Department of Economics

SOAS University of London

Working Paper Series

ISSN 1753 - 5816

Please cite this paper as:

van Huellen, Sophie and Duo Qin. (2016), "Compulsory Schooling and the Returns to Education: A Re-examination", SOAS Department of Economics Working Paper Series, No. 199, The School of Oriental and African Studies.

No. 199

Compulsory Schooling and the Returns to Education: A Re-examination

by

Sophie van Huellen and Duo Qin

(September, 2016)

Department of Economics
School of Oriental and African Studies
London
WC1H 0XG
Phone: + 44 (0)20 7898 4730
Fax: 020 7898 4759
E-mail: economics@soas.ac.uk
http://www.soas.ac.uk/economics/

Design and layout: O.G. Dávila

Compulsory Schooling and the Returns to Education: A Re-examination

Sophie van Huellen* and Duo Qin**

Abstract

We re-examine the effect of compulsory school law on education in the US pioneered by Angrist and Krueger (1991). We show that the standard instrumental variable approach of the education variable not only yields empirically inconsistent estimates, but is conceptually confused. The confusion arises from the rejection of the key causal variable as a valid conditional variable. By route of a causally explicit model design we are able to identify the circumstances under which the formerly rejected variable can yield valid inference values. Our investigation demonstrates the importance of building data-consistent models over estimator choice in successful research designs.

Keywords: instrumental variables, randomisation, research design, returns to education, treatment effect

JEL classification: C26, C52, H75, I21, I26, J24, N32

* Department of Economics, SOAS, University of London. Email: sv8@soas.ac.uk ** Department of Economics, SOAS, University of London. Email: dq1@soas.ac.uk

1. Introduction

Over the past century, compulsory school law (CSL) was introduced in virtually every middle and high income country. Such law sets either the minimum number of years of school attendance or the minimum age for leaving school or both. The introduction of the CSL in the US is believed to have greatly contributing to the "High School Movement" between 1910–40 (Goldin and Katz 2007), by which the rate of US high school graduates more than quintupled (Goldin 1998). The precise effect of the CSL with regards to educational attainment as well as income has been subject to empirical investigations in labour economics and microeconomics since the 1990s.

These investigations are pioneered by Angrist and Krueger (1991) and Acemoglu and Angrist (2001), who use CSL indicators as instruments to 'randomise' latent ability across educational attainment groups. As CSL treatment can be conceptualised as natural experiment, the instrumental variable (IV) treatment of education corrects for the assumed bias in the ordinary least square (OLS) estimator; see Angrist and Pischke (2009, Ch. 4) and also Harmon et al. (2003). The influence of these two seminal papers reaches far beyond the estimation of CSL effects. The empirical strategy is now common practice in research on returns to education as well as programme evaluation modelling, e.g. Harmon et al. (2003), Ludwig et al. (2012, 2013), and the two papers have entered the standard economics curriculum as evident from their appearance in two popular textbooks by Angrist and Pischke (2009, 2015).

Despite the far reaching influence of these studies, the estimation of CSL effect on income is subject to controversy. The controversy is evoked by two interlinked developments in the literature: a) a shift in the interpretation of the CSL instrumentalised returns to schooling coefficient despite identical modelling choice, and b) contradictory empirical results which vary significantly with the choice of CSL indicators. Angrist and Krueger (1991), who approximate CSL with quarter of birth dummies, interpret their results as unbiased estimates of returns to education, and find that the IV estimates are not statistically different from the 'untreated' education estimates obtained via OLS.[1] Acemoglu and Angrist (2001) construct alternative CSL indicators based on labour law that produce IV estimates which, although significantly different from OLS estimates, are insignificant or negative.[2] Acemoglu and Angrist (2001) no longer interpret the IV estimates as unbiased returns to schooling, but as the causal effect of CSL on earnings via schooling. These CSL indicators are further refined by Stephens Jr. and Yang (2014). Their results verify the finding of largely insignificant IV estimates by Acemoglu and Angrist (2001) and Stephen Jr. and Yang (2014, p.1789) conclude that there is 'no evidence of benefits to additional schooling' due to CSL. The result of insignificant IV estimates potentially encouraged this shift in interpretation. The conclusion of no CSL effect is counterintuitive, but zero returns to schooling are economically untenable.

Despite the careful refinement of CSL indicators, IV estimates presented by Stephen Jr. and Yang (2014) lack consistency and robustness relative to their OLS counterparts. The IV estimates fail to converge in large samples, are highly sensitive to the inclusion of additional control variables and standard errors remain large despite large sample size. For instance, the inclusion of interaction terms, which allow for regional differences in year of birth effects, leads to large changes in the IV estimates whereas the OLS estimates remain virtually invariant (see Figure 1).

[1] E.g. column (5) versus (6) in Table 4, (7) versus (8) in Table 5, (1) versus (2) and (5) versus (6) in Table 6 in Angrist and Krueger (1991). More evidence in Hoogerheide and van Dijk (2006, Table 5) and in Harmon et al. (2003, Section 5).
[2] See Angrist and Pischke (2015, Table 6.3) for a summary.

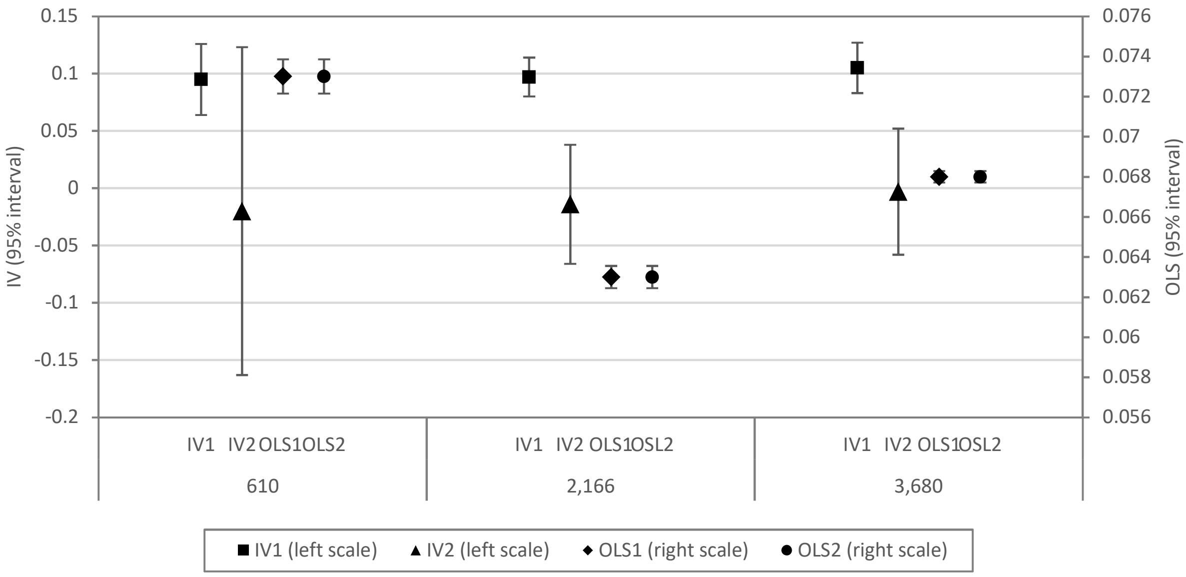

FIGURE 1. OSL AND IV ESTIMATOR CONSISTENCY.

Notes: IV1 and OLS1 are IV and OLS estimates without regional control variables and IV2 and OLS2 are IV and OLS estimates with regional control variables included. The x-axis provides the sample size and the y-axes coefficient values (left axis for IV estimates and right axis for OLS estimates). The bars indicate the 95% confidence interval. Source: SY, Table 1.

The lack of consistency and robustness of attendance law based IV-estimates led Angrist and Pischke (2015, p.227) to discard the Acemoglu and Angrist (2001) indicator choice as 'a failed research design', while presenting the Angrist and Krueger (1991) quarterly birth dummy choice as a success. The judgment is made despite the obvious fact that attendance law indicators are a better reflection of the CSL than quarterly birth dummies. The finding that IV estimates diverge further away from values that are economically sensible as indicators become a better approximation of the true CSL is surprising. Although Stephen Jr. and Yang (2014) use a further improved set of attendance law instruments and control variables to rescue the formerly failed research design, the rescue is unsuccessful as evident from Figure 1.

This paper probes into the causes of the 'failed research design' through a careful re-examination of two data sets, one from Angrist and Krueger (1991) (AK hereafter) and the other from Stephens Jr. and Yang (2014) (SY hereafter).[3] The two datasets are useful for cross-

[3] The data used by SY is an extended version of the data and indicators used by Acemoglu and Angrist (2001).

validation as they are both created from the 1980 US census but with different indicators and choice of control variables. The paper commences by addressing two questions: 1) Are the instruments used in SY statistically valid for cases where the resulting IV estimates are significantly different from their OLS counterpart? 2) Can those valid IV estimates, presented in AK and SY, outperform their OLS counterpart in terms of consistency and robustness? Based on extensive experiments presented in Section 2, the answer to both questions is negative. This puts into doubt the usefulness of the CSL-randomisation route.

