Signaling and Employer Learning with Instruments*

Gaurab Aryal† Manudeep Bhuller‡ Fabian Lange§

March 25, 2019

Abstract

We show how we can use instrument to disentangle social and private returns to education, when schooling can increase productivity *and* also be used to signal productivity. Within the employer learning framework of Farber and Gibbons [1996] and Altonji and Pierret [2001], we show that instrumental variable (IV) estimates of the returns to education among experienced workers identify the social returns to education. What the IV identifies among less experienced workers depends on whether it is hidden from or observed by employers. If the IV is hidden then it identifies private returns to education, but if it is transparent then it identifies social returns to education. We use local variation in compulsory schooling laws across multiple cohorts in Norway as an IV to estimate our model. Our preferred estimates indicate that the social return to an additional year of education is 6.3%, and the private internal rate of return, aggregating the returns over the life-cycle, is 7.9%. Thus 80% of the private return to education can be attributed to education raising productivity and 20% to education signaling worker's ability. Extending our framework to allow for productive externalities, we find evidence indicating large external returns to education, which suggests that social returns could exceed private returns.

Keywords: signaling, human capital, employer learning, instruments

JEL code: J24, J31, D83

---

*We thank Emily Nix, Leora Friedberg 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 and University of British Columbia for helpful comments.

†Department of Economics, University of Virginia, aryalg@virginia.edu

‡Department of Economics, University of Oslo; and Statistics Norway, manudeep.bhuller@econ.uio.no

§Department of Economics, McGill University, fabolange@gmail.com

1 Introduction

Two competing models can explain positive returns to education. According to the human capital model [Becker, 1962] education increases skills that are valued by employers. The model of job-market signaling [Spence, 1973] instead posits that education signals differences in innate skills among workers to potential employers. Both models imply a positive relationship between education and earnings.

From workers' perspective, what matters for their education decisions are the *private returns* to education, i.e., the increase in individual earnings resulting from an additional year of schooling, and it is irrelevant whether education increases earnings because of human capital acquisition or because of signaling. Signaling, however, is socially inefficient as workers expend valuable resources to signal their productivity – resources that otherwise could be spent on other productive activities. Thus signaling creates a wedge between private returns and *social returns* to education, i.e., the effect of education on productivity, by raising the private returns above the social returns, which leads to over investment in education.[1]

Education policy thus requires empirical guidance on the relative contribution of signaling and human capital acquisition to the total returns to education. It is, however, difficult to separately identify these two effects for at least two reasons. First, both models postulate that observable earnings reflect payments for latent skills, which the signaling model specifically assumes is employee's private information and thus likely unobserved by researchers. Second, both models rely on similar behavioral restrictions – workers choose schooling to maximize the present value of earnings, and employers maximize profits by employing workers as long as their wages do not exceed their marginal productivity.

The models thus do not provide separate behavioral restrictions on observed wages and

[1]This assumes the absence of other external effects associated with education. We also consider productive externalities beyond the employer-employee relationship, like in Acemoglu and Angrist [2000] and Moretti [2004] and provide evidence on external returns to education in Section 6.1. Education might also entail various non-production and non-pecuniary benefits, e.g., reducing crime or improving public health [Lange and Topel, 2006; Lochner, 2011; Oreopoulos and Salvanes, 2011], which we abstract from. More comprehensive measures of social returns should account for such spillovers.

education that can be used to select the "right" model. These difficulties have long been recognized in the literature; see, e.g., Lange and Topel [2006].[2]

One way forward is to make a priori assumption that employers learn their employees' abilities over time. In two influential papers, Farber and Gibbons [1996] and Altonji and Pierret [2001] (henceforth FG and AP, respectively) propose a model of employer learning, and use it to test employers use schooling to infer unobserved ability. The key identifying assumption underlying their approach is that the researchers observe a correlate of unobserved ability that is not available to the employers. FG propose the Armed Forces Qualification Test (AFQT) score available in the NLSY1979 to be such a *hidden* correlate. They show that the increasing association between the AFQT score and earnings over the life-cycle, is consistent with the assumption that employers learn about workers' abilities over time. AP show that the partial correlation (after controlling for AFQT) between earnings and schooling declines over the life-cycle, which is consistent with statistical discrimination based on schooling.

Following their lead, Lange [2007] provides evidence that employers learn, and they learn quickly. However, he shows that the contribution of signaling *cannot* be point-identified within the employer learning model *even* when a hidden correlate of ability is available. Using the first-order condition that characterizes the schooling decision, Lange [2007] next identifies an upper bound for the contribution of signaling to the private returns over the life-cycle, and estimates it to be less than 25% of the private returns to schooling.

We contribute to this literature by asking how instrumental variables can be used to estimate the model of employer learning developed by FG, AP, and Lange [2007]. We show that, under the standard conditions of instrument validity and employer learning used in the literature, the IV estimates over the life-cycle allow us to point-identify the causal effect of education on productivity *separately* from the signaling returns to education. This allows us to answer the question of how much do the private and social returns to education differ

[2] Also see Tyler et al. [2000]; Bedard [2001]; Fang [2006]; Hopkins [2012]; Clark and Martorell [2014]; Feng and Graetz [2017] and Arteaga [2018].

from each other due to signaling. Unlike Lange [2007], however, we neither have to rely on the assumption that we can measure schooling costs that enter the optimality (first-order) condition for schooling that he exploits nor do we use any optimality condition. Equally important is the fact that, unlike FG and AP, our identification strategy *does not* rely on the assumption that we have access to a hidden correlate of ability.

Our approach rests on considering the following question: "What do IV estimates of returns to education identify within the employer learning framework?" Our first result is that any conventional IV estimate of returns to education on earnings measured at sufficiently high levels of work experience identifies the causal effect of schooling on productivity. This result implies that, as long as the structure of our employer learning model is maintained, having access to an IV and a long panel dataset with earnings over the workers' careers, one can identify the productivity effect of schooling without a hidden correlate of ability or any behavioral assumptions. This interpretation of 'long-run' IV estimates follows directly from the (limit) result in the employer learning model that wages eventually converge to the true productivity because employers eventually learn workers' productivity. An IV that identifies the causal effect of schooling on 'long-run' earnings thus also identifies the causal effect of schooling on productivity.

Our second result is to show that, what 'short-run' IV estimates identify depends on what employers know about the IV. To that end, we make a distinction between a *hidden* IV and a *transparent* IV. We say that an IV is hidden if it is unobserved by the employers (e.g., eligibility rules for student aid or schooling laws that depend on personal characteristics that are often unobserved by the employers), and that it is transparent if it is observed by the employers and priced into the wage equation (e.g., labor market conditions). We show that if the IV is hidden then the 'short-run' IV estimates identify the private returns to education at each stage in the life-cycle, and if the IV is transparent then the 'short-run' IV estimates identify the social returns to education.

For our third result, we propose a way to use IV to identify the speed with which employers learn workers' productivity. We show that the estimates of the speed of learning and the 'short-run' and 'long-run' IV estimates of returns to education based on a hidden IV are sufficient to separately point-identify how much human capital and signaling contribute to earnings over the life-cycle.

To implement these ideas we use a unique dataset consisting of the population of Norwegian males born between 1950 and 1980, with earnings and employment histories between 1967 and 2014. We also observe an ability correlate – a cognitive test administered by the Norwegian military and taken by male conscripts around the age of 18, and a hidden instrument based on local variation in compulsory schooling laws across many birth cohorts. Next, we provide a summary of our main empirical findings using this dataset.

First, we compare our OLS estimates of the association between earnings and schooling and cognitive test scores with those uncovered by previous studies for the NLSY. The patterns uncovered 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%. Controlling for the interaction between the ability score and experience, we find that the coefficients on years of schooling decline rapidly from about 10% to 3% within the first 20 years. Likewise, in the NLSY these coefficients decline from about 9% to about 6%.[3]

Second, 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 15% in the first year following graduation, and then decline, rapidly at first and then slowly, until they stabilize to 6%-7%, after approximately 20 years of work experience.

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

These findings are consistent with the hypothesis that employers use past performance to learn about workers' productivity, and the assumption that our IV is hidden. We use these estimates to determine the speed of employer learning, and like Lange [2007], our estimate suggests that employers learn fast.

Third, we quantify the contribution of signaling and human capital acquisition to the returns to education. Our analysis reveal a productivity effect of education of 6.3% and a private internal rate of return in lifetime earnings, discounted to the time of schooling choice, of 7.9%. These estimates suggest that 80% of the private returns to schooling over the life-cycle represent a productivity-enhancing effect of schooling while only the remaining 20% are due to signaling. Thus, we find a modest role for signaling in explaining the positive returns to education estimated in our data.

Finally, building on Acemoglu and Angrist [2000] and Moretti [2004], we extend our framework to embed productive externalities that go beyond the employer-employee relationship. Using the same dataset and equipped with an additional instrument, we provide IV estimates of the returns to individual schooling *and* the returns to average schooling in an individual's local labor market. An increase in average schooling in an individual's local labor market by an additional year is estimated to increase log-earnings by around 15%, conditioning on individual schooling. These estimates indicate large external returns to education, which suggests that the social returns to education could actually 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 presents our main results. We consider two extensions of our model in Section 6, including the evidence on external returns to education, and 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. Under perfect competition, workers are paid their expected output, conditional on information available to employers. Let worker i's productivity be

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

where S is the years of schooling, Q is a correlate of ability observed by employers, and A is ability possibly correlated with the employer-observed correlates (S, Q). An example of a Q could be learning foreign languages that are mentioned in their résumé and can easily be verified. The function $H(t)$ captures how log-productivity varies with experience t.[4] While we can 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.

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. 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 $\mathcal{E}_{it}$ to be 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.[6]

The assumption that (S_i, Q_i, A_i) are joint normal random variables, implies that the conditional expectation of A_i given the information at $t = 0$, i.e., $\mathbb{E}[A_i|\mathcal{E}_{i0}] = \mathbb{E}[A_i|S, Q]$, is

[4] In our empirical implementation, we allow the experience profile $H(t, X_i)$ to vary flexibly with individual characteristics X_i, including a full set of dummies for birth cohort and municipality (see Section 4.4)

[5] In Section 6.2 we study the identification of a model when schooling and experience are nonseparable.

[6] We consider a model with symmetric information about workers' ability and past output across all employers. An important literature [Kahn, 2013; Schönberg, 2007; Waldman, 1984] analyzes how asymmetric information across the current employer and the other firms affects the labor market.

linear in S and Q. Thus,

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

where $\varepsilon_{A|S,Q} := A_i - \mathbb{E}[A_i|S,Q]$. The wage in period t is 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

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

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 use the Kalman filter to represent the process by which employers update their expectations $\mathbb{E}[A_i|\mathcal{E}_{it}]$. It results in the following simple form

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

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 $\theta_t = \frac{1-K_1}{1+(t-1)K_1}$ with $K_1 = \frac{\sigma_0^2}{\sigma_0^2+\sigma_\varepsilon^2} \in [0,1]$. 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 received up to t.

