NBER WORKING PAPER SERIES

SIGNALING AND EMPLOYER LEARNING WITH INSTRUMENTS

Gaurab Aryal
Manudeep Bhuller
Fabian Lange

Working Paper 25885
http://www.nber.org/papers/w25885

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2019

This project received generous financial support from the Research Council of Norway through grants 194339, 250516, 267428 and 275123. Fabian Lange received support through the Social Sciences and Humanities Research Council of Canada and the Canada Research Chair program. We thank Leora Friedberg, John Pepper, Emily Nix and participants at the 2018 Cowles Conference in Honor of Joseph Altonji and seminar participants at University of Bergen, Ohio State University, Norges Bank, University of Southern California, University of Calgary, Goethe University Frankfurt, University of Alberta, University of British Columbia, University of Sussex, University of Delaware, University of Waterloo, and Harris School of Public Policy for helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Signaling and Employer Learning with Instruments
Gaurab Aryal, Manudeep Bhuller, and Fabian Lange
NBER Working Paper No. 25885
May 2019
JEL No. J01,J24,J3

ABSTRACT

The social and the private returns to education differ when education can increase productivity and also be used to signal productivity. We show how instrumental variables can be used to separately identify and estimate the social and private returns to education within the employer learning framework of Farber and Gibbons (1996) and Altonji and Pierret (2001). What an instrumental variable identifies depends crucially on whether the instrument is hidden from or observed by the employers. If the instrument is hidden, it identifies the private returns to education, but if the instrument is observed by employers, it identifies the social returns to education. Interestingly, however, among experienced workers the instrument identifies the social returns to education, regardless of whether or not it is hidden. We operationalize this approach using local variation in compulsory schooling laws across multiple cohorts in Norway. Our preferred estimates indicate that the social return to an additional year of education is 5%, and the private internal rate of return, aggregating the returns over the life-cycle, is 7.2%. Thus, 70% of the private returns to education can be attributed to education raising productivity and 30% to education signaling workers' ability.

Gaurab Aryal
Department of Economics
University of Virginia
Monroe Hall
248 McCormick Rd
Charlottesville, VA 22903
ga6x@virginia.edu

Manudeep Bhuller
University of Oslo
Department of Economics
Post Box 1095 Blindern
0317 Oslo
Norway
manudeep.bhuller@econ.uio.no

Fabian Lange
Department of Economics
McGill University
855 Sherbrooke Street West
Montreal QC H3A, 2T7
and NBER
fabian.lange@mcgill.ca

1 Introduction

Two competing models rationalize the positive relation between earnings and education that is universally found in data. According to the human capital model of Becker [1962], education increases skills valued by employers. By contrast, the job-market signaling model of Spence [1973] posits that education signals differences in innate skills/abilities among workers. Signaling, however, is inherently socially inefficient because workers expend valuable resources just to signal their productivity. Thus, the signaling aspect of education creates a wedge between the private returns and the social returns to education.[1]

Education policy requires empirical guidance on how large this wedge is.[2] The difficulties in separating signaling from human capital models, however, have long been recognized in the literature; see Lange and Topel [2006] and the references therein. One way forward is to recognize that if workers use education to signal their abilities at the start of their careers, then employers might also update their beliefs to learn workers' abilities over time.

In two influential papers, Farber and Gibbons [1996] and Altonji and Pierret [2001] (henceforth FG and AP, respectively) propose an empirical model of employer learning and use it to test whether employers use schooling to infer unobserved ability. The key identifying assumption underlying their approach is that researchers observe a correlate of unobserved ability, but this correlate is unobserved by employers. FG propose the Armed Forces Qualification Test (AFQT) score in the NLSY1979, to be such a correlate. Then they derive testable implications based on how the correlation between earnings from schooling and AFQT changes over the life-cycle. They also show that data from the NLSY1979 are consistent with employer learning (FG) and with statistical discrimination using schooling (AP).

[1] There are other reasons why the private and social returns might differ from each other. For instance, in Section 6.2, we consider productive externalities beyond the employer-employee relationship, as in Acemoglu and Angrist [2000] and Moretti [2004], and provide evidence on external returns to education. Education might also entail various non-production and non-pecuniary benefits, e.g., reducing crime, improving public health [Lange and Topel, 2006; Lochner, 2011; Oreopoulos and Salvanes, 2011], that we abstract from in this paper. Comprehensive measures of social returns should aim to account for such spillovers. More broadly, education policy might also be concerned with fiscal externalities and motivated by distributional impacts.

[2] For more on identifying the signaling and the human capital model, see Tyler et al. [2000]; Bedard [2001]; Fang [2006]; Hopkins [2012]; Clark and Martorell [2014]; Feng and Graetz [2017] and Arteaga [2018].

Following their lead, Lange [2007] shows that employers learn fast. Using the first-order condition that characterizes schooling decisions, Lange [2007] provides an upper bound on the contribution of signaling to the private returns over the life-cycle. He estimates the bound to be 25% of the private returns to schooling.[3]

We contribute to this literature by asking what instrumental variable (IV) estimates of the returns to education identify within the same employer learning framework as in FG, AP, and Lange [2007]. We show that the IV estimates of returns to education, over workers' life-cycle, allow us to point-identify the private and social returns to education. Unlike Lange [2007], however, our identification strategy does not exploit the first-order condition for schooling choices and thus does not require specifying the costs and benefits of schooling. Equally important, unlike FG and AP, our identification strategy does not require access to a correlate of ability unobserved by employers.

We present several identification results. First, we show that any conventional IV estimate of returns to education on earnings, measured at sufficiently high level of work experience, identifies the causal effect of schooling on productivity. This result implies that access to an IV and a repeated cross-section of earnings across workers' careers are enough to identify the productivity effect of schooling. This interpretation of "long-run" IV estimates follows directly from a (limit) result in our employer learning model that wages eventually converge to the true productivity because employers eventually learn workers' productivity.

Our second set of results illustrate how central the assumptions about employers' knowledge about the instruments are for interpreting the IV estimates. We distinguish between a *hidden* IV and a *transparent* IV. We say that an IV is hidden if it is unobserved by employers, and that it is transparent if it is observed by the employers and priced into the wages. We show that if the IV is hidden, it identifies the private returns to education all across the life-cycle, and if the IV is transparent, it identifies the social returns to education. For our third result, we propose a method that uses IV estimates to identify the speed at which

[3] To identify this bound Lange [2007] makes strong behavioral assumptions on schooling decisions, including the assumption that the costs of schooling are observed.

employers learn workers' productivity.

In summary, within the FG and AP framework of employer learning, a hidden IV is sufficient to (i) point-identify the relative contributions of human capital and signaling in the returns to earnings over the life-cycle; and (ii) estimate the speed of learning. Thus, any additional information – either because we have access to multiple instruments or because of the availability of an ability correlate like AFQT – allows testing and/or relaxing functional form assumptions in the employer learning model.

To implement these ideas we use a unique dataset consisting of all Norwegian males born between 1950 and 1980, with earnings and employment histories between 1967 and 2014. We also observe an ability correlate in the form of a cognitive test administered by the Norwegian military that is taken by male conscripts around the age of 18. This test score is not directly observed by employers. We also have access to a hidden IV based on local variation in the implementation of compulsory schooling reform across many birth cohorts.

Using these data, we examine how the IV returns to schooling vary over the life-cycle, and interpret this through the lens of our employer learning model. The returns to schooling start high at around 15% in the first year following graduation, and then decline, rapidly at first and then slowly, until they stabilize to about 5-6%, after approximately 20 years of work experience. These findings are consistent with the hypothesis that employers use past performance to learn about workers' productivity, and our assumption that the IV (staggered implementation of compulsory school reform) is hidden from the employers. Like Lange [2007] we find that employers learn fast.

Second, we quantify the contribution of signaling and human capital acquisition to the returns to education. Our analysis reveals a productivity effect of education of 5% and a private internal rate of return in lifetime earnings, discounted to the time of schooling choice, of 7.2%. These estimates suggest that 70% of the private returns to schooling, over the life-cycle, represents a productivity-enhancing effect of education and the remaining 30% represents the signaling contribution of education. Thus, we find a non-negligible role for signaling in explaining the positive returns to education estimated in our data.

Third, we compare our OLS estimates of returns to schooling and cognitive test scores with estimates from previous studies that use the NLSY data. The patterns we uncover in the Norwegian data are strikingly similar to those found by FG, AP, Lange [2007], and Arcidiacono et al. [2010] for the NLSY. In Norway, the estimated return to one standard deviation increase in the ability score increases from near zero in the first few years in the labor market to about 7% after around 15 years of experience. The experience pattern in the NLSY with respect to the AFQT is similar, except that the return to a standard deviation increase in the AFQT score converges to approximately 14% after 15 years. Controlling for the interaction between the ability score and experience, we find that OLS estimates of the coefficients on years of schooling decline rapidly from about 10% to about 3% within the first 20 years. Likewise, in the NLSY, the returns decline from about 9% to 6%.[4]

Finally, we consider two extensions of our model. In the first extension, we allow workers' productivity to be non-separable in schooling and experience, and show that a hidden IV still identifies the private and social returns and that these returns converge with enough work experience also in the non-separable model. Thus, our identification results are not entirely driven by the log-additive functional form, although estimating the returns and the speed of learning in a non-separable model is nontrivial and beyond the scope of our paper.

In the second extension, we extend our framework to embed productive externalities that go beyond the employer-employee relationship, building on Acemoglu and Angrist [2000], Moretti [2004] and Lange and Topel [2006]. Using the same dataset we provide IV estimates of the returns to individual schooling *and* the returns to average schooling in an individual's local labor market. We find that, after conditioning on individual schooling, an increase in average schooling in an individual's local labor market by an additional year increases individual log-earnings by around 25%. These estimates indicate large external returns to

[4] We also find that the association between test score and log-earnings increases with experience, but only for those with a high school degree or less, which is similar to Arcidiacono et al. [2010]. For those with more than a college degree, the returns from one standard deviation increase in the ability score remain constant at around 6-7% across all years of experience, which is consistent with a college degree revealing ability.

education, suggesting that the social returns to education could exceed the private returns.

The rest of our paper proceeds as follows. Section 2 describes the model of employer learning as developed by FG, AP, and Lange [2007] and defines the private and the social returns to education within this structure. Section 3 then discusses identification of the private and the social returns to education in the model of employer learning using instrumental variables. Section 4 presents the data and our empirical setting. Section 5 contains our main empirical findings. We consider two extensions of our model in Section 6. Our first extension considers the identification of a general model that is non-separable in schooling and experience. The second extension considers the evidence of local spillovers due to education and presents evidence regarding these spillovers. We conclude in Section 7.

2 Model of Employer Learning

In this section, we present the model of employer learning in a perfectly competitive labor market, first proposed by FG and AP. Let worker i's productivity be given by

$$\chi_{it} = \exp\left(\beta_{ws} S_i + \beta_{wq} Q_i + A_i + H(t) + \varepsilon_{it}\right) \equiv \exp(\psi_{it}), \tag{1}$$

where S is the years of schooling, Q is a correlate of ability observed by employers but unobserved by researchers, and A is ability unobserved (to employers and researchers) and possibly correlated with the employer-observed correlates (S, Q). An example of a Q could be knowledge of foreign languages, which is typically mentioned in job applicants' résumés and that can easily be verified. The function $H(t)$ captures how log-productivity varies with experience t. While we allow $H(t)$ to be a nonparametric function of t, we assume that it does not depend on either schooling or ability.[5] Finally, ε_t represents time-varying noise in the production process that is independent of all other variables.

[5] When estimating (Section 4.4) we use $H(t, X_i)$ which allows the experience profile to vary flexibly with individual characteristics that include a full set of dummies for birth cohort and municipality, X_i. In Section 6.1 we study the identification of an extension of our model where schooling and experience are nonseparable.

To model employer learning we follow Lange [2007] and assume that $\varepsilon_{it} \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$ and $(S_i, Q_i, A_i) \overset{i.i.d}{\sim} \mathcal{N}(\boldsymbol{\mu}, \Sigma)$, across workers and across time.[6] Let $\sigma_0^2 = Var(A_i|S_i, Q_i)$ be the conditional variance of A_i given (S_i, Q_i). Besides knowing (S_i, Q_i), every period employers also observe total output (χ_{it}), which is equivalent to observing a signal $\xi_{it} := A_i + \varepsilon_{it}$ about i's productivity. If we let $\mathcal{E}_{it}$ denote employers' information about i in period t, then $\mathcal{E}_{it} = (S_i, Q_i, \xi_i^t)$ with $\xi_i^t = \{\xi_{i\tau}\}_{\tau<t}$ as the history of all past signals.[7]

Assuming (S_i, Q_i, A_i) are jointly normal random variables implies that the conditional expectation of A_i given information at $t = 0$, $\mathbb{E}[A_i|\mathcal{E}_{i0}] = \mathbb{E}[A_i|S, Q]$, is linear in (S, Q)

$$A_i = \phi_{A|S} S_i + \phi_{A|Q} Q_i + \varepsilon_{A|S,Q}, \tag{2}$$

where $\varepsilon_{A|S,Q} := A_i - \mathbb{E}[A_i|S, Q]$. Under perfect competition workers are paid their expected output, conditional on the information available to the employers. The wage in period t is then equal to the expected productivity conditional on $\mathcal{E}_{it}$, so that $W_{it} = \mathbb{E}[\chi_{it}|\mathcal{E}_{it}] = \mathbb{E}[\chi_{it}|S_i, Q_i, \xi_i^t]$. Taking the expectation of the log of (1) and using the fact that $\exp(A_i + \varepsilon_{it})$ is log-normal with conditional variance $v_t = Var(A_i + \varepsilon_{it}|\mathcal{E}_{it})$, we get

$$\ln W_{it} \quad = \quad \beta_{ws} S_i + \beta_{wq} Q_i + \tilde{H}(t) + \mathbb{E}[A_i|\mathcal{E}_{it}], \tag{3}$$

where $\tilde{H}(t) \equiv H(t) + \frac{1}{2} v_t$ collects the terms that vary only with t but not across the realizations of ξ_i^t. For notational simplicity, we suppress $\tilde{H}(t)$ until our empirical implementation.

We can use the Kalman filter to represent the process by which employers update their expectations $\mathbb{E}[A_i|\mathcal{E}_{it}]$. It allows us to write the expectation of ability in a simple form as

$$\mathbb{E}[A_i|\mathcal{E}_{it}] = \theta_t \mathbb{E}[A_i|S, Q] + (1 - \theta_t)\bar{\xi}_i^t, \tag{4}$$

[6]For much of what follows, S need not be Gaussian, but it simplifies the exposition of the argument. Without it, we would work with a linear projection of A_i on S_i and Q_i instead of the Equation (2).