Section 3 traces the failed research design to confused causal model specification. In particular, it explains that the IV approach amounts to the premise of the education variable being an invalid conditional variable; a premise which has been overwhelmingly rejected in Section 2. Careful consideration of the causal links between the variables of interest reveals the logical necessity of decomposing the outcome effects into two parts: returns to schooling and the average treatment effect (ATE) of the CSL via schooling. The decomposition enables the translation of the possible moderation effect of CSL on schooling to a parametric shift in a multivariate model setting.

Section 4 applies this approach of decomposition to the two data sets. For the outcome effects, we find a virtually invariant returns to schooling estimate of 0.06, and smaller ATE of the CSL estimates between 1-5% if using school law indicators and 0.2- 0.9% if using quarter of birth indicators. However, considering imperfections of CSL indicators identified in Section 2, estimates of the CSL effect are far less robust than those of the returns to schooling parameter. Moreover, we find little evidence for a CSL-induced parametric shift in the returns to schooling parameter using the available CSL indicators.

The methodological implications of our findings are discussed in Section 5. We pinpoint the failure of the research design to the choice of the IV approach instead of inappropriate indicators. The IV approach is a detour into deadlock, as it effectively denies direct translation of causal postulates of interest into statistically conditional relationships. We argue that a successful research design has to be focused on model design, i.e. on how to faithfully translate the causal postulates into testable conditional hypotheses in data-consistent models.

2. Where Does the Research Design Fail?

Let us start from the basic model design of AK and SY. Briefly, denote education by s and outcome by y, the returns to education can be expressed by:

(1) $$y = \alpha + \beta s + \eta.$$

AK and SY assume $cov(s\eta) \neq 0$, based on the argument that η contains omitted variables which are not directly observable but collinear with s, such as innate ability. The OLS is thus an inconsistent estimator for (1) and the inconsistency can be treated by IV estimators using the CSL as a key instrument. The resulting IV estimator amounts to appending (1) by:

(2) $$s = \pi L + I_j' \boldsymbol{\gamma}_j + \varepsilon$$

where L represents the CSL and I_j other IVs.[4] Equation (2) can be seen as the first stage of the two-stage OLS (2SLS) estimator to produce an IV estimate β^L, with $\beta^L \neq \beta$.

Since the CSL is latent in the cross-section samples used by AK and SY, it has to be approximated by observable indicators, $\mathcal{L}$. Quarterly birth dummies are chosen by AK (hereafter denoted as $\mathcal{L}_{AK}$). The indicator choice is based on the insight that the CSL requires a minimum age which has to be reached before students can drop out of school. Those born in the first quarters of the year reach this age sooner than those born in later quarters and hence are less constrained by the law than their peers. Accordingly, AK define three birth dummies for those born in the first ($\mathcal{L}_{AK}^1$), second ($\mathcal{L}_{AK}^2$) and third ($\mathcal{L}_{AK}^3$) quarter of the year.

SY, with reference to Acemoglu and Angrist (2001), propose two alternative indicators based on state school and labour law. These capture required years of schooling ($\mathcal{L}_{SY1}$) and

[4] See equation (2) of SY and equation (1) of AK.

compulsory attendance ($\mathcal{L}_{SY2}$).[5] As in AK the indicators compose of three dummies. $\mathcal{L}_{SY1}^{1}$, $\mathcal{L}_{SY1}^{2}$, and $\mathcal{L}_{SY1}^{3}$ capture those with minimum of 7 or below, 8, and 9 or above required years of schooling and $\mathcal{L}_{SY2}^{1}$, $\mathcal{L}_{SY2}^{2}$, and $\mathcal{L}_{SY2}^{3}$ capture those with 8 or below, 9, and 10 or above years of compulsory school attendance.

The choice of CSL indicators alters estimation results considerably and SY, in contrast to AK, succeed in finding $\beta^{L} \neq \beta$. However, a close scrutiny through replication of SY's Tables 1 and A2 suggests that their CSL indicators are largely invalid instruments. Column (1) in T1B of our Table 1 is the only exception, with no rejection of Sargan's null of valid overidentifying restrictions and rejection of Hausman's null of OLS estimator consistency relative to IV. Although the validity of instruments is confirmed for column (2) in T1A of Table 1, the IV estimates remain insignificant.

TABLE 1—SARGAN AND HAUSMAN TEST FOR INSTRUMENTS USED BY SY

	T1A ($\mathcal{L}_{SY1}$)				T1B ($\mathcal{L}_{SY2}$)			
White males	aged 40-49		aged 25-54		aged 40-49		aged 25-54	
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
β (OLS)[1]	0.073**	0.073**	0.063**	0.063**	0.073**	0.073**	0.063**	0.063**
β^{L} (2SLS)[1]	0.095**	-0.020	0.097**	-0.014	0.142**	0.092**	0.140**	0.086**
Tests:								
Sargan[2]	0.99	4.65	17.99	7.51	0.64	0.83	12.75	17.57
(p-value)	(0.6088)	(0.0977)	(0.0001)	(0.0234)	(0.7271)	(0.6589)	(0.0017)	(0.0002)
Hausman	3.80	9.67	43.24	36.32	16.33	0.53	150.29	3.28
(p-value)	(0.0512)	(0.0019)	(0.0000)	(0.0000)	(0.0001)	(0.4671)	(0.0000)	(0.0701)
Fixed effects:								
State of birth	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year of birth	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region x Yob	No	Yes	No	Yes	No	Yes	No	Yes
Additional controls	None	None	Age quartic, census yr	Age quartic, census yr	None	None	Age quartic, census yr	Age quartic, census yr

Notes: ** denotes significance at the 1% level and * at the 5% level; [1] robust and cluster adjusted standard errors are used; [2] Wooldridge's extension of Sargan's test of overidentifying restrictions is performed. Source: SY Tables 1 and A2 – Sargan and Hausman added through replication.

A closer look at Table 1 reveals that larger differences between IV and OLS estimates occur predominantly when regional interaction terms with year of birth indicators are included (columns (2) and (4) of T1A and T1B). At the same time, the inclusion of interaction terms invalidates the claim of endogeneity if using $\mathcal{L}_{SY}$ indicators and leads to insignificant IV

[5] See SY for a detailed definition of the indicators.

estimates if using $\mathcal{L}_{SY}$ indicators. The sensitivity of IV estimates to regional factors is acknowledged by SY and reiterated by Hogerheide and van Dijk (2006, Table 5).[6] This raises the question whether SY's IVs solely represent the CSL treatment. Since regional variation effects cannot be disentangled from those IVs, the interpretation of insignificant coefficient estimates as zero CSL effect is disputable.

The insignificance and empirical inconsistency of IV estimates, already identified in Figure 1, could be caused by a negligible share of 'compliers' in the full sample; a point made by Oreopoulos (2006a) in the context of the CSL effect when using minimum years of schooling indicators. Since most people remain in school beyond the required years, the great majority of the sample belongs to a sub-population for which the ATE of the CSL is expected to be zero. In other words, the CSL is potentially binding only for school leavers, but by and large not for those who have continued education beyond the compulsory years of schooling. Using $\mathcal{L}_{SY2}$ indicators, only 4.11 per cent of the 1930-39 born cohort complies to the law. The share of compliers is even smaller for the later born cohort with 2.31 per cent. Using $\mathcal{L}_{SY}$ indicators instead, the share of complies is similarly small with 4.18 and 2.53 per cent in the 1930s and 1940s birth cohort respectively (see Table 1A, Appendix).

Aiming for consistent results, we replicate AK Tables V and VI and SY Tables 1 and A2, but use sub-samples by levels of educational attainment in order to separate 'always takers' from 'compliers'. Those who receive 12 or less years of schooling are allocated to the *School* sub-sample and those with more years of education are allocated to the *Higher* sub-sample, see. Further, the tails of the two sub-samples, *Higher* and *School*, are cut to investigate whether dissimilarities between the sub-samples arise due to outliers (cf. Figure 1A, Appendix).

[6] CSL indicators based on quarter of birth dummies face similar problems and Bound and Jaeger (2000) and Carneiro and Heckman (2002) show an entanglement of indicators with social status.

