Online Appendix:

Signaling and Employer Learning with Instruments

Gaurab Aryal* Manudeep Bhuller† Fabian Lange‡

January 2022

*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; and NBER, fabolange@gmail.com

A Additional Results

Table A.1: First-Stage Estimates on Years of Schooling – Controlling for Differential Trends.

	A. Full Sample		B. Hidden IV Sample		C. Transparent IV Sample	
	Baseline (1)	Trends (2)	Baseline (3)	Trends (4)	Baseline (5)	Trends (6)
Instrument:						
Exposure to Reform	0.237***	0.209***	0.228***	0.209***	0.240***	0.199***
	(0.025)	(0.034)	(0.034)	(0.040)	(0.032)	(0.038)
Controls:						
Municipality Fixed Effects	✓	✓	✓	✓	✓	✓
Cohort Fixed Effects	✓	✓	✓	✓	✓	✓
Municipality-Specific Trends		✓		✓		✓
Number of Observations	14,746,755		8,697,979		6,048,776	

Note: The full sample (panel A) consists of Norwegian males born in 1950-1980 observed any time in earnings data over years 1967-2014 with years of potential experience between 0 and 30 years and annual earnings above 1 SGA threshold (N=14,746,755). The hidden IV sample (panel B) 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), while the transparent IV sample (panel C) retains only individuals who grew up in the municipality with the largest population size in each labor market (N=6,048,776). All estimations include fixed effects for birth cohort and childhood municipality. The trends specifications in columns (2), (4) and (6) control 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.
* p < 0.10, ** < 0.05, *** p < 0.01.

Table A.2: IV Estimates of the Speed of Employer Learning, Initial and Limit Returns to Schooling, and the Signaling Value Contribution – Controlling for Differential Trends.

Model Specifications:	A. Experience-Varying Returns to Skill				B. Experience-Invariant Returns to Skill	
	Sequential Estimation		*Joint Estimation*			
	(1) Baseline	(2) Trends	(3) Baseline	(4) Trends	(5) Baseline	(6) Trends
Parameters of Interest:						
Speed of Learning κ	0.505***	0.506***	0.550***	0.592***	0.532***	0.565***
	(0.107)	(0.152)	(0.126)	(0.205)	(0.058)	(0.055)
Initial Return $b_0^{IV^{\mathfrak{h}}}$	0.198***	0.209***	0.195***	0.205***	0.192***	0.204***
	(0.015)	(0.023)	(0.012)	(0.017)	(0.010)	(0.010)
Limit Return $b_\infty^{IV^{\mathfrak{h}}}$	0.055***	0.049***	0.055***	0.054***	0.050***	0.045***
	(0.006)	(0.010)	(0.008)	(0.015)	(0.003)	(0.003)
Average Return $b^{IV^{\mathfrak{h}}}$	0.064***	0.054***	0.066***	0.060***	0.067***	0.062***
	(0.006)	(0.007)	(0.004)	(0.004)	(0.005)	(0.005)
Average Return $b^{IV^{\mathfrak{t}}}$	0.088***	0.058***	0.089***	0.060***	–	–
	(0.004)	(0.005)	(0.002)	(0.002)		
Internal Rate of Return (IRR):						
Private IRR	0.079	0.076	0.076	0.075	0.072	0.066
Social IRR	0.054	0.048	0.055	0.054	0.050	0.044
Signaling Value	31.6%	36.8%	27.6%	28.0%	30.6%	33.3%
Controls:						
Municipality Fixed Effects	✓	✓	✓	✓	✓	✓
Cohort Fixed Effects	✓	✓	✓	✓	✓	✓
Differential Trends		✓		✓		✓

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), partitioned in a hidden IV sample (N=8,697,979) and a transparent IV sample (N=6,048,776) as discussed in Section III.C. Parameter estimates for the model specification with experience-varying returns to skill (panel A) are based on a combination of the hidden and the transparent IV estimates plotted in Figure 3(a)-(b). In the sequential estimation approach (columns (1)-(2)), we first estimate the $\hat{\lambda}_t$ profile based on the transparent IV estimates using equation (11) under location normalization $\lambda_0 = 1$, and then insert $\hat{\lambda}_t$ in equation (10) and solve for the model parameters using the non-linear least squares estimation. In the joint estimation approach (columns (3)-(4)), we jointly solve for the model parameters and λ_t using equations (10)-(11) by non-linear least squares estimation. Parameters estimates for the model specification with experience-invariant returns to skill (panel B) are based on the hidden IV estimates plotted in Figure 3(a) and constructed using non-linear least squares estimation of equation (12). Columns (2), (4) and (6) rely on IV estimates that control 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.
* $p < 0.10$, ** < 0.05, *** $p < 0.01$.

Table A.3: Alternative Hidden IV Estimates of the Speed of Learning, Initial and Limit Returns to Schooling.

	A. Baseline	B. Sensitivity to Unbalanced Panel of Worker Careers			C. Sensitivity to Alternative Earnings Thresholds			D. Sensitivity to Alternative Definitions of Hidden IV Sample		
		Years of Experience $t \leq 10$	Birth Cohorts 1950–65	Birth Cohorts 1940–80	Earnings $> .5$ SGA	Earnings $> .75$ SGA	Earnings > 1.5 SGA	Population $< 5,000$	Population $< 10,000$	Population $< 25,000$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Parameters:										
Speed of Learning κ	0.532***	0.605***	0.753***	0.485***	0.594***	0.498***	0.656***	0.690***	0.621***	0.562***
	(0.058)	(0.065)	(0.079)	(0.049)	(0.070)	(0.062)	(0.041)	(0.099)	(0.063)	(0.070)
Initial Return $b_0^{IV^\natural}$	0.192***	0.193***	0.159***	0.161***	0.231***	0.198***	0.189***	0.174***	0.214***	0.185***
	(0.010)	(0.008)	(0.010)	(0.007)	(0.015)	(0.011)	(0.007)	(0.015)	(0.012)	(0.010)
Limit Return $b_\infty^{IV^\natural}$	0.050***	0.056***	0.041***	0.042***	0.054***	0.049***	0.050***	0.042***	0.046***	0.059***
	(0.003)	(0.005)	(0.002)	(0.002)	(0.004)	(0.004)	(0.002)	(0.004)	(0.003)	(0.003)
Controls:										
Municipality FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cohort FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: The baseline hidden IV sample in column (1) 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, excluding 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). In column (2), we restrict estimations to years of potential experience between 0 and 10 years for the same sample as in column (1). In column (3), we restrict the sample in column (1) to males born 1950-1965 (N=5,326,336), while in column (4), we instead expand this sample to males born 1950-1980 and also include observations with missing IQ data (N=13,172,738). In column (5), we expand the sample in column (1) to include individuals with annual earnings above 50% of the SGA threshold (N=8,918,180), while in column (6), we similarly include individuals with annual earnings above 75% of the SGA threshold (N=8,803,482). In column (7), we instead restrict the sample in column (1) to include individuals with annual earnings above 150% of the SGA threshold (N=8,486,015). In columns (8)-(10), we change the definition of hidden IV sample to individuals who grew up in municipalities with, respectively, less than 5,000 (N=4,105,767), 10,000 (N=6,249,821) or 25,000 (N=7,617,157) residents in 1960, besides excluding individuals who grew up in the municipality with the largest population size in each labor market region as in our baseline.
* $p < 0.10$, ** < 0.05, *** $p < 0.01$.