[7]We assume that all employers have symmetric information about workers' ability and past outputs. For assessment of how to test asymmetry among current and potential employers and its effect on labor market outcomes, see, e.g., Kahn [2013], Schönberg [2007] and Waldman [1984], among others.

where $\bar{\xi}_i^t = \frac{1}{t}\sum_{\tau<t}\xi_{it}$ is the average of signals up to period t and θ_t is the weight on the initial signal (S, Q) with $\theta_t = \frac{1-\kappa}{1+(t-1)\kappa}$ with $\kappa = \frac{\sigma_0^2}{\sigma_0^2+\sigma_\varepsilon^2} \in [0, 1]$. In particular, equation (4) shows that the conditional expectation of ability at time t is the weighted average of the expectation at $t = 0$, before any additional information about productivity has been received, and the average of all additional signals up to period t received by the employers.

The weight θ_t declines with experience (t) because with time, observed measures of productivity become better predictors of productivity than the correlates (S, Q), which were the only information available at $t = 0$. The rate at which θ_t declines, however, depends on the parameter κ that Lange [2007] refers to as the "speed of learning." The speed of learning governs how quickly information about productivity accumulates in the market, which depends on the information contained in the signals. In particular, if the signal-to-noise ratio is high, i.e., when the variance of noise (ε) in production is small, so that $\sigma_\varepsilon^2/\sigma_0^2$ is small, then κ will be close to 1, and the market quickly learns the ability A. But, irrespective of κ, after a sufficiently long work experience employers will put all weights on the new information, i.e., $\lim_{t\to\infty}\theta_t = 0$, and the initial productivity correlates will become less important determinants of earnings.

Social and Private Returns to Education

Next, we define *social returns* and *private returns* to education. To that end, note that the coefficient β_{ws} in Equation (1) is not the causal effect of education on productivity, but it is instead only the "partial" causal effect of schooling, holding the employer-observed ability correlate, Q, and unobserved ability, A, fixed. Schooling, however, can causally affect both Q and A, so the "total" causal effect of schooling on productivity also includes the (indirect) effect on productivity mediated through (Q, A). We refer to this total causal effect as the *social returns* to education. Education moreover also affects wages through employers' expectations about the ability of a worker, and these expectations can change over the life cycle. We define *private returns* to be the expected earnings from an additional year of schooling evaluated at the beginning of a life-cycle.

To formalize these two measures of returns, we need new notations and a simplifying assumption. For a random variable Y, let $\delta^{Y|S}$ denote the causal effect of S on Y and let $\tilde{Y}$ denote the part of Y that is not caused by schooling S but may correlate with S. Furthermore, let there be a linear causal relationship between S and (Q, A), i.e.,

$$Q_i = \delta^{Q|S} S_i + \tilde{Q}_i; \quad \text{and} \quad A_i = \delta^{A|S} S_i + \tilde{A}_i. \tag{5}$$

Then, substituting (Q, A) from the above equations into Equation (1), we obtain

$$\psi_{it} = \underbrace{\left(\beta_{ws} + \beta_{wq}\delta^{Q|S} + \delta^{A|S}\right)}_{:=\delta^{\psi|S}} S_i + \underbrace{\beta_{wq}\tilde{Q}_i + \tilde{A}_i + \varepsilon_{it}}_{:=u_{it}} \equiv \delta^{\psi|S} \times S_i + u_{it}. \tag{6}$$

The first term, $\delta^{\psi|S}$, in (6) is the total causal effect of schooling on productivity, or the social returns to education, and it captures the causal effect (direct and indirect) on other ability components (Q, A). Thus (6) shows that an extra year of schooling increases Q by $\delta^{Q|S}$ units, which in turn raises productivity by β_{wq}, and it also raises ability A by $\delta^{A|S}$ units.

Consider now the private returns to education. Schooling can affect expected log-earnings at t in three different ways: (i) directly, because employers use schooling to form expectations about productivity; (ii) indirectly, because schooling may impact Q observed by employers (as shown above in Equation (5)); and (iii) through learning, because schooling affects productivity, which employers learn over time by observing workers' outputs.

Substituting (2) and (4) in (3), and using $\bar{\xi}_i^t = \frac{1}{t}\Sigma_{\tau<t}(A_i + \varepsilon_{it}) = A_i + \bar{\varepsilon}_i^t$ and that $\mathbb{E}(A_i|S_i, Q_i)$ is linear and separable in S_i and Q_i we get the following wage equation:

$$\ln W_{it} = \left(\beta_{ws} + \theta_t\phi_{A|S}\right) S_i + \left(\beta_{wq} + \theta_t\phi_{A|Q}\right) Q_i + (1 - \theta_t)\left(A_i + \bar{\varepsilon}_i^t\right).$$

Furthermore, using the causal relationships from (5) in the above equation, we obtain

$$\begin{aligned}
\ln W_{it} &= \left(\beta_{ws} + \theta_t \phi_{A|S}\right) S_i + \left(\beta_{wq} + \theta_t \phi_{A|Q}\right) \left(\delta^{Q|S} S_i + \tilde{Q}_i\right) + (1-\theta_t) \left(\delta^{A|S} S_i + \tilde{A}_i + \bar{\varepsilon}_i^t\right) \\
&= \underbrace{\left(\beta_{ws} + \beta_{wq}\delta^{Q|S} + \delta^{A|S} + \theta_t(\phi_{A|S} + \phi_{A|Q}\delta^{Q|S} - \delta^{A|S})\right)}_{:=\delta_t^{W|S}} S_i \\
&\quad + \underbrace{\left(\beta_{wq} + \theta_t \phi_{A|Q}\right) \tilde{Q}_i + (1-\theta_t) \left(\tilde{A}_i + \bar{\varepsilon}_i^t\right)}_{:=\tilde{u}_{it}} = \delta_t^{W|S} S_i + \tilde{u}_{it}. \qquad (7)
\end{aligned}$$

The coefficient of schooling, $\delta_t^{W|S}$, in Equation (7) is the private returns to education. Comparing this coefficient with the coefficient in Equation (6), we get the following relationship:

$$\underbrace{\delta_t^{W|S}}_{\texttt{private returns}} = \underbrace{\delta^{\psi|S}}_{\texttt{social returns}} + \underbrace{\theta_t}_{\texttt{weight}} \underbrace{\left(\phi_{A|S} + \phi_{A|Q}\delta^{Q|S} - \delta^{A|S}\right)}_{\texttt{adjustment term}}. \qquad (8)$$

Thus the private returns $\delta_t^{W|S}$ differs from the social return $\delta^{\psi|S}$ if the effect of schooling on expected A based on the information available to firms, which is captured by $(\phi_{A|S} + \phi_{A|Q}\delta^{Q|S})$, differs from the causal effect of schooling on unobserved ability $\delta^{A|S}$. The signaling literature assumes that this "adjustment term" is non-negative, so that education has signaling value, which in turn implies that $\delta_t^{W|S} \geq \delta^{\psi|S}$. This wedge between private and social returns, however, disappears with work experience, i.e., $\lim_{t\to\infty} \delta_t^{W|S} = \delta^{\psi|S}$ because $\lim_{t\to\infty} \theta_t = 0$.

3 Identification

In this section, we study the identification of the social and private returns to education. We show how commonly used estimators (e.g., ordinary least squares and instrumental variables) are related to the private returns to education, the social returns to education, and the speed of learning defined above. We begin by considering least-squares projections of log-earnings on education over the life-cycle and then consider how one might proceed if additional information is available in the form of: (i) a correlate of ability not observed by the employers (e.g., AFQT score); and/or (ii) an instrument for schooling. We show that what IV identifies depends crucially on whether the IV is observed by employers.

3.1 Bias in the OLS

Begin by considering the regression of log-earnings on years of schooling for any given level of experience. Using Equation (7), we can derive the probability limit of the OLS estimate of the coefficient for schooling, evaluated at experience t, to be

$$\texttt{plim}\left(\hat{b}_{OLS,t}\right) = \underbrace{\delta_t^{W|S}}_{\texttt{private returns}} + \underbrace{\left(\beta_{wq} + \theta_t \phi_{A|Q}\right) \frac{\mathrm{cov}(\tilde{Q}, S)}{\mathrm{var}(S)} + (1-\theta_t) \frac{\mathrm{cov}(\tilde{A}, S)}{\mathrm{var}(S)}}_{\texttt{omitted variable bias}} . \tag{9}$$

The OLS estimate of the schooling coefficient is a biased estimate of the experience-specific private return $\delta_t^{W|S}$ because the omitted ability components $(\tilde{Q}, \tilde{A})$ correlate with, but are not *caused* by, schooling. This omitted ability bias is the main reason why researchers rely on IV(s) to identify the returns to education. The magnitude of bias depends on the speed of learning, κ, which determines the weight θ_t employers put at time t on the initial signal.

Now, let us consider what happens to this bias as workers accumulate work experience, i.e., as $t \to \infty$. Taking the limit in (9) and using $\lim_{t\to\infty} \delta_t^{W|S} = \delta^{\psi|S}$ from (8) we get

$$\texttt{plim}\ \left(\lim_{t\to\infty} \hat{b}_{OLS,t}\right) = \underbrace{\delta^{\psi|S}}_{\texttt{social returns}} + \underbrace{\frac{\mathrm{cov}\left(\beta_{wq}\tilde{Q} + \tilde{A}, S\right)}{\mathrm{var}\left(S\right)}}_{\texttt{remaining bias}} . \tag{10}$$

Thus, even after employers observe a long history of outputs, the bias does not disappear. We conclude that the OLS does not identify the private or the social returns to education.

3.2 Exploiting a Hidden Correlate of Ability

Now suppose that we have access to a correlate of ability, denoted by Z, and suppose Z is unobserved by employers. We refer to this as a "hidden" correlate of ability. Furthermore, suppose that $A_i = \beta_{Az} Z_i + \eta_i, \eta_i \perp Z_i$ so that η_i represents the productivity component observed by neither researchers nor employers.[8] Substituting A_i in Equation (1) gives

$$\chi_{it} = \exp\left(\beta_{ws} S_i + \beta_{wq} Q_i + \beta_{Az} Z_i + \eta_i + H(t) + \varepsilon_{it}\right) = \exp(\psi_{it}). \tag{11}$$

Following the steps from Section B in [Lange, 2007], we can show that

$$\mathbb{E}\left[\ln W_{it} | S, Z, t\right] = \theta_t \mathbb{E}\left[\ln W_{i0} | S, Z\right] + (1-\theta_t)\mathbb{E}\left[\ln W_{i\infty} | S, Z\right], \tag{12}$$

where W_{i0} is the wage received in period $t = 0$ and $W_{i\infty}$ is the wage received at $t \to \infty$, when enough information has been revealed so that worker productivity is known in the market. The linearity of (12) allows us to estimate the speed of learning, κ, by projecting log-earnings on (S, Z) across different work experience levels, t, because the weight θ_t depends only on κ. Thus, the regression coefficients of log-earnings on (S, Z) converge from their $t = 0$ value to their $t = \infty$ value at a rate that depends only on κ, thereby identifying κ.

The projection coefficients obtained from estimating (12) across different experience levels, however, do not identify the causal effect of S or Z on productivity. These coefficients are biased (even when $t \to \infty$) because (S, Z) can be correlated with the omitted variables (Q, η). Thus, while we can identify κ if we have a hidden correlate of ability, we cannot identify the private or the social returns to education without additional information.

3.3 Instrumental Variables

Next, suppose that we have access to a binary instrument variable $D_i \in \{0, 1\}$. In other words, suppose D_i satisfies the following standard assumptions for a valid IV.

Assumption 1. *Instrumental Variables*

1. *(Conditional Independence): $u_{it} \perp D_i | S_i$, where u_{it} is defined in (6).*

[8] β_{wq} in Equation (1) accounts for variation in productivity with Q. Therefore omitting Q in the projection of A_i on Z_i simply represents a normalization.

2. *(First Stage):* $\mathbb{E}[S_i|D_i = 0] \neq \mathbb{E}[S_i|D_i = 1]$.

Under Assumption 1, for a binary instrument D_i, in period t we get

$$\texttt{plim}\ \hat{b}_{IV,t} \quad := \quad \frac{\mathbb{E}\left[\ln W_{it}|D_i = 1, t\right] - \mathbb{E}\left[\ln W_{it}|D_i = 0, t\right]}{\mathbb{E}\left[S_i|D_i = 1, t\right] - \mathbb{E}\left[S_i|D_i = 0, t\right]}. \tag{13}$$

Furthermore, using the fact that S is constant across t and $\lim_{t\to\infty} \ln W_{it} = \psi_i$, we get

$$\texttt{plim}\ \left(\lim_{t\longrightarrow\infty} \hat{b}_{IV,t}\right) = \text{plim} \quad \hat{b}_{IV,t\longrightarrow\infty} = \frac{\mathbb{E}\left[\psi_i|D_i = 1\right] - \mathbb{E}\left[\psi_i|D_i = 0\right]}{\mathbb{E}\left[S_i|D_i = 1\right] - \mathbb{E}\left[S_i|D_i = 0\right]} = \delta^{\psi|S},$$

where the second equality follows from Assumption 1-(1), which implies that the part of the productivity, ψ_i, not caused by schooling, S, is orthogonal to the instrument, D. Thus, as $t \longrightarrow \infty$ the IV identifies the causal effect of schooling on productivity. In other words, the IV estimate of returns to education at sufficiently high levels of experience is a consistent estimator of the causal effect of schooling on productivity.

Note that this identification strategy is valid *irrespective* of what the employers know about D. Heuristically, in the long run everything about a worker's ability is revealed to the employers, and thus knowledge of the instrument itself has become irrelevant for wage setting. For intermediate work experience (i.e., $t < \infty$), however, what the IV identifies depends on whether or not D is hidden from the employers. To determine how the information of employers affects the interpretation of the IV estimates, we distinguish between *hidden* and *transparent* instruments next.

Hidden Instrument

We begin with D being unobserved by the employers, i.e., when D is a hidden IV.

Assumption 2. *(Hidden Instrument) For all* $i, D_i \notin \mathcal{E}_{it}$ *which implies* $\ln W_{it} \perp D_i|(S_i, Q_i, \xi_i^t)$.