The weight (θ_t) declines with experience (t) because with time, observed measures of productivity become a better indicator of productivity than the correlates (S,Q), which were the only available information at $t=0$. The rate at which this weight decreases depends on the parameter K_1 that Lange [2007] refers to as the "speed of learning." The speed of learning governs how quickly information about productivity accumulates in the market, and it 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 K_1 will be close to 1 and the market learns quickly A. But, irrespective of K_1, after sufficiently long work experience the employer will put all weights on 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. Note that the coefficient β_{ws} in Equation (1) is not the causal effect of education on productivity, but only the "partial" causal effect of schooling, holding the employer-observed ability correlate, Q, and unobserved ability, A, fixed. Schooling, however, might also cause (Q, A), so the "total" causal effect of schooling on productivity also includes any (indirect) effect on productivity mediated through (Q, A). We refer to this total causal effect as the *social returns* to education. Education also affects the wages through employers' expectations about the ability of a worker, which change with time. We define *private returns* to be the expected earnings from an additional year of schooling evaluated at the beginning of a life-cycle. Thus, if education is used as a signaling device then the private returns will exceed the social returns.

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

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

Then, substituting (Q, A) from the above equations into (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, which is also the social returns to education. It captures the causal effect (direct and indirect) on other ability components (Q, A): an extra year of schooling increases Q by $\delta^{Q|S}$ units, which in turn raises productivity by β_{wq}. Schooling also raises ability (A) by $\delta^{A|S}$ units that are of value in the labor market, but are unobserved by the employer.

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 causes changes in Q observed by employers (see Equation (5)); and (iii) because schooling affects productivity, which employers ultimately learn by observing workers outputs over time.

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

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

The coefficient of S in (7), $\delta_t^{W|S}$, is the private returns to education. Comparing $\delta_t^{W|S}$ from (7) and $\delta_t^{\psi|S}$ from (6), we get the following relationship between social and private returns:

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

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 for education to have signaling value, this "adjustment term" must be non-negative, thus $\delta_t^{W|S} \geq \delta^{\psi|S}$. The wedge between private returns and social returns disappear with work experience, i.e., $\lim_{t\to\infty} \delta_t^{W|S} = \delta^{\psi|S}$, because as mentioned above, $\lim_{t\to\infty} \theta_t = 0$.

3 Identification

In this section, we study the identification of the private and social returns to education, and the employers' speed of learning defined above. We first consider the case when only education, experience, and earnings are observed. Then we consider the identification problem if we have additional information in the form of: (i) a hidden correlate of ability; and (ii) an instrument for schooling is available, which might or might not be observed by employers.

3.1 Bias in the OLS

Let us first consider the coefficients from regressing log earnings on years of schooling, adding controls for a full set of experience dummies and interaction terms between years of schooling and experience dummies. Using Equation (7), we can derive the probability limit of the OLS coefficient estimate for years of schooling evaluated at experience t to be

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

Inspecting Equation (9) reveals that the OLS coefficient estimate on years of schooling: (i) has a bias term stemming from omitting ability $(\tilde{Q},\tilde{A})$ correlated with, but not *caused* by, schooling; (ii) the size of this bias depends on the weight θ_t employers put on the initial signal at each t, and thus also the speed of employer learning; and (iii) absent this bias term, the OLS coefficient estimate on years of schooling will equal the private return $\delta_t^{W|S}$ at each experience t. This omitted ability bias is the main reason why researchers opt for instruments to identify the causal effects of schooling on log-earnings.

Assuming that we have access to a long panel dataset, let's also consider what happens to the probability limit of the OLS coefficient estimate on years of schooling as $t \to \infty$. Taking the limit of (9) with respect to t, and using the fact that $\lim_{t\to\infty} \delta_t^{W|S} = \delta^{\psi|S}$ from (8) we get

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

We also note that having access to a long panel dataset with information on earnings, education and experience is not sufficient to 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 hidden correlate of ability, denoted by Z, which is unobserved by employers. Furthermore, suppose that $A_i = \beta_{Az} Z_i + \eta_i$, so that η_i represents the productivity component that is observed neither by researchers nor employers. Substituting for 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 [Lange, 2007, Section B] 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 productivity is observed in the market. The linearity of the conditional expectation in (12) allows us to estimate the weight (θ_t) and the speed of learning (K_1) by projecting log wages on (S, Z) across different work experience levels t. The regression coefficients of log wages on (S, Z) converge from their $t = 0$ value to the regression coefficients in $\mathbb{E}[\ln W_{i\infty}|S_i, Z_i]$, at a rate that depends on K_1; thereby identifying K_1.

The projection coefficients obtained from estimation of (12) across experience, 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, as shown in Lange [2007], while we can identify K_1 assuming that 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 variable $D_i \in \{0, 1\}$, which satisfies the standard assumptions that are necessary to qualify as a valid instrumental variable.

Assumption 1. *Instrumental Variables*

1. *(Conditional Independence): $u_{it} \perp D_i|S_i$, where u_{it} is defined in (6).*
2. *(First Stage): $\mathbb{E}[S_i|D_i = 0] \neq \mathbb{E}[S_i|D_i = 1]$.*
3. *(Monotonicity): $S_i(D_i = 1) \geq S_i(D_i = 0)$ for all i.*

Under Assumption 1, following Imbens and Angrist [1994] we can show that in period t

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

Further, using the invariance of S with respect to t and that $\lim_{t\to\infty} \ln W_{it} = \psi_i$, gives

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

where the second equality follows from the fact that the part of ψ_i that is not caused by S_i is orthogonal to D_i, i.e., if $\psi_i = \delta^{\psi|S} S_i + \tilde{\psi}_i$, Assumption 1 implies that $\tilde{\psi}_i \perp D_i|S_i$. This means that as $t \to \infty$, the IV identifies the causal effect of schooling on productivity. Thus, the IV estimates of returns to education at sufficiently high levels of experience provide consistent estimates of the effects of schooling on productivity for the compliers, i.e., individuals who are induced to choose more years of schooling on account of the IV.

Note that the identification argument is valid *irrespective* of what employers know about the instrument. This is because employers ultimately base compensation on signals observed over the career of the worker only, so we can identify the social returns to education from the instrument directly by focusing on the returns to education in the long run. For $t < \infty$, however, what IV identifies under Assumption 1 depends on whether the employers know D or not. To show 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 the case when employers do not observe D. We refer to such instruments as *hidden* instruments.

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

Note that Assumption 2 is conceptually different from Assumption 1-1. Assumption 1-1 asserts that the instrument is conditionally independent of the determinants of productivity that are not caused by schooling. Assumption 2 captures the idea that given the information $\mathcal{E}$, wages do not depend on the instrument D. When the instrument is hidden and Assumption 2 holds, then $\ln W_{it} = \mathbb{E}\left[\psi_i|\mathcal{E}_{it}\right] = \mathbb{E}\left[\psi_i|\mathcal{E}_{it}, D_i\right]$.

In many settings, Assumption 2 is a natural assumption. The clearest examples relate to field experiments providing subsidies or information that induces higher school enrollment. In these cases, whether a student is in the control or treatment group is typically not known to potential employers. Some examples of hidden instrument 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]. 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 is similar to what we use in our empirical application.

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 instrument D_i affects $\ln W_{it}$ only indirectly by affecting (S_i, Q_i, ξ_i^t) that make 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 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

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

Comparing (14) with the private returns in (8) we observe that the hidden IV identifies private returns to education at t. We also obtain $\texttt{plim}\ (\hat{b}_{IV,t})$ across t, which we can use with the convergence of $b_{IV,t}$ from $b_{IV,t=0}$ to $b_{IV,t\to\infty}$ to identify the speed of learning K_1. Thus, hidden instruments identify the private returns to schooling at each t as well as the parameter K_1 governing the learning process.

Transparent Instrument

We define an instrument to be *transparent* if it is factored in by employers when setting the wages. Thus if the instrument 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 information set employers have about i in t.

Assumption 3. *(Transparent Instrument) Employers observe D_i so that* $\ln W_{it} = \mathbb{E}[\psi_{it}|D_i, S_i, Q_i, \xi^t]$.

By Assumption 1, we have that transparent instruments satisfy the exclusion restriction with respect to productivity ψ. The instrument is thus orthogonal to unobserved ability and thus can plausibly account for ability bias. Assumption 3, however, implies that the instrument is used in the 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 might be thought to be transparent are: (i) tuitions at two and four years 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, then it violates the exclusion restriction for wages and thus it 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). 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 productivity effect of schooling. Therefore if 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 binary transparent instrument identifies the 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 that are utilized in our analysis. Then, we describe the Norwegian compulsory schooling reform that we utilize as a source of exogenous variation in education attainment to construct IV estimates of the returns to education in log-earnings at each year of experience. Finally, we discuss the empirical specifications that are 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 *every* Norwegian male from 1967 to 2014. The variables available to us include individual demographic information (e.g., cohort of birth and childhood municipality of residence) and socio-economic data (e.g., years of schooling and annual earnings). Importantly, the dataset also includes unique personal identifiers which allow us to follow individuals' earnings across time. The personal identifiers also allow us to merge information from Statistics Norway's registry databases with data from the Norwegian Armed Forces that provide us with IQ test scores for male conscripts.

The Norwegian earnings data have several advantages over those available in most other countries. First, there is no attrition from the original sample beyond natural attrition due to either death or out-migration. Second, our earnings data pertain to all individuals, and is 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 experience.[7] 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 years of schooling, childhood municipality of residence, IQ test score, or indicators for 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.[8] To avoid variation in earnings across labor market experience due to the intensity of part-time work, we focus solely on full-time workers defined as having annual labor earnings (adjusted for wage inflation) above the substantial gainful activity (henceforth, SGA) threshold in the Norwegian Social Security System, which in 2015 levels amounted to 10,650 USD.[9] Restricting the sample to males with full-time work for three consecutive years, we retain 718,237 individuals (i.e., most males are recorded having a three-year full-time work spell at least once) and a panel data set comprising of 14,746,755 person-year observations (i.e., on average an individual is observed with full-time work for 20.5 years), which we label as the

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

[8] Note that this measure of income does *not* include income from self-employment, capital income or unconditional cash transfers such as social economic assistance, housing assistance, child allowance, etc.