Table A.4: IV Estimates of Years of Schooling on Standardized IQ Test Scores.

	A. Full Sample		B. Hidden IV Sample		C. Transparent IV Sample	
	Baseline	Trends	Baseline	Trends	Baseline	Trends
	(1)	(2)	(3)	(4)	(5)	(6)
Reduced Form:						
Exposure to Reform	0.042***	0.036***	0.037***	0.035***	0.047***	0.037**
	(0.009)	(0.013)	(0.010)	(0.011)	(0.014)	(0.016)
IV Estimates:						
Years of Schooling at Age 18	0.265***	0.235***	0.229***	0.219**	0.318***	0.258**
	(0.055)	(0.074)	(0.060)	(0.068)	(0.075)	(0.099)
Municipality Fixed Effects	✓	✓	✓	✓	✓	✓
Cohort Fixed Effects	✓	✓	✓	✓	✓	✓
Municipality-Specific Trends		✓		✓		✓

*Note:*The full sample (panel A) consists of Norwegian males born in 1950-1980 observed any time in earnings data over years 1967-2014 with years of potential experience between 0 and 30 years and annual earnings above 1 SGA threshold (N=14,746,755). The hidden IV sample (panel B) 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), while the transparent IV sample (panel C) retains only individuals who grew up in the municipality with the largest population size in each labor market (N=6,048,776). All estimations include fixed effects for birth cohort and childhood municipality. The trends specifications in columns (2), (4) and (6) further also controls for municipality-specific linear 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.

* p < 0.10, ** < 0.05, *** p < 0.01.

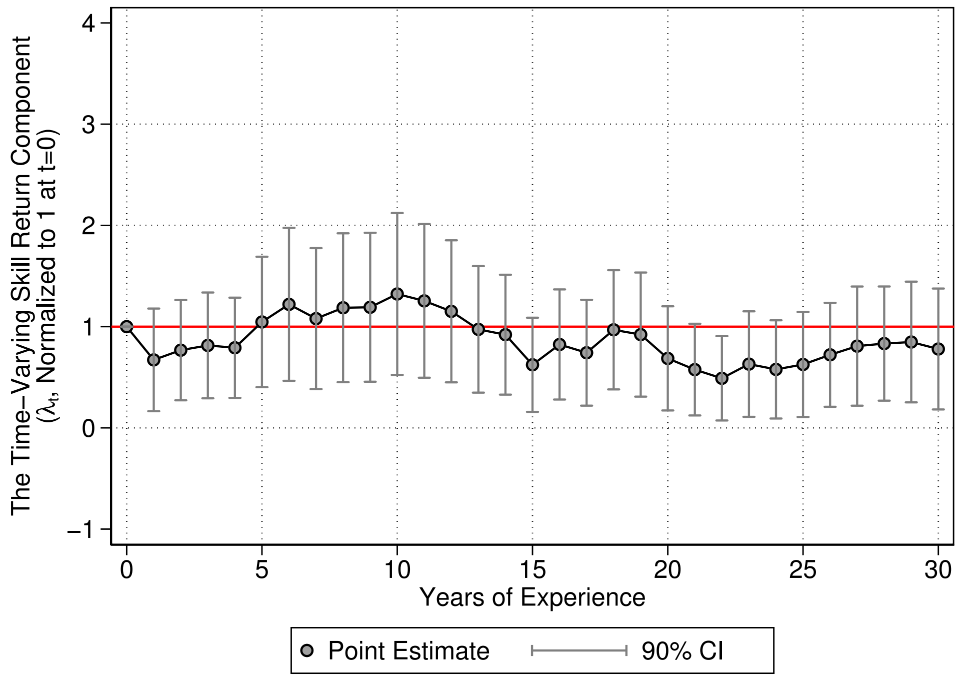
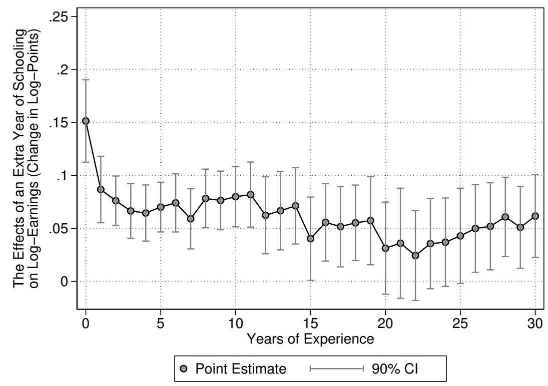


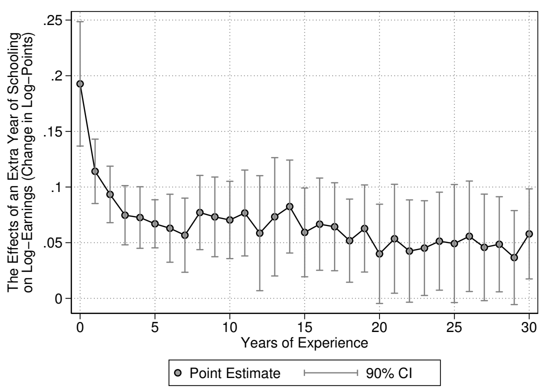
Figure A.1: The Experience-Varying Component of Returns to Skill – λ_t.

Note: The $\hat{\lambda}_t$ estimates are constructed using the IV estimates displayed in Figure 3(b) for the transparent IV sample. We use the formula $\hat{\lambda}_t = (\hat{b}_t^{IV^t}/\hat{b}_0^{IV^t})$ for $t > 0$ and $\lambda_0 = 1$ (location normalization), and employ the delta-method to construct standard errors. The 90% confidence intervals corresponding to each point estimate are displayed as vertical bars. The joint test for the hypothesis that for all $t \in \mathbb{T}_{>0}$ the ratio $\hat{\lambda}_t = 1$ provides an F-statistic of 0.9, which means that we cannot statistically reject the hypothesis of constant social returns. The transparent IV 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 who who grew up in the municipality with the largest population size in each labor market (N=6,048,776).