Note that Assumption 2 is conceptually different from Assumption 1-(1). The latter assumption asserts that the IV is conditionally independent of the determinants of productivity not caused by schooling, whereas Assumption 2 captures the idea that given the information, $\mathcal{E}_t$, wages do not depend on the instrument D, so $\ln W_{it} = \mathbb{E}\left[\psi_i|\mathcal{E}_{it}\right] = \mathbb{E}\left[\psi_i|\mathcal{E}_{it}, D_i\right]$.

In many settings, Assumption 2 is a natural assumption. The clearest examples relate to field experiments that provide subsidies or information that induce higher school enrollment. In these cases, whether a student is in the control or treatment group is typically not known to the (potential) employers. Some examples of hidden instruments from the empirical literature in quasi-experimental settings include (i) the interaction of draft lottery number and year of birth in Angrist and Krueger [1992]; (ii) the interaction of a policy intervention, family background and season of birth in Pons and Gonzalo [2002]; (iii) parents' education and number of siblings in Taber [2001]; and (iv) the elimination of student aid programs interacted with an indicator for a deceased father in Dynarski [2003]. Besides these, many studies also exploit interactions of birth year and location of birth with locally implemented policy reforms, e.g., Duflo [2001] and Meghir and Palme [2005], which are similar to our IV.

Let Δ_D denote the difference from $D = 1$ to $D = 0$. Then the numerator in the definition of $\texttt{plim}\ \hat{b}_{IV,t}$ shown in Equation (13) for a binary, hidden instrument D_i is

$$\Delta_D \mathbb{E}\left[\ln W_{it}|D_i, t\right] = \Delta_D \mathbb{E}\left[\beta_{ws} S_i + \beta_{wQ} Q_i + \mathbb{E}\left[A_i|S_i, Q_i, \xi_i^t\right]|D_i, t\right],$$

where $\ln W_{it}$ does not directly depend on D_i because it is not used by the employers in the wage setting. The IV, D_i, affects $\ln W_{it}$ only indirectly by affecting (S_i, Q_i, ξ_i^t) that makes up the information $\mathcal{E}_{it}$ used by employers to infer productivity. From Assumption 1 we get $\mathbb{E}\left[Q_i|D_i, S_i\right] = \delta^{Q|S} S_i$. Using that with Equations (2) and (5) and simplifying further gives

$$\begin{aligned}
\mathbb{E}\left[\ln W_{it}|D_i, t\right] &= \left(\left(\beta_{ws} + \beta_{wQ}\delta^{Q|S}\right) + \theta_t\left(\phi_{A|S} + \phi_{A|Q}\delta^{Q|S}\right) + (1-\theta_t)\,\delta^{A|S}\right)\mathbb{E}\left[S_i|D_i\right], \\
&= \left(\delta^{\psi|S} + \theta_t\left(\phi_{A|S} + \phi_{A|Q}\delta^{Q|S} - \delta^{A|S}\right)\right)\mathbb{E}\left[S_i|D_i\right].
\end{aligned}$$

Then, taking the probability limit, we get

$$\texttt{plim}\ \hat{b}_{IV,t} \quad = \quad \frac{\Delta_D \mathbb{E}\left[\ln W_{it}|D_i\right]}{\Delta_D \mathbb{E}\left[S_i|D_i\right]} = \delta^{\psi|S} + \theta_t \left(\phi_{A|S} + \phi_{A|Q}\delta^{Q|S} - \delta^{A|S}\right). \tag{14}$$

Comparing Equation (14) with the private returns defined in Equation (8), we can conclude that, for every work experience level t, the hidden IV identifies the private returns to education. Besides the private returns, the hidden IV also identifies the speed of learning. To see that, note that we can identify $\texttt{plim}\ \hat{b}_{IV,t}$ across t, which together with the convergence of $b_{IV,t}$ from $b_{IV,t=0}$ to $b_{IV,t\to\infty}$ identify the speed of learning κ.

Transparent Instrument

We say that an instrumental variable is *transparent* if it is known to the employers and is thus "priced-in" the wages. In other words, if the IV, D, is transparent, it is included in the information set of the employers, but it is still a valid IV because it satisfies Assumption 1. Let $\tilde{\mathcal{E}}_{it} := \mathcal{E}_{it} \cup \{D_i\}$ be the new set of information employers have about i in t.

Assumption 3. *(Transparent Instrument) Employers observe D_i so that $\ln W_{it} = \mathbb{E}[\psi_{it}|\tilde{\mathcal{E}}_{it}]$.*

By Assumption 1, we have that transparent instruments satisfy the exclusion restriction with respect to productivity ψ. Assumption 3, however, implies that the instrument is used in wage setting and thus will not be orthogonal to wages conditional on schooling and other controls. Some examples of instruments used in the literature that are more likely to be transparent than not are (i) tuitions at two- and four-year state colleges in Kane and Rouse [1995]; (ii) a dummy for being a male aged 19-22 from Ontario in Lemieux and Card [2001]; (iii) local labor market conditions in Cameron and Heckman [1998]; Cameron and Taber [2004] and Carneiro et al. [2011]; (iv) change in minimum school-leaving age in the U.K. from 14 to 15 in Oreopoulos [2006]; and (v) the distance to the college in Card [1993], Kane and Rouse [1995], Kling [2001] and Cameron and Taber [2004].

So if D is transparent, it violates the exclusion restriction for wages and thus does not estimate the causal effect of schooling on individual wages (which is the private return), but it estimates the social returns (the effect on productivity). To see the intuition as to how transparent IV identifies the social returns to education, consider two workers $i \neq j$, who have the same abilities and past outputs but different realizations of the instrument. Suppose $D_i = 1$ but $D_j = 0$ and $S_i > S_j$. D is transparent, so employers can deduce that $S_i > S_j$ because of D and and not because of A. So if $\ln W_i \geq \ln W_j$ then this wage difference can be attributed to the productivity effect of schooling. Therefore if the employers are informed about the instrument, the IV estimate of returns to education is a consistent estimate of the productivity effect of education on earnings, i.e.,

$$\begin{aligned} \mathbb{E}\left[\ln W_{it}|S_i, D_i\right] &= \mathbb{E}\left[\delta^{\psi|S} S_i + \tilde{\psi}|S_i, D_i\right] = \delta^{\psi|S} \times S_i; \\ \mathbb{E}\left[\ln W_{it}|D_i\right] &= \delta^{\psi|S} \mathbb{E}\left[S_i|D_i\right]. \end{aligned}$$

Hence the Wald estimator for a transparent IV, D, identifies the social returns to education at all t, i.e., $\texttt{plim}\ \hat{b}_{IV,t} = \delta^{\psi|S}$.

4 Data and Empirical Setting

In this section, we first describe our data sources, sample construction and the key variables utilized in our analysis. Then we describe the Norwegian compulsory schooling reform that we utilize as a source of exogenous variation in educational attainment to construct IV estimates of the returns to education in log-earnings at each year of experience. Finally, we discuss the empirical specifications motivated by the discussion in Section 3.

4.1 Data Sources and Sample Construction

Our empirical analysis uses several registry databases maintained by Statistics Norway. These databases allow us to construct a rich longitudinal dataset containing records for all Norwegian males from 1967 to 2014. We observe demographic information (e.g., cohort of birth and childhood municipality of residence) and socio-economic information (e.g., years of schooling and annual earnings) for these individuals. Importantly, the dataset also includes a unique personal identifier which allows us to follow individuals' earnings across time. The personal identifier also allows us to merge information on IQ test scores for males from the Norwegian Armed Forces to our dataset.

The Norwegian earnings data have several advantages over those available in most other countries. First, there is no attrition from the original sample other than natural attrition due to either death or out-migration. Second, our earnings data pertain to all individuals, and are not limited to some sectors or occupations. Third, we can construct long earnings histories that allow us estimate the returns to education at each year of labor market experience.

We restrict our sample to Norwegian males born between 1950 and 1980, including several cohorts with earnings observed over a wide-range of labor market experiences.[9] We restrict the sample to males because the military IQ test scores are not available for females. We further exclude immigrants as well as Norwegian males with missing information on either of the following variables, including years of schooling, childhood municipality of residence, IQ test score, or exposure to the compulsory schooling reform. Applying these restrictions we retain a sample consisting of 732,163 Norwegian males born between 1950 and 1980.

Our primary outcome variable is the natural logarithm of pre-tax annual labor-earnings.[10] To avoid variation in earnings across labor market experience due to the intensity of part-time work, we focus only on full-time workers who are defined as having annual labor earnings (adjusted for wage inflation) above the substantial gainful activity threshold (henceforth, SGA) as defined by the Norwegian Social Security System. In 2015, the SGA threshold was USD 10,650.[11] Restricting the sample to full-time employed males, we retain $718,237$

[9] In our annual income panel data from 1967 to 2014, we observe the oldest cohort (1950) between ages 17 and 64 and the youngest cohort (1980) up to age 34.

[10] We exclude income from self-employment, capital income or unconditional cash transfers such as social economic assistance, housing assistance, child allowance, etc.

[11] The earnings data are top-coded only at the very high earnings levels, and less than 3% of observations have right-censored earnings in any given year.

individuals– thus most males are recorded having a full-time employment spell at least once–and a panel data set comprising $14,746,755$ person-year observations. On average an individual is thus observed working full-time for 20.5 years. This sample is utilized in the empirical part of our analysis. Note that this sample is unbalanced: we have earnings for $579,984$ individuals in the 1^{st} year of experience and for $190,900$ individuals in the 30^{th} year.

4.2 Measures of Schooling and IQ Test Scores

The first key regressor of interest is the years of schooling corresponding to the highest level of completed education. This variable is taken from Statistics Norway's Education Register and it is based on the educational attainment reports submitted by educational establishments directly to Statistics Norway, which minimizes the chance of misreporting. Our second regressor of interest is the IQ test score accessed from the Norwegian Armed Forces. In Norway, military service was compulsory for all able males in the birth cohorts we study. Before each male entered the service, his medical and psychological suitability was assessed. Most eligible Norwegian males in our sample took this test around their 18^{th} birthday. The IQ test score we use is a composite unweighted mean from three speeded tests–arithmetics, word similarities, and figures.[12]

Figure 1 displays the average and the conditional density of IQ for each year of schooling between 9 and 21 years. This figure illustrates two striking patterns in our data worth noting. First, the measures of IQ and schooling are strongly correlated, with a correlation of almost 0.5. Second, sharp increases in the average IQ score occur around the entry years of high school (10 years) and college (13/14 years), with more gradual increases at later stages of schooling. This pattern could be due to substantial ability-related (psychic) costs for enrolling in high school or selective entry requirements enforced in the entry to higher education in Norway.[13]

Arguably, Norway is an interesting setting to assess employer learning and the signaling

[12]The arithmetic test mirrors the test in the Wechsler Adult Intelligence Scale (WAIS), the word test is similar to the vocabulary test in WAIS, and the figures test is comparable to the Raven Progressive Matrix

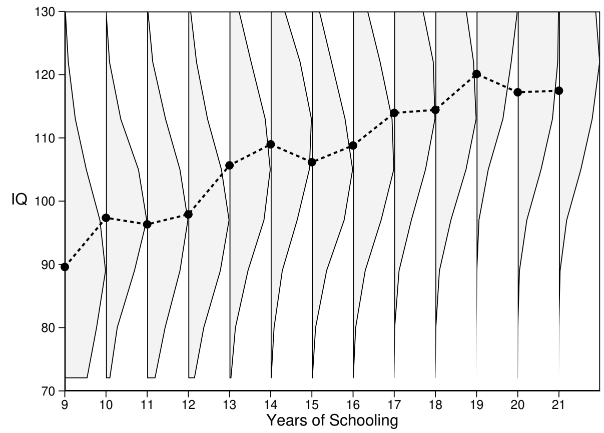


Figure 1: The Conditional Probability Density of IQ Test Scores on Years of Schooling.
Note: The sample consists of Norwegian males born 1950-1980 observed in earnings data over years 1967-2014 with years of potential experience between 0 and 30 years with annual earnings above 1 SGA threshold (N=14,746,755). The IQ test score along the y-axis is standardized to have a mean of 100 and a standard deviation of 15. The black dotted line plots the average IQ test score by individuals' years of schooling, while the shaded areas plot the conditional probability density of IQ.

value of education for several reasons. First, the strong correlation between schooling and ability test scores in our data suggests that schooling may predict ability, satisfying a necessary condition for schooling to have a signaling value. Second, employers cannot request the test scores from the military conscription and applicants don't (voluntarily) disclose this information in their job applications. It is thus reasonable to assume that employers do not observe the ability test scores from military conscription. Using the correlation between the military IQ test scores and earnings across experience, researchers can thus infer the process of employer learning. As discussed above, we allow for the possibility that other correlates (as captured by Q in Section 2) of applicants' ability could be revealed in the job application process. Finally, most cohorts in our sample entered the labor market before the arrival of online recruitment tools in the early 2000s, which might have altered the way in which employers tended to screen or recruit workers.

test. See Sundet et al. [2004] and Thrane [1977] for details.

[13] As documented in Kirkeboen et al. [2016], Norway has a system where access to public higher education is based on merit, and it is administered through a centralized admissions process. Students with higher GPAs from high school can thus more easily select into fields with high demand, and these students may also have higher IQ test scores in military conscription.

4.3 The Compulsory Schooling Reform

Between 1960 and 1975 Norway enacted a compulsory schooling reform that increased the minimum required schooling from 7 to 9 years. This reform was implemented by different municipalities–the lowest level of local administration–in different years. Thus, for more than a decade, Norwegian schools were divided into two separate systems, where the length of compulsory schooling depended on the birth year and the municipality of residence at age 14, which we refer to as the childhood municipality. We use the timing differences across municipalities, induced by the staggered implementation of the reform, as our instrumental variable for school years. For more on the reform see Black et al. [2005].[14]

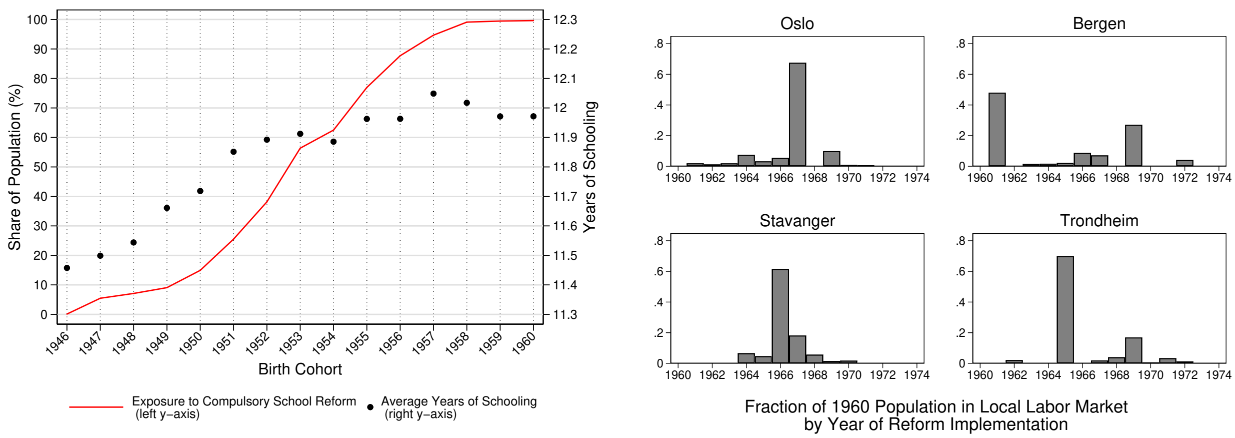
Historical records provide information about the year in which the reform was implemented for 672 out of the 732 municipalities in 1960. This information is missing for the remaining 60 municipalities [Monstad et al., 2008]. As shown in Figure 2, there is considerable variation in the fraction of birth cohort exposed to the reform (Figure 2-(a)) and in the timing of reform even within local labor markets (Figure 2-(b)). In particular, panel (a) shows that nobody born before 1946 was subjected to 9 years of compulsory schooling law, whereas everyone born after 1960 was affected by the new law.