The experiment confirms our conjecture of zero ATE by the CSL treatment for the *Higher* sub-sample, revealing a striking inconstancy of β^L between the *School* and *Higher* sub-sample (Table 2). The IV estimates are significant and similar to the OLS estimates for most of the *School* sub-sample while insignificant for the *Higher* sub-sample.[7] The only IV estimates that are significantly different from their OLS counterparts with valid instruments across the two cohorts are based on $\mathcal{L}_{SY1}$ indicators for the *School* sub-sample. The OLS estimates, while significant and positive throughout all sub-samples and cohorts, vary across sub-samples, with returns to education decreasing for those attaining 13 to 15 years of education.[8]

TABLE 2—ESTIMATION OF β VIA SUB-SAMPLING ON EDUCATIONAL ATTAINMENT

	1930-39 Born				1940-49 Born			
Years of	School		Higher		School		Higher	
schooling	≤ 12	7-12	13-15	≥ 13	≤ 12	7-12	13-15	≥ 13
				SY				
β	.0613**	.0587**	.0546**	.0935**	.0612**	.0596**	.0273**	.0641**
[95% CI][1]	[.0602 .0625]	[.0573 .0601]	[.0476 .0615]	[.0912 .0957]	[.0599 .0624]	[.0581 .0611]	[.0223 .0322]	[.0625 .0657]
β^L (with $\mathcal{L}_{SY1}$)	.0942**	.0860*	-.2577	-.0218	.1008**	.2072**	.4449	.2292
[95% CI][1]	[.0623 .1261]	[.0094 .1626]	[-.809 .2937]	[-.585 .541]	[.0710 .1305]	[.0558 .3586]	[-.101 .991]	[-.124 .583]
Sargan	.66 (0.7198)	.17 (0.9173)	2.43 (0.2963)	5.44 (0.0660)	2.07 (0.3549)	3.59 (0.1661)	8.58 (0.0137)	22.7 (0.0000)
Hausman	4.07 (0.0437)	.49 (0.4859)	1.43 (0.2325)	.17 (0.6805)	7.02 (0.0081)	4.63 (0.0315)	2.93 (0.0867)	.98 (0.3230)
β^L (with $\mathcal{L}_{SY2}$)	.1596**	.1411*	-.5833	.0831	.0836**	.0671	.3829	.0168
[95% CI][1]	[.1182 .2010]	[.0193 .2630]	[-1.76 .591]	[-.341 .5073]	[.0499 .1172]	[-.030 .164]	[-.203 .969]	[-.267 .3004]
Sargan	.21 (0.8990)	11.7 (0.0028)	3.06 (0.2171)	.56 (0.7563)	7.27 (0.0264)	4.23 (0.1209)	2.33 (0.3123)	9.23 (0.0099)
Hausman	24.6 (0.0000)	1.86 (0.1732)	1.79 (0.1807)	.002 (0.9621)	1.72 (0.1896)	.02 (0.8799)	1.75 (0.1862)	.11 (0.7419)
				AK				
β	.0565**	.0583**	.0442**	.0591**	.0701**	.0734**	.0153**	.0405**
[95% CI]	[.0551 .0578]	[.0564 .0601]	[.0371 .0515]	[.0575 .0607]	[.0686 .0716]	[.0715 .0753]	[.0105 .0201]	[.0393 .0416]
β^L	.0864**	.1107**	.0295	-.0446	.0067	-.0040	.0606	.2612**
[95% CI]	[.0357 .1372]	[.0288 .1926]	[-.309 .368]	[-.147 .058]	[-.051 .0641]	[-.078 .0696]	[-.268 .3888]	[.1580 .3644]
Sargan	24.4 (0.7088)	24.2 (0.7217)	33.4 (0.2633)	30.1 (0.4097)	60.8 (0.0005)	59.6 (0.0007)	38.0 (0.1230)	33.5 (0.2587)
Hausman	1.35 (0.2447)	1.60 (0.2054)	.01 (0.9320)	4.44 (0.0352)	4.84 (0.0279)	4.37 (0.0367)	.07 (0.7863)	27.5 (0.0000)

Notes: SY data follows model specification (1) Tables 1 and A2; AK data follows Tables V and VI columns (5) and (6) model specifications; "." indicates negative R-square in the 2SLS regression; ** denotes significance at the 1% level and * at the 5% level; 95% CI is 95 per cent confidence interval; (.) are p-values; [1] standard errors cluster adjusted. The discrepancy between OLS estimates obtained with SY and AK data sets is due to differences in the construction of the education variable; see Table 4 and discussion.

The results of the above experiments show us a severe lack of consistency in the IV estimates. Since this finding is contrary to what is asserted, we conduct a further experiment in order to find out whether consistency can be restored for the *School* sub-sample for which CSL

[7] The only exception is the later born cohort with AK's model specification where all return to education estimators are insignificant or diagnostics reveal problems with the model design.

[8] This effect is more pronounced for the later born cohort, potentially due to educational inflation. These data patterns are undetectable by the CSL-based IV method, since instruments narrowly target school goers but not those attaining higher education and large standard errors hide significant difference across cohorts.

indicators $\mathcal{L}_{SY}$ are valid. Following SY's Table 1 experiment, we stepwise enlarge the sample size by widening our focus group for the *School* sub-sample only: white males ages 40–49, white males ages 30–49, white males ages 25–54, males ages 25–54, males and females ages 25–54 (Table 3).[9] While IV estimates are relatively more invariant as sample sizes increase, there is no visible sign of convergence. In stark contrast, the OLS estimates show clear signs of convergence with increased sample size and remain remarkably invariant with regards to additional explanatory variables added in columns (4–5). A slight decrease in standard errors with an increase in sample size is discernible for IV estimates, but there is no sign of them converging to a single value. Further, Sargan and Hausman tests clearly reject the validity of the IVs for the more comprehensive samples in columns (3–5).

TABLE 3—ESTIMATION OF β WITH INCREASING SAMPLE SIZE USING SY SCHOOL SUB-SAMPLE DATA

	(1) White Male 40-49	(2) White Male 30-49	(3) White Male 25-54	(4) Male 25-54	(5) All 25-54
β	.0613**	.0613**	.0606**	.0608**	.0618**
[95% CI]	[.0602 .0625]	[.0604 .0622]	[.0599 .0613]	[.0601 .0614]	[.0613 .0624]
β^L (with $\mathcal{L}_{SY1}$)	.0970**	.0903**	.0683**	.0699**	.1126**
[95% CI]	[.0647 .1293]	[.0712 .1095]	[.0537 .0828]	[.0582 .0816]	[.1027 .1226]
Sargan	1.20 (0.5501)	0.48 (0.7856)	5.19 (0.0747)	4.03 (0.1337)	74.6 (0.0000)
Hausman	4.77 (0.0290)	8.98 (0.0027)	1.07 (0.3004)	2.37 (0.1241)	103. (0.0000)
β^L (with $\mathcal{L}_{SY2}$)	.1592**	.1064**	.1038**	.1010**	.1633**
[95% CI]	[.1177 .2007]	[.0838 .1289]	[.0866 .1210]	[.0892 .1127]	[.1529 .1737]
Sargan	0.10 (0.9497)	2.08 (0.3532)	18.2 (0.0001)	15.3 (0.0005)	77.6 (0.0000)
Hausman	24.3 (0.0000)	15.8 (0.0001)	25.0 (0.0000)	45.9 (0.0000)	409. (0.0000)
Sample Size	399,428	829,304	1,283,824	1,471,196	2,618,852

Notes: Based on SY model specifications Table 1 column (3) with race and gender added as the sample becomes more inclusive in columns 4-5; ** denotes significance at the 1% level and * at the 5% level; 95% CI is 95 per cent confidence interval, cluster adjusted standard errors; (.) are p-values.

Experimenting with the research design in SY and AK, we find, contrary to what is expected, that the OLS estimates outperform the IV estimates in terms of robustness, consistency and invariance, regardless the choice of CSL indicators. Further, the conclusion of no CSL effect on earnings via schooling by SY cannot be maintained because of the entanglement of the CSL effect with regional effects and the negligible share of law compliers in the samples.

[9] AK datasets do not provide for the experiment since other age groups and genders are excluded.

3. Model Specification of Schooling Effects under CSL treatment: An anatomy

The above empirical finding indicates the presence of serious misfit between the research design and the data sets. Let us now delve into the research design as expressed by equations (1) and (2). If we combine the two equations and take into consideration the shift in the interpretation of β from the OVB-free returns to schooling to the CSL effect via schooling, we see that the IV approach actually implies a rejection of (1) in favour of:

(3) $$y = \alpha + \beta^L s^L + \eta^L$$

where s^L denotes the fitted s from (2) and superscript L denotes the latent variable representing the CSL. The expected $\beta^L \neq \beta$ entails $s^L \not\approx s$, which leads to the shifted interpretation. Hence, the essence of the IV approach is to assert s^L, instead of s, as the valid conditional variable, see Qin (2015, 2016). This assertion is de facto rejected since the empirical evidence of the previous section goes overwhelmingly against the IV estimates in favour of their OLS counterpart.

In order to understand what has caused the rejection, let us scrutinise the model of schooling effects under the CSL treatment. Since the CSL effects on income via schooling is a sequential event, an intuitive way to represent the event is a directed acyclic graph (DAG); see the left panel in Figure 2.

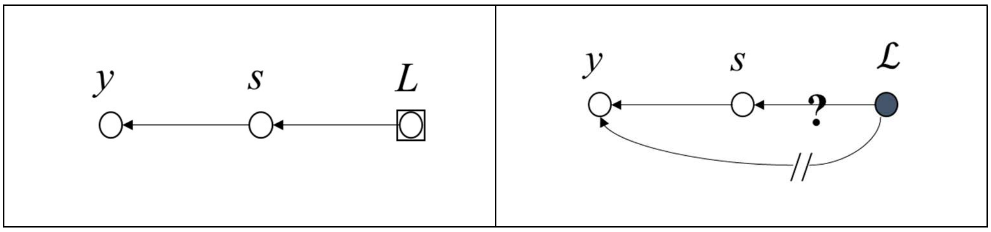

FIGURE 2. DAGS OF RETURNS TO SCHOOLING UNDER THE CSL TREATMENT

Notes: y denotes earnings, s, schooling, L, the CSL and $\mathcal{L}$ its observable indicator. A node inside a square indicates a latent variable; a solid node denotes a dummy/binary variable.