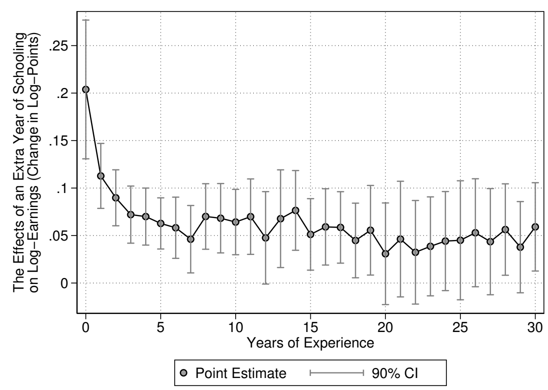
(a) Full Sample



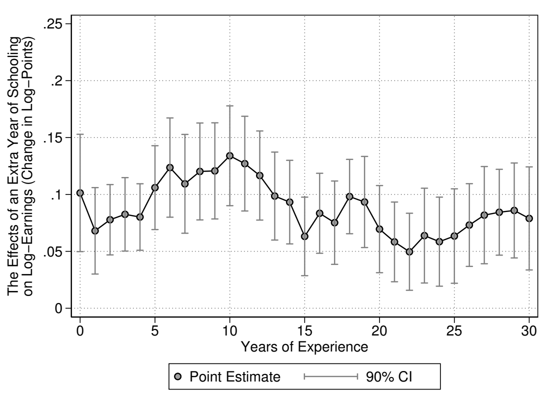
(b) Full Sample w/ Differential Trends



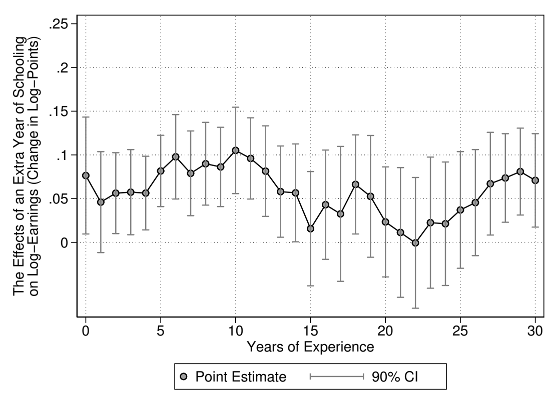
(c) Hidden IV Sample



(d) Hidden IV w/ Differential Trends



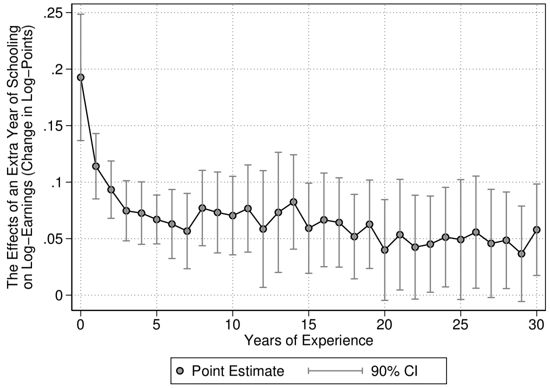
(e) Transparent IV Sample



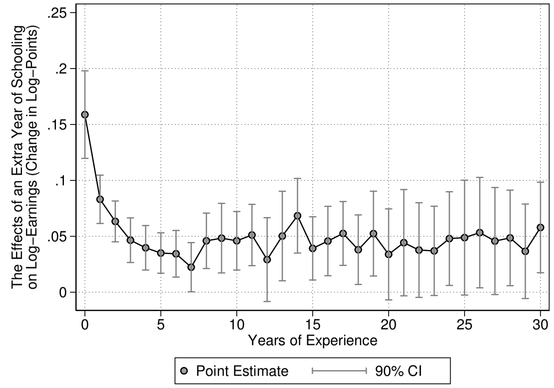
(f) Transparent IV w/ Differential Trends

Figure A.2: IV Estimates of the Returns to Schooling – Controlling for Differential Trends.

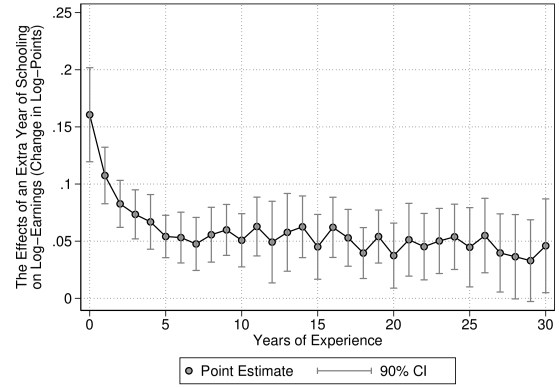
Note: See notes below Figure 3 for details on each estimation. Plots (b), (d) and (f) control for linear and quadratic municipality-specific trends estimated using data on all pre-reform cohorts born 1930 or later and extrapolated to post-reform cohorts.