Figure 2-(b) shows that there is considerable variation even within the four largest local labor markets (the four biggest metropolitan areas in Norway). For instance, the municipality of Oslo city, which accounted for two-thirds of the population in the Oslo labor market region in 1960, implemented the reform in 1967, whereas the timing of the reform varied between 1961 and 1971 across the remaining population living in one of the other 39 municipalities.[15]

As discussed in Section 3, to separately identify the private and the social returns to education, the instrument should satisfy the standard IV assumptions (Assumption 1) and also be a hidden instrument (Assumption 2). An implication of the latter assumption in our

[14] This compulsory schooling reform in Norway has been used previously, albeit in different contexts, by Monstad et al. [2008]; Aakvik et al. [2010]; Machin et al. [2012], and Bhuller et al. [2017].

[15] We use the classification of Norway into 160 local labor markets based on geographic commuting patterns constructed by Gundersen and Juvkam [2013]. On average each market has 5 municipalities.



(a) Exposure across Birth Cohorts

(b) Timing within Local Labor Markets

Figure 2: Compulsory School Reform Across Birth Cohorts and Local Labor Markets.
Note: The red line in plot (a) shows the cohort-specific share of population exposed to the compulsory school reform, while the black dots indicate the average years of schooling for Norwegian male cohorts born 1946-1960. Plot (b) shows the fraction of 1960 population in the four biggest local labor markets (concentrated around the four major cities) by the year of reform implementation. Using the 1960 classification of municipalities, there were 40 municipalities in the Oslo region, 27 municipalities in the Trondheim region, and 25 municipalities each in the Bergen and Stavanger regions. The variation in the timing of reform within local labor markets (LLMs) is due to variation in the timing of reform across municipalities within LLMs.

setting is that employers are not informed about the interaction between a worker's birth cohort and the timing of compulsory school reform in the worker's municipality of childhood.

For two reasons we think this assumption is reasonable in our setting. First, in contrast to compulsory schooling laws legislated centrally in many countries or by the states in the U.S. states, the timing of the implementation of the Norwegian compulsory school reform was decentralized and decided at the local municipal level. This decentralized implementation is consistent with our data, e.g., Figure 2-(b) that displays substantial variation in the timing of the reform even within local labor markets. Within local labor markets, there are high rates of commuting and mobility. This means that to know whether or not an individual was treated, an employer not only would have to know the exact date of implementation for each municipality, but would have to determine the childhood municipality of each worker (or job applicant). While it might be easier to discern the place of residence and birth year, from the CV, say, determining the childhood municipality would be difficult and expensive, if not impossible.

Second, even if employers had information on each applicant's birth year and childhood municipality, retrieving information on exposure to compulsory school reform for each applicant would still be onerous and costly. The information on the timing of compulsory reform was until recently not readily available in online public databases.[16] Therefore, for the 1946-1960 cohorts, graduating in an era long before the internet, this information would not have been easily traceable for employers.

Even though we do not directly test the hidden instrument assumption, to substantiate the identifying assumption that our IV is indeed hidden, we also restrict our analytical sample in Section 5 by *excluding* workers who grew up in the municipality with the largest population in each local labor market. Heuristically, by focusing on the subset of remaining workers, for whom it is plausible to assume that the employers are uninformed about the timing of reform in their childhood municipality and consequently their reform exposure status, we argue that the hidden instrument assumption is likely to be satisfied in our setting. This restricted – and our preferred – sample retains 422,749 individuals and 8,697,979 person-year observations, which is 59% of the full sample. For completeness, we also present results from the IV analysis for the full sample retaining individuals from the main municipality.

4.4 Empirical Specifications

4.4.1 Instrumental Variable Specification

Our first empirical specification projects log-earnings on schooling and control variables X at each experience t:

$$\ln W_{it} = \alpha_t^{(1)} + \beta_{s,t}^{(1)} S_i + \tau_t^{(1)} X_i + \omega_{i,t}^{(1)}, \tag{15}$$

where $\ln W_{it}$ and S_i represent log-earnings and schooling respectively, X_i is a vector of control variables, including a full set of dummies for birth cohort and childhood municipality. We discuss the reasons for including these control variables shortly below. The data allow us to estimate Equation (15) separately for each t, and thus we can allow the work experience

[16] Previously, Black et al. [2005] and Monstad et al. [2008] tracked various historical documents and databases to construct information on the timing of reform for 672 out of 732 municipalities.

to “interact” with individual characteristics X_i flexibly. In particular, in Equation (15) we have specified $H(t, X_i) = \alpha_t^{(1)} + \tau_t^{(1)} X_i$, where coefficients $\alpha_t^{(1)}$ and $\tau_t^{(1)}$ are t-specific, and thus flexibly capture both common experience profile and its interactions with X_i.

Our parameter of interest is $\beta_{s,t}^{(1)}$, which we estimate using 2SLS. Under Assumption 2, $\beta_{s,t}$ converges to the social returns to education as $t \to \infty$, and for any small (finite) t the IV consistently estimates of the private returns at t. Thus we can estimate the social returns to education as the limit of $\beta_{s,t}$. And we can use the rate at which $\beta_{s,t}^{(1)}$ converges to $\beta_{s,\infty}$ to estimate the speed of learning, κ.

The IV model consists of the second-stage Equation (15) and the first-stage equation

$$S_i = \mu + \lambda D_i + \rho X_i + \varphi_i, \tag{16}$$

where the binary instrument $D_i \in \{0, 1\}$ is equal to 1 if the individual was exposed to the new schooling law, and 0 otherwise and X is as before a full set of dummies for birth cohort and childhood municipality. An individual i is coded to be exposed if the reform had been implemented in i's childhood municipality of residence by the time he had turned 14.[17]

We estimate the system of Equations (16) and (15) by 2SLS, separately for each year of experience, t.[18] We use the childhood municipality indicators to control for unobservable determinants of earnings or schooling fixed at the municipality level, and use the birth cohort indicators to control for aggregate changes in schooling and earnings across cohorts (e.g., due to technical change). We assume that conditional on X, D satisfies Assumption 1, and as discussed in Section 4.3, that D satisfies the hidden IV Assumption 2.[19]

[17] At that time the school starting age in Norway was 7 years, and before the reform the critical age at which a pupil would be required to take two additional years of schooling was 14 years. Cohorts with ages 14 years or less at the time of school reform would be required to take the two additional years, while all cohorts aged above 14 at the time the new law went into effect would not.

[18] Unlike Equation (15), there is no experience subscript t attached to the λ coefficient on our instrument D in the first-stage equation because both compulsory schooling reform exposure status D and schooling S are time-invariant variables. However, with an unbalanced panel and separate estimations by experience, the first-stage estimates of λ will be allowed to vary by t. In practice, estimates λ are very stable across the experience range that we consider despite differences in the sample composition by experience.

[19] The timing of reform is also uncorrelated with baseline municipality characteristics [Bhuller et al., 2017].

Following Bhuller et al. [2017], we also test the stability of our first-stage and IV estimates to the inclusion of extrapolated linear and quadratic municipality-specific trends in education attainment and lifetime earnings estimated using data on pre-reform cohorts as additional controls. We refer to estimates based on Equations (16) and (15) as obtained from the baseline specification, and estimates that we get after further controlling for municipality-specific trends as coming from the trends specification.

4.4.2 OLS Specification Using a Hidden Correlate

As discussed above, we can estimate the speed of learning κ using the IV estimates as well as the estimates relying on the IQ test score as a hidden correlate (Section 3.2). This latter approach requires projecting log-earnings on schooling S, IQ score, Z, and other control variables, X, at different work experience level t:

$$\ln W_{it} = \alpha_t^{(2)} + \beta_{s,t}^{(2)} S_i + \beta_{z,t}^{(2)} Z_i + \tau_t^{(2)} X_i + \omega_{i,t}^{(2)}. \tag{17}$$

Under the assumptions that schooling does not independently enter $H(t, X)$ and that Z is unobserved in the market, we can use the OLS estimates of $\{\beta_{z,t}^{(2)}, \beta_{s,t}^{(2)}\}$ to obtain two estimates of the speed of learning κ. See Lange [2007] for further details.

It is well known that log-earnings tend to be nonlinear in schooling. Thus, we cannot simply compare the OLS estimates and IV estimates that we get from Equations (15) and (17). Comparing OLS and the IV estimates in the presence of non-linearities can be misleading simply because the OLS and the IV estimates weigh different marginal returns to schooling differently. We can, however, construct weighted OLS estimates that are comparable to the IV estimates by first estimating the fully non-linear model in OLS and then weighting the marginal returns using the weights that correspond to the IV estimator. This re-weighting procedure ensures that the OLS estimates are obtained from the same support of schooling distribution as the IV estimates and thus allows us to compare estimates of the speed of learning across estimators in the presence of non-linearities. We refer to these re-weighted OLS estimates as IV-weighted OLS estimates and denote them by $(\beta^{(3)}_{s,t}, \beta^{(3)}_{z,t})$.[20]

5 Main Results

This section contains our main empirical results. To begin, we present the IV estimates of returns to education over work experience and use these estimates to determine the speed of learning. We then use the same IV estimates for our main contribution, which is to provide estimates of the private and social returns to education. The gap between these two returns represents our estimate of the contribution of signaling to the return to education. Finally, we present OLS estimates that use the IQ test score as a hidden correlate of ability.

5.1 IV Estimates of the Returns to Education

Table 1 column (1) displays the first-stage estimate of the effect of our compulsory schooling reform instrument on years of schooling, as defined in Equation (16), for the full sample. This estimate indicates that exposure to compulsory schooling reform increased completed schooling by 0.237 years. The partial F-statistic is approximately 88, which means that weak instrument bias is not a concern for our analysis.

As described above, how we interpret the IV estimates depends crucially on whether the IV is hidden or transparent. We are more confident that the IV is hidden when we restrict ourselves to the variation across small municipalities that surround the core of large urban agglomerations. Our preferred estimates therefore derive variation from a sample that excludes those born in the largest municipalities in the different labor markets. These estimates are in column (3), and we can see that the effect of our IV on education is unchanged. We repeat these two estimation exercises including municipality-specific trends (columns 2 and 4), and find that although the absolute effect is smaller, the conclusion does not change.

[20]We follow Angrist and Imbens [1995]; Løken et al. [2012] and Mogstad and Wiswall [2016], and provide additional details on the re-weighting procedure in the Appendix Section A.1.

Table 1: First-Stage Estimates on Years of Schooling.

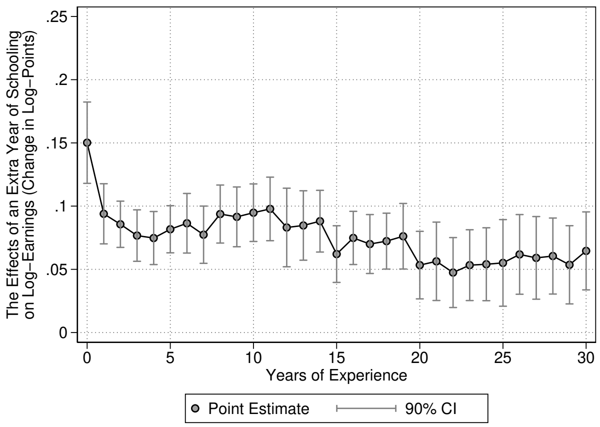
	Full Sample		Preferred Sample	
	Baseline Specification (1)	Trends Specification (2)	Baseline Specification (3)	Trends Specification (4)
Instrument:				
Exposure to Compulsory Schooling Reform	0.237^{***}	0.209^{***}	0.228^{***}	0.209^{***}
	(0.025)	(0.034)	(0.034)	(0.040)
Municipality Fixed Effects	✓	✓	✓	✓
Cohort Fixed Effects	✓	✓	✓	✓
Municipality-Specific Trends		✓		✓
F-statistic (instrument)	87.7	37.9	45.7	27.7
Sample Mean Years of Schooling	12.36	12.36	12.27	12.27
Standard Deviation Years of Schooling	2.50	2.50	2.46	2.46

Note: The full estimation sample consists of Norwegian males born in 1950-1980 observed any time in earnings data over years 1967-2014 with years of potential experience between 0 and 30 years and annual earnings above 1 SGA threshold (N=14,746,755). The restricted estimation sample further drops individuals who grew up in the municipality with the largest population size in each of the 160 labor market regions in Norway (N=8,697,979). All estimations include fixed effects for birth cohort and childhood municipality. The trends specifications in columns (2) and (4) further also controls for linear and quadratic municipality-specific trends estimated using data on all pre-reform cohorts born 1930 or later and extrapolated to all post-reform cohorts, separately for each municipality. Standard errors are clustered at the local labor market region (160 groups). * $p < 0.10$, ** < 0.05, *** $p < 0.01$.