The DAG in the left panel of Figure 2 shows us that, when the effect of L forms the focal causal interest, s takes the role of an intermediate variable, or a mediator, but when s is the causal variable of interest, L takes the role of a moderator exclusive for s. Whatever the causal variable of interest, $f(y,s,L)$, the joint density of the three variables can be recursively factorised and reduced as below:

(4) $$f(y,s,L) = f(y|s,L)f(s|L)f(L) = f(y|s,L)f(s|L),$$

since $f(L) = 1$ when retrospective cross-section data samples are used. It should be noted that the CSL treatment implies a rule of intervention,[10] namely $y \perp L|s$, so that we can further factorise the conditional density in (4) as:

(5) $$f(y,s|L) = f(y|s)f(s|L).$$

The sequential nature of the ATE of L on y via s is expressed by the conditional expectation of (5), $E(y,s|L) = E(y|s)E(s|L)$. In a linear model setting, the ATE, denoted by β_{yL}, can be derived from a chain of two simple regressions corresponding to $E(y|s)$ and $E(s|L)$ respectively:[11]

(6) $$\begin{matrix} y = \alpha_y + \beta_{ys}s + \varepsilon_y \\ s = \alpha_s + \beta_{sL}L + \varepsilon_s \end{matrix} \implies \beta_{yL} = \beta_{ys}\beta_{sL}.$$

Model (6) tells us that $\beta_{yL} \neq \beta_{ys}$ holds in general unless $\beta_{sL} = 1$ can be verified, which is highly unlikely in view of the available findings, e.g. see Goldin and Katz (2011). Hence, we should expect that $\beta_{yL} \ll \beta_{ys}$.

If β_{yL} forms the only parameter of substantive interest, the chain route appears a long way round to estimate it, because the parameter can be estimated directly from:

(7) $$y = a_y + \beta_{yL}L + \epsilon_y.$$

However, in the event that indicators, $\mathcal{L}$, are used to approximate the latent L, i.e.:

(7') $$y = a_y + b_{y\mathcal{L}}\mathcal{L} + e_y,$$

[10] Refers to Rule 2 on the external intervention in (Pearl 2009, Section 3.4).
[11] See Cox and Wermuth (2004) for a general discussion on the parametric chain representation. We adopt their method of subscript-based parametric notation in order to highlight the consequence of different causal chain specifications on the parameters of regressors.

this direct route may result in $b_{y\mathcal{L}} \neq \beta_{ys}\beta_{s\mathcal{L}}$. The defectiveness of CSL indicators due to entanglement with regional factors and other controls leading to indirect effects has already been identified in the previous section. In other words, $\mathcal{L}$ may fail the rule of intervention such that $\beta_{y\mathcal{L}.s} \neq 0$ from the following regression:

(8) $$y = \alpha_y + \beta_{ys.\mathcal{L}} s + \beta_{y\mathcal{L}.s}\mathcal{L} + \varepsilon_y.$$

This situation is illustrated in the modified DAG in the right panel of Figure 2. Two consequences follow. First, the chain route of (6) is more reliable than (7') for estimating the ATE of the CSL via schooling; second, a test of $\beta_{y\mathcal{L}.s} = 0$ using (8) can be exploited as an additional criterion for CSL indicator selection purpose.

The advantage of the chain route becomes even more evident when the presence of control variables, denoted by *Z*, is taken into consideration. Although Z is chosen primarily from consideration of $cov(sZ) \neq 0$, some of the control variables are likely to correlate with CSL indicators, such as age and regional dummies in the two data sets by AK and SY. The DAGs with Z included are shown in Figure 3. The correlation between *Z* and CLS indicators complicates the estimation of the CSL effect. Extend (6) by *Z*:

(9) $$\begin{aligned} y &= \alpha_y + \beta_{ys.Z} s + Z'\boldsymbol{\beta}_{yZ.s} + \varepsilon_y \\ s &= \alpha_s + \beta_{sL} L + \varepsilon_s \\ Z &= \boldsymbol{\alpha}_Z + L\boldsymbol{\beta}_{ZL} + \boldsymbol{\varepsilon}_Z \end{aligned}$$

The corresponding chain representation of parameters becomes:

(10) $$\beta_{yL} = \beta_{ys.Z}\beta_{sL} + \boldsymbol{\beta}'_{yZ.s}\boldsymbol{\beta}_{ZL} = \beta_{yL_s} + \beta_{yL_Z}.$$

Now, only the first component, β_{yL_s}, in (10) corresponds to the ATE of the CSL via schooling. The simple regression model of (7') is no longer fit for the purpose.

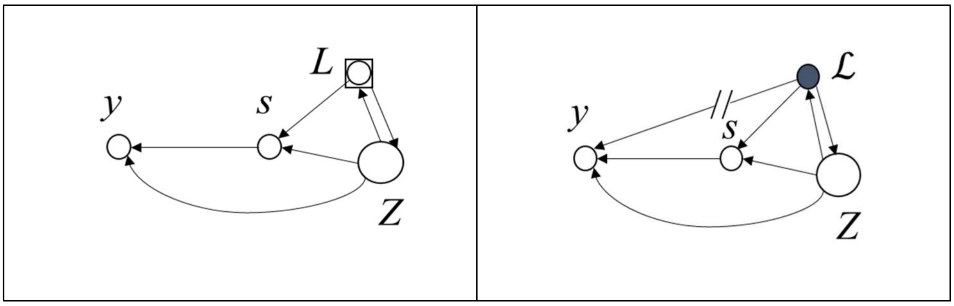

FIGURE 3. DAGS AUGMENTED WITH Z

Notes: A node inside a square indicates a latent variable; a solid node denotes a dummy/binary variable.

Model (9) differs from model (3) in two essential aspects: (i) s is taken as a valid conditional variable in the first equation of (9), and (ii) the lower equations are part of the chain representation rather than an instrumental step. The two income effects can thus be estimated separately. Moreover, the OVB-based argument underlying the IV-treatment actually applies to the CSL effect, rather than to the schooling effect, $\beta_{ys.Z}$, in (9) because of the inclusion of *Z*. In other words, β_{yL} or β_{yL_s} defines OVB or part of it if viewed from β_{yL} of (7). The measurement-errors based argument is far more relevant to CSL indicators than to s in the present context.[12] In fact, it is virtually impossible to find conclusive proof of $cov(s\varepsilon_y) \neq 0$ in (9) because the choice of this multivariate regression implies that no further information is available to downsize the error term or divide it into individual regressors. Consequently, only the argument of selection bias remains pertinent to the estimation of the schooling effect. Specifically, the CSL treatment could alter the population composition of educated workers, as compared to that of the pre-treatment population, e.g. through a diluted concentration level of 'aptitude' (see Angrist and Pischke, 2009, Chapter 4), so much so that the post-treatment schooling effect becomes significantly different from the pre-treatment one.

[12] The inapplicability of the measurement-errors based arguments in the present context can also be seen from the fact that almost no signs of expected OLS attenuations caused by measurement-error concerns can be found in AK or SY, namely that the OLS estimates should be statistically insignificant and smaller in magnitude than the IV estimates, e.g. Durbin (1954).

The IV remedy of this selection-bias induced effect is to maintain $\beta^L \neq \beta$ from the perspective of retrospective data. As pointed out earlier, the IV approach amounts to rejecting the causal validity of s and substituting it with s^L. Figure 4 illustrates the situation in two DAGs. Apart from the difficulty of lacking empirical means to identify a unique s^L (the right panel), the IV substitution confines L to an instrumental role in producing s^L and thus blocks the route to a chain model specification, as illustrated in the left panel of Figure 2.

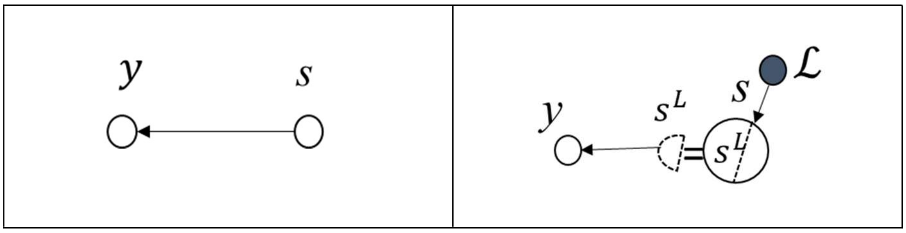

FIGURE 4. DAGS OF MODEL (1) (LEFT) VERSUS MODEL (3) (RIGHT)

Notes: Dotted lines indicate non-uniqueness; dissimilarity of s^L from s is shown by a semicircle; the 'identity' sign differentiates the first stage of the 2SLS, (2), from a chain model specification with respect to L.

