Speed of Learning κ	0.271	0.298	0.001	0.131
	(0.003)	(0.002)	(0.000)	(0.125)
Initial value b_0	0.119	0.010	0.112	0.011
	(0.008)	(0.015)	(0.000)	(0.016)
Limit value b_∞	0.065	0.148	-2.659	0.088
	(0.005)	(0.002)	(0.175)	(0.017)

Table A3: The Speed of Employer Learning with classic potential experience

Note: The reported parameters are estimated by nonlinear least squares using the coefficient estimates on schooling and Armed Forces Qualification Test (AFQT) score at different experience levels. Both samples consist of individuals who accumulate less than or equal to 17 years of labor market experience. The labor market experience is defined as age-years of education-7. The standard errors are obtained by bootstrapping with 5,000 repetitions.

B Scholarly Literature

This appendix reviews the literature on employer learning and statistical discrimination following and based on Farber and Gibbons (1996), Altonji and Pierret (2001), and Lange (2007).

B.1 US Evidence based on the NLSY 1979 and NLSY 1997

Much of the literature relies on a single data source, the NLSY 1979. In particular, the rich tapestry of information in the NLSY 1979 lends itself to explore heterogeneity in employer learning. Arcidiacono et al. (2010) investigate heterogeneity across education levels. They document evidence for employer learning subsequent to graduating from high school among those who do not continue with higher education. By contrast, much of the discovery process about skills of college graduates is completed by the time these graduate from college. Mansour (2012) studies the heterogeneity in learning across occupations. His findings reveal that employer learning varies across workers' initial occupations. He relates differences in the growth rate of wage residual variance across occupations to differences in the occupation-specific speed of employer learning and finds that these are related, as predicted by the model. Light and McGee (2015) study whether the employer learning process depends on skill type and on the importance of skill for the occupation. By merging data from O*NET with the NLSY 1979, they document that employer learning occurs for each skill type, for college and high school graduates, and for white- and blue-collar workers. Also, they show that employer learning does not vary across workers or skill types. There are some indicators of non-cognitive skills in the NLSY 1979, and these are used by Petre (2018) to study employer learning about non-cognitive skills. The findings suggest that employers reward self-esteem, internal control, and schooling initially and learn about cognitive ability and workers' motivation as workers gain experience in the labor market.

Castex and Kogan Dechter (2014) use both the 1979 and the 1997 waves of the NLSY and study the changing role of education and ability in wage determination. Their analysis suggests that returns to formal education increased between the 1980s and 2000s while the return to cognitive ability decreased during the same time period. They also estimate the empirical employer learning model developed by Altonji and Pierret (2001) and while they find evidence in support of employer learning for the NLSY 1979 cohort, they do not find that the Mincerian returns to AFQT and schooling change with experience in the NLSY 1997 cohort. In this, they differ from the present paper which documents that the return to schooling declines while return to the AFQT score increases with experience in both the NLSY 1979 and the NLSY 1997.[14] Our results do however suggest a lower speed of employer learning in the more recent NLSY cohort.

[14]The differences can largely be accounted for by the fact that we use significantly more data and estimate a non-linear specification. Castex and Kogan Dechter (2014) use data from the NLSY 1997 up to 2008 which restricts them to ages 19-28. Our estimates are based on individuals 15-39 in the NLSY 1997, which is particularly valuable for obtaining data on those with Bachelor's degrees. In addition, Castex and Kogan Dechter (2014) drop observations not enrolled in school based on the year school enrolment variable, a variable that is frequently missing. We follow the same strategy, but fill in the school enrolment using monthly schooling histories when the enrollment variables are missing. Moving to a non-linear specification as in 1 is particularly important when estimating the employer learning model across a wide age range.

B.2 International Evidence following the Hidden Correlate approach

The predictions of the employer learning model have also been explored in non-US data. Evidence for employer learning is found in Germany (Bauer and Haisken-DeNew, 2001), Switzerland (Falter, 2006), Sweden (Hensvik and Skans, 2016), Canada (Pan, 2005), and Britain (Galindo-Rueda, 2003). The evidence from Denmark is mixed (Lesner, 2018) in that evidence for learning is found when using fathers income and years of schooling as a hidden correlate, but not when using birth-weight. Interestingly, Bauer and Haisken-DeNew (2001) report evidence for employer learning only among low-wage, blue-collar workers and similarly Galindo-Rueda (2003) report that learning is particularly strong for blue-collar workers. Combined with the evidence from Arcidiacono et al. (2010), this suggests that employer learning might be particularly strong among those with relatively low levels of education. Lesner (2018) develops an employer learning model which incorporates screening discrimination, stereotyping, and prejudiced beliefs.