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

estimation sample utilized in this analysis. It should be noted however that this sample is unbalanced; we have earnings for 579,984 individuals in the first year of potential experience and for 190,900 individuals in the 30th year of potential experience.

4.2 Measures of Schooling and IQ Test Scores

The first key regressor of interest is the number of years of schooling corresponding to the highest level of completed education. This variable is taken from Statistics Norway's Education Register and is based on the educational attainment reports submitted by educational establishments directly to Statistics Norway, thereby minimizing measurement error due to 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 entering the service, each male conscript's medical and psychological suitability were assessed. The majority of eligible Norwegian males took this examination around their 18^{th} birthday. The IQ test score we have access to is a composite unweighted mean from three speeded tests – arithmetics, word similarities, and figures.[10]

Figure 1 plots the average ability and the conditional density of IQ for each year of schooling between 7 and 21 years. This figure illustrates two striking patterns of our data that are 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 and selective entry requirements enforced in the entry to higher education in Norway.[11]

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

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

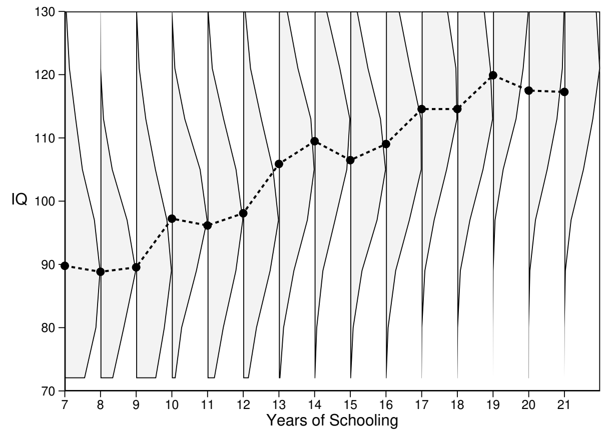


Figure 1: The Conditional Probability Density of IQ Test Scores on Years of Schooling.
Note: The sample consists of Norwegian males born 1950-1980 observed in earnings data over years 1967-2014 with years of potential experience between 0 and 30 years with annual earnings above 1 SGA threshold (N=14,746,755). 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.

Arguably, Norway is an interesting setting to assess employer learning and the signaling value of education for several reasons. First, the strong correlation between schooling and ability test scores in our data suggests that schooling has the potential to function as a predictor of ability, satisfying a necessary condition for schooling to have a signaling value. Second, employers cannot request access to test scores from the military conscription nor is it common to (voluntarily) disclose this information in job applications. We therefore believe that it is reasonable to assume that the ability test scores from military conscriptions are unobserved by employers, such that a researcher can infer about the process of employer learning from the correlation between the military IQ test scores and earnings across experience. As discussed above, we do however allow for the possibility that other correlates applicants' ability could be revealed in the job application process and the model can encompass such correlates (as captured by Q_i in Section 2). Finally, most cohorts in our sample entered the labor market before the arrival of online recruitment tools in the early 2000s, which could have altered the way in which employers tended to screen workers.

4.3 The Compulsory Schooling Reform

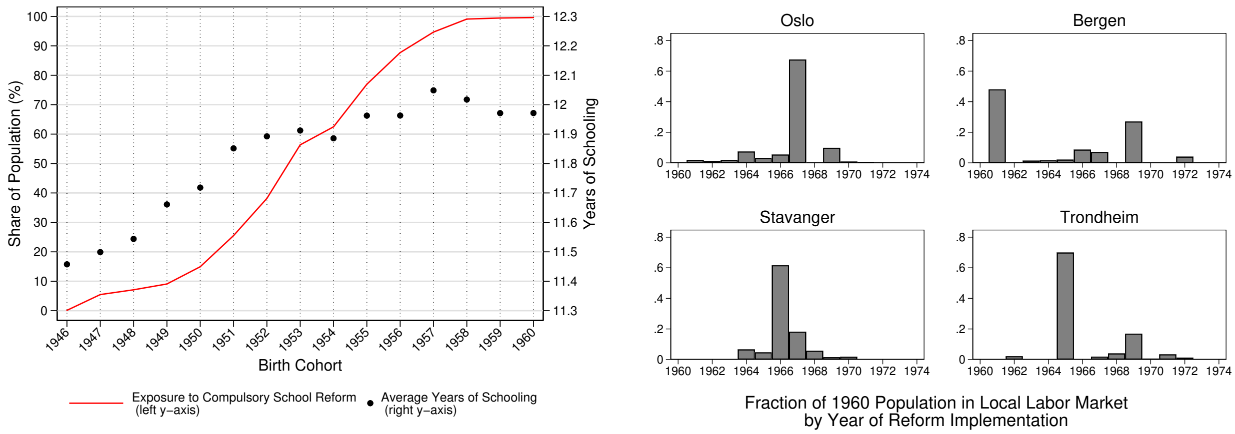
Between 1960 and 1975 Norway enacted a compulsory schooling reform that increased the minimum required schooling from 7 to 9 years. This increase was implemented in different years in the different municipalities (the lowest level of local administration). Thus, for more than a decade, Norwegian schools were divided into two separate systems, where the length of compulsory schooling depended on the year in which an individual was born 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 source of exogenous variation in educational attainment. For further details about the reform see Black et al. [2005].[12]

Historical records provide information about the year in which the reform was implemented for 672 out of the 732 municipalities that existed in 1960. For the remaining 60 municipalities this information is unknown; see Monstad et al. [2008]. As shown in Figure 2, there is a considerable variation in the overall fraction of each birth cohort that was exposed to the reform (see Figure 2-(a)) and also in the timing of reform even within local labor markets (see Figure 2-(b)). In particular, panel (a) shows that nobody born before 1946 was subjected to 9 years of compulsory schooling law, whereas all individuals born after 1960 were 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 2/3 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 1/3 of the population living in one of the other 39 municipalities in this labor market.[13]

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

[13]We utilize a classification of Norway in 160 local labor markets based on geographic commuting patterns constructed by Gundersen and Juvkam [2013]; on average each market comprises of 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.

As discussed in Section 3, for the purpose of identifying both the private and the social returns to education, the instrument needs to satisfy not just the standard IV assumptions (Assumption 1), but also the hidden instrument assumption (Assumption 2). An implication of the latter assumption in our setting is that employers must be uninformed about the interaction between a worker's birth cohort and the timing of compulsory school reform in the worker's municipality of childhood. There are two reasons why we think this assumption is reasonable in our setting.

First, in contrast to compulsory schooling laws legislated centrally in many countries or states (e.g., the U.S. states), the timing of the Norwegian compulsory school reform was decentralized and decided at the local municipal level. This is consistent with our data, e.g., Figure 2-(b) shows substantial variation in the timing of the reform also within local labor markets. Within local labor markets, there are high rates of commuting and mobility. This means that in order to know whether or not an individual was treated, an employer would have to not only know the exact date of implementation for each municipality but would also 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, the task of 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.[14] Therefore, it is reasonable to assume that for cohorts 1946-1960, graduating in an era long before internet, this information would not have been easily traceable for employers.

Even though we cannot directly test the hidden instrument assumption, we perform a robustness analysis in Section 5 in which we further restrict our sample to a subset workers by *excluding* workers who grew up in the municipality with the largest population in each local labor market. 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 sample retains 422,749 individuals and 8,697,979 person-year observations, which is 59% of the full sample.

4.4 Empirical Specifications

Our first empirical specification projects log-earnings on schooling and control variables 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 $\ln W_{it}$ represents log-earnings, S_i represents schooling, X_i represents all control variables, including a full set of dummies for birth cohort and childhood municipality. The reasons for adding these control variables will be discussed below. Note however that since our empirical design and a long panel dataset enables us to estimate Equation (15) sepa-

[14] 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.

rately for each t, we can model the role of work experience in interaction with individual characteristics X_i in a very flexible manner. Comparing Equation (15) with Equation (3) from our model, one can see that 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 absorb both a common experience profile and its interactions with X_i flexibly.

Our primary focus is on $\beta_{s,t}^{(1)}$, which we estimate using the IV approach described below. Under Assumption 2 we can use these estimates to obtain the social returns to education for $\beta_{s,\infty} = \beta_{s,T} = \lim_{t \longrightarrow T} \beta_{s,t}^{(1)}$, while the IV estimates for any experience $t < T$ provide consistent estimates of the private returns at each experience level t. In addition, we can estimate the speed of learning by using the rate of convergence of $\beta_{s,t}^{(1)}$ to its limit $\beta_{s,\infty}$.

Next, we compare estimates of the speed of learning K_1 obtained using our IV estimates with estimates obtained using the IQ test score as a hidden correlate. This latter approach requires projecting log-earnings on schooling (S), IQ test score (Z), and controls (X) at different experience levels (t):

$$\ln W_{it} = \alpha_t^{(2)} + \beta_{s,t}^{(2)} S_i + \beta_{z,t}^{(2)} Z_i + \tau_t^{(2)} X_i + \omega_{i,t}^{(2)}. \tag{16}$$

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 K_1. See Lange [2007] for further details.

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

$$S_i = \mu + \lambda D_i + \rho X_i + \kappa_i, \tag{17}$$

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. The reform exposure indicator is coded based on whether the reform had been implemented in i's childhood municipality of residence by the time he had turned 14 and thus required to complete at least 9 years of schooling.[15] As earlier, X includes a full set of dummies for birth cohort and childhood municipality.

We estimate the system of Equations (17) and (15) by 2SLS, separately for each year of experience (t).[16] We use the childhood municipality indicators to control for unobservable determinants of earnings or schooling that are fixed at the municipality level, and use the birth cohort indicators to control for the aggregate changes in schooling and earnings across cohorts (e.g., due to technical change). We assume that conditional on controls X, our instrument satisfies the standard IV assumptions (Assumption 1). As discussed in Section 4.3, we also find it plausible to assume that our instrument satisfies the hidden instrument assumption (Assumption 2), especially for the restricted sample of workers.

As in 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.[17] In the following, we will refer to estimates based on Equations (17) 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.

Table 1 displays the first-stage estimates of years of schooling on our instrumental variable for the full estimation sample and the restricted estimation sample which excludes workers who grew up in the largest municipality in each local labor market. The results in column (1) obtained for the full sample using the baseline specification indicate that the exposure to compulsory schooling reform increased completed years of schooling by 0.237, which is

[15] The school starting age in Norway at that time was 7 years and the pre-reform system required 7 years of compulsory schooling, which meant that the critical age at which a pupil would be required to take two additional years of schooling is 14 years. Cohorts aged 14 or below 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.

[16] Unlike Equation (15), there is no experience subscript t attached to the λ coefficient on our instrument variable 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.