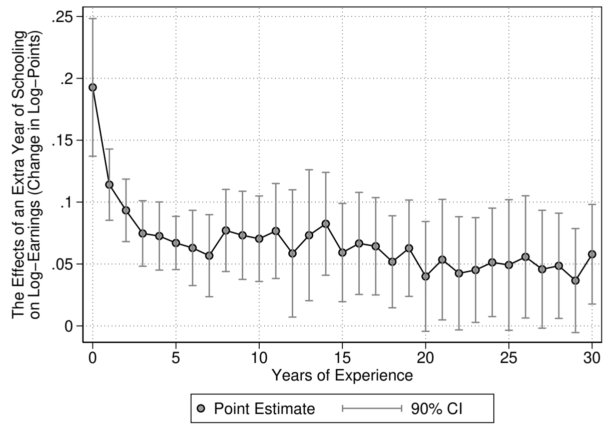
We now turn our attention to the second-stage IV estimates in Equation (15). Figure 3-(a) displays the IV estimates for the full sample, and these coefficients represent the private returns to schooling, at each year of work experience. Similarly, Figure 3-(b) displays the IV estimates for the restricted sample, and Figures 3-(c) and 3-(d) display the IV estimates for each of these two samples with municipality-specific trends, respectively.

All four panels exhibit point estimates that suggest high initial returns to schooling, followed by a relatively steep decline during the first 5 years of work. Then, the returns gradually stabilize and approach 5-6% for those with 15 years or more of work experience. These patterns are consistent with employers learning about workers' ability. Moreover, these estimates also indicate that employers did not fully price in the variation in schooling that is induced by the variation in compulsory schooling reform exposure, across cohorts and municipalities, which is consistent with our hidden IV assumption.

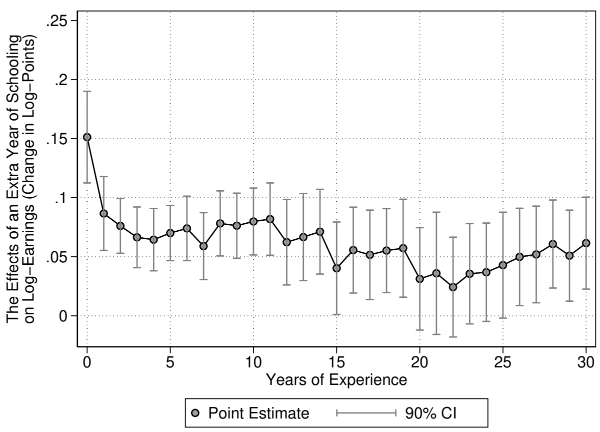
Using these IV estimates of returns to schooling, i.e., $\{\beta_{s,t}^{(1)}\}_{t=0}^{T}$, we can determine the



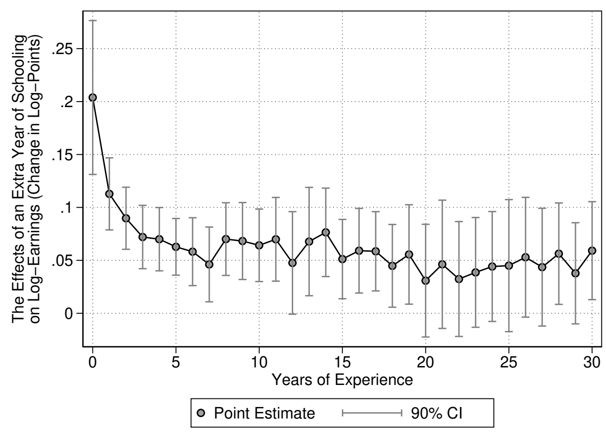
(a) Full Sample–Baseline Specification



(b) Preferred Sample–Baseline Specification



(c) Full Sample–Trends Specification



(d) Preferred Sample–Trends Specification

Figure 3: IV Estimates of the Returns to Schooling by Year of Experience.

Note: The full estimation sample consists of Norwegian males born 1950-1980 observed in earnings data over years 1967-2014 with years of potential experience between 0 and 30 years and annual earnings above 1 SGA threshold (N=14,746,755). The restricted estimation sample further drops those growing up in the municipality with largest population size in each the 160 labor market regions in Norway (N=8,697,979). Plots (a) and (b) display IV estimates from separate estimations of Equation (15) for each year of experience using the two samples, while plots (c) and (d) further control for municipality-specific trends. All estimations include fixed effects for birth cohort and childhood municipality. Standard errors are clustered at the local labor market region (160 groups). The 90% confidence intervals corresponding to each point estimate are displayed as vertical bars.

speed of employer learning. To see this, note that we can express the IV estimate as $\beta^{(1)}_{s,t} = \theta_t \times b^{(1)}_{s,0} + (1-\theta_t) \times b^{(1)}_{s,\infty}$, where $b^{(1)}_{s,0}$ is the private returns to education at $t = 0$, $b^{(1)}_{s,\infty}$ is the social returns to education and θ_t is the weight defined in Equation (4). Using the IV estimates for all t, we can estimate the RHS parameters using the non-linear least squares method. Heuristically, we can "solve" for $\{b^{(1)}_{s,0}, b^{(1)}_{s,\infty}, \theta_t\}$ from $\{\beta^{(1)}_{s,t}\}^T_{t=0}$, and once we know θ_t we can determine the speed of learning parameter κ.

Table 2: IV Estimates of the Speed of Employer Learning, Initial Value and Limit Value.

	Full Sample		Preferred Sample	
	(1) Years of Schooling	(2) Years of Schooling	(3) Years of Schooling	(4) Years of Schooling
Speed of Learning κ	0.447***	0.490***	0.532***	0.565***
	(0.127)	(0.110)	(0.058)	(0.055)
Initial Value $b_{s,0}$	0.145***	0.148***	0.192***	0.204***
	(0.013)	(0.014)	(0.010)	(0.010)
Limit Value $b_{s,\infty}$	0.063***	0.048***	0.050***	0.045***
	(0.004)	(0.004)	(0.003)	(0.003)
Weight θ_t on Initial Signal:				
at $t = 5$	19.8%	17.2%	15.0%	13.3%
at $t = 10$	11.0%	9.4%	8.1%	7.1%
at $t = 15$	7.6%	6.5%	5.5%	4.9%
Municipality Fixed Effects	✓	✓	✓	✓
Cohort Fixed Effects	✓	✓	✓	✓
Municipality-Specific Trends		✓		✓

Note: The full estimation sample consists of Norwegian males born 1950-1980 observed in earnings data over years 1967-2014 with years of potential experience between 0 and 30 years and annual earnings above 1 SGA threshold (N=14,746,755). The estimates plotted in Figure 3(a) for the full estimation sample are used to construct the corresponding IV estimates of speed of learning, initial value and limit value in columns (1)-(2). The estimates in columns (3)-(4) are similarly based on the estimates plotted in Figure 3(b) for a restricted estimation sample in which the municipality with largest population size in each of the 160 labor market regions in Norway is dropped (N=8,697,979).
* $p < 0.10$, ** < 0.05, *** $p < 0.01$.

Table 2 displays the estimates of these parameters obtained using the coefficient estimates shown in Figure 3. Comparing the estimates across columns (1)-(4), we can see that the estimates are robust with respect to sample restrictions and controls for municipality-specific trends. In particular, we cannot reject that the equality of the speed of learning estimates across columns (1)-(4). The point estimates of κ are between 0.447 and 0.565, which imply very rapid learning on the part of employers. More precisely, our preferred estimate of the speed of employer learning at 0.532 in Table 2-(3) implies that already after the first five years of employment, employers put only 15% weight on the initial signal they received from the worker, and after 15 years of employment history this weight further declines to 5.5%.

5.2 The Signaling Value of Education

Next, we use the estimates from Table 2 to determine the signaling value of education. Our employer learning model implies that the limit return to schooling $b^{(1)}_{s,\infty}$ is the social returns to education. The experience-specific IV estimates directly represent the private returns to education. Using estimates of model parameters in Table 2, we can also construct estimates of the private returns at each t. In Figure 4 we display the private and social returns based on the estimates from Table 2-(3), as well as the experience-specific IV estimates. The scatter plot displays the IV estimates obtained from the preferred sample in Figure 3-(c), and the horizontal solid line is the estimated social return to education $b^{(1)}_{s,\infty}$ from Table 2-(3) at 5%.

In order to determine the signaling value of education we also need the private *internal rate of return* (IRR) for an additional year of schooling. The private IRR is defined as the discount rate that equates present discounted value of earnings over the career for different choices of schooling. Using the experience-specific IV estimates of the private returns to education (the scatter plot in Figure 4), we estimate the private IRR to be 7.2%.[21] The private IRR is 2.2 percentage points greater than the social returns to schooling at 5%. From these estimates, we conclude that 70% of the private return to education can be attributed to education raising the productivity of workers and 30% to the signaling value of education.

Alternatively, we can use the estimates of $b^{(1)}_{s,0} = 0.192$, $b^{(1)}_{s,\infty} = 0.05$ and $\kappa = 0.532$ from Table 2-(3) directly to calculate the private IRR at each t, corresponding to the black dotted line in Figure 4. Imposing this learning process and assuming a career length of 40 years, we obtain an estimate of the private IRR of 7.2%. This estimate of the private IRR is identical to the estimate we obtained using the experience-specific IV estimates, and so in both cases we calculate that 30% of the private return to education can be attributed to signaling. For robustness to the speed of employer learning in this calculation, we repeat the exercise using $\kappa = 0.447$ from Table 2-(1), corresponding to the gray dotted line in Figure 4. If employer learning is slower, then our model implies a higher private IRR for the same earnings profiles,

[21]For $t > 31$ and beyond retirement age, we assume that experience-specific IV estimate also equal 5%.

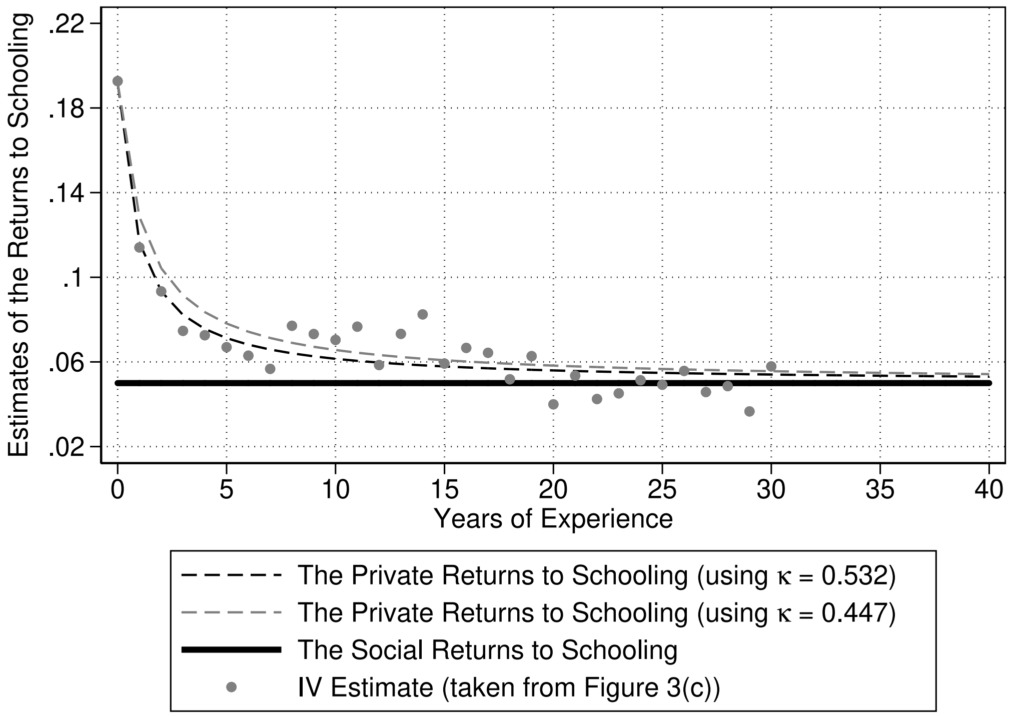


Figure 4: The Private and Social Returns to Schooling.
Note: This figure is constructed using the estimates presented in Table 2, columns (1) and (4), and displayed in Figure 3-(a).

but the social return does not change because it does not depend on the speed of learning. Indeed, we then estimate the private IRR to be at 7.7%, which exceeds social returns by 2.7 percentage points and implies a signaling contribution of 35% in private returns.

As an additional evidence that education has direct effect on productivity, we also estimated specifications with our standardized IQ test score as the dependent variable and years of schooling, instrumented using the compulsory schooling reform, as the main independent variable. These IV estimates showed strong effects of schooling on IQ, with an additional year of schooling at age 18 causing an 1/4 of a standard deviation increase in IQ. [22] This evidence supports the hypothesis that schooling increases productive skills and that most of the private return to education is attributed to productivity enhancing effects of education.

We can also compare the estimate of the social returns to education of 5% with a standard

[22]These additional results are available upon request. Brinch and Galloway [2012] have previously documented these results for Norway, and Carlsson et al. [2015] also documented similar results for Sweden.

Mincer returns to education – that is the (time constant) coefficient on years of schooling in a non-interacted specification controlling for a flexible experience profile. This comparison indicates how much the social returns differ from the observed average differences in earnings in the population and is of interest since the Mincer coefficient is a very commonly used indicator of the value of education. We find that the Mincer coefficient is at 6.8%, exceeding the social returns by 1.8 percentage points.

5.3 OLS Estimates Using a Hidden Correlate of Ability

Next, we present results from the OLS specification that uses a correlate of ability that is observed by us but not by the employers. We begin by presenting IV-weighted OLS estimates of Equation (17), for each year of work experience, using the standardized IQ test score as the hidden correlate of ability, and after controlling for municipality and cohort fixed effects. Figures 5-(a) and 5-(b) display the IV-weighted OLS estimates of returns to schooling, $\beta_{s,t}^{(3)}$, and returns to IQ, $\beta_{z,t}^{(3)}$, respectively. The estimates of $\beta_{s,t}^{(3)}$ decline rapidly in the first few years before stabilizing, and the estimates of $\beta_{z,t}^{(3)}$ increase with experience, rapidly at first, and then slowly until stabilizing after 15 years.

Comparing Figures 5-(a) and 3-(a) we can see that the IV-weighted OLS and the IV estimates of returns to schooling reflect a similar pattern although these two estimators use different sources of variation. It is also noteworthy that the patterns in the returns to schooling and IQ over the workers' careers are surprisingly similarly to those found in the NLSY using the AFQT score. For instance, Figure 1 in Lange [2007] indicates that the returns to schooling decline early in the career and the returns to IQ score increase rapidly before converging to stable, long-run, values after a few years.

Relatedly, Arcidiacono et al. [2010] also use the NLSY data and find that the returns to the AFQT score increases with experience for those with a high school degree or less, and for those with a college degree the returns to the AFQT score are constant over the life-cycle. This led them to conclude that a college degree has a direct role in revealing ability. Such a mechanism could be at play if employers are better informed about the differences in cognitive ability among those with and without a college degree, possibly because they observe transcripts, field of study, reference letters and students have additional work experience, e.g., internships.