Consequently, the IV approach is unsuitable for tackling the question whether the CSL treatment has indeed resulted in a compositional shift to such an extent that it has caused a parametric shift in the returns to schooling, since β of (1) is already rejected *a priori* as an inconsistent parameter. Under the chain model specification, the only feasible way to tackle this question is to carefully divide the available samples into two parts — an L-treated part versus a CSL unaffected part — so as to investigate whether there exists a parametric difference: $\beta_{ys_L.Z} \neq \beta_{y\widetilde{s_L}.Z}$, where s_L denotes schooling of the L-treated part, and $\widetilde{s_L}$ the treatment unaffected part. It should be noted that even if the inequality is supported by data, the evidence alone is insufficient for rejecting s as a valid conditional variable for y, e.g. see Engle et al. (1983). Conversely, evidence against such a parametric shift does not imply $\beta_{yL_s} = 0$ unless $\beta_{sL} = 0$ can be verified.

4. CSL Treatment in A Multiple Model Framework

Following the chain model approach outlined previously, this section starts from estimating the two outcome effects on earnings: (a) the schooling effect or returns to schooling, $\beta_{ys.Z}$, and (b) the CSL effect or ATE of the CSL via schooling, β_{yL_s}. It then turns to investigating the possible presence of a L-treatment induced parametric shift in the schooling effect, $\beta_{ys_L.Z} \neq \beta_{y\widetilde{s_L}.Z}$.

4.1. Sub Estimating the Schooling Effect: $\boldsymbol{\beta_{ys.Z}}$

The presentation of varying returns to schooling estimates, $\beta_{ys.Z}$, by AK and SY, despite the use of almost identical samples, indicates model specification problems (Table 2). Therefore, we proceed with the question of how to specify Z in order to find an empirically adequate specification of (9), which is as parsimonious as possible and also can align the returns to schooling estimates by AK and SY data respectively. This is achieved through firstly, unification of the education variable and secondly, a parsimonious model specification.

Towards a unification of the education variable, the AK education variable is capped at 17 years to resemble the SY education variable. The unification is found to play a vital role in aligning the returns to schooling estimates across the two date sets. As for the experiments reported in Table 2, we rely on AK's division between those born in the 1930s and 1940s respectively using observations from the 1980 census. Towards a more parsimonious model, year of birth dummies included by both AK and SY are replaced with quadratic age (*age2*).[13] Regional dummies for individual states are replaced by a single variable distinguishing between four regions for SY and nine regions for AK data (*region*). Considering a possible regional effect on school quality, variables capturing school quality (*pupilt, term*, *reltwage*)

[13] Coefficients on year of birth dummies are found to decline with years revealing non-linearity. These patterns can be almost perfectly replicated with a quadratic age variable. See also Murphy and Welch (1992) for the non-linear relationship between experience and wage earnings.

suggested by Card and Krueger (1992a, 1992b) are used by SY and included in our model as well.

A notable pattern of parameter inconstancy in Table 2 is the variation in the returns to schooling estimates with the level of education. A variable is thus added to capture these variations, known as 'sheepskin effects'. Sheepskin effects as well as education inflation are well documented phenomena in the literature[14] and clearly discernible in the AK and SY data; see Figure 1A and Table 1A, Appendix. This shift in the population education composition also explains the finding by Goldin and Katz (2000). In order to control for these data patterns across cohorts, a dummy (uni) turning one for those who obtained a university degree (15 or more years of schooling) is added.

The key results of this model search are reported in Table 4, alongside with those from the 'Original' models by SY and AK. We refer to our more parsimonious specifications as 'Alternative' in the table. An alignment of return to schooling estimates across datasets is achieved with the 'Alternative' model specification, which outperforms the 'Original' specification in terms of model fit. In particular, the OLS estimates point to a constant $\beta_{ys.Z}$ of roughly 0.06 across data sets and cohorts. Our returns to schooling estimates are in line with findings by Pischke and Wachter (2008) and Acemoglu and Angrist (2001), who report estimates of 0.061 and 0.075 respectively. This finding indicates that the risk of OVB for $\beta_{ys.Z}$ comes mainly from inadequately specified Z in the 'Original' models.

[14] See, for instance Angrist (1995), Murphy and Welch (1992), Trostel (2005) and Clark and Martorell (2014).

TABLE 4—PARSIMONIOUS SPECIFICATION OF (9)

	Original				Alternative			
	SY		AK		SY		AK	
	1930-1939	1940-1949	1930-1939	1940-1949	1930-1939	1940-1949	1930-1939	1940-1949
$\beta_{ys,Z}$	0.0751**	0.0622**	0.0630**	0.0519**	0.0600**	.0643**	.0576**	.0648**
[95% CI]	[.074 .077]	[.0612 .063]	[.062 .064]	[.051 .053]	[.058 .062]	[.063 .066]	[.057 .059]	[.064 .066]
AIC	714262.9	1034376	594994.7	858645.2	705271.2	1018112	594343.4	858594.8
Adj.-R2	0.0119	-0.0232	0.1745	0.1354	0.1217	0.0968	0.1761	0.1355
Z	age2 age3 age4 yob31-yob39/ yob41-yob49 sob1-sob55		ageq, ageq2, race, married, smsa, neweng midatl, enocent, wnocent, soatl, esocent, wsocent, mt, yr20-yr28		age2, mar, emp, jail, handcap, pupilt, term, reltwage, uni, region		age2, mar, race, smsa, uni, region	

Notes: 1980 census, data for SY white male with positive weekly earnings, data for AK male with positive weekly earnings; ** denotes significance at the 1% level and * at the 5% level; AIC is Akaike information criteria; 95% CI is the 95 per cent confidence interval; 95% CI based on cluster adjusted standard errors in SY data; *pupilt* is the pupil-teacher ratio, *term* is the length of the school term, and *reltwage* is the average teacher salaries (see Card and Krueger (1992b) for more information on the creation of the variables); *mar, emp, jail,* and *handcap* are dummy variables capturing marital status, employment status, prior convictions and disabilities.

4.2. Estimating the ATE of the CSL via Schooling: β_{yL_s}

Given the imperfections of CSL indicators identified in Section 2, we conduct two simple experiments to further test the appropriateness of the indicator choice before continuing with the estimation of β_{yL_s}. Since CSL is only binding for school leavers, we would expect the ATE to be insignificant or at least smaller for those with higher education than for those without. Following this reasoning, we estimate the middle equation of (9) using subsample groups by educational attainment as in Table 2, with the expectation that $\hat{\beta}_{sL} \neq 0$ for *School* and $\hat{\beta}_{sL} = 0$ for *Higher*.

It is shown in Table 5 that, although $\hat{\beta}_{sL}$ tends to be larger for the *School* subsample than for the *Higher* sub-sample, none of the indicators confirms the hypothesis of $\hat{\beta}_{sL} = 0$ for *Higher*. Noticeably, the size of those $\hat{\beta}_{sL} \neq 0$ in the first cohort has almost doubled that of the second cohort in the case of SY indicators. This shift appears to reflect a general shift towards more years of education. As seen from Table 1A (Appendix), the share of those attaining less or equal the minimum years of schooling is halved in the later cohort.

TABLE 5—β_{sL} IN (9) VIA SUB-SAMPLING ON EDUCATIONAL ATTAINMENT

1930-1939	SY								AK			
	$\mathcal{L}_{SY1}$				$\mathcal{L}_{SY2}$				$\mathcal{L}_{AK}$			
	School		Higher		School		Higher		School		Higher	
	Coef.	t-stat[1]	Coef.	t-stat[1]	Coef.	t-stat[1]	Coef.	t-stat[1]	Coef.	t-stat[2]	Coef.	t-stat[2]
$\beta_{s\mathcal{L}_1}$	0.38*	2.50	0.06	0.69	0.18	1.84	-0.10**	-4.33	-0.12**	-8.74	0.02	1.17
$\beta_{s\mathcal{L}_2}$	0.35*	2.49	0.07	0.91	-0.02	-0.23	-0.05*	-1.98	-0.11**	-8.43	0.05**	3.10
$\beta_{s\mathcal{L}_3}$	0.18	1.25	0.06	0.75	0.39**	3.99	-0.12**	-4.27	-0.03*	-2.39	-0.00	-0.04
1940-1949	$\mathcal{L}_{SY1}$				$\mathcal{L}_{SY2}$				$\mathcal{L}_{AK}$			
	School		Higher		School		Higher		School		Higher	
	Coef.	t-stat[1]	Coef.	t-stat[1]	Coef.	t-stat[1]	Coef.	t-stat[1]	Coef.	t-stat[2]	Coef.	t-stat[2]
$\beta_{s\mathcal{L}_1}$	0.70**	11.34	0.13**	3.47	0.24**	4.59	-0.05	-1.57	-0.10**	-9.89	0.04**	3.71
$\beta_{s\mathcal{L}_2}$	0.69**	10.61	0.25**	7.63	-0.03	-0.57	0.03	1.00	-0.08**	-7.82	0.06**	5.09
$\beta_{s\mathcal{L}_3}$	0.42**	6.25	0.19**	5.75	0.21**	3.40	-0.06*	-2.11	-0.02*	-2.45	0.03*	2.25

Notes: 1980 census, data for SY white male with positive weekly earnings, data for AK male with positive weekly earnings; [1] robust cluster adjusted standard errors; [2] robust standard errors; ** denotes significance at the 1% level and * at the 5% level.