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

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

	Full Sample		Restricted 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 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, *** $p < 0.01$.

about 10 % increase of a standard deviation. Notably, this instrument is strong with a partial F-statistic about 88, implying that weak instrument bias is not a concern for our analysis. Comparing the results in columns (1)-(4), we find the first-stage estimates to be fairly stable across the two samples and the two specifications. We note that we get some reduction in precision once we include controls for municipality-specific trends.

A complication arises when we compare the OLS and IV estimates because earnings are not necessarily log-linear in schooling. It is well known that when the true model is non-linear, OLS and IV estimates of the linear specification identify different weighted averages of the marginal effects of schooling, which differ across the support of schooling on log earnings.[18] Comparisons between the OLS and the IV estimates in the presence of non-linearities can therefore be misleading simply because the OLS and the IV estimates weight

[18] 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].

different parts of the support of schooling differently. It is, however, possible to construct OLS estimates that are comparable to the IV estimates by re-weighting margin-specific OLS estimates. We refer to these as the IV-weighted OLS estimates and denote them by $\beta_{s,t}^{(3)}$:

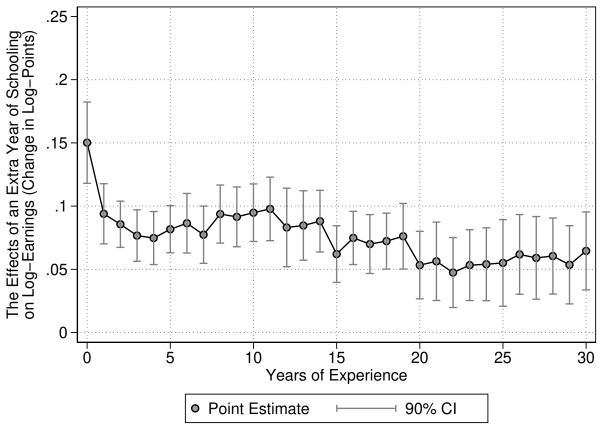
$$\begin{aligned} \ln W_{it} &= \alpha_t^{(3)} + \sum_{s=8}^{21} \gamma_{s,t}^{(3)} \times \mathbb{1}\left(S_i \geq s\right) + \beta_{z,t}^{(3)} \times Z_{it} + \tau_t^{(3)} \times X_i + \omega_{it}^{(3)}; \\ \beta_{s,t}^{(3)} &= \sum_{s=8}^{21} \gamma_{s,t}^{(3)} \times \pi_s; \qquad \pi_s = \frac{\operatorname{cov}\left(\mathbb{1}\left(S_i \geq s\right), D_i\right)}{\operatorname{cov}\left(S_i, D_i\right)}, \end{aligned} \tag{18}$$

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

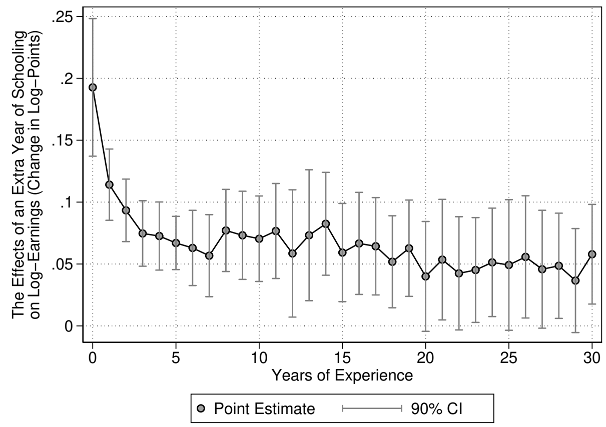
5 Main Results

In this section we present our estimation results. We begin by presenting the IV estimates and then comparing those estimates with the findings that rely on the standardized IQ score. Figure 3-(a) displays the IV estimates based on Equation (15) for the full sample. These coefficients represent the private returns to schooling, for every year of experience. In Figure 3-(b) we restrict the sample to workers growing up in municipalities other than the municipality representing the largest population in each local labor market. We further also display the IV estimates for each of the two samples controlling for municipality-specific trends in Figure 3-(c) and Figure 3-(d), respectively.

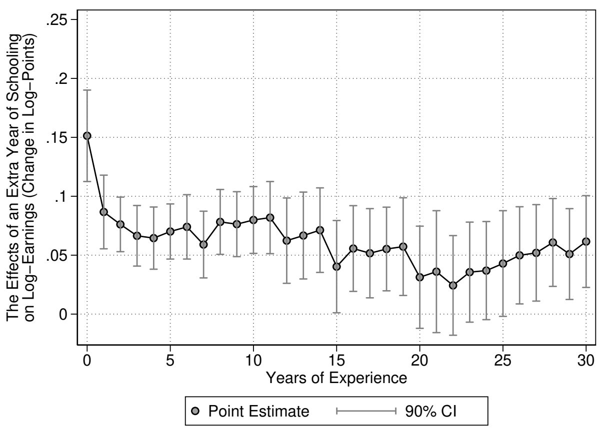
All four panels exhibit point estimates that suggest high initial returns to schooling, followed by a relatively steep decline in the first 5 years of experience. Then the returns



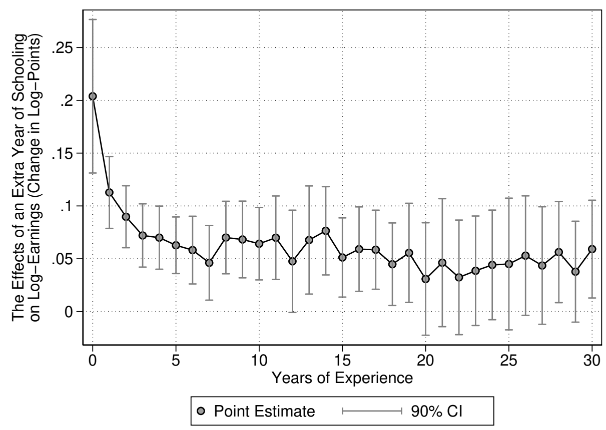
(a) Full Sample – Baseline Specification



(b) Restricted Sample – Baseline Specification



(c) Full Sample – Trends Specification



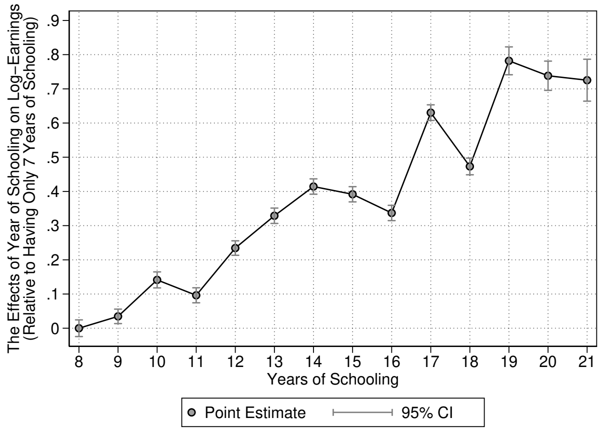
(d) Restricted Sample – Trends Specification

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

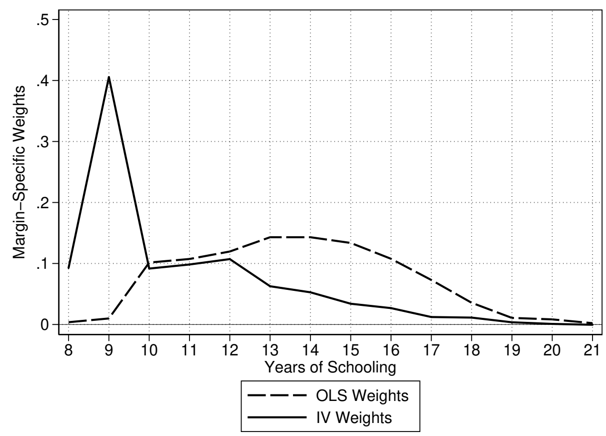
Note: The full estimation sample consists of Norwegian males born 1950-1980 observed in earnings data over years 1967-2014 with years of potential experience between 0 and 30 years and annual earnings above 1 SGA threshold (N=14,746,755). The restricted estimation sample further drops those growing up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

gradually stabilize after few years and approach 5-6% returns beyond 15 years of work experience. These estimates are consistent with employers gradually learning about workers' ability, and indicate that employers did not fully price in the variation in schooling induced by variation in compulsory schooling reform exposure across cohorts and municipalities.

Next, we compare these IV estimates with the OLS estimates of Equation (16) for each year of work experience, using the standardized IQ test score as a hidden correlate of ability



(a) Non-Linear Returns to Schooling



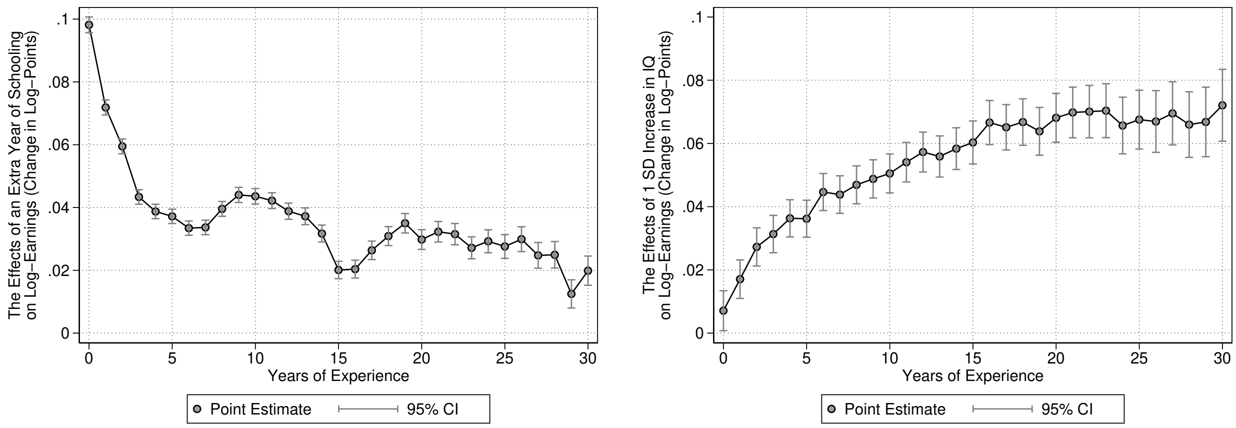
(b) Margin-Specific OLS and IV Weights

Figure 4: 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.

and controlling for municipal and cohort fixed effects. As argued above, the relationship between earnings and schooling is typically not log-linear. To compare OLS with the IV estimates we use the specification in Equation (18). Figure 4-(a) displays 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, illustrating the non-linear relationship between years of schooling and log-earnings in our data. Figure 4-(b) further displays the margin-specific IV weights π_s that are used to obtain estimates $\beta^{(3)}_{s,t} = \sum_s \gamma_{s,t}\pi_s$. We also display the margin-specific weights for a linear OLS and as expected these differ substantially from the IV weights.