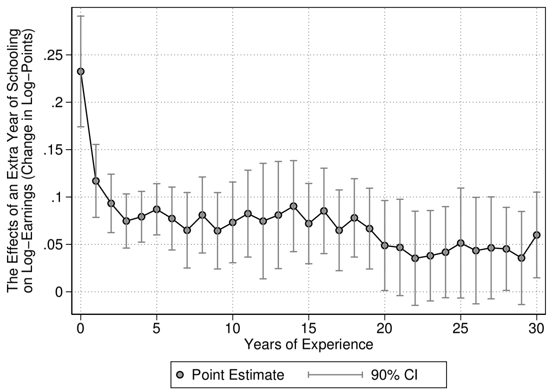
(a) Baseline Hidden IV



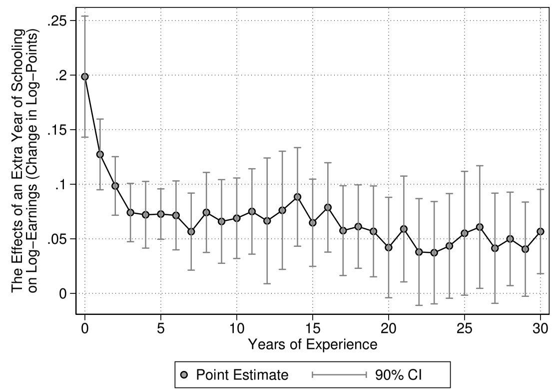
(b) Birth Cohorts 1950–65



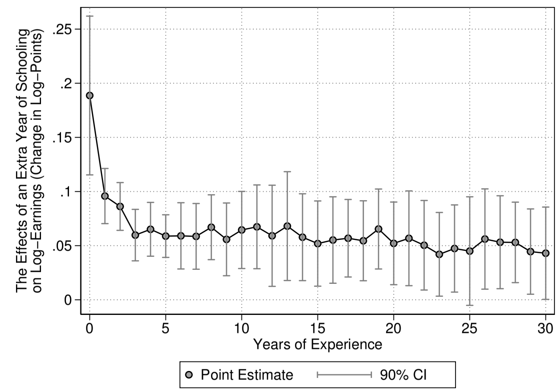
(c) Birth Cohorts 1940–80



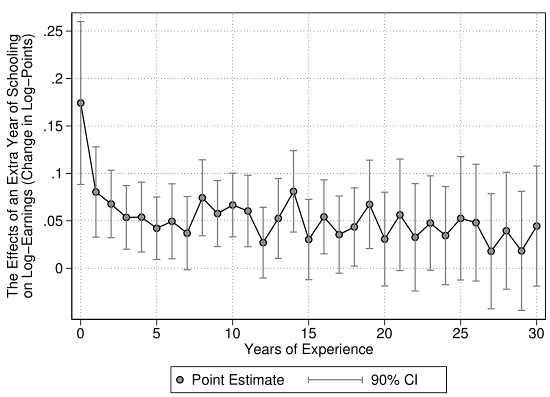
(d) Annual Earnings > .5 SGA



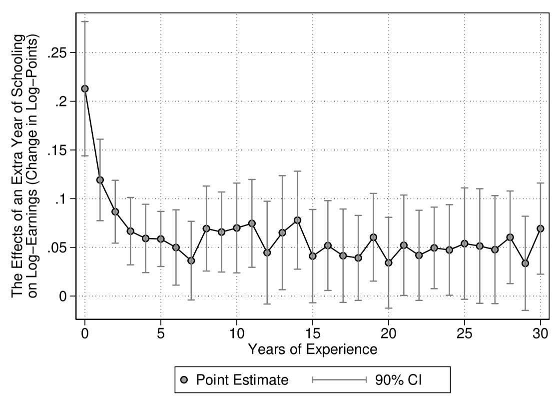
(e) Annual Earnings > .75 SGA



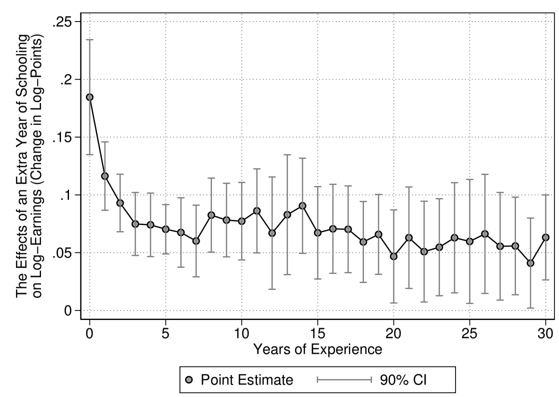
(f) Annual Earnings > 1.5 SGA



(g) Population < 5,000



(h) Population < 10,000



(i) Population < 25,000

Figure A.3: Alternative Hidden IV Estimates of the Returns to Schooling.

Note: See notes below Table A.3 for details on each estimation.

B Identification with Partially-Transparent Instrument

In this section, we extend the identification with partially-transparent instrument variables. Let $D^{\mathfrak{p}} \in \{0,1\}$ denote the partially-transparent IV such that a $\rho \in [0,1]$ fraction of workers be "exposed" to a transparent IV, and $(1-\rho)$ fraction to a hidden IV. Here, the fraction ρ is unobserved to the researchers.

Experience-Invariant Returns to Skill. $D^{\mathfrak{p}}$ satisfies Assumption 1, and with experience-invariant returns to skill, the conditional mean of log wages at t given $D^{\mathfrak{p}}$ is

$$\begin{aligned}\mathbb{E}\left[\ln W_{it}|D_i^{\mathfrak{p}},t\right] &= \rho \times \mathbb{E}\left[\ln W_{it}|D_i^{\mathfrak{t}},t\right] + (1-\rho)\times \mathbb{E}\left[\ln W_{it}|D_i^{\mathfrak{h}},t\right] \\ &= \rho \times \delta^{\psi|S} \times \mathbb{E}\left[S_i|D_i^{\mathfrak{t}}\right] + (1-\rho)\times\left(\delta^{\psi|S}+\theta_t\phi_{A|S}\right)\mathbb{E}\left[S_i|D_i^{\mathfrak{h}}\right], \qquad \text{(B.1)}\end{aligned}$$

which is the weighted average of the conditional log wage under transparent and hidden IV. For notational simplicity, and without loss of generality, we assume that the first-stage effect of $D^{\mathfrak{p}}$ on schooling does not depend on whether it is hidden or transparent, i.e., $\mathbb{E}[S_i|D_i^{\mathfrak{h}}=1] - \mathbb{E}[S_i|D_i^{\mathfrak{h}}=0] = \mathbb{E}\left[S_i|D_i^{\mathfrak{t}}=1\right] - \mathbb{E}\left[S_i|D_i^{\mathfrak{t}}=0\right]$. Then from equation (8), with $\lambda_t = 1$

$$\begin{aligned}\texttt{plim}\ \hat{b}_t^{IV^{\mathfrak{p}}} &= \frac{\mathbb{E}\left[\ln W_{it}|D_i^{\mathfrak{p}}=1,t\right] - \mathbb{E}\left[\ln W_{it}|D_i^{\mathfrak{p}}=0,t\right]}{\mathbb{E}\left[S_i|D_i^{\mathfrak{p}}=1,t\right] - \mathbb{E}\left[S_i|D_i^{\mathfrak{p}}=0,t\right]} \\ &= \delta^{\psi|S}\rho + \left(\delta^{\psi|S}+\theta_t\phi_{A|S}\right)(1-\rho) \\ &= \delta^{\psi|S} + \theta_t\phi_{A|S}(1-\rho), \qquad \text{(B.2)}\end{aligned}$$

where the second equality follows from (B.1). Thus, a partially-transparent IV identifies a *lower bound* on the private return to education. To identify the social return we have to rely on the information at $t=\infty$, because $\lim_{t\to\infty}\theta_t = 0$, and from (B.2) we get $\texttt{plim}\ \left(\lim_{t\to\infty}\hat{b}_t^{IV^{\mathfrak{p}}}\right) = \delta^{\psi|S}$. Then, subtracting $\texttt{plim}\ \left(\lim_{t\to\infty}\hat{b}_t^{IV^{\mathfrak{p}}}\right)$ from (B.2) evaluated at two finite experience levels, $t \neq t'$, and taking their ratios identify $\theta_t/\theta_{t'}$, identifies the speed of learning parameter κ and with it the lower bound of the signaling value.

Experience-Varying Returns to Skill. When returns to skill vary with experience, access to a partially-transparent IV is insufficient to bound the returns to signaling because both λ_t and θ_t vary with t. However, if we have access to a transparent IV and a partially-transparent IV, then under Assumptions 1-4, we can identify the lower bound of the returns to signaling. Then we can use equation (11) to identify $\{\lambda_t : t \in \mathbb{T}\}$, and the rest of the identification strategy follows the same steps as with the experience-invariant returns to skill.

Furthermore, if we have a hidden IV and a partially-transparent IV, we can point-identify private returns to education and the returns to signaling. As we show next, for this identification result, we rely on the homogeneity of social returns across the two IV samples.

We begin with observation that as with (B.2), $D^{\mathfrak{p}}$ identifies a mixture of social and private returns, i.e., from equation (8) we get $\texttt{plim}\ \hat{b}_t^{IV^{\mathfrak{p}}} = \lambda_t \times \left(\delta^{\psi|S} + \theta_t \phi_{A|S}(1-\rho)\right)$. Simplifying further and using $b_0^{IV^{\mathfrak{h}}} = \delta^{\psi|S} + \phi_{A|S}$ and $b_\infty^{IV^{\mathfrak{h}}} = \delta^{\psi|S}$ gives

$$\begin{aligned} \texttt{plim}\ \hat{b}_t^{IV^{\mathfrak{p}}} &= \lambda_t \times \left(\theta_t \times b_0^{IV^{\mathfrak{h}}} + (1-\theta_t) \times b_\infty^{IV^{\mathfrak{h}}}\right) - \lambda_t \times \theta_t \times \left(\rho \times \phi_{A|S}\right) \\ &= \texttt{plim}\ \hat{b}_t^{IV^{\mathfrak{h}}} - \lambda_t \times \theta_t \times \rho \times (b_0^{IV^{\mathfrak{h}}} - b_0^{IV^{\mathfrak{t}}}), \end{aligned} \tag{B.3}$$

where the last equality follows from equation (13). Next, we make the following assumption.