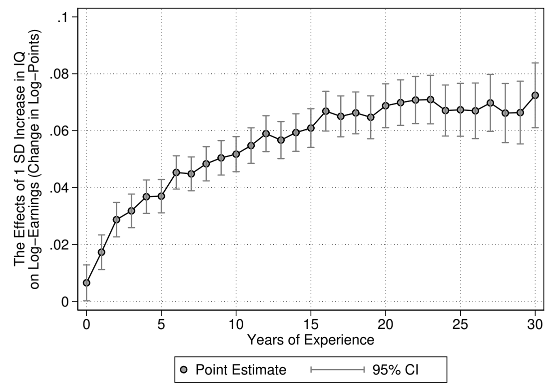
When we perform the same analysis as Arcidiacono et al. [2010], i.e., split our estimation sample in two groups –one with at most a high school degree and other with a college degree– we find similar results in our data. Figures 5-(c) and 5-(d) display the OLS estimates of returns to schooling and IQ, respectively, for the first sample and Figures 5-(e) and 5-(f) show the corresponding estimates for the second sample. We can see that the returns to IQ increase with experience but only for those with at most high school degree, and for those with a college degree the returns to IQ are constant at around 6-7% across all years of experience. This pattern is consistent with a college degree revealing a worker's ability also in the Norwegian labor market.

Next, we compare the IV-weighted OLS estimates to the OLS estimates we obtained from separate estimations for the two groups of workers differentiated by their education attainment. Comparing the estimates of returns to schooling in Figures 5-(a) and 5-(c) and returns to IQ in Figures 5-(b) and 5-(d), we confirm that the IV-weighted OLS estimates are similar to the OLS estimates for workers with a high school degree or less. This is reassuring because the IV-weighted OLS estimates must put substantially more weight on marginal returns in the lower end of the schooling distribution, as shown in Appendix Section A.1. In contrast, estimates for college educated workers display a very different pattern.

Using the IV-weighted OLS estimates of returns to schooling and IQ displayed in Figures 5 (a)-(b), i.e., $\{\beta_{s,t}^{(3)}, \beta_{z,t}^{(3)}\}_{t=0}^{T}$, we can construct additional estimates of the speed of employer learning. As before, from Equation (12) we know that $\{\beta_{s,t}^{(3)}, \beta_{z,t}^{(3)}\}_{t=0}^{T}$ satisfy

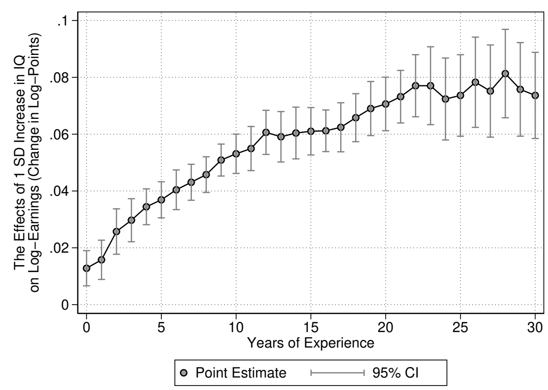
$$\beta_{s,t}^{(3)} = \theta_t b_{s,0}^{(3)} + (1-\theta_t)\, b_{s,\infty}^{(3)} \quad \text{and} \quad \beta_{z,t}^{(3)} = \theta_t b_{z,0}^{(3)} + (1-\theta_t)\, b_{z,\infty}^{(3)},$$

(a) Returns to Schooling: IV-Weighted OLS

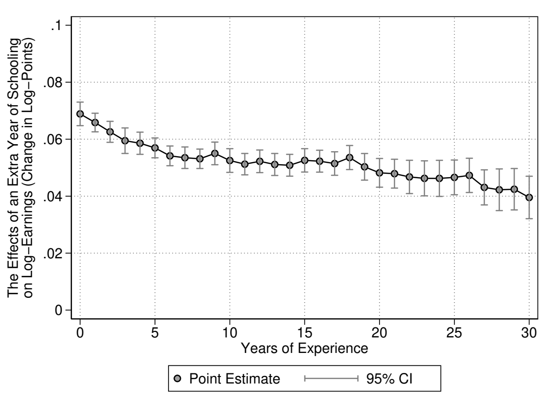


(b) Returns to IQ: IV-Weighted OLS

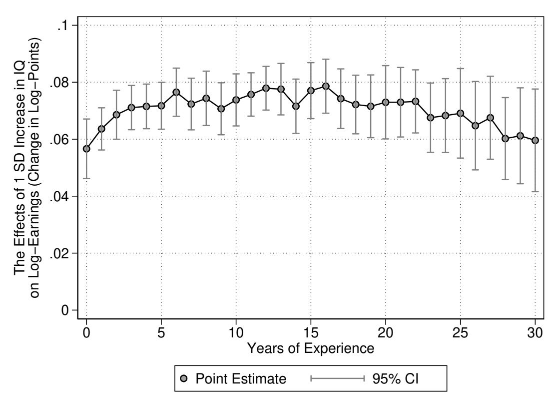
(c) Returns to Schooling: High School or Below



(d) Returns to IQ: High School or Below



(e) Returns to Schooling: College or University



(f) Returns to IQ: College or University

Figure 5: OLS Estimates of the Returns to Schooling and IQ.

Note: The estimation sample consists of Norwegian males born 1950-1980 observed in earnings data over years 1967-2014 with years of potential experience between 0 and 30 years and annual earnings above 1 SGA threshold (N=14,746,755).

where $\{b_{s,0}^{(3)}, b_{z,0}^{(3)}\}$ are the projection coefficients of log-earnings on schooling and ability at the start of a career, and $\{b_{s,\infty}^{(3)}, b_{z,\infty}^{(3)}\}$ are the projection coefficients that would be observed

Table 3: OLS Estimates of the Speed of Employer Learning, Initial Value and Limit Value.

	Two Values of κ		One Value of κ	
	(1) Years of Schooling	(2) IQ Test Score	(3) Years of Schooling	(4) IQ Test Score
	A. Full Sample – IV-Weighted OLS			
Speed of Learning κ	0.386***	0.127***	0.214***	
	(0.045)	(0.023)	(0.029)	
Initial Value $(b_{s,0}, b_{z,0})$	0.096***	0.007**	0.084***	-0.001
	(0.004)	(0.004)	(0.004)	(0.004)
Limit Value $(b_{s,\infty}, b_{z,\infty})$	0.024***	0.086***	0.019***	0.077***
	(0.002)	(0.004)	(0.002)	(0.003)
	B. Compulsory/High School Sample – Standard OLS			
Speed of Learning κ	0.333***	0.065***	0.104***	
	(0.034)	(0.010)	(0.015)	
Initial Value $(b_{s,0}, b_{z,0})$	0.091***	0.012***	0.074***	0.006**
	(0.003)	(0.002)	(0.003)	(0.003)
Limit Value $(b_{s,\infty}, b_{z,\infty})$	0.030***	0.111***	0.017***	0.097***
	(0.001)	(0.006)	(0.003)	(0.004)
	C. College/University Sample – Standard OLS			
Speed of Learning κ	0.061	0.788	0.115	
	(0.056)	(0.603)	(0.080)	
Initial Value $(b_{s,0}, b_{z,0})$	0.067***	0.056***	0.069***	0.068***
	(0.003)	(0.004)	(0.003)	(0.002)
Limit Value $(b_{s,\infty}, b_{z,\infty})$	0.033***	0.071***	0.039***	0.072***
	(0.008)	(0.001)	(0.004)	(0.002)
Municipality Fixed Effects	✓	✓	✓	✓
Cohort Fixed Effects	✓	✓	✓	✓

Note: The estimation sample consists of Norwegian males born 1950-1980 observed in earnings data over years 1967-2014 with years of potential experience between 0 and 30 years and annual earnings above 1 SGA threshold (N=14,746,755). The estimates of speed of learning, initial values of returns to schooling and IQ, and limit values of returns to schooling are obtained from non-linear least squares estimations on the experience-specific returns to schooling and IQ presented in Figure 5.
* p < 0.10, ** < 0.05, *** p < 0.01.

once productivity of individuals was fully revealed in the market. Similarly, using estimates displayed in Figures 5 (c)-(f), we can also construct separate estimates of speed of learning for workers with a high school degree or less and for workers with a college degree.[23]

Table 3 displays estimates of the speed of employer learning, κ, the initial returns to

[23] Unlike the IV estimates, the parameters $\{b^{(3)}_{s,0}, b^{(3)}_{z,0}, b^{(3)}_{s,\infty}, b^{(3)}_{z,\infty}\}$ lack a meaningful interpretation.

schooling and IQ, $(b_{s,0}^{(1)}, b_{z,0}^{(1)})$, and the limit returns to schooling and IQ, $(b_{s,\infty}^{(1)}, b_{z,\infty}^{(1)})$, that we obtained using non-linear least squares for each of the three sets of OLS estimates of returns to schooling and IQ displayed in Figure 5. Estimates in each panel in Table 3 correspond to one of the three sets of estimates in Figure 5. As noted by Lange [2007], κ is over-identified when a hidden correlate of ability is available, so we can construct two different estimates of κ based on the OLS estimates of returns to schooling and the OLS estimates of returns to IQ, respectively. If the same learning process drives how schooling and IQ coefficients evolve with worker experience then these two estimates of κ should be identical.

Alternatively, we can restrict the estimate of κ using the returns to schooling to be the same as the κ using the returns to IQ. We implement both methods and present the estimation results in Table 3. Estimates that allow for differential learning are in columns (1)-(2) and the estimates that impose a common learning process are in columns (3)-(4).

Consistent with what we found in Figure 5, here too we find similar estimates of employer learning, initial returns to schooling and IQ and limit returns to schooling and IQ across panels A and B in Table 3. As earlier, in panel A we show results using the IV-weighted OLS estimates, while in panel B we use OLS estimates for workers with a high school degree or below. Moreover, the estimates of returns to schooling and speed of employer learning in panels A-B, column (1), which allow for a differential learning process across schooling and IQ, are also comparable to the corresponding IV estimates we presented in Table 2, column (1). At conventional levels, we cannot reject the equality of the speed of learning κ across these OLS estimates and the IV estimates presented earlier.

There are however two striking differences among the estimates in Table 3. First, comparing estimates of the speed of learning κ across columns (1)-(2), we can reject the assumption of a common learning process for schooling and IQ over worker experience. This result suggests that a standard assumption made in the OLS approach that uses a hidden correlate of ability to identify the speed of employer learning is violated in our context. Second, we do not find any evidence of employer learning for workers with a college degree as shown in panel C in Table 3, unlike the results shown in panels A-B.

6 Extensions

The remainder of this paper considers two extensions we view as particularly important in charting the way ahead for the literature on employer learning and returns to education. First, we relax some of the strong functional form restrictions that the standard employer learning framework imposes on how earnings vary with schooling and ability over work experience. We provide additional results regarding identification of the private and the social returns to education in a more general setting than has been considered in the empirical literature. In that context, access to multiple instruments and hidden correlates (or both) holds the promise that some of the common functional form assumptions in the literature can be relaxed. Second, we consider spillovers from education that manifest across employers. Externalities from signaling arise because signaling affects how the economic surplus is shared between employers and workers. If education, however, also imparts productive spillovers across employers, education might increase the economic pie over and above the private returns. We adapt our baseline model to allow for local productivity spillovers of this type– and present some results that suggest that such spillovers might indeed be quite substantial.

6.1 Non-Separability between Education and Experience

The basic framework in the employer learning literature assumes that log-earnings are additively separable in education, unobserved ability, and experience. We next investigate whether we can relax this assumption and accommodate a more general functional form which allows an interaction between schooling and experience.

Thus, we assume that productivity is given by

$$\chi_{it} = H(S_i, t, Q_i) + A_i + \varepsilon_{it},$$

where $H(S,t,Q)$ is an unknown (to researchers) function that captures the effect of schooling, work experience and correlate of ability. While this model allows the marginal effect of work experience to vary with schooling, and vice versa, we still maintain the assumption that productivity is additively separable in A.[24] This simplifying assumption allows us to keep the employers' learning process tractable. We still maintain all previous assumptions about the model primitives, including the assumption that employers know $H(\cdot,\cdot,\cdot)$.

Following the steps that lead to (6) and (7), we can express productivity and wages as

$$\begin{aligned} \psi_{it} &= H(S_i,t,Q_i) + \delta^{A|S} S_i + \tilde{A}_i + \varepsilon_{it} \\ W_{it} &= H(S_i,t,Q_i) + \mathbb{E}(A_i|\mathcal{E}_{it}), \end{aligned}$$

respectively. The social returns to education defined to be the first partial derivative of ψ with respect to schooling generalizes (6), and can be expressed, after using (5), as

$$\underbrace{\frac{\partial \psi_{it}}{\partial S}}_{\texttt{social returns}} = H_1(S_i,t,\delta^{Q|S}S_i + \tilde{Q}_i) + H_3(S_i,t,\delta^{Q|S}S_i + \tilde{Q}_i)\delta^{Q|S} + \delta^{A|S}, \tag{18}$$

where $H_\iota(\cdot,\cdot,\cdot)$ denotes the partial derivative of $H(\cdot,\cdot,\cdot)$ with respect to its ι^{th} argument. Similarly, private returns to education are the marginal effect of schooling on wages, i.e.,

$$\begin{aligned} \underbrace{\frac{\partial W_{it}}{\partial S}}_{\texttt{private returns}} &= H_1(S_i,t,\delta^{Q|S}S_i + \tilde{Q}_i) + H_3(S_i,t,\delta^{Q|S}S_i + \tilde{Q}_i)\delta^{Q|S} + \delta^{A|S} \\ &\quad + \theta_t \left(\phi_{A|S} + \phi_{A|Q}\delta^{Q|S} - \delta^{A|S}\right). \end{aligned} \tag{19}$$

Comparing (18) and (19) we find the same relationship between social returns and private returns analogous to (8), i.e., $\frac{\partial W_{it}}{\partial S} = \frac{\partial \psi_{it}}{\partial S} + \theta_t \left(\phi_{A|S} + \phi_{A|Q}\delta^{Q|S} - \delta^{A|S}\right)$. Thus, with long enough work experience $\lim_{t\to\infty}\theta_t = 0$, and the returns converge, i.e., $\lim_{t\to\infty}\frac{\partial \ln W_{it}}{\partial S} = \frac{\partial \psi_{it}}{\partial S}$.