Figures 5-(a) and 5-(b) display the estimates of $\beta^{(3)}_{s,t}$ and $\beta^{(3)}_{z,t}$, respectively. Comparing Figures 5-(a) and 3-(a) shows that the OLS and the IV estimates of returns to schooling reflect a similar pattern across years of experience, despite the fact that these estimators use different sources of variation. The point estimates of both $\beta^{(1)}_{s,t}$ and $\beta^{(3)}_{s,t}$ decline rapidly within the first few years and then stabilize after some years of experience. Furthermore, the point estimates of $\beta^{(3)}_{z,t}$ increase by experience, rapidly at first, and then stabilize after

15 years of work experience.



(a) The Returns to Schooling

(b) The Returns to IQ

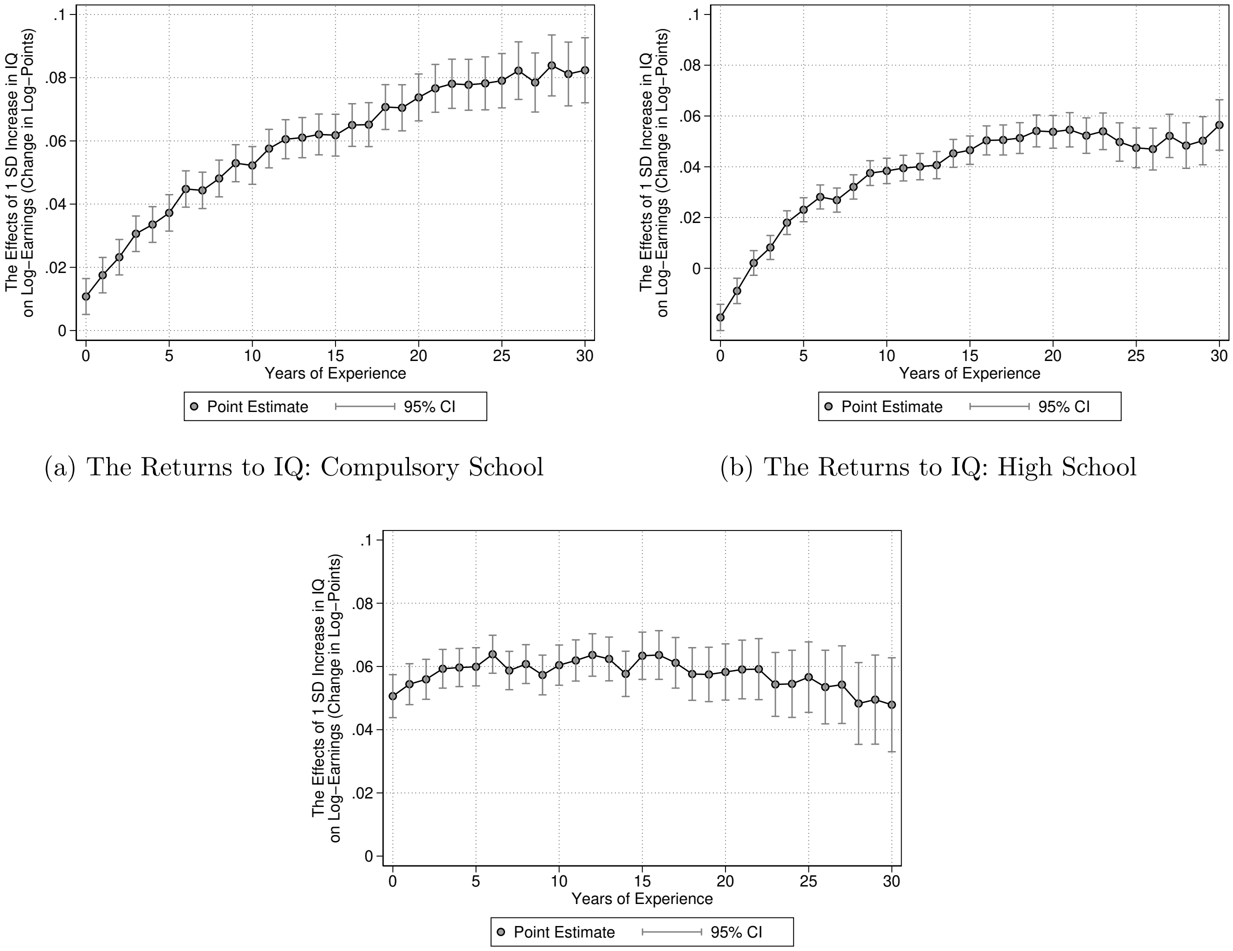
Figure 5: IV-Weighted OLS Estimates of the Returns to Schooling and IQ.

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). Plots (a) and (b) display IV-weighted OLS estimates of returns to schooling and IQ from separate estimations of Equation (18) for each year of experience. All estimations include fixed effects for birth cohort and childhood municipality. Standard errors are clustered at the local labor market region (160 groups). The 95% confidence intervals corresponding to each point estimate are shown as vertical bars.

At this point, we note that the patterns in the returns to schooling and IQ over the workers' careers measured in our data are surprisingly similarly to those found in the NLSY using the AFQT score. The fast decline in the returns to schooling and the increase in the returns to the IQ test score early in the career and the convergence to long-run values after a few years are present in both datasets (also see Figure 1 in Lange [2007]).

We also find patterns in the variation of the interaction between the AFQT and experience across the education distribution that are similar to the ones reported by Arcidiacono et al. [2010]. They report that the increase in the returns to the AFQT score with experience is absent for those with more than a high school degree. In Figure 6 we display the returns to IQ estimated separately for three groups differentiated by their education attainment. Figure 6-(a) shows the returns to IQ over the life-cycle among those with 7-9 years of schooling, Figure 6-(b) for those with 9 years of schooling up to receiving a high school degree, and Figure 6-(c) for those with more than a high school degree.

We find that the interaction between experience and the IQ test score is driven entirely



(a) The Returns to IQ: Compulsory School

(b) The Returns to IQ: High School

(c) The Returns to IQ: College/University

Figure 6: OLS Estimates of the Returns to IQ by Level of Education.
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).

by those with less than a college degree, which is similar to what Arcidiacono et al. [2010] found. Like Arcidiacono et al. [2010], we also conjecture that employers are better informed about differences in cognitive ability among those with post-secondary education, possibly because they observe the grades, choice of field of study, reference letters, and because college-educated individuals get internships that reveals information about their ability.

Returning to the IV and OLS estimates from Figures 3 and 5, we find through visual inspection that these estimates are consistent with each other. This variation in the coefficients on schooling and IQ over the workers' careers is consistent with employers learning about ability differences across individuals with relatively little education. Next, we estimate how rapidly this learning process unfolds using three different sets of estimates: either of the estimates $\{\beta^{(3)}_{s,t}, \beta^{(3)}_{z,t}\}^T_{t=0}$ from the OLS specification or the IV estimate of $\{\beta^{(1)}_{s,t}\}^T_{t=0}$.

From Equation (12), we know that the coefficients $\{\beta^{(3)}_{s,t}, \beta^{(3)}_{z,t}\}^T_{t=0}$ are given by

$$\begin{aligned}
\beta^{(3)}_{s,t} &= \theta_t b^{(3)}_{s,0} + (1-\theta_t)\, b^{(3)}_{s,\infty}; \\
\beta^{(3)}_{z,t} &= \theta_t b^{(3)}_{z,0} + (1-\theta_t)\, b^{(3)}_{z,\infty},
\end{aligned}$$

where $\{b^{(3)}_{s,0}, b^{(3)}_{z,0}\}$ are the projection coefficients of log-earnings on schooling and ability at the onset of the individuals career and $\{b^{(3)}_{s,\infty}, b^{(3)}_{z,\infty}\}$ are the projection coefficients that would be observed once productivity of individuals was fully revealed in the market. The parameters $\{b^{(3)}_{s,0}, b^{(3)}_{z,0}, b^{(3)}_{s,\infty}, b^{(3)}_{z,\infty}\}$, however, do not have a straightforward interpretation in terms of social or private returns. The IV estimates of $\{\beta^{(1)}_{s,t}\}^T_{t=0}$ can also be expressed as

$$\beta^{(1)}_{s,t} = \theta_t b^{(1)}_{s,0} + (1-\theta_t)\, b^{(1)}_{s,\infty},$$

where the weights $\{\theta_t\}$ are the same as before. The parameters $\{b^{(1)}_{s,0}, b^{(1)}_{s,\infty}\}$, however, do have interesting economic interpretations. The parameter $b^{(1)}_{s,0}$ represents the private returns to education at $t = 0$ and the parameter $b^{(1)}_{s,\infty}$ represents the social returns to education. As shown above, the IV estimates of $\beta^{(1)}_{s,t}$ represent the private returns at time t.

Table 2 displays the estimates of the parameters that enter these weighted averages for all three sets of estimates obtained by fitting the parameter estimates in Figures 3 and 5 using non-linear least squares. We observe that the estimates of the speed-of-learning parameter, K_1, using schooling coefficients are similar across the OLS and the IV specifications, but they are different from the estimate of K_1 that is obtained from using the standardize IQ score. Although the IV estimate of K_1 is imprecisely measured, its 95% confidence interval does not overlap with the estimate using the IQ, suggesting that a simple learning model can not explain the full dynamics in the education and the IQ coefficients over workers' experience.

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

	A. IV-Weighted OLS Estimates			
	Two Values of K_1		One Value of K_1	
	(1) Years of Schooling	(2) IQ Test Score	(3) Years of Schooling	(4) IQ Test Score
Speed of Learning K_1	0.377^{***}	0.116^{***}	0.207^{***}	
	(0.046)	(0.023)	(0.029)	
Initial Value b_0	0.098^{***}	0.008^{**}	0.084^{***}	-0.001
	(0.005)	(0.004)	(0.004)	(0.004)
Limit Value b_∞	0.023^{***}	0.087^{***}	0.017^{***}	0.077^{***}
	(0.002)	(0.005)	(0.002)	(0.003)
Municipality Fixed Effects	✓	✓	✓	✓
Cohort Fixed Effects	✓	✓	✓	✓

	B. IV Estimates			
	Full Sample		Restricted Sample	
	(1) Years of Schooling	(2) Years of Schooling	(3) Years of Schooling	(4) Years of Schooling
Speed of Learning K_1	0.447^{***}	0.490^{***}	0.532^{***}	0.565^{***}
	(0.127)	(0.110)	(0.058)	(0.055)
Initial Value b_0	0.145^{***}	0.148^{***}	0.192^{***}	0.204^{***}
	(0.013)	(0.014)	(0.010)	(0.010)
Limit Value b_∞	0.063^{***}	0.048^{***}	0.050^{***}	0.045^{***}
	(0.004)	(0.004)	(0.003)	(0.003)
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 estimated experience-specific coefficients of schooling and IQ on log-earnings using the full estimation sample presented in Figure 5 are used to construct the IV-weighted OLS estimates of speed of learning, initial values og limit values in panel A, columns (1)-(4). 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 og limit value in panel B, columns (1)-(2). The IV estimates in panel B, columns (3)-(5), are based on the estimates plotted in Figure 3(b) for a restricted estimation sample in which the municipality with largest population size in each the 160 labor market regions in Norway is dropped (N=8,697,979).
* $p < 0.10$, ** < 0.05, *** $p < 0.01$.