Assumption 4′ (Homogeneity of Social Returns) Let the social returns $\lambda_t \times \delta^{\psi|S}$ be homogenous across the hidden IV sample and partially-transparent IV sample at each $t \in \mathbb{T}$.

Although Assumption 4′ is stronger than Assumption 4 it has a testable implication. In particular, it implies that the hidden IV estimate is always larger than the partially-transparent IV estimate at every $t > 0$.[1] Suppose $D^{\mathfrak{h}}$ and $D^{\mathfrak{p}}$ satisfy Assumptions 1-3 and

[1]For instance, this assumption is rejected in our sample. As we can see from the estimates in Figure 3, for some intermediate t, the social returns estimated from transparent IV sample (Figure 3(b)) is larger than the private returns estimated from hidden IV sample (Figure 3(a)).

4′. Evaluating (B.3) at $t = 0$ and using $\theta_0 = 1$ and $\lambda_0 = 1$ give

$$\texttt{plim}\ \hat{b}_0^{IV^{\mathfrak{h}}} - \texttt{plim}\ \hat{b}_0^{IV^{\mathfrak{p}}} = \rho \times (b_0^{IV^{\mathfrak{h}}} - b_0^{IV^{\mathfrak{t}}}), \tag{B.4}$$

which identifies ρ up to $(b_0^{IV^{\mathfrak{h}}} - b_0^{IV^{\mathfrak{t}}})$. Substituting (B.4) in (B.3) gives $\lambda_t \times \theta_t = \frac{\texttt{plim}\ \hat{b}_t^{IV^{\mathfrak{h}}} - \texttt{plim}\ \hat{b}_t^{IV^{\mathfrak{p}}}}{\texttt{plim}\ \hat{b}_0^{IV^{\mathfrak{h}}} - \texttt{plim}\ \hat{b}_0^{IV^{\mathfrak{p}}}}$, and substituting this in $\texttt{plim}\ \hat{b}_t^{IV^{\mathfrak{h}}}$ for $1 \leq t \leq \infty$, and simplifying gives

$$\texttt{plim}\ \hat{b}_t^{IV^{\mathfrak{h}}} \times \left(\frac{\texttt{plim}\ \hat{b}_t^{IV^{\mathfrak{h}}} - \texttt{plim}\ \hat{b}_t^{IV^{\mathfrak{p}}}}{\texttt{plim}\ \hat{b}_0^{IV^{\mathfrak{h}}} - \texttt{plim}\ \hat{b}_0^{IV^{\mathfrak{p}}}} \right)^{-1} = b_0^{IV^{\mathfrak{h}}} + b_\infty^{IV^{\mathfrak{h}}} \times \left(\frac{1 - \theta_t}{\theta_t} \right). \tag{B.5}$$

Assumption 4′ implies that $\texttt{plim}\ \hat{b}_t^{IV^{\mathfrak{h}}} > \texttt{plim}\ \hat{b}_t^{IV^{\mathfrak{p}}}$ and $\texttt{plim}\ \hat{b}_0^{IV^{\mathfrak{h}}} \geq \texttt{plim}\ \hat{b}_0^{IV^{\mathfrak{p}}}$ for $t > 0$. Thus, with a sufficiently large panel, we can use the NLLS method to estimate the right-hand side parameters of (B.5) and from that the speed of learning parameter κ.

C Identification with Heterogeneous Returns

In this section, we extend the employer learning model to allow heterogeneous returns to education and determine conditions under which we can use IV to identify key model parameters. We use the binary potential outcomes framework of Neyman-Rubin. Let schooling takes two values, $S_i \in \{0, 1\}$, where $S_i = 1$ (respectively, 0) denotes a higher (respectively, lower) level of schooling. Worker i is characterized by a vector of potential outcomes $\{\psi_{0,i}, \psi_{1,i}\}$, where $\psi_{S,i}$ is the experience-invariant component of i's productivity, which subsumes A_i in equation (1).

As in (1), let $\varepsilon_{i,t} \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$ are mean-zero "noise" in the production process that are independent of the model primitives. Then the realized productivity $\psi_{i,t}$ at time t is:

$$\psi_{i,t} = S_i \times [\psi_{1,i} + \varepsilon_{1,i,t}] + (1 - S_i) \times [\psi_{0,i} + \varepsilon_{0,i,t}] + H(t). \tag{C.1}$$

For notational ease, we suppress $H(t)$ in the following.[2] Worker i knows his potential outcomes $\{\psi_{0,i}, \psi_{1,i}\}$, but employers only observe (S_i, ψ_i^t), where $\psi_i^t := \{\psi_{i,\tau}\}_{\tau<t}$ are observed only for $S_i = S$. Note that observing $\psi_{i,t}$, conditional on S_i is informationally equivalent to observing $\xi_{S_i,i,t} = \psi_{S_i,i} + \varepsilon_{S_i,i,t}$. We can thus denote the employers' information set by $\mathcal{E}_{i,t}^{S_i} = (S_i, \xi_{S_i,i}^t)$, where $\xi_{S_i,i}^t = \{\xi_{S_i,i,\tau}\}_{\tau<t}$. Wages are set equal to the expected productivity, conditional on information $\mathcal{E}_{i,t}^{S_i}$. Let $W_{S_i,i,t} := \mathbb{E}\left[\psi_{1,i}\middle|\mathcal{E}_{i,t}^{S_i}\right]$ denotes potential wage outcomes for different S_i. Then we can write the wage equation as

$$W_{i,t} = \mathbb{E}\left[\psi_{i,t}\middle|\mathcal{E}_{i,t}^{S_i}\right] = S_i \times W_{1,i,t} + (1 - S_i) \times W_{0,i,t}, \tag{C.2}$$

where the second equality follows from (C.1) and the independence and zero-mean properties of $\varepsilon_{S_i,i,t}$.

Then, the social returns and the private returns to schooling for i are, respectively

$$\delta_i^{\psi|S_i} := \psi_{1,i} - \psi_{0,i}; \tag{C.3}$$

$$\delta_{i,t}^{W|S_i} := W_{1,i,t} - W_{0,i,t} = \mathbb{E}\left[\psi_{1,i}|\mathcal{E}_{i,t}^1\right] - \mathbb{E}\left[\psi_{0,i}|\mathcal{E}_{i,t}^0\right]. \tag{C.4}$$

Note that both the social returns $\delta_i^{\psi|S}$ and the private returns $\delta_{i,t}^{W|S}$ are individual-specific. The *average social returns* and *average private returns* are then the population averages of (C.3) and (C.4), respectively, while measures such as *social returns for the treated* and *private returns for the treated* are averages across the corresponding populations.