In a second step, we test whether the rule of intervention $\beta_{y\mathcal{L}.Zs} = 0$ holds for the different CSL indicators by estimation of (8) with additional controls *Z*. In reference to earlier experiments, we conduct the test for the *School* sub-sample in addition to the full sample estimation. It is shown in Table 6 that the condition $\beta_{y\mathcal{L}.Zs} = 0$ is validated for SY's $\mathcal{L}_{SY1}$ indicator across cohorts and also for AK's $\mathcal{L}_{AK}$ indicator for the early born cohort. But it is violated without exception if using $\mathcal{L}_{SY2}$ as CSL indicator. Where conditional independence is rejected in Table 6, we have also failed to confirm $\hat{\beta}_{sL} = 0$ for the *Higher* sub-sample (see Table 5) and rejected instruments as invalid (see Table 2). In cases like this, we should be cautious with the estimate of β_{yL_s} via the chain representation of (10).

TABLE 6—TEST FOR THE RULE OF INTERVENTION $\beta_{y\mathcal{L}.Zs} = 0$ USING (8) EXTENDED BY Z

	SY								**AK**			
	1930-39				**1940-49**				**1930-39**		**1940-49**	
	Full		**School**		**Full**		**School**		**Full**	**School**	**Full**	**School**
	$\mathcal{L}_{SY1}$	$\mathcal{L}_{SY2}$	$\mathcal{L}_{SY1}$	$\mathcal{L}_{SY2}$	$\mathcal{L}_{SY1}$	$\mathcal{L}_{SY2}$	$\mathcal{L}_{SY1}$	$\mathcal{L}_{SY2}$	$\mathcal{L}_{AK}$		$\mathcal{L}_{AK}$	
$\mathcal{L}1$	-.001	.041**	.007	.041**	.014	.041**	.007	.046**	-.007*	-.008	-.012**	-.005
$\mathcal{L}2$	.011	.028**	.029	.030**	.016	.047**	.020	.052**	-.004	-.009*	.012**	.010**
$\mathcal{L}3$	.021	.060**	.033	.063**	.033**	.078**	.031*	.082**	.001	-.003	.012**	.017**

Notes: 1980 census, data for SY white male with positive weekly earnings, data for AK male with positive weekly earnings; ** denotes significance at the 1% level and * at the 5% level.; robust cluster adjusted (for SY) standard errors are used; Z as specified in 'Alternative' in Table 4.

Table 7 provides β_{yL_s} estimated via (10). Where conditional independence was verified, the chain approximation yields significant ATE estimates that confirm our expectation of a larger effect of schooling on earnings than that of the CSL, i.e. $\beta_{ys.Z} \gg \beta_{yL_s}$. Direct ATE estimates $\beta_{y\mathcal{L}}$ obtained via (7') exceed estimates obtained via chain approximation for the later born cohort (see Table 2A, Appendix). The effect is indicative of positive indirect CSL effects through control variables Z in later years. Further, chain approximations using SY indicators are much more varied across cohorts than across sub-samples, due mainly to the varying estimates of β_{sL} in Table 5. The estimated ATE almost doubles for the later born cohort from 1–3 to 3–5 per cent using $\mathcal{L}_{\mathrm{SY1}}$ indicators. The ATE estimates using AK indicators are relatively constant across both sub-samples and cohorts. It should be noted that the negative sign here actually implies a positive ATE because people born in the first three quarters $\mathcal{L}_{AK}^{1}$, $\mathcal{L}_{AK}^{2}$ and $\mathcal{L}_{AK}^{3}$ are associated with less years of schooling as compared to those born in the fourth quarter. The CSL effect is strongest for those born in the first quarter and weakens with the second and third quarter born consecutively.

TABLE 7—ESTIMATED ATE OF CSL, $\boldsymbol{\beta_{y\mathcal{L}_s}}$, USING CHAIN MODEL (9)

	SY								AK			
	1930-39				1940-49				1930-39		1940-49	
	Full		School		Full		School		Full	School	Full	School
	$\mathcal{L}_{SY1}$	$\mathcal{L}_{SY2}$	$\mathcal{L}_{SY1}$	$\mathcal{L}_{SY2}$	$\mathcal{L}_{SY1}$	$\mathcal{L}_{SY2}$	$\mathcal{L}_{SY1}$	$\mathcal{L}_{SY2}$	$\mathcal{L}_{AK}$		$\mathcal{L}_{AK}$	
$\mathcal{L}1$	.025**	.011	.022*	.011*	.054**	.016*	.048**	.016**	-.009**	-.007**	-.007**	-.007**
$\mathcal{L}2$	.005	-.003	.021*	-.001	.033**	-.004	.045**	-.002	-.006**	-.006**	-.004**	-.005**
$\mathcal{L}3$	.004	.012	.011	.023**	.027**	.002	.028**	.014**	-.002*	-.002*	-.003**	-.002**

Notes: See Tables 4 and 5 for $\beta_{ys.Z}$ and β_{sL} estimates respectively. 1980 census, data for SY white male with positive weekly earnings, data for AK male with positive weekly earnings; SY adjusted for clustering of observation and robust standard errors are used; ** denotes significance at the 1% level and * at the 5% level. The p-values for $\beta_{y\mathcal{L}_s}$ are based on X^2 statistics estimated following Weesie (1999).

Our ATE estimates are in rough agreement with findings reported in the literature. For instance, Pischke and Wachter (2008) report the ATE of CSL in Germany to be between 0.012 and 0.017 for different datasets. Results presented by Oreopoulos (2006b) for Canada suggest a larger ATE between 0.031 and 0.107, which is similar to our estimates using $\mathcal{L}_{SY1}$. Acemoglu and Angrist (2001) report the ATE of CSL to be between 0.008 and 0.009. A moderate positive impact of CSL on schooling was further confirmed by Lleras-Muney (2002), Oreopoulos (2006a), Goldin and Katz (2011).

4.3. Testing for a CSL Induced Shift in the Schooling Effect: $\boldsymbol{\beta_{ys_L.Z} \neq \beta_{y\tilde{s}_L.Z}}$

If the introduction of the CSL has altered the ability composition of workers, the post treatment schooling effect, $\beta_{ys_L.Z}$, might differ from the schooling effect before treatment,

$\beta_{y\tilde{s}_L.Z}$, namely $\beta_{ys_L.Z} \neq \beta_{y\tilde{s}_L.Z}$. To investigate if such a parameter shift has occurred, break-point Chow tests are conducted. The test requires careful sub-sample division conditional on the L-treatment and also different choices of CSL indicators.

For the AK law indicators, the treated group is defined as those born in the first and second quarters of the year, while the untreated group is defined as those born in the remaining quarters. Since the CSL is binding for only a minority of the treated group — as evident from the negligible share of compliers (Table 1A, Appendix)[15]— the ability composition is unlikely to render a significant parameter shift using the full sample. Sub-sample division minimises defectors and always takers in the treated group of the *School* sub-sample. A parametric shift should hence be discernible for the *School* but not the *Higher* sub-sample. As can be seen from Table 8,[16] despite the carful sub-sample division, the null hypothesis of no break point cannot be rejected at the 5 per cent level in all experimental settings.

TABLE 8—TEST FOR $\mathcal{L}_{AK}$ INDUCED PARAMETRIC SHIFT

	1930-1939			1940-1949		
	Full	**School (7-12)**	**Higher (13-15)**	**Full**	**School (7-12)**	**Higher (13-15)**
Chow Test (Treated-Full)[1]	0.65	0.72	0.19	0.76	0.25	0.18
	(0.4184)	(0.3949)	(0.6652)	(0.3842)	(0.6137)	(0.6673)
Chow Test (Untreated-Full)[2]	0.56	0.63	0.19	1.05	0.17	0.18
	(0.4549)	(0.4286)	(0.6617)	(0.3053)	(0.6769)	(0.6673)

Notes: 1980 census, male with positive weekly earnings, p-values in (.). [1] The treated sub-sample is defined as those born in the first and second quarters. [2] Untreated sub-sample is defined as those born in the third and fourth quarters.

For the SY indicators the sub-sample division is slightly more complicated. $\mathcal{L}_{SY}$ indicators capture the minimum years of schooling and school attendance required by the respective state's labour and education law, but only few states had no law in place over the sample periods, which results in a rather small sub-sample for the untreated. Further, the nature of the CSL indicators enables us to minimise defectors and always takers even further in the treated group of the *School* sub-sample. The treated group is re-defined as those among the *School* sub-sample who received treatment under a particular law and leave school right after

[15] Although Table 1A (Appendix) is based on SY data, the patterns are likely to be identical for the AK data.
[16] See Table 3A (Appendix) for full results.

the minimum years of schooling required by the law are completed. The sub-sample is denominated *School Binding*. Although the treated group does not fully capture those compliant to the law, the group fully encompasses those compliant while minimising the number of those defecting in the group under the given information.

As can be seen from Table 9,[17] using $\mathcal{L}_{SY1}$ indicators, the break-point Chow tests provide no evidence for a treatment induced parametric shift in the returns to education parameter at the 1 per cent significance level and some evidence at 5 per cent significance level for the *School Binding* sub-sample. Using $\mathcal{L}_{SY2}$ as indicator, there is some evidence for a parametric effect for the later born cohort. The effect is absent from the *Higher* sub-sample as expected, but detectable at the 5 per cent level for the *Full* sample and *School* sub-sample. However, given measurement errors identified for the $\mathcal{L}_{SY2}$ indicators, the evidence is too weak to conclude on a parametric shift.