[24]It would be equally important to relax the assumption that ability and experience are additively separable. For now, we find it natural to examine whether it is possible to relax the assumption that schooling and experience are additive because our instrumental variable approach relies on inducing varation in schooling.

Now, let us consider the identification of causal effect of education using a binary IV $D_i \in \{0,1\}$ that satisfies Assumption 1. Following the same steps as before, we find that

$$\texttt{plim}\ \hat{b}_{IV,t} \quad = \quad \frac{\mathbb{E}\left[\ln W_{it}|D_i=1,t\right]-\mathbb{E}\left[\ln W_{it}|D_i=0,t\right]}{\mathbb{E}\left[S_i|D_i=1,t\right]-\mathbb{E}\left[S_i|D_i=0,t\right]}. \tag{20}$$

So, with experience $t\to\infty$ we can again identify the social returns to education as

$$\texttt{plim}\ \left(\lim_{t\to\infty}\hat{b}_{IV,t}\right)=\texttt{plim}\ \hat{b}_{IV,t\to\infty}=\frac{\mathbb{E}\left[\psi_{it}|D_i=1\right]-\mathbb{E}\left[\psi_{it}|D_i=0\right]}{\mathbb{E}\left[S_i|D_i=1\right]-\mathbb{E}\left[S_i|D_i=0\right]}=\lim_{t\to\infty}\mathbb{E}\left[\frac{\partial\psi_{it}}{\partial S}\right].$$

Thus, for long enough experience, IVs identify the (experience-specific) social returns to education. Now, let us consider those with limited years of experience ($t<\infty$). When D is a hidden instrument, $W_{it}=\mathbb{E}\left[\psi_i|\mathcal{E}_{it}\right]=\mathbb{E}\left[\psi_i|\mathcal{E}_{it},D_i\right]$. Then using $\mathbb{E}\left[Q_i|D_i,S_i\right]=\delta^{Q|S}S_i$ from Assumption 1 and Equations (2) and (5) and after some simplification we get

$$\begin{aligned}\mathbb{E}\left[W_{it}|D_i,t\right] &= \mathbb{E}[H(S_i,t,\delta^{Q|S}S_i+\tilde{Q}_i)|D_i]+\mathbb{E}\left[\mathbb{E}\left[A_i|S_i,Q_i,\xi_i^t\right]|D_i\right]\\ &= \mathbb{E}[H(S_i,t,\delta^{Q|S}S_i+\tilde{Q}_i)|D_i]+\left(\theta_t\left(\phi_{A|S}+\phi_{A|Q}\delta^{Q|S}\right)+(1-\theta_t)\,\delta^{A|S}\right)\mathbb{E}\left[S_i|D_i\right].\end{aligned}$$

Then, evaluating the LHS at $D=1$ and subtracting its value at $D=0$ gives $\Delta_D\mathbb{E}\left[W_{it}|D_i,t\right]=\Delta_D\mathbb{E}(H(S_i,t,\delta^{Q|S}S_i+\tilde{Q}_i)|D_i)+\delta^{A|S}\Delta_D\mathbb{E}\left[S_i|D_i\right]+\theta_t\left(\phi_{A|S}+\phi_{A|Q}\delta^{Q|S}-\delta^{A|S}\right)\Delta_D\mathbb{E}\left[S_i|D_i\right].$ Then substituting this expression in (20) and taking the probability limit gives

$$\begin{aligned}\texttt{plim}\ \hat{b}_{IV,t} &= \frac{\Delta_D\mathbb{E}\left[\ln W_{it}|D_i\right]}{\Delta_D\mathbb{E}\left[S_i|D_i\right]}=\frac{\Delta_D\mathbb{E}(H(S_i,t,\delta^{Q|S}S_i+\tilde{Q}_i)|D_i)}{\Delta_D\mathbb{E}\left[S_i|D_i\right]}+\delta^{A|S}\\ &\quad+\theta_t\left(\phi_{A|S}+\phi_{A|Q}\delta^{Q|S}-\delta^{A|S}\right)\\ &= \mathbb{E}[H_1(S_i,t,\delta^{Q|S}S_i+\tilde{Q}_i)]+\mathbb{E}[H_3(S_i,t,\delta^{Q|S}S_i+\tilde{Q}_i)]\delta^{Q|S}+\delta^{A|S}\\ &\quad+\theta_t\left(\phi_{A|S}+\phi_{A|Q}\delta^{Q|S}-\delta^{A|S}\right),\end{aligned}$$

which is the expected private returns to education, as defined for each t in (19). Thus with a hidden instrument we identify the private returns to education. Following the same logic, we conclude that if D is transparent, it identifies the average of social returns to education.

In this section, we have only considered the problem of identifying the private and social returns to education within our employer learning framework. The estimation of $H(\cdot, \cdot, \cdot)$ itself represents a difficult problem that we leave for future research. The difficulties arise in part because our instrument is binary, so we cannot rely on the continuous nonparametric IV methods of Newey and Powell [2003]. Recently, Torgovitsky [2017] has proposed a minimum distance estimator for a similar problem that relies on the identification results from Torgovitsky [2015]; also see D'Haultfœuille and Février [2015] for a related approach.

6.2 Productive Externalities

We next consider local productivity spillovers due to education. In other words, we study if, all else equal, working in a labor market that has a more educated work force leads to higher earnings. If there are such spillovers, it stands to reason that workers and employers will choose to locate to those markets with higher aggregate education, shoring up the land value and the rent, until they are indifferent between all markets. Thus, to estimate the local spillovers properly, we must enrich our basic model and incorporate spatial constraints. For this purpose, we use Rosen [1979] and Roback [1982]'s framework of spatial equilibrium with mobile firms and workers, as described in Lange and Topel [2006].[25]

For brevity and following the derivations in Lange and Topel [2006], we get the following specification that is standard and widely used in the literature:

$$\ln W_{it} = \alpha_t + \beta_t S_i + \overline{\beta}\ \overline{S}_{\ell y} + \tau_t X_i + \tilde{\omega}_{it}, \tag{21}$$

where $\overline{S}_{\ell y}$ is the average years of schooling across all individuals aged 16 to 67 in market ℓ as individual i who worked full-time (earnings $>$ 1 SGA) in the calendar year y. The rest of the variables are defined as in Equation (15), except the error $\tilde{\omega}_{it}$, which includes a variable

[25] The Rosen-Roback framework has been adapted by Rauch [1993], Acemoglu and Angrist [2000], Moretti [2004], and Lange and Topel [2006] to study productive externalities of education using wage regressions.

that is correlated with $\overline{S}_\ell$. Thus $\overline{S}_\ell$ is also endogenous and we need an additional IV.[26]

A natural choice for an IV is the fraction of workers in market ℓ in year y who had been exposed to the compulsory schooling reform. Let $\overline{D}_{\ell y} \in [0,1]$ denote such a fraction. The reason why $\overline{D}_{\ell y}$ is a valid instrument is similar to the reason why D_i is a valid instrument for S_i, i.e., $\overline{D}_{\ell y}$ is positively correlated with $\overline{S}_{\ell y}$, and it does not directly affect the local wages except through its effect on average education in market ℓ and year y. Thus the first-stage equations for individual schooling and aggregate schooling are given by

$$S_i = \mu + \lambda_1 D_i + \bar{\lambda}_1 \overline{D}_{\ell y} + \rho X_i + \varphi_i; \tag{22}$$

$$\overline{S}_{\ell y} = \bar{\mu} + \lambda_2 D_i + \bar{\lambda}_2 \overline{D}_{\ell y} + \bar{\rho} X_i + \bar{\varphi}_{iy}. \tag{23}$$

It is important to note that there is variation in the instrument $\overline{D}_{\ell y}$ even after conditioning on D_i, birth cohort and childhood municipality. This variation is crucial for the identification in our setting with two endogenous regressors $(\overline{S}_{\ell y}, S_i)$ in Equation (21). To understand the source of this variation, consider two cohorts (c', c'') from the same local labor market, both exposed to the reform, i.e., $D_i = 1$ for all i in cohorts c' and c'', but suppose cohort c' was the first cohort in that labor market to be exposed to the reform, such that any later cohort $c > c'$ (including c'') will also be exposed to the reform. Conditional on $D_i = 1$, there will be variation in $\overline{D}_{\ell y}$ across c' and c'', since the two cohorts enter the labor market at different times and will thus be exposed to different individuals in the labor market. Specifically, when cohort c' enters the labor market, none of the earlier cohorts in this local labor market would have been subjected to the compulsory schooling reform, while when c'' enters the labor market, workers from c'' will be exposed to *some* earlier cohorts $c \in \{c', \ldots, c''-1\}$ that were also exposed to the reform. Thus, conditional on D_i, there will still be variation in $\overline{D}_{\ell y}$ that we can rely for the identification. Using the staggered reform, we can still control

[26] Lange and Topel [2006] show that the $\overline{\beta}$ parameter is proportional to the external return of education, where the factor of proportionality depends on the parameters of preferences and the production function in the Rosen-Roback model. They calibrate the factor of proportionality to be, approximately, 0.8, which means that the estimate of $\overline{\beta}$ underestimates the external return of education by approximately 20%.

Table 4: Returns to Individual Schooling and Average Schooling on Log-Earnings.

	Outcome Equation:		First-Stage Equations:	
	(1)		(2)	(3)
Outcome Variable:	*Log-Earnings*	Endogenous Variables:	*Individual Schooling*	*Average Schooling*
Endogenous Variables:		Instruments:		
Individual Schooling	0.033^{**}	*Individual CSR Exposed*	0.210^{***}	0.023^{***}
	(0.015)		(0.025)	(0.005)
Average Schooling in LLM	0.251^{***}	*LLM Fraction CSR Exposure*	1.150^{***}	0.732^{***}
	(0.091)		(0.295)	(0.177)
SW F-statistic			51.7	42.2
Cohort Fixed Effects	✓		✓	✓
Municipality Fixed Effects	✓		✓	✓

Note: CSR= Compulsory Schooling Reform. LLM=Local Labor Market. The sample consists of Norwegian males born 1950-1980 observed in earnings data over years 1967-2014 with years of potential experience between 0 and 30 years and annual earnings above 1 SGA threshold (N=14,758,689). All estimations include fixed effects for birth cohort and childhood municipality. Standard errors are clustered at the local labor market region (160 groups).
* $p < 0.10$, $** < 0.05$, $***p < 0.01$.

for a full set of dummies X_i for birth cohort and childhood municipality.

As earlier, one can estimate this extended IV model separately for each year of experience t. For simplicity in the exposition of results and since our primary goal here is to construct an estimate of the productivity spillovers $\overline{\beta}$, we jointly estimate the system of Equations (21), (22) and (23) across all experience level t using 2SLS, while constraining the effect of S_i and $\overline{S}_{\ell y}$ on log-earnings to be constant across t.

Estimation results from this model are provided in Table 4. All standard errors are clustered by the local labor market region. The results from the first-stage regressions in columns (2)-(3) indicate that the individual exposure to compulsory schooling reform and the fraction of workers in a local labor market exposed to this reform strongly affect individual schooling and average schooling. The second-stage estimates in column (1) suggests that the external effect of an additional year of average schooling (across all workers in the labor market) on individual earnings is 25.1%, after correcting for the spatial effect mentioned in footnote 26. This estimate suggests quite substantial external returns to education.

In comparison, using the U.S. data, Rauch [1993] estimates the externality to be between

2.8% and 5.1%, and Acemoglu and Angrist [2000] find the externality to be between 1% and 3.5%. Moretti [2004] finds that for each percentage point increase in the share of college-educated workers, average individual earnings increase by 0.6% to 1.2% over and above the private returns. To translate his estimates into returns per year of schooling, note that a one percentage point increase in college attainment amounts to approximately 1/25 year of average years of schooling (under the assumption that it takes 4 years to get a college degree). These estimates thus suggest external effects between $6 \times (25/0.8)\% = 18\%$ and $1.2 \times (25/0.8)\% = 36\%$ per year of schooling. Our estimates are towards the upper end but within the range of estimates reported in the literature. Our estimates are also consistent with those obtained from cross-country evidence by Lange and Topel [2006]. That being said, the scope of our analysis was to provide a suggestive estimate on productive spillovers and this may not be taken as definitive evidence on externalities in the labor market.

7 Conclusion

Education policy hinges on the estimates of private and social returns to education, but these returns are notoriously difficult to disentangle. In this paper, we determine conditions under which instrumental variables allow us to separately identify the private and social returns to education, within the context of an employer learning model.

We distinguish between hidden and transparent IVs where the former are unobserved by employers and thus not directly priced in the wages, while the latter are observed by the employers and correctly factored in the wages. We show that hidden IV identifies the private returns to education. If log-earnings profiles are additively separable in experience and schooling, they also allow identifying the social returns to education. Transparent instruments by contrast identify the social returns to education throughout the life-cycle. Building on this distinction between hidden and transparent instruments, we propose a strategy to identify the returns to education that can be attributed to job market signaling.

Using data from Norway we estimate that the causal effect of schooling on productivity, i.e., the social return to schooling, is 5% and the private return is 7.2%. The difference between the two is attributable to the signaling value of education. In other words, we estimate that 70% of the total private returns to education accrues to human capital and 30% accrues to signaling. Our estimates also suggest that employers learn workers' ability quickly. We also provide evidence examining earnings across local labor markets that suggest large external returns to education that manifest beyond the employer-employee relationship.

We conclude this paper by pointing out a few shortcomings of the employer learning literature following Farber and Gibbons [1996] and Altonji and Pierret [2001]. The standard specification assumes that the log-earnings profiles are additively separable in schooling, ability, and experience. Such an assumption naturally emerges from various formulations of the human capital model (e.g., the Ben-Porath model). Moreover, the data patterns that are taken as evidence of employer learning [Lange, 2007] are also compatible with other calibrations of human capital models [Kaymak, 2014]. We have taken some steps (see Section 6.1) towards addressing these shortcomings. We, however, believe that more work is needed and can be done in this area, especially when researchers have access to instruments and hidden correlates at the same time. We leave this work for future research.

References

Aakvik, Arild, Kjell G. Salvanes, and Kjell Vaage. 2010. Measuring Heterogeneity in the Returns to Education in Norway Using Educational Reforms. *European Economic Review* 54, no. 4:483–500.

Acemoglu, Daron and Joshua Angrist. 2000. How Large Are Human-Capital Externalities? Evidence from Compulsory Schooling Laws. *NBER Macroeconomics Annual* 15:9–59.