Identification Using Instrumental Variables

To understand what a binary IV identifies, we proceed analogously to Imbens and Angrist [1994]. Let $S_i(D_i)$ denote potential schooling, which is a function of $D_i \in \{0, 1\}$, and define compliers as $\mathbb{C} \equiv \{i|S_i(1) = 1, \text{and } S_i(0) = 0\}$ and defiers as $\mathbb{D} \equiv \{i|S_i(1) = 0, \text{and } S_i(0) = 1\}$.

[2] Note that productivity $\psi_{i,t}$ is expressed in levels and not in logs, because with this level of generality it is easier to work in levels. We continue to maintain the assumption that S and $H(t)$ are additively separable.

Similarly, we can define always-takers to be $\mathbb{A} \equiv \{i | S_i\,(1) = 1, \text{and } \; S_i\,(1) = 1\}$ and never-takers to be $\mathbb{N} \equiv \{i | S_i\,(1) = 0, \text{and } \; S_i\,(0) = 0\}$. Then, as before, the Wald estimator gives

$$\texttt{plim}\ \hat{b}_t^{IV} \quad := \quad \frac{\mathbb{E}\,[W_{i,t}|D_i = 1] - \mathbb{E}\,[W_{i,t}|D_i = 0]}{\mathbb{E}\,[S_i|D_i = 1] - \mathbb{E}\,[S_i|D_i = 0]}. \tag{C.5}$$

As D satisfies the monotonicity condition, $\Pr(\mathbb{D}) = 0$, (C.5)'s denominator simplifies to

$$\begin{aligned}
\mathbb{E}\,[S_i|D_i = 1] - \mathbb{E}\,[S_i|D_i = 0] \quad &= \quad (\mathbb{E}\,[S_i|D_i = 1, \mathbb{A}] - \mathbb{E}\,[S_i|D_i = 0, \mathbb{A}]) \times \Pr(\mathbb{A}) \\
&+ \quad (\mathbb{E}\,[S_i|D_i = 1, \mathbb{N}] - \mathbb{E}\,[S_i|D_i = 0, \mathbb{N}]) \times \Pr(\mathbb{N}) \\
&+ \quad (\mathbb{E}\,[S_i|D_i = 1, \mathbb{C}] - \mathbb{E}\,[S_i|D_i = 0, \mathbb{C}]) \times \Pr(\mathbb{C}) \\
&+ \quad (\mathbb{E}\,[S_i|D_i = 1, \mathbb{D}] - \mathbb{E}\,[S_i|D_i = 0, \mathbb{D}]) \times \Pr(\mathbb{D}) = \Pr(\mathbb{C}). \quad \text{(C.6)}
\end{aligned}$$

As before, $D_i^{\mathfrak{h}} \in \{0, 1\}$ denotes a hidden IV and $D_i^{\mathfrak{t}} \in \{0, 1\}$ a transparent IV. With a hidden IV, we also know that i's wage conditional on employer information $\mathcal{E}_{i,t}^S$ does not depend on the IV itself. Thus, for $D_i^{\mathfrak{h}}$, (C.5)'s numerator can be expressed as

$$\begin{aligned}
\mathbb{E}[W_{i,t}|D_i^{\mathfrak{h}} = 1] - \mathbb{E}[W_{i,t}|D_i^{\mathfrak{h}} = 0] \quad &= \quad \Big(\mathbb{E}[W_{i,t}|D_i^{\mathfrak{h}} = 1, \mathbb{C}] - \mathbb{E}[W_{i,t}|D_i^{\mathfrak{h}} = 0, \mathbb{C}]\Big) \times \Pr(\mathbb{C}) \\
&= \quad (\mathbb{E}[W_{1,i,t}|\mathbb{C}] - \mathbb{E}[W_{0,i,t}|\mathbb{C}]) \times \Pr(\mathbb{C}) = \mathbb{E}[\delta_{i,t}^{W|S}|\mathbb{C}] \times \Pr(\mathbb{C}). \quad \text{(C.7)}
\end{aligned}$$

The first equality follows from the law of total expectation, the second from the definition of a complier and substituting for the potential outcomes from (C.2) and using the properties of a hidden IV, and the last from the definition of private returns in (C.4). Using (C.6) and (C.7) in (C.5) with a hidden IV, gives $\texttt{plim}\ \hat{b}_t^{IV^{\mathfrak{h}}} = \mathbb{E}\left[\delta_{i,t}^{W|S}|\mathbb{C}\right]$. Thus, a binary hidden IV identifies the *average private returns* to education among compliers.

Next, we consider the identification with transparent IV, $D_i^{\mathfrak{t}} \in \{0, 1\}$. Wages equal expected productivity given employer information $(\mathcal{E}_{i,t}^S, D_i^{\mathfrak{t}})$, i.e., $W_{i,t} = \mathbb{E}[\psi_{i,t}|\mathcal{E}_{i,t}^S, D_i^{\mathfrak{t}}]$ and from the law of total expectation we get $\mathbb{E}\,[W_{i,t}|D_i^{\mathfrak{t}}] = \mathbb{E}[\mathbb{E}[\psi_{i,t}|\mathcal{E}_{i,t}^{S_i}, D_i^{\mathfrak{t}}]|D_i^{\mathfrak{t}}] = \mathbb{E}[\psi_{i,t}|D_i^{\mathfrak{t}}]$.

Conditional on $D_i^{\mathfrak{t}}$, the average wages equals the average product, and hence

$$\begin{aligned}\mathbb{E}\left[W_{i,t}|D_i^{\mathfrak{t}}=1\right]-\mathbb{E}[W_{i,t}|D_i^{\mathfrak{t}}=0] &= \mathbb{E}[\psi_{i,t}|D_i^{\mathfrak{t}}=1]-\mathbb{E}[\psi_{i,t}|D_i^{\mathfrak{t}}=0] \\ &= \mathbb{E}[\psi_{1,i}-\psi_{0,i}|\mathbb{C}]\times\Pr\left(\mathbb{C}\right)=\mathbb{E}[\delta_i^{\psi|S}|\mathbb{C}]\times\Pr\left(\mathbb{C}\right). \qquad \text{(C.8)}\end{aligned}$$

Using (C.6) and (C.8) in (C.5) gives $\texttt{plim}\ \hat{b}^{IV^{\mathfrak{t}}}=\mathbb{E}[\delta_i^{\psi|S}|\mathbb{C}]$. Thus, a binary transparent IV identifies the *average social returns* to education among compliers. Therefore, with heterogeneous returns, a transparent IV identifies the average social returns and a hidden IV identifies the average private returns to education for the compliers.