TABLE 9—TEST FOR $\mathcal{L}_{SY}$ INDUCED PARAMETRIC SHIFT IN $\beta_{yS_L.Z}$

	1930-1939							
	Full		**School (7-12y)**		**Higher (13-15y)**		**School Binding[3]**	
	$\mathcal{L}_{SY1}$	$\mathcal{L}_{SY2}$	$\mathcal{L}_{SY1}$	$\mathcal{L}_{SY2}$	$\mathcal{L}_{SY1}$	$\mathcal{L}_{SY2}$	$\mathcal{L}_{SY1}$	$\mathcal{L}_{SY2}$
Chow Test	0.01	1.80	0.01	0.06	0.46	0.09	1.26	2.59
(Treated-Full)[1]	(0.9172)	(0.1797)	(0.9236)	(0.8056)	(0.4990)	(0.7693)	(0.2607)	(0.1076)
Chow Test	0.02	0.37	0.03	0.10	0.38	0.20	4.45*	0.19
(Untreated-Full)[2]	(0.8848)	(0.5417)	(0.8574)	(0.7556)	(0.5395)	(0.6569)	(0.0350)	(0.6596)
	1940-1949							
	Full		**School (7-12y)**		**Higher (13-15y)**		**School Binding[3]**	
	$\mathcal{L}_{SY1}$	$\mathcal{L}_{SY2}$	$\mathcal{L}_{SY1}$	$\mathcal{L}_{SY2}$	$\mathcal{L}_{SY1}$	$\mathcal{L}_{SY2}$	$\mathcal{L}_{SY1}$	$\mathcal{L}_{SY2}$
Chow Test	2.24	2.28	1.49	1.96	0.02	0.00	0.00	14.05**
(Treated-Full)[1]	(0.1342)	(0.1312)	(0.2220)	(0.1617)	(0.8847)	(0.9896)	(0.9580)	(0.0002)
Chow Test	5.04	4.23*	3.25	6.06*	0.05	0.02	0.26	2.41
(Untreated-Full)[2]	(0.0248)	(0.0396)	(0.0713)	(0.0138)	(0.8195)	(0.8966)	(0.6115)	(0.1206)

Notes: 1980 census, white male with positive weekly earnings; p-values in (.). [1] The treated sub-sample is defined as those born in a state with some school law in place, that is, minimum years of schooling unequal zero. [2] Untreated sub-sample is defined as those born in a state with no school law in place, that is, minimum years of schooling equal to zero. [3] Definition of treated and untreated change for this experiment. The treated sub-sample is defined as those who drop out after the minimum years of schooling and the untreated sub-sample comprises of the remaining observations.

Overall, the evidence for the presence of an L-treatment induced parametric shift is weak and rattled with deficiencies in the CSL indicators. If the substantive interest is with a data-consistent estimate of the returns to schooling parameter, data patterns originating from

[17] See Table 4A (Appendix) for full results.

sheepskin effects and educational inflation are found to be of far greater concern than an L-treatment induced parametric shifts (see Tables 2 and 4).

5. What have we learnt?

Angrist and Pischke (2015), in their analysis of the failed research design in Acemoglu and Amgrist (2001), ascribe the failure to inappropriate CSL indicators, while maintaining the IV approach as appropriate. In contrast, this paper elucidates that the failure is not with the choice of instruments, but the choice and conceptualisation of the IV approach. A re-assessment of the approach in Section 3 leads us to identify several cognitive flaws. The IV treatment is originally motivated by the desire for consistent estimation in the (potential) presence of OVB, measurement error and/or self-selection bias linked to innate ability when using observational data. However, a careful reconsideration of the causal claims via DAGs, shows that the relevance of these concerns crucially depends on the parameter of interest associated with the key conditional variable.

If the parameter of interest is the ATE of the CSL via schooling, concerns regarding selection bias and measurement error in the schooling variable are irrelevant, since s merely acts as a mediator for L. As evident from Tables 5–7, measurement error in CSL indicators is actually a major concern. CSL indicators fail to consistently target school goers, fail the rule of intervention and result in non-robust ATE estimates. These weaknesses in the CSL indicators call for a more careful indicator selection if the key parameter of interest is the law treatment effect.

If the parameter of interest is returns to schooling, L acts as an exclusive moderator for s. The parameter would not be affected by L unless the moderating effect induces a parametric shift. We find little evidence of such a shift using the CSL indicators of AK and SY (Tables 8 and 9), despite the shift being implicitly asserted in the IV estimation and apparently verified by SY (Table 2 and Figure 1). However, the evidence by SY is shaky as we have shown how empirically inconsistent their IV estimates are (Section 2). The apparent parametric shift in SY can be explained by the fact that the CSL dummy based IV estimation effectively implies reweighting the samples towards those with primary and secondary school education, leading to a substantial loss of sample information (best seen from Figure 4). The reweighting is justified by the need for post-treatment quasi-randomisation due to selection bias. However, it would be a serious misunderstanding to associate both concepts – randomisation and selection bias – directly with the data samples here since census data, used by both AK and SY, are randomly collected by design.

It is shown in Section 3 that the essence of the IV approach is to reject s as a valid conditional variable. This rejection is refuted by cross validation in Section 2 and by the empirical evidence presented in Section 4. The implicit alteration of model (1) to model (3) by the rejection results in a conflation of the underlying causal relationships, leading to interpretational difficulties of the IV estimates. This conceptually ambiguous interpretation of the IV estimates highlights the importance of a clear translation of causal relations into a well-designed statistical model, which allows for an assessment of the circumstances under which the conditional variable of interest is valid.

Using DAGs, we reinstate clearly defined causal relationships by explicit model design. We show that, if the causal effect of the key conditional variable is compounded by another variable – be it a moderator or a mediator – a multivariate model specification is essential in order to verify the conditional variable of interest. Further, a clear differentiation between moderator and mediator effects is imperative. In a multivariate model setting concerns arising over measurement error, selection bias and OVB need to be carefully assessed with regards to individual parameters. For instance, in the case analysed in this paper, our model design reveals that the greater risk of OVB with respect to the returns to schooling estimate comes from inadequate specification of Z rather than the possible omission of L, opposite to what underlies the IV-based research design.

Our results demonstrate that the IV method is a detour into deadlock for the estimation of both the returns to schooling and the ATE of the CSL. Empirical evidence overwhelmingly supports the rehabilitation of s as the valid conditional variable in a multivariate model setting. Our causally explicit model design allows for a careful analysis of the model setting within which relatively data-coherent inference can be made about the conditional status of s. The investigation tells us that a successful research design has to be focused on appropriate model design, i.e. a faithful translation of causal postulates into data-consistent statistical models, rather than on the choice of estimators.

References

Acemoglu, Daron, and Joshua D. Angrist. 2001. "How Large are Human Externalities? Evidence from Compulsory Schooling Laws." *NBER Macroeconomics Annual, 15* 9-74.

Angrist, Joshua D. 1995. "The Economic Returns to Schooling in the West Bank and Gaza Strip." *The American Economic Review, 85 (5)* 1065-1087.

Angrist, Joshua D., and Alan B. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics, 106 (4)* 979-1014.

Angrist, Joshua D., and Joern-Steffen Pischke. 2015. *Mastering 'Metrics: The Path from Cause to Effect.* Princeton, New Jersey and Woodstock, Oxfordshire: Princeton University Press.

—. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton, New Jersey and Woodstock, Oxfordshire: Princeton University Press.

Bound, John, and David A. Jaeger. 2000. “Do Compulsory School Attendance Laws Alone Explain the Association between Quarter of Birth and Earnings?” *Worker Well-Being, 19* 83-108.

Card, David, and Alan B. Krueger. 1992a. “Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States.” *Journal of Political Economy, 100 (1)* 1-40.

Card, David, and Alan B. Krueger. 1992b. “School Quality and Black-White Relative Earnings: A Direct Assessment.” *Quarterly Journal of Economics, 107 (1)* 151-200.

Carneiro, Pedro, and James J. Heckman. 2002. “The Evidence on Credit Constraints in Post-Secondary Schooling.” *The Economic Journal, 112* 989-1018.

Clark, Damon, and Paco Martorell. 2014. “The Signaling Value of a High School Diploma.” *Journal of Political Economy, 122 (2)* 282-318.

Cox, D.R., and Nanny Wermuth. 2004. “Causality: a Statistical View.” *International Statistical Review, 72 (3)* 285–305.

Durbin, J. 1954. “Errors in Variables.” *eview of the International Statistical Institute, 22 (1)* 23-32.

Engle, Robert F., David F. Hendry, and Jean-Francois Richard. 1983. “Exogeneity.” *Econometrica, 51 (2)* 277-304.

Goldin, Claudia. 1998. “America's graduation from high school: The evolution and spread of secondary schooling in the twentieth century.” *Journal of Economic History, 58 (2)* 345-374.

Goldin, Claudia, and Lawrence F. Katz. 2000. “Education and Income in the Early 20th Century: Evidence from the Prairies.” *The Journal of Economic History, 60 (3)* 782-818.