Altonji, Joseph G. and Charles R. Pierret. 2001. Employer Learning and Statisical Discrimination. *Quarterly Journal of Economics* 116, no. 1:313–350.

Angrist, Joshua D. and Guido W. Imbens. 1995. Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity. *Journal of the American Statistical Association* 90:431–442.

Angrist, Joshua D. and Alan B. Krueger. 1992. Estimating the Payoff to Schooling Using the Vietnam-Era Draft Lottery. *NBER Working Paper 4067* .

———. 1999. Empirical Strategies in Labor Economics. In *Handbook of Labor Economics*, eds. Orley Ashenfelter and David Card. Amsterdam: North-Holland, 1277–1366.

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, no. 4:76–104.

Arteaga, Carolina. 2018. The Effect of Human Capital on Earnings: Evidence from a Reform in Columbia's Top University. *Journal of Public Economics* 157:212–225.

Becker, Gary S. 1962. Investment in Human Capital: A Theoretical Analysis. *Journal of Political Economy* 70, no. 5:9–49.

Bedard, Kelly. 2001. Human Capital versus Signaling Models: University Access and High School Dropouts. *Journal of Political Economy* 109, no. 4:749–775.

Bhuller, Manudeep, Magne Mogstad, and Kjell G. Salvanes. 2017. Life-Cycle Earnings, Education Premiums, and Internal Rate of Return. *Journal of Labor Economics* 35, no. 4:993–1030.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2005. Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital. *American Economic Review* 95, no. 1:437–449.

Brinch, Christian N. and Taryn A. Galloway. 2012. Schooling in adolescence raises IQ scores. *Proceedings of the National Academy of Sciences of the United States of America* 109, no. 2:425–430.

Cameron, Stephen V. and James J. Heckman. 1998. Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males. *Journal of Political Economy* 106, no. 2:262–333.

Cameron, Stephen V. and Christopher R. Taber. 2004. Estimation of Educational Borrowing Constraints Using Returns to Schooling. *Journal of Political Economy* 112, no. 1:132–182.

Card, David. 1993. Using Geographic Variation in College Proximity to Estimate the Return to Schooling. *NBER Working Paper 4483* .

Carlsson, Magnus, Gordon B. Dahl, Björn Öckert, and Dan-Olof Rooth. 2015. The Effect of Schooling on Cognitive Skills. *Review of Economics and Statistics* 97, no. 3:533–547.

Carneiro, Pedro, James J. Heckman, and Edward J. Vytlacil. 2011. Estimating Marginal Returns to Education. *American Economic Review* 101, no. 6:2754–2781.

Clark, Damon and Paco Martorell. 2014. The Signaling Value of a High School Diploma. *Journal of Political Economy* 122, no. 2:282–318.

D'Haultfœuille, Xavier and Philippe Février. 2015. Identification of Nonseparable Triangular Models with Discrete Instruments. *Econometrica* 83, no. 3:1199–1210.

Duflo, Esther. 2001. Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. *American Economic Review* 91, no. 4:795–813.

Dynarski, Susan M. 2003. Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion. *American Economic Review* 93, no. 1:279–288.

Fang, Hanming. 2006. Disentangling the College Wage Premium: Estimating a Model with Endogenous Education Choices. *International Economic Review* 47, no. 4:1151–1185.

Farber, Henry S. and Robert Gibbons. 1996. Learning and Wage Dynamics. *Quarterly Journal of Economics* 111, no. 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.

Gundersen, Frantz and Dag Juvkam. 2013. Inndelinger i senterstruktur, sentralitet og BA-regioner. *NIBR - rapport 2013:1 (in Norwegian)* .

Heckman, James J., Sergio Urzua, and Edward J. Vytlacil. 2006. Understanding Instrumental Variables in Models with Essential Heterogeneity. *Review of Economic Studies* 88, no. 3:389–432.

Hopkins, Edward. 2012. Job Market Signaling of Relative Position, or Becker Married to Spence. *Journal of the European Economic Association* 10, no. 2:290–322.

Kahn, Lisa B. 2013. Asymmetric Information between Employers. *American Economic Journal: Applied Economics* 5, no. 4:165–205.

Kane, Thomas J. and Cecilia Elena Rouse. 1995. Labor-Market Returns to Two- and Four-Year Colleges. *American Economic Review* 85, no. 3:600–614.

Kaymak, Baris. 2014. Postschooling training investment and employer learning. *Journal of Human Capital* 8, no. 3:318–349.

Kirkeboen, Lars J., Edwin Leuven, and Magne Mogstad. 2016. Field of Study, Earnings, and Self-Selection. *Quarterly Journal of Economics* 131, no. 3:1057–1111.

Kling, Jeffrey R. 2001. Interpreting Instrumental Variables Estimates of the Returns to Schooling. *Journal of Business and Economic Statistics* 19, no. 3:358–364.

Lange, Fabian. 2007. The Speed of Employer Learning. *Journal of Labor Economics* 25, no. 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.

Lemieux, Thomas and David Card. 2001. Education, Earnings, and the 'Canadian G.I. Bill'. *Canadian Journal of Economics* 34, no. 2:313–344.

Lochner, Lance. 2011. Chapter 2 - Nonproduction Benefits of Education: Crime, Health, and Good Citizenship. In *Handbook of the Economics of Education*, vol. 4, eds. Eric A. Hanushek, Stephen Machin, and Ludger Woessmann. Elsevier, 183–282.

Løken, Katrine V., Magne Mogstad, and Matthew Wiswall. 2012. What Linear Estimators Miss: The Effects of Family Income on Child Outcomes. *American Economic Journal: Applied Economics* 4, no. 2:1–35.

Machin, Stephen, Kjell G. Salvanes, and Panu Pelkonen. 2012. Education and Mobility. *Journal of the European Economic Association* 10, no. 2:417–450.

Meghir, Costas and Mårten Palme. 2005. Educational Reform, Ability, and Family Background. *American Economic Review* 95, no. 1:414–424.

Mogstad, Magne and Matthew Wiswall. 2016. Testing the Quantity–Quality Model of Fertility: Estimation Using Unrestricted Family Size Models. *Quantitative Economics* 7, no. 1:157–192.

Monstad, Karin, Carol Propper, and Kjell G. Salvanes. 2008. Education and Fertility: Evidence from a Natural Experiment. *Scandinavian Journal of Economics* 110, no. 4:827–852.

Moretti, Enrico. 2004. Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Sectional Data. *Journal of Econometrics* 121, no. 1-2:175–212.

Newey, Whitney K. and James L. Powell. 2003. Instrumental Variable Estimation of Non-parametric Models. *Econometrica* 71, no. 5:1565–1578.

Oreopoulos, P. 2006. Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter. *American Economic Review* 96:152–175.

Oreopoulos, Philip and Kjell G. Salvanes. 2011. Priceless: The Nonpecuniary Benefits of Schooling. *Journal of Economic Perspectives* 25, no. 1:159–184.

Pons, Empar and Maria Teresa Gonzalo. 2002. Returns to Schooling in Spain: How Reliable Are IV Estimates? *Labour* 16, no. 4:747–770.

Rauch, James. 1993. Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities. *Journal of Urban Economics* 34, no. 4:380–400.

Roback, Jennifer. 1982. Wages, Rents, and the Quality of Life. *Journal of Political Economy* 90, no. 6:1257–1278.

Rosen, Sherwin. 1979. Wages-based indexes of urban quality of life. In *Current issues in urban economics*, eds. P. Mieszkowkski and M. Straszheim. John Hopkins University Press.

Schönberg, Uta. 2007. Testing for Asymmeric Employer Learning. *Journal of Labor Economics* 25, no. 4:651–691.

Spence, Michael. 1973. Job Market Signaling. *Quarterly Journal of Economics* 87, no. 3:355–374.

Sundet, Jon Martin, Dag G. Barlaug, and Tore M. Torjussen. 2004. The End of the Flynn Effect? A Study of Secular Trends in Mean Intelligence Test Scores of Norwegian Conscripts During Half a Century. *Intelligence* 32:349–362.

Taber, Christopher R. 2001. The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability? *Review of Economic Studies* 68, no. 3:665–691.

Thrane, Vidkunn C. 1977. Evneprøving av utskrivingspliktige i Norge 1950-53. *Arbeidsrapport nr. 26, INAS (in Norwegian)* .

Torgovitsky, Alexander. 2015. Identification of Nonseparable Models Using Instruments with Small Support. *Econometrica* 83, no. 3:1185–1197.

———. 2017. Minimum Distance from Independence Estimation of Nonseparable Instrumental Variables Models. *Journal of Econometrics* 199, no. 1:35–48.

Tyler, John H., Richard J. Murnane, and John B. Willett. 2000. Estimating the Labor Market Signaling Value of the GED. *Quarterly Journal of Economics* 115, no. 2:431–468.

Waldman, Michael. 1984. Job Assignments, Signalling, and Efficiency. *The RAND Journal of Economics* 15, no. 2:255–267.

Appendix

A.1 Non-Linear Returns to Schooling and IV Weights

When the true relationship between log-earnings and schooling is non-linear, the marginal effects of schooling on log-earnings differ across the support of schooling distribution. In such settings, comparisons of OLS and IV estimates are complicated because linear OLS and IV estimators typically identify different weighted averages of the marginal effects of schooling.[27] It is, however, possible to re-weight margin-specific OLS estimates and construct IV-weighted OLS estimates that are comparable to the IV estimates.

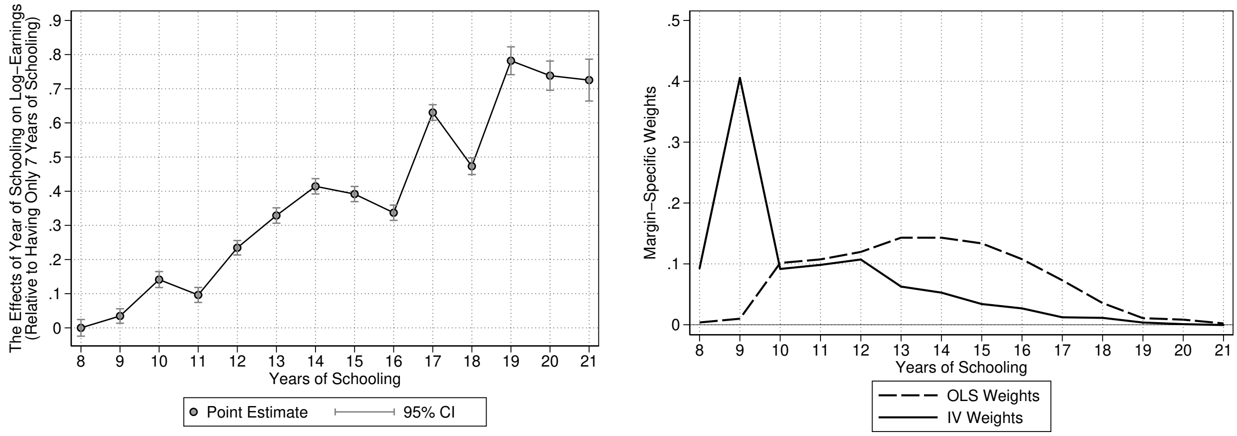
Let $\beta_{s,t}^{(3)}$ denote such IV-weighted OLS estimates. Consider the following non-linear relationship between log-earnings and schooling:

$$
\begin{aligned}
\ln W_{it} &= \alpha_t^{(3)} + \sum_{s=8}^{21} \gamma_{s,t}^{(3)} \times \mathbb{1}\left(S_i \geq s\right) + \beta_{z,t}^{(3)} Z_{it} + \tau_t^{(3)} X_i + \omega_{it}^{(3)}; \\
\beta_{s,t}^{(3)} &= \sum_{s=8}^{21} \gamma_{s,t}^{(3)} \times \pi_s; \qquad \pi_s = \frac{\operatorname{cov}\left(\mathbb{1}\left(S_i \geq s\right), D_i\right)}{\operatorname{cov}\left(S_i, D_i\right)},
\end{aligned}
\tag{A.1}
$$

where $\mathbb{1}\left(S_i \geq s\right)$ is an indicator for having at least s years of schooling, and $D_i \in \{0, 1\}$ is a binary instrument which equals 1 if individual i was exposed to the compulsory schooling reform. The parameter $\beta_{s,t}^{(3)}$ is a weighted sum of margin-specific OLS estimates $\gamma_{s,t}^{(3)}$ from a non-linear relationship between schooling and log-earnings, using weights π_s that mimic the variation exploited by the IV estimator. Intuitively, the IV estimates emphasize the marginal effects of schooling for those that are most affected by the instrument.

Using the specification in Equation (A.1), we can thus construct IV-weighted OLS estimates of returns to education. There are two components that determine the differences between these estimates and the standard linear OLS estimates. One is the extent of non-linearity in the margin-specific OLS estimates $\gamma_{s,t}^{(3)}$, and the other one is the differences

[27] See, e.g., discussions in Angrist and Imbens [1995]; Angrist and Krueger [1999]; Heckman et al. [2006]; Løken et al. [2012] and Mogstad and Wiswall [2016].



(a) Non-Linear Returns to Schooling

(b) Margin-Specific OLS and IV Weights

Figure A.1: Non-Linear Returns to Schooling and Margin-Specific OLS and IV Weights.
Note: Panel (a) plots OLS estimates of returns to schooling at 10-20 years of experience from a specification with dummies for each year of schooling, controlling for cohort and childhood municipality fixed effects, and flexible time trends. The estimation sample consists of Norwegian males born 1950-1980 observed in earnings data over years 1967-2014 with years of experience between 10 and 20 years and annual earnings above 1 SGA threshold. The estimates show the returns to each year of schooling relative to 7 years of compulsory schooling. Panel (b) plots the margin-specific OLS and IV weights at each year of schooling.

between the IV weights π_s and the corresponding OLS regression weights. In Figure A.1-(a), we display the (stacked average of) estimates $\gamma_{s,t}$ at experiences $t = (10, .., 20)$ for each additional year of schooling relative to 7 years of compulsory schooling, which illustrates the non-linear relationship between years of schooling and log-earnings in our data.

Next, in Figure A.1-(b), we display the margin-specific IV weights π_s that are used to obtain estimates $\beta_{s,t}^{(3)} = \sum_s \gamma_{s,t}\pi_s$. We also display the margin-specific weights for a standard linear OLS regression, and as expected these weights differ substantially from the IV weights. In particular, the IV places substantially more weight on the marginal effects of schooling in the lower end of the schooling distribution. This is consistent with the compulsory schooling reform instrument triggering changes in schooling attainment mainly at the lower end of the schooling distribution.