We also note that estimates of K_1, $b_{s,0}^{(1)}$ and $b_{s,\infty}^{(1)}$ in Table 2-B are quite similar across columns (1)-(4) that are based on IV estimates for the full sample and the restricted sample, as well as for specifications with and without controls for municipality-specific trends, respectively.

Given our assumptions, we can construct estimates of the social and private returns to education based on the results reported in Table 2. As shown above, the limit $b^{(1)}_{s,\infty}$ estimates the social returns to education, while the returns for fewer years of experience represent the private returns to education. Figure 7 displays the private and social returns based on the estimates from Table 2. The scatter plot displays the estimate obtained from the full sample (as in Figure 3-(a)) and the horizontal solid line represents the social returns.

To construct this figure we use the values $b^{(1)}_{s,0} = 0.145$ and $b^{(1)}_{s,\infty} = 0.063$ from column (1) in Panel B of Table 2. The latter is an estimate of the social returns to education and the former represents an estimate of the private returns to education prior to any learning. Given the large spread in estimates for K_1, we show results using the two estimates that bracket the value of K_1 in Table 2. The value $K_1 = 0.447$ (reported in column (1) in Panel B) implies a very rapid learning, while the value $K_1 = 0.116$ (reported in column (2) in Panel A) implies a much slower learning process. As expected from our model (see Equation (8)), we can see that the speed of learning (K_1) substantially affects the wedge between private returns and the social returns to education for medium and long experiences.

Given these estimates, we can calculate the internal private rate of return for an additional year of schooling. This equates the present discounted value of earnings over the career assuming a career length of 40 years. The private internal rate of return associated with $K_1 = 0.447$ and the lower broken curve in Figure 7 is 7.9% which is 1.6 percentage points higher than the social returns to schooling at 6.3%. The internal rate of return associated with slower learning ($K_1 = 0.116$), and the upper broken line in Figure 7 is at 10.6%, which exceeds the social returns by 4.3 percentage points. Finally, using the experience-specific IV estimates estimates of the private returns to education (the scatter plot), we obtain an internal rate of return of 8.1 %.[19]

As discussed earlier, the wedge between the private and the social returns to education informs us about the signaling value of education. Using an estimate of social returns to

[19]After experience 31, when increasing rates of individuals start retiring, we use 6.3%.

education at 6.3% and the lowest implied estimate of private internal rate of return at 7.9% (associated with $K_1 = 0.447$), we can conclude that 80% of the private return to education can be attributed to education raising the productivity of workers and 20% to the signaling value of education. Using other estimates of the private internal rates of return that are calculated at lower speeds of employer learning would indicate that a higher fraction of the private returns to education can be attributed to the signaling value of education.

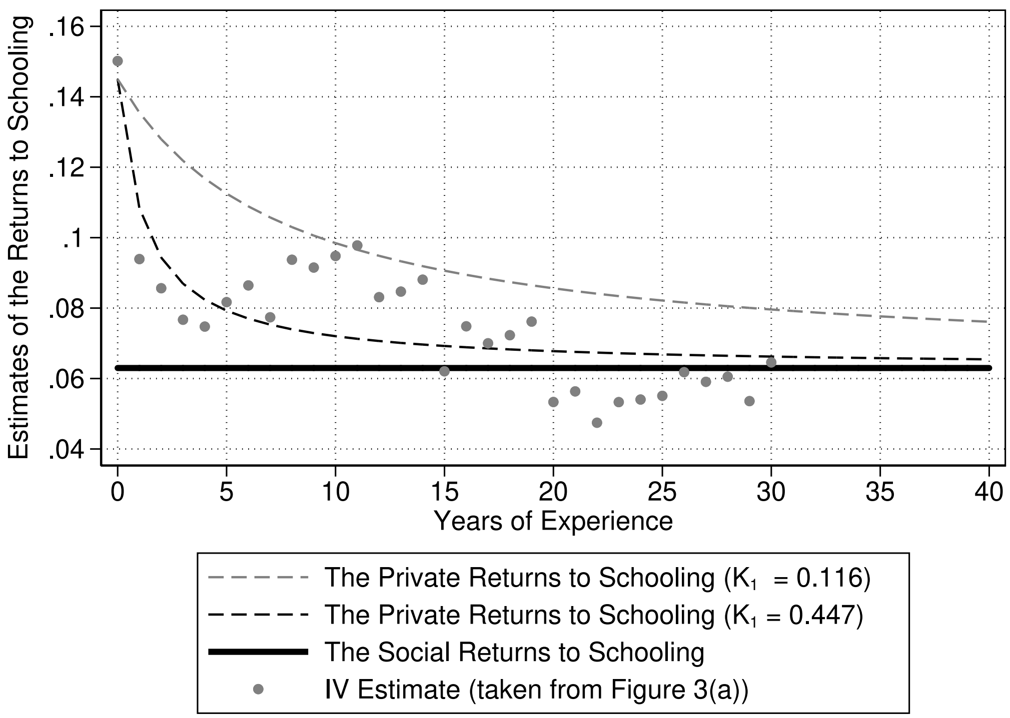


Figure 7: The Private and Social Returns to Schooling.

Finally, we can also compare the estimate of the social returns to education of 6.3% with a standard Mincer returns to education – that is the coefficient on years of schooling in a non-interacted specification controlling for a flexible experience profile. This comparison provides an indication on 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 estimated using the full sample is at 6.8%, exceeding the social returns by about half a percentage point.

6 Extensions

6.1 Productive Externalities

So far, we have abstracted from externalities of education that manifest across employers. We consider such productive externalities next. For this we follow the approach developed by Acemoglu and Angrist [2000] and Moretti [2004] and discussed in Lange and Topel [2006].

To that end, we begin by defining the total effective units of labor, χ_ℓ, employed in market ℓ to be the sum of all effective individual labor, χ_{it}, that was defined in Equation (1). Suppressing the index t, we can write the total effective labor as $\chi_\ell = \sum_{i\in\ell} \chi_i$. Further, suppose that the aggregate output in market ℓ, which we denote by Q_ℓ, depends on the aggregate labor χ_ℓ, capital $\mathbf{K}_\ell$ and location specific total factor productivity (TFP) Ψ_ℓ. In particular, suppose Q_ℓ is given by a Cobb-Douglas production function, i.e., $Q_\ell = \Psi_\ell \mathbf{K}_\ell^a \chi_\ell^{1-a}$, $a \in (0, 1)$.

To simplify our analysis and to focus on education, we further assume that the capital is freely mobile across all markets and the rental rate of capital, which we denote by r, is the same across all markets. Thus, the wage per effective unit of labor, is given by

$$W_\ell = \tau(a, r) \times \Psi_\ell^{1/1-a}, \tag{19}$$

where $\tau(a, r)$ is a function of (a, r) that does not vary across locations, and can thus be suppressed for our analysis.[20] To capture education externalities we posit that the TFP in market ℓ is a function of a measure of aggregate education in this market, and is given by

$$\Psi_\ell = \exp\left(\vartheta \times \overline{S}_\ell\right) \tilde{\Psi}_\ell, \tag{20}$$

where $\overline{S}_\ell$ is the aggregate schooling and ϑ captures the strength of the local production externality of education, and $\tilde{\Psi}_\ell$ represents differences in total factor productivity across

[20] This wage function is appropriate for markets across which there is no migration. With freely mobile labor and capital market the wage function will change; see Roback [1982] and Lange and Topel [2006].

locations not caused by educational attainment of the population in that location. In our empirical application we use average years of schooling in ℓ to measure $\overline{S}_\ell$.

In this environment, wages are equal to their expected marginal product and are given by $\mathbb{E}[W_\ell \times \chi_{i,t}|\mathcal{E}_{it}] = W_\ell \times \mathbb{E}[\chi_{i,t}|\mathcal{E}_{it}]$, where the expectation $\mathbb{E}[\chi_{i,t}|\mathcal{E}_{it}]$ is defined in Section 2. Substituting (20) in (19) and taking logs, we can express log-earnings as a function of location specific average schooling as well as the expectation of $\ln(\chi_{i,t})$,

$$\ln W_{it} = \beta_0 + 1/(1-a) \times \vartheta \times \overline{S}_\ell + 1/(1-a) \times \ln(\tilde{\Psi}_\ell) + \mathbb{E}[\ln(\chi_{i,t})|\mathcal{E}_{it}], \tag{21}$$

which forms the basis of our estimation of the externality of education. In other words, (21) suggests that we can augment our empirical specification in Equation (15) by including a measure of aggregate schooling $\overline{S}_\ell$ as an additional regressor. But that introduces $1/(1-a)\times \ln(\tilde{\Psi}_\ell)$ in the unobservable, which can correlate with $\overline{S}_\ell$, because skill biased technical change can lead to location specific skill upgrading as well as productivity improvements. This endogeneity could lead to an overestimation of the effect of $\overline{S}_\ell$ on log-earnings. We thus require additional instrument(s) for $\overline{S}_\ell$, which we introduce below.

Equation (21) implies the following estimation equation:

$$\ln W_{it} = \alpha_t + \beta_t S_i + \overline{\beta}\ \overline{S}_{\ell y} + \tau_t X_i + \tilde{\omega}_{it}, \tag{22}$$

where $\overline{S}_{\ell y}$ is the average years of schooling across all individuals aged 16 to 67 in the same market ℓ as individual i who worked full-time (earnings $>$ 1 SGA) in the calendar year y, and the remaining variables are defined exactly as before.[21] A natural choice of an instrument for $\overline{S}_{\ell y}$ is the fraction of workers in market ℓ in year y who had been exposed to the compulsory

[21] Comparing the coefficient of average schooling in (21) and that in (22), we see that the causal effect of average schooling on wages exceeds the production externality ϑ by a factor of $1/(1-a)$. Therefore, any estimate of the coefficient on average wages from earnings data needs to be adjusted by the labor share to obtain an estimate of the production externality. Using OECD data from 1995–2012, we estimate an average labor share in Norway to be 0.566, and use this share to adjust the coefficient estimate for average schooling in our estimations. The corresponding average labor share for the U.S. is at 0.66 using the same data source.