Speed of Learning. Next, we consider identifying the speed of learning by determining how quickly the market learns workers' ability. The speed of learning will depend on schooling and the selection between schooling and unobserved ability. Let the unobserved components of productivity, conditional on schooling S, follow a Normal distribution, i.e., $\psi_{S,i}\big|S\sim\mathcal{N}\left(\mu_{\psi,S},\sigma_\psi^2\right)$. Using the potential wage outcomes defined in (C.2), and the Kalman property as in equation (5), we can express wages as

$$W_{S_i,i,t}=\mathbb{E}\left[\psi_{S_i,i}|\mathcal{E}_{i,t}^{S_i}\right]=\theta_t\times\mu_{\psi,S}+(1-\theta_t)\times\bar{\xi}_{S_i,i}^t, \qquad \text{(C.9)}$$

where, $\theta_t=\frac{1-\kappa}{1+(t-1)\kappa}$, $\kappa=\frac{\sigma_\psi^2}{\sigma_\psi^2+\sigma_\varepsilon^2}$ is the speed of learning, and $\bar{\xi}_{S_i,i}^t=\frac{1}{t}\sum_{\tau<t}\xi_{S_i,i,\tau}$ is the average of signals up to period t. Conditional expectation of (C.9) for $(S_i,\psi_{S_i,i})$ gives

$$\begin{aligned}\mathbb{E}[W_{S_i,i,t}|S_i,\psi_{S_i,i}] &= \theta_t\times\mu_{\psi,S}+(1-\theta_t)\times\mathbb{E}\left[\bar{\xi}_{S_i,i}^t|S_i,\tilde{\psi}_{S_i,i}\right]=\theta_t\times\mu_{\psi,S}+(1-\theta_t)\times\psi_{S,i} \\ &= \mu_{\psi,S}+(1-\theta_t)\times\left(\psi_{S_i,i}-\mu_{\psi,S}\right). \qquad \text{(C.10)}\end{aligned}$$

Thus, the wage of a worker with schooling S_i at with t years of work experience is the sum of the average productivity $\mu_{\psi,S}$ and the deviation of i's productivity from its mean $(\psi_{S_i,i}-\mu_{\psi,S})$ augmented by employers' learning $(1-\theta_t)$. So, at the start, i.e., $t=0$, i's wage is the average productivity because $\theta_0=1$, and as information about i's ability is accumulated in the market, i's wage becomes more responsive to i's true productivity as $\lim_{t\to\infty}\theta_t = 0$. Using (C.4), (C.7) and (C.10) in (C.5) the hidden IV estimate becomes

$$\begin{aligned}
\texttt{plim}\ \ \hat{b}_t^{IV^{\mathfrak{h}}} &= \mathbb{E}\left[(\mu_{\psi,1}-\mu_{\psi,0}) + (1-\theta_t)(\psi_{1,i}-\mu_{\psi,1}) - (1-\theta_t)(\psi_{0,i}-\mu_{\psi,0})\,\middle|\mathbb{C}\right] \\
&= \underbrace{\mathbb{E}\left[(\mu_{\psi,1}-\mu_{\psi,0})\,|\mathbb{C}\right]}_{:=\Upsilon} + (1-\theta_t)\left(\underbrace{\mathbb{E}\left[(\psi_{1,i}-\mu_{\psi,1})\,|\mathbb{C}\right]}_{:=\Upsilon_1} - \underbrace{\mathbb{E}\left[(\psi_{0,i}-\mu_{\psi,0})\,|\mathbb{C}\right]}_{:=\Upsilon_0}\right) \\
&:= \Upsilon + (1-\theta_t)(\Upsilon_1 - \Upsilon_0). \qquad \text{(C.11)}
\end{aligned}$$

Furthermore, because private returns are greater than social returns, it follows from (C.3) and (C.4) that $\Upsilon_1 < \Upsilon_0$, and T is sufficiently large. Then we can use (C.11) to identify θ_t and the speed of learning parameter κ. For instance, at $t = 0$ we can get $\texttt{plim}\ \ \hat{b}_0^{IV^{\mathfrak{h}}} = \Upsilon$ and at $t \to \infty$ we get $\texttt{plim}\ \ \hat{b}_\infty^{IV^{\mathfrak{h}}} = \Upsilon + (\Upsilon_1 - \Upsilon_0)$. So, for $0 < t < \infty$ we identify $\theta_t = \frac{\texttt{plim}\ \ \hat{b}_t^{IV^{\mathfrak{h}}} - \texttt{plim}\ \ \hat{b}_0^{IV^{\mathfrak{h}}}}{\texttt{plim}\ \ \hat{b}_\infty^{IV^{\mathfrak{h}}} - \texttt{plim}\ \ \hat{b}_0^{IV^{\mathfrak{h}}}}$.

Experience-Varying Returns to Skills. We end this section by briefly considering the identification with experience-varying returns to skill. First note that with experience-varying returns to skill, realized productivity (C.1) becomes

$$\psi_{i,t} = \lambda_t \times (S_i \times [\psi_{1,i} + \varepsilon_{1,i,t}] + (1 - S_i) \times [\psi_{0,i} + \varepsilon_{0,i,t}]) + H(t). \qquad \text{(C.12)}$$

Once we re-define social and private returns to include the effect of λ_t, following the same identification strategy as above, it follows that the hidden IV identifies the average private returns and transparent IV identifies the average social returns, among the compliers, at each work experience $t \in \mathbb{T}$. Suppose, as before, we have access to a hidden IV and a transparent IV and suppose $\{\lambda_t : t \in \mathbb{T}\}$ satisfy Assumption 4.

The hidden IV and the transparent IV estimates the (average) private and (average) social returns for different complier groups. Under Assumption 4, however, we can combine the two sets of estimates to identify the model with experience-varying returns for hidden IV compliers, even though the transparent IV estimates relate to a different set of compliers.

In particular, following the same step as in (C.11) we get

$$\texttt{plim}\ \hat{b}_t^{IV^{\mathfrak{h}}} = \lambda_t \Upsilon + \lambda_t (1-\theta_t)(\Upsilon_1 - \Upsilon_0); \tag{C.13}$$

$$\texttt{plim}\ \hat{b}_t^{IV^{\mathfrak{t}}} = \lambda_t \Upsilon + \lambda_t (\Upsilon_1 - \Upsilon_0). \tag{C.14}$$

Evaluating (C.13) at $t = 0$, and using $\lambda_0 = 1$, identifies $\texttt{plim}\ \hat{b}_0^{IV^{\mathfrak{h}}} = \Upsilon$. Substituting it in (C.14) at $t = 0$ identifies $\texttt{plim}\ \hat{b}_0^{IV^{\mathfrak{t}}} - \texttt{plim}\ \hat{b}_0^{IV^{\mathfrak{h}}} = (\Upsilon_1 - \Upsilon_0)$. Then from the transparent IV (C.14) we can identify $\{\lambda_t : t > 0\}$. Then using these variables in (C.13) we identify κ.

D Identification with Employer-Observed Correlate

In this section, we extend our primary model to allow employers to observe a correlate of ability Q that the researcher does not observe. Throughout this section, we maintain all other assumptions from our model. Worker i's log-productivity for $t \in \mathbb{T}$ is given by

$$\psi_{it} := \ln \chi_{it} = \lambda_t \times (\beta_{ws} S_i + \beta_{wq} Q_i + A_i + \varepsilon_{it}) + H(t), \tag{D.1}$$

where Q is a correlate of ability observed by employers and is possibly correlated with A. An example of a Q could be knowledge of a foreign language, which is typically mentioned in job applicants' résumés, and can be verified by the employers.