Strobl (2004) studies the employer learning hypothesis in Ghana and finds evidence in support for employer learning but only among workers without occupational training and with employment through formal channels. Wang and Li (2020) finds evidence in support of employer learning using Chinese data - and this study finds stronger evidence for advantaged workers such as males, college graduates, highly skilled workers, high-wage earners, and workers in large businesses or the urban labor market.

Comi and Grasseni (2021) exploit an exogenous change in the number of apprenticeship contracts in Italy to study the relationships between training during apprenticeship and employer learning about worker ability. Their findings show that employers use apprenticeships to learn about worker ability and use this learned information to screen workers when offering continued employment subsequent to the apprenticeship. They also show apprentices that are revealed to be of high ability receive more training, potentially putting into question the assumption in Farber and Gibbons (1996), Altonji and Pierret (2001), and Lange (2007) that productivity growth with experience is unrelated to unobserved ability.

Two papers that fail to find evidence for employer learning are Cheung (2010) and Zhangaliyeva et al. (2019). Cheung (2010) reports that ability of Australian males aged 19–26 years is perfectly observed at the time of labor market entry. Similarly, Zhangaliyeva et al. (2019) find no evidence supporting the employer learning hypothesis in Russia.

Broecke (2015) estimate the specification proposed by Altonji and Pierret (2001) using data from the Program of International Assessment of Adult Competencies for 22 OECD countries. For 18 out of 22 countries they find a positive interaction between their skill measure and skill. Out of these, 8 are significant at the 5% level. The results reported by Broecke (2015) also report the interaction between years of schooling and experience with and without controlling for their skill measure interacted with education. For 19 of the 22 countries this interaction declines once controlling for skill*experience, while it increases for the remaining 3.[15] Our reading of Broecke (2015) is that the patterns in the interactions between experience and ability and schooling that Altonji and Pierret (2001) propose as evidence of employer learning and statistical discrimination are generally found in the international data. At times, there is too little power to reject the absence of these patterns, but only

[15] The decline is significant for 6 countries for which the schooling-experience interaction declines and for 1 for which it increases.

rarely do studies report rejections of the employer learning hypothesis.

B.3 Evidence based on Diploma Effects

A related literature considers what has been termed the "sheepskin effect" that is excess correlation of earnings with degree completion exceeding the Mincerian returns observed for non-degree years. Hungerford and Solon (1987) and Layard and Psacharopoulos (1974) are the starting point for this literature. Hungerford and Solon (1987) estimate a standard Mincer earning equation but allow for excess returns during years of schooling commonly associated with degree completion (S=8,12,16). Returns associated with these degree years exceed those for non-degree years, a finding that might be considered evidence for signaling. However, as the authors themselves acknowledge, this robust and common finding (see also Kane and Rouse (1993); Jaeger and Page (1996); Chatterji et al. (2003); Ehrmantraut et al. (2020)) is open to alternative explanations such as dynamic selection (see eg. Lange and Topel (2006)). As few students drop out in the final year of schooling foregoing the degree, the concern is that degree effects are estimated using a small number of students who are systematically selected on negative realizations during the final year of a degree program.

One way to address this selection concern is to compare individuals that differ only by small differences in their performance on an exam that result in them receiving a degree or not. This approach is taken by Clark and Martorell (2014) and Mazrekaj and Cabus (2019). They show that average earnings between workers who barely passed and barely failed high school exit exams are not significantly different from each other, suggesting that the signaling value of a high school diploma is close to zero.

The other set of papers studies the signaling value of degree class among students with similar lengths of study. The researchers in this line of the literature extensively use a Regression Discontinuity (RD) design and provide evidence in support of the signaling value of degree class (Khoo and Ost, 2018; Feng and Graetz, 2017; Di Pietro, 2017; Freier et al., 2015).[16] Graetz (2021) contributes to the debate about explanations of returns to the diploma or similar credentials, by pointing out that a positive RD estimate is evidence for information frictions in the labor market. Also, his analysis suggests that the rate at which return to diploma or credentials decreases with experience can be attributed to the speed of employer learning.