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. In particular, $\overline{D}_{\ell y}$ is positively correlated with $\overline{S}_{\ell y}$ but it does not directly affect the local wages except through its effect on average equation in market ℓ and year y.

Importantly, there will be variation in the instrument $\overline{D}_{\ell y}$ also conditional on D_i, which is critical in our setting as we now have two endogenous regressors $(\overline{S}_{\ell y}, S_i)$ in Equation (22). To see this, let's compare 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 and so any later cohort $c > c'$ (including c'') will eventually 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 at $t = 0$, none of the earlier cohorts in his local labor market would have been subject to the 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 be some variation in $\overline{D}_{\ell y}$ that we can exploit in our setting. Using the staggered reform, we can moreover still control for a full set of dummies X_i for birth cohort and childhood municipality as before.

That being said, one concern in using $\overline{D}_{\ell y}$ as an IV is that the variation in average schooling this provides is limited to the range of cohorts born 1946–1960 (see Figure 2-(a)). Specifically, no cohort born before 1946 was subject to the reform, while all cohorts born after 1960 were exposed. To address this issue of limited identifying variation in average schooling over time stemming directly from the compulsory schooling reform, we use also another instrument for average schooling utilizing information on the presence of a high school in the municipality an individual lived in at age 16. As earlier, by conditioning on birth cohort and childhood municipality, this instrument effectively utilizing openings (or closings) of high schools over time. This variable has substantial temporal variation; while around 80% of the 1950 birth cohort had access to a local high school at age 16, this fraction reached 90% for 1960 birth cohort, 95% for 1970 birth cohort, and higher for later cohorts.

Let $\overline{D}_i^* \in \{0,1\}$ be a dummy variable that is equal to 1 if there was at least one high-school in the municipality where i lived when he was sixteen years, and 0 otherwise, and similarly, let $\overline{D}_{\ell y}^* \in [0,1]$ be the fraction of workers in market ℓ in year y who lived in a municipality with at least one high school when they were sixteen years old. Together with the other instruments described above, we propose to use the following first-stage equations:

$$S_i = \mu_t + \lambda_{1t} D_i + \bar{\lambda}_{1t} \overline{D}_{\ell y} + \lambda_{1t}^* D_i^* + \bar{\lambda}_{1t}^* \overline{D}_{\ell y}^* + \rho_t X_i + \kappa_{it}; \tag{23}$$

$$\overline{S}_{\ell y} = \bar{\mu}_t + \lambda_{2t} D_i + \bar{\lambda}_{2t} \overline{D}_{\ell y} + \lambda_{2t}^* D_i^* + \bar{\lambda}_{2t}^* \overline{D}_{\ell y}^* + \bar{\rho}_t X_i + \bar{\kappa}_{it}. \tag{24}$$

We included D_i^* and $\overline{D}_{\ell y}^*$ as instruments for S_i for symmetry. As we have an unbalanced panel, we estimate the baseline IV model for each year of experience t separately. For simplicity in the exposition of results, we *also* estimate the system of Equations (22), (23) and (24) jointly across all experience level t using 2SLS, while constraining the effect of individual schooling S_i and average schooling $\overline{S}_{\ell y}$ on log-earnings to be constant across all t. In that case, we also impose the same constraints on the first-stage coefficients.

Estimation results from the constrained model are provided in Table 3 and estimation results from the unconstrained model that allows the coefficients for individual schooling to vary by t are presented in in Figure 8. All standard errors are clustered by local labor market. The results from the first-stage regressions indicate that the exposure to compulsory schooling reform and fraction of workers exposed to this reform strongly affect the individual schooling and average schooling. As mentioned above, these effects capture the reform induced variation in schooling for cohorts 1946–1960. The second instrument also has a strong positive effect on average schooling but its effect on individual schooling is weak. Over all the Sanderson and Windmeijer [2016]'s multivariate $F-$test of excluded instruments suggest that the although our instruments are not weak, they are not very strong either.

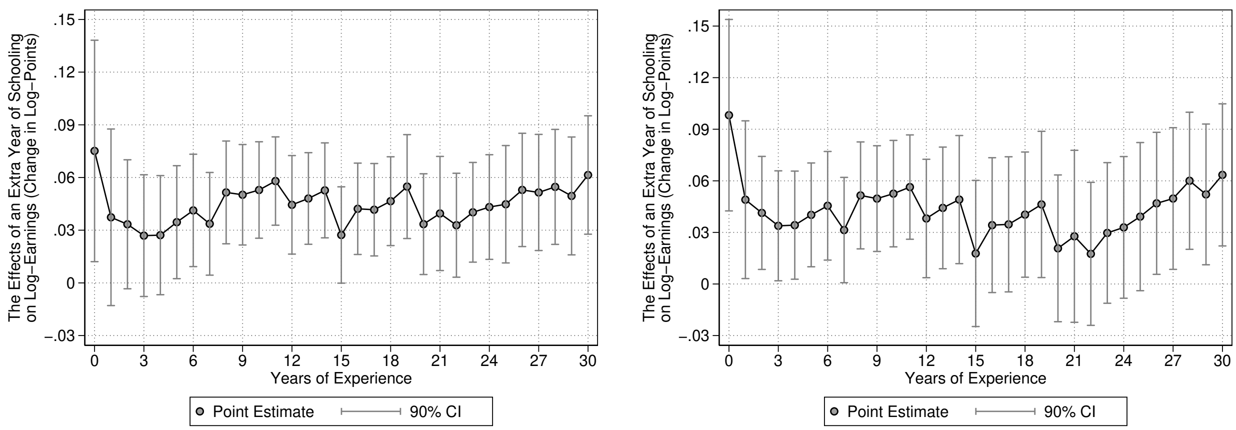
Table 3: Returns to Individual Schooling and Average Schooling on Log-Earnings.

	A. Returns to Schooling Estimates	
	(1)	(2)
Outcome Variable:	*Log-Earnings*	*Log-Earnings*
Endogenous Variables:		
Individual Years of Schooling	0.045**	0.042**
	(0.015)	(0.019)
Average Years of Schooling in LLM	0.156***	0.127**
	(0.052)	(0.056)

	B. First-Stages Estimates			
	(1)	(2)	(3)	(4)
Outcome Variable:	*Individual Schooling*	*Average Schooling*	*Individual Schooling*	*Average Schooling*
Instruments:				
Individual Exposed to the CSR	0.209***	0.022***	0.186***	0.017***
	(0.025)	(0.005)	(0.033)	(0.006)
Individual Had Access to a High School	0.077*	0.008	0.044	-0.005
	(0.040)	(0.006)	(0.034)	(0.055)
Fraction in LLM Exposed to the CSR	0.960***	0.660***	0.689***	0.548***
	(0.275)	(0.166)	(0.240)	(0.148)
Fraction in LLM with Access to a High School	0.553**	0.247**	0.398**	0.192**
	(0.205)	(0.102)	(0.177)	(0.082)
Municipality Fixed Effects	✓	✓	✓	✓
Cohort Fixed Effects	✓	✓	✓	✓
Municipality-Specifc Trends			✓	✓
SW multivariate F-statistic (instruments)	19.8	16.3	12.0	14.0

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 a full set of fixed effects for birth cohort and childhood municipality. Panel A, column (2) and Panel B, columns (3)-(4), further control for linear and quadratic municipality-specific trends. Standard errors are clustered at the local labor market region (160 groups).
* p < 0.10, ** < 0.05, *** p < 0.01.

Notably, the second-stage estimates of the constrained model shown in Table 3 suggests that the externality effect of an additional year of average schooling (across all workers in the labor market) on individual earnings is 15.6%, and controlling for municipality-specific trends this effect decreases to 12.7%. These estimates suggest sizable external returns of schooling. Moreover, estimates of the external returns to education from the unconstrained model are nearly identical to the estimates discussed above. Estimating Equations (22)-(24), we find



(a) Specification with Externalities

(b) Specification with Externalities and Trends

Figure 8: IV Estimates of the Returns to Individual Schooling with Externalities.

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=14,758,689). All estimations include fixed effects for birth cohort and childhood municipality. Standard errors are clustered at the local labor market region (160 groups). The coefficient estimate (and the corresponding clustered-robust standard error) on average schooling obtained from an IV estimation of the specification with productive externalities (corresponding to plot (a)) is at 0.157*** (0.054); with municipality-specific trends the estimate is at 0.126** (0.057).

that the externality effect of an additional year of average schooling on individual earnings is at 15.7% (with controls for municipality-specific trends we get an estimate of 12.6%). In comparison, in the U.S. data, Rauch [1993] estimates the externality to be 7%, and Acemoglu and Angrist [2000] suggest the effect to be between 1% and 3.5%. Although not directly comparable to our estimates, Moretti [2004] finds that for each percentage point increase in the share of college educated workers average individual earnings increases by 0.6% to 1.2% over and above the private returns. We must note that an increase in *average* schooling by an additional year is a very large change compared to historical changes; Figure 2-(a) shows that average years of schooling increased from 11.5 to 12 across cohorts 1946–1960.[22]

Finally, IV estimates of the returns to individual schooling from the unconstrained model are displayed in Figure 8. We see that these estimates are close to the average effect of 4.5% we reported in Table 3 based on the constrained model. Compared to our previous estimates

[22]It is also worth pointing out that the above model is incomplete in that it does not account for migration of workers and firms across locations. Embedding this structure in the Roback [1982] spatial equilibrium model and calibrating the consumption share of housing and the shares of labor and housing in production costs for the U.S., Lange and Topel [2006] find that the increase in average log-wages caused by average schooling across locations actually *underestimates* the social return by about 20%.

without externalities, the estimates of returns to individual schooling with externalities are slightly lower in magnitude and we see less of a decline in the coefficients in the early years. One reason for this disparity could be that unlike D, it is less clear that $\overline{D}$ and $\overline{D}^*$ are hidden instruments. It is likely that employers know more about the aggregate schooling reform exposure or high school availability in the local labor market than what they know about the individual workers. If these IVs are instead transparent then the estimates in Figure 8 would measure the social returns to schooling, which we previously estimated to be at 6.3%. We must recognize however that these IV estimates of returns to individual schooling are less precise, making it is challenging to perform precise inference regarding the shape of the earnings-schooling-experience profile.