To model employer learning in addition to $\varepsilon_{it} \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$ let $(S_i, Q_i, A_i) \overset{i.i.d}{\sim} \mathcal{N}(\boldsymbol{\mu}, \Sigma)$, across workers and across time. The joint normality assumption allows us to express A as a linear function of (S, Q)

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

where $\varepsilon_{A_i|S_i,Q_i} := A_i - \mathbb{E}[A_i|S_i, Q_i]$. Under perfect competition, log wages is

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

where, as before, $\tilde{H}(t) \equiv H(t) + \frac{1}{2}v_t$ 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 the empirical analysis.

The normality assumptions also allow us to use the Kalman filter to write the conditional expectation of ability $\mathbb{E}[A_i|\mathcal{E}_{it}]$ in linear form as

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

where $\bar{\xi}_i^t = \frac{1}{t}\sum_{\tau<t}\xi_{i\tau}$ is the average of signals up to period t and $\theta_t = \frac{1-\kappa}{1+(t-1)\kappa} \in [0,1]$ is the weight on the initial signal (S_i, Q_i) with $\kappa = \frac{\sigma_0^2}{\sigma_0^2+\sigma_\varepsilon^2} \in [0,1]$. Next, we define the social and private returns to education. Recall the notation that for Y, $\delta^{Y|S}$ denotes the causal effect of S on Y and $\tilde{Y}$ denote the part of Y that is not caused by schooling S but may correlate with S. Using these notations and assumptions for $Y = Q$ and $Y = A$, we get, respectively,

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

Then, substituting (D.5) into (D.1), we obtain

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

The coefficient ($\delta_t^{\psi|S}$) in (D.6) is the total causal effect of schooling on productivity –the social return to education– and it captures the direct and indirect effect on other ability components, i.e., (Q, A). Thus (D.6) shows that an extra year of S increases Q by $\delta^{Q|S}$ and A by $\delta^{A|S}$ units, and in turn, they increase the productivity by $\lambda_t\beta_{wq}$ and λ_t, respectively.

Consider now the private returns to education. Substituting (D.2) and (D.4) in (D.3), and using $\bar{\xi}_i^t := \frac{1}{t}\Sigma_{\tau<t}(A_i + \varepsilon_{i\tau}) = A_i + \bar{\varepsilon}_i^t$, and the fact that $\mathbb{E}(A_i|S_i, Q_i)$ is linear and separable in S_i and Q_i, we can write the log-earnings as $\ln W_{it} = \lambda_t \times \left(\beta_{ws} + \theta_t\phi_{A|S}\right)S_i +$

$\lambda_t \times \left(\beta_{wq} + \theta_t \phi_{A|Q}\right) Q_i + \lambda_t \times (1-\theta_t)\left(A_i + \overline{\varepsilon}_i^t\right)$. Then, using (D.5) to replace Q and A gives

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

The coefficient of schooling ($\delta_t^{W|S}$) in (D.7) is the private return to education. Comparing this coefficient with the coefficient of schooling $\delta_t^{\psi|S}$ in equation (3), gives:

$$\delta_t^{W|S} = \delta_t^{\psi|S} + \theta_t \times \lambda_t \times \left(\phi_{A|S} + \phi_{A|Q}\delta^{Q|S} - \delta^{A|S}\right). \quad \text{(D.8)}$$

Once we have augmented the definition of the social returns and the adjustment term to capture the effect of Q in (D.8), the rest of the identification results applies verbatim. In particular, when the returns to skill is experience-invariant and the hidden IV, $D^{\natural}$, satisfies the assumption $\ln W_{it} \perp D_i^{\natural} | (S_i, Q_i, \xi_i^t)$ then it identifies the private returns to education, i.e.,

$$\texttt{plim}\ \hat{b}_t^{IV^{\natural}} = \frac{\mathbb{E}\left[\ln W_{it} | D_i^{\natural} = 1, t\right] - \mathbb{E}\left[\ln W_{it} | D_i^{\natural} = 0, t\right]}{\mathbb{E}\left[S_i | D_i^{\natural} = 1\right] - \mathbb{E}\left[S_i | D_i^{\natural} = 0\right]} = \delta^{\psi|S} + \theta_t \left(\phi_{A|S} + \phi_{A|Q}\delta^{Q|S} - \delta^{A|S}\right). \quad \text{(D.9)}$$

Comparing (D.9) with the private returns defined in (D.8), we can conclude that, for every work experience level t, the hidden IV identifies the private returns to education, i.e., $\texttt{plim}\ \hat{b}_t^{IV^{\natural}} = \delta_t^{W|S}$. Hidden IV also identifies the speed of learning. Likewise, the Wald estimator for a transparent IV, $D^{\mathfrak{t}}$, identifies the social returns to education at all t, i.e.,

$$\texttt{plim}\ \hat{b}_t^{IV^{\mathfrak{t}}} = \frac{\mathbb{E}\left[\ln W_{it} | D_i^{\mathfrak{t}} = 1, t\right] - \mathbb{E}\left[\ln W_{it} | D_i^{\mathfrak{t}} = 0, t\right]}{\mathbb{E}\left[S_i | D_i^{\mathfrak{t}} = 1\right] - \mathbb{E}\left[S_i | D_i^{\mathfrak{t}} = 0\right]} = \delta^{\psi|S}. \quad \text{(D.10)}$$

Next, we consider the case when the returns to skill vary with experience. As before, a hidden IV identifies the private returns to education, and a transparent IV identifies the experience-varying social returns to education. Formally, following the same steps as in

(D.9), $D^{\mathfrak{h}}$ and $D^{\mathfrak{t}}$ at t, respectively, identify the private and social returns as

$$\texttt{plim}\ \hat{b}_t^{IV^{\mathfrak{h}}} = \frac{\mathbb{E}\left[\ln W_{it}|D_i^{\mathfrak{h}}=1,t\right] - \mathbb{E}\left[\ln W_{it}|D_i^{\mathfrak{h}}=0,t\right]}{\mathbb{E}\left[S_i|D_i^{\mathfrak{h}}=1,t\right] - \mathbb{E}\left[S_i|D_i^{\mathfrak{h}}=0,t\right]} = \lambda_t\big(\delta^{\psi|S} + \theta_t(\phi_{A|S} + \phi_{A|Q}\delta^{Q|S} - \delta^{A|S})\big)\ ; \tag{D.11}$$

$$\texttt{plim}\ \hat{b}_t^{IV^{\mathfrak{t}}} = \frac{\mathbb{E}\left[\ln W_{it}|D_i^{\mathfrak{t}}=1,t\right] - \mathbb{E}\left[\ln W_{it}|D_i^{\mathfrak{t}}=0,t\right]}{\mathbb{E}\left[S_i|D_i^{\mathfrak{t}}=1,t\right] - \mathbb{E}\left[S_i|D_i^{\mathfrak{t}}=0,t\right]} = \lambda_t \times \delta^{\psi|S} := \delta_t^{\psi|S}. \tag{D.12}$$

Note that with access to both IVs that satisfy Assumption 4, we can identify $\{\lambda_t : t \in \mathbb{T}\}$.

References

Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica*, 62(2): 467–475.