Researchers in this line of the literature also study the signaling value of college quality (a coarse measure of productivity) and to what extent employers discriminate on the basis of college quality among students with similar lengths of study. These papers mostly exploit the discontinuity around the college admission cutoff to study the effects of admission to more selective colleges on labor market wages over the life cycle. MacLeod et al. (2017) use data from Colombia and provide evidence that employers use college reputation to make inferences about workers' productivity. In line with the employer learning hypothesis, they show that the introduction of a new measure of individual skill decreases the return to college reputation. They further document that earnings growth over the life cycle is positively correlated with her college's reputation, suggesting that beyond its signaling value colleges add to skill. Also, Barrera-Osorio and Bayona-Rodríguez (2019) find evidence in support of the statistical discrimination on the basis of college reputation which disappears

[16] A closely related literature examines the role of GPA in labor market outcomes for individuals with otherwise similar credentials and provides evidence in support of the signaling value of GPA (eg., see Hansen et al. (2021)). Also, the signaling value of high-school grades depends on whether they are obtained in local or centralized exams (Schwerdt and Woessmann, 2017). The awards given to top-performer students on an exam can possess signaling value and influence early labor market outcomes (Busso et al., 2022).

as employers learn about workers' productivity, in Columbia. Bordón and Braga (2020) investigate to what extent college reputation signals workers' ability in Chile using a regression discontinuity design. Their findings suggest that employers use college reputation as a signal of true productivity at the wage determination upon labor market entry and the college reputation becomes gradually less important for employers as workers gain experience in the labor market. Moreover, their finding suggests that the role of ability (measured by Prueba de Aptitud Académica (PAA), the Chilean university selection test) in wage determination increase over the life cycle.

Araki et al. (2016) use personnel data from two large Japanese manufacturers on white-collar university graduates in Japan and study the extent to which employers discriminate on the basis of college prestige and whether they learn about workers' true productivity over the life cycle. Their findings posit that performance evaluations (as a measure of true productivity) become a more important determinant for promotion than educational credentials as workers gain experience in the labor market. Moreover, they estimate the speed of employer learning and find that observing workers' performance on the job allows employers to learn quickly about workers' ability. Employers' expectation errors decline by half after around 3 to 6 years of labor market experience. The signaling value of college reputation and its evolution over the life cycle is also documented for some other countries such as Israel (Lang and Siniver, 2011) and Australia (Carroll et al., 2019).

B.4 A brief detour: the literature on asymmetric employer learning

Our paper is situated within the symmetric or common employer learning framework. A rich literature explores asymmetric employer learning which allows for incumbent employers to have an informational advantage. Here we briefly touch upon a few papers in this literature.

Schönberg (2007) uses the NLSY 1979 and tests whether current employers are more informed about workers' ability compared to outside firms (i.e, asymmetric learning). Like others, she uses the AFQT score as a hidden correlate. Her findings suggest that employer learning is mostly symmetric, especially for workers without a post-secondary education. Using the same data, Kim and Usui (2021) reaches a similar conclusion suggesting that employer learning is public (symmetric) for high school graduates while it is private (asymmetric) for college graduates. Also using the NSLY 1979, Pinkston (2009) likewise explores symmetric and asymmetric learning - he concludes that employer learning is mostly asymmetric.

Kahn (2013) tests the asymmetric employer learning model investigating job transitions in the NLSY 1979. Her findings suggest that learning about workers' AFQT scores is asymmetric across incumbent and outside firms. Specifically, she finds that incumbent firms learn three times as rapidly about worker ability than do outside firms. Once more using the NLSY 1979, Ge et al. (2021) study the implication of asymmetric learning and statistical discrimination on the basis of race using the data from the NLSY 1979. They find that employer learning is mostly asymmetric (increase in return to ability over the life cycle is stronger when one uses job tenure as opposed to experience) for non-college-educated workers and employers statistically discriminate against black workers, which is in line with the findings in Arcidiacono et al. (2010).