6.2 Non-Separability between Education and Experience

Going back to our original model without externalities, we note that we have assumed an additively separable structure. So the returns to schooling are constant over the life-cycle. We next investigate whether we can relax this assumption and accomodate a more general functional form which allows for an interaction between schooling and experience. To that end, we begin with an assumption that the 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. 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 same steps leading to (6) and (7) we can express productivity and wages, respectively, as

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

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

$$\underbrace{\frac{\partial \psi_{it}}{\partial S}}_{\texttt{social returns}} = H_1(S_i, t, \delta^{Q|S} S_i + \tilde{Q}_i) + H_3(S_i, t, \delta^{Q|S} S_i + \tilde{Q}_i)\delta^{Q|S} + \delta^{A|S}, \tag{25}$$

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 schooling is the marginal effect of schooling on wages, i.e.,

$$\begin{aligned}\underbrace{\frac{\partial W_{it}}{\partial S}}_{\texttt{private returns}} &= H_1(S_i, t, \delta^{Q|S} S_i + \tilde{Q}_i) + H_3(S_i, t, \delta^{Q|S} S_i + \tilde{Q}_i)\delta^{Q|S} + \delta^{A|S} \\ &\quad + \theta_t \left(\phi_{A|S} + \phi_{A|Q}\delta^{Q|S} - \delta^{A|S}\right). \end{aligned} \tag{26}$$

Comparing (25) and (26) 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 they converge: $\lim_{t\to\infty}\frac{\partial \ln W_{it}}{\partial S} = \frac{\partial \psi_{it}}{\partial S}$.

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

$$\texttt{plim}\ \hat{b}_{IV,t} = \frac{\mathbb{E}\left[\ln W_{it}|D_i = 1, t\right] - \mathbb{E}\left[\ln W_{it}|D_i = 0, t\right]}{\mathbb{E}\left[S_i|D_i = 1, t\right] - \mathbb{E}\left[S_i|D_i = 0, t\right]}. \tag{27}$$

So, with experience $t \to \infty$ we can again identify the social returns to schooling 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 schooling. Now, let us consider those with limited years of experience ($t < \infty$). When D is a hidden instrument then $W_{it} = \mathbb{E}\left[\psi_i|\mathcal{E}_{it}\right] = \mathbb{E}\left[\psi_i|\mathcal{E}_{it}, D_i\right]$. 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 (27) 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 schooling, as defined for each t in (26). Thus with a hidden instrument we identify the private returns to schooling. Following the same logic, we conclude that if D is transparent then it identifies the average of social returns to schooling.

In this section, we have only considered the problem of identifying private and social returns within our employer learning framework. The estimation of $H(\cdot, \cdot, \cdot)$, however, is a more difficult problem specially because our excluded IV is binary, so we cannot rely on 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]. We leave the estimation of $H(\cdot)$ and the speed of learning, K_1, for future research.

7 Conclusion

Optimal policy design with respect to education hinges on estimates of private and social returns to education, yet these are notoriously difficult to measure. In this paper, we examine conditions under which instrumental variables allow estimating the private and social returns to education within the context of the employer learning model. We distinguish between hidden and transparent instruments where the former are instruments that are not observed by employers and thus not directly priced in the wages, while the former are instruments that employers correctly price. Hidden instruments identify the private returns to education. If log earnings profiles are additive separable in experience and schooling, then 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 method to identify the returns to schooling 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 6.3% and the private returns is 7.9%. The difference is attributable to the signaling value of education. In particular, we estimate that 80% of the total private returns to education accrues to human capital accumulation and the remaining 20% can be attributed to signaling. Our estimates also suggest that employers learn the unobserved skill very quickly. Furthermore, we provide evidence examining earnings across local labor markets that suggest large external returns to education that manifest beyond the employer-employee relationship.

We finish this paper by pointing towards what we believe to be the largest shortcoming 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 additive separability naturally emerges from some formulations of the Human Capital model (such as versions of the Ben-Porath model). However, the patterns in the data traditionally taken as evidence for employer learning (e.g., by Lange [2007]) are also compatible with standard calibrations of Human Capital models [Kaymak, 2014].

We have taken some steps (see section Section 6.2) towards addressing this shortcoming of the employer learning literature. However, we believe that more is possible, especially if researchers have access to instruments and hidden correlates at the same time. We leave this work to 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. 21

Acemoglu, Daron and Joshua Angrist. 2000. How Large are Human-Capital Externalities? Evidence from Compulsory Schooling Laws. *NBER Macroeconomics Annual* 15:9–59. 2, 6, 36, 41

Altonji, Joseph G. and Charles R. Pierret. 2001. Employer Learning and Statisical Discrimination. *Quarterly Journal of Economics* 116, no. 1:313–350. 1, 3, 45

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. 26

Angrist, Joshua D. and Alan B. Krueger. 1992. Estimating the Payoff to Schooling Using the Vietnam-Era Draft Lottery. *NBER Working Paper 4067* . 15

———. 1999. Empirical Strategies in Labor Economics. In *Handbook of Labor Economics*, eds. Orley Ashenfelter and David Card. Amsterdam: North-Holland, 1277–1366. 26

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. 5, 30, 31

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. 3

Becker, Gary S. 1962. Investment in Human Capital: A Theoretical Analysis. *Journal of Political Economy* 70, no. 5:9–49. 2

Bedard, Kelly. 2001. Human Capital versus Signaling Models: University Access and High School Dropouts. *Journal of Political Economy* 109, no. 4:749–775. 3

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. 21, 25

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. 21, 23

Cameron, Stephen V. and James J. Heckman. 1998. Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Makes. *Journal of Political Economy* 106, no. 2:262–333. 16

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. 16

Card, David. 1993. Using Geographic Variation in College Proximity to Estimate the Return to Schooling. *NBER Working Paper 4483* . 16

Carneiro, Pedro, James J. Heckman, and Edward J. Vytlacil. 2011. Estimating Marginal Returns to Education. *American Economic Review* 101, no. 6:2754–2781. 16

Clark, Damon and Paco Martorell. 2014. The Signaling Value of a High School Diploma. *Journal of Political Economy* 122, no. 2:282–318. 3

D'Haultfœuille, Xavier and Philippe Février. 2015. Identification of Nonseparable Triangular Models with Discrete Instruments. *Econometrica* 83, no. 3:1199–1210. 44

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. 15

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. 15

Fang, Hanming. 2006. Disentangling the College Wage Premium: Estimating A Model with Endogenous Education Choices. *International Economic Review* 47, no. 4:1151–1185. 3

Farber, Henry S. and Robert Gibbons. 1996. Learning and Wage Dynamics. *Quarterly Journal of Economics* 111, no. 4:1007–1047. 1, 3, 45

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. 3

Gundersen, Frantz and Dag Juvkam. 2013. Inndelinger i senterstruktur, sentralitet og BA-regioner. *NIBR - rapport 2013:1 (in Norwegian)* . 21

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. 26

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. 3

Imbens, Guido W. and Joshua D. Angrist. 1994. Identification and Estimation of Local Average Treatment Effects. *Econometrica* 62, no. 2:465–475. 13

Kahn, Lisa B. 2013. Asymmetric Information between Employers. *American Economic Journal: Applied Economics* 5, no. 4:165–205. 7

Kane, Thomas J. and Cecilia Elena Rouse. 1995. Labor-Market Returns to Two- and Four-Year College. *American Economic Review* 85, no. 3:600–614. 16

Kaymak, Baris. 2014. Postschooling training investment and employer learning. *Journal of Human Capital* 8, no. 3:318–349. 46

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. 19

Kling, Jeffrey R. 2001. Interpreting Instrumental Variables Estimates of the Returns to Schooling. *Journal of Business and Economic Statistics* 19, no. 3:358–364. 16

Lange, Fabian. 2007. The Speed of Employer Learning. *Journal of Labor Economics* 25, no. 1:1–35. 3, 4, 5, 6, 7, 8, 12, 13, 24, 30, 46

Lange, Fabian and Robert Topel. 2006. The Social Value of Education and Human Capital. *Handbook of the Economics of Education* 1:459–509. 2, 3, 36, 41

Lemieux, Thomas and David Card. 2001. Education, Earnings, and the 'Canadian G.I. Bill'. *Canadian Journal of Economics* 34, no. 2:313–344. 16

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

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. 26

Machin, Stephen, Kjell G. Salvanes, and Panu Pelkonen. 2012. Education and Mobility. *Journal of the European Economic Association* 10, no. 2:417–450. 21

Meghir, Costas and Mårten Palme. 2005. Educational Reform, Ability, and Family Background. *American Economic Review* 95, no. 1:414–424. 15

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. 26

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. 21, 23

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. 2, 6, 36, 41

Newey, Whitney K. and James L. Powell. 2003. Instrumental Variable Estimation of Nonparametric Models. *Econometrica* 71, no. 5:1565–1578. 44

Oreopoulos, P. 2006. Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws really Matter. *American Economic Review* 96:152–175. 16

Oreopoulos, Philip and Kjell G. Salvanes. 2011. Priceless: The Nonpecuniary Benefits of Schooling. *Journal of Economic Perspectives* 25, no. 1:159–184. 2

Pons, Empar and Maria Teresa Gonzalo. 2002. Returns to Schooling in Spain: How Reliable are IV Estimates? *Labour* 16, no. 4:747–770. 15

Rauch, James. 1993. Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities. *Journal of Urban Economics* 34, no. 4:380–400. 41

Roback, Jennifer. 1982. Wages, Rents, and the Quality of Life. *Journal of Political Economy* 90, no. 6:1257–1278. 36, 41

Sanderson, Elanor and Frank Windmeijer. 2016. A weak instrument f-test in linear iv models with multiple endogenous variables. *Journal of Econometrics* 190:212–221. 39

Schönberg, Uta. 2007. Testing for Asymmeric Employer Learning. *Journal of Labor Economics* 25, no. 4:651–691. 7

Spence, Michael. 1973. Job Market Signaling. *Quarterly Journal of Economics* 87, no. 3:355–374. 2

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. 19

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. 15

Thrane, Vidkunn C. 1977. Evneprøving av utskrivingspliktige i Norge 1950-53. *Arbeidsrapport nr. 26, INAS (in Norwegian)* . 19

Torgovitsky, Alexander. 2015. Identification of Nonseparable Models Using Instruments with Small Support. *Econometrica* 83, no. 3:1185–1197. 44

———. 2017. Minimum Distance from Independence Estimation of Nonseparable Instrumental Variables Models. *Journal of Econometrics* 199, no. 1:35–48. 44

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. 3

Waldman, Michael. 1984. Job Assignments, Signalling, and Efficiency. *The RAND Journal of Economics* 15, no. 2:255–267. 7