Goldin, Claudia, and Lawrence F. Katz. 2007. "Long-Run Changes in the U.S. Wage Structure: Narrowing, Widening, Polarizing." *NBER Working Paper Series, No. 13568.*

Goldin, Claudia, and Lawrence F. Katz. 2011. "Mass Secondary Schooling and the State: The Role of State Compulsion in the High School Movement." In *Understanding Long Run Economic Growth*, by D. Costa and N. Lamoreaux, 275-310. Chicago: University of Chicago Press.

Harmon, Colm, Hessel Oosterbeek, and Ian Walker. 2003. "The Returns to Education: Microeconomics." *Journal of Economic Surveys, 17 (2)* 115-155.

Hoogerheide, Lennart, and Herman K. van Dijk. 2006. "A Reconsideration of the Angrist-Krueger Analysis on Returns to Education." *Econometric Institute report EI 2006-15.*

Lleras-Muney, Adriana. 2002. "Were Compulsory Attendance and Child Labor Laws Effective? An Analysis from 1915 to 1939." *The Journal of Law and Economics, 45 (2).*

Ludwig, Jens, Greg J. Duncan, Lisa A. Gennetian, Lawrence F. Katz, Ronald C. Kessler, Jeffrey R. Kling, and Lisa Sanbonmatsu. 2013. "Long-Term Neighborhood Effect on Low-Income Families: Evidence from Moving to Opportunity." *American Economic Review, 103 (3)* 226-231.

Ludwig, Jens, Greg J. Duncan, Lisa A. Gennetian, Lawrence F. Katz, Ronald C. Kessler, Jeffrey R. Kling, and Lisa Sanbonmatsu. 2012. "Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults." *Science, 337 (6101)* 1505-1510.

Murphy, Kevin M., and Finis Welch. 1992. "The Structure of Wages." *The Quarterly Journal of Economics, 107 (1)* 285-326.

Oreopoulos, Philip. 2006a. "Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter." *American Economic Review, 96 (1)* 152-175.

Oreopoulos, Philip. 2006b. “The Compelling Effects of Compulsory Schooling: Evidence from Canada.” *The Canadian Journal of Economics, 39 (1)* 22-52.

Pearl, Judea. 2009. *Causality: Models, Reasoning, and Inference (2nd edition).* Cambridge, US: Cambridge University Press.

Pischke, Jörn-Steffen, and Till von Wachter. 2008. “Zero Returns to Compulsory Schooling in Germany: Evidence and Interpretation.” *The Review of Economics and Statistics, 90 (3)* 592-598.

Qin, Duo. 2016. “Let’s Take the Bias Out of Econometrics.” *SOAS, University of London: mimeo.*

Qin, Duo. 2015. “Resurgence of the Endogeneity-Backed Instrumental Variable Methods.” *Economics: The Open-Access, Open-Assessment E-Journal, 9 (7)* 1-35.

Stephens Jr., Melvin, and Dou-Yan Yang. 2014. “Compulsory Education and the Benefits of Schooling.” *American Economic Review, 104 (6)* 1777-1792.

Trostel, Philip A. 2005. “Nonlinearity in the Return to Education.” *Journal of Applied Economics, 8 (1)* 191-202.

Weesie, Jeroen. 1999. “Seemingly Unrelated Estimation And The Cluster-adjusted Sandwich Estimator.” *Stata Technical Bulletin, 52, 34–47. Reprinted in Stata Technical Bulletin Reprints, 9* 231–248.

Acknowledgments

This research was partly funded by an internal research fund from the Faculty of Law and Social Sciences, SOAS, University of London.

Appendix

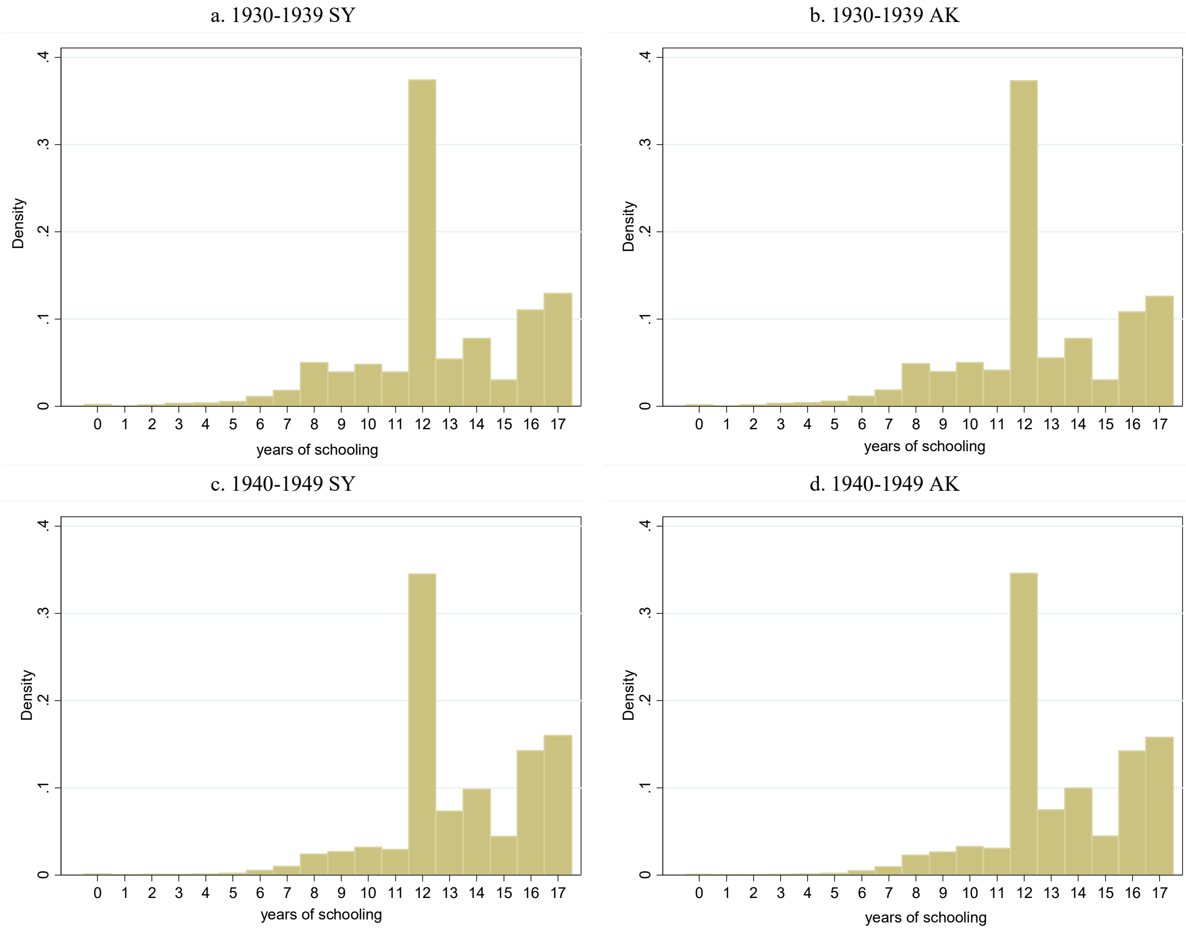

FIGURE 1A. YEARS FOR SCHOOLING DENSITY

Notes: AK education variable is capped at 17 years of schooling for comparability between the AK and SY datasets.

TABLE 1A—COMPOSITION OF CSL COMPLIERS, DEFECTORS AND ALWAYS TAKERS FOR $\mathcal{L}_{SY}$

Composition of CSL Compliers for SY $\mathcal{L}_{SY1}$						
	$\mathcal{L}_{SY1}^1 = 1$	$\mathcal{L}_{SY1}^2 = 1$	$\mathcal{L}_{SY1}^3 = 1$	Total	Years of Schooling	Untreated
1930-1939						
Equal	1.37%	5.48%	4.10%	4.11%	$\mathcal{L}_{SY1}^1$: <7	3.75%
Less	2.09%	4.00%	10.19%	6.97%	$\mathcal{L}_{SY1}^2$: <8	5.88%
More	96.45%	90.57%	85.71%	88.50%	$\mathcal{L}_{SY1}^3$: <9	12.28%
N	54,992	116,112	193,730	366,381	N (% of total)	1,547 (0.42%)
1940-1949						
Equal	0.45%	2.23%	2.90%	2.31%	$\mathcal{L}_{SY1}^1$: <7	3.20%
Less	0.69%	1.55%	5.40%	3.75%	$\mathcal{L}_{SY1}^2$: <8	5.28%
More	98.86%	96.22%	91.70%	91.71%	$\mathcal{L}_{SY1}^3$: <9	8.87%
N	86,252	112,481	337,979	548,870	N (% of total)	12,158 (2.22%)
Composition of CSL Compliers for SY $\mathcal{L}_{SY2}$						
	$\mathcal{L}_{SY2}^1 = 1$	$\mathcal{L}_{SY2}^2 = 1$	$\mathcal{L}_{SY2}^3 = 1$	Total	Years of Schooling	Untreated
1930-1939						
Equal	5.23%	4.11%	5.03%	4.18%	$\mathcal{L}_{SY2}^1$: <8	5.09%
Less	3.65%	10.67%	10.52%	7.44%	$\mathcal{L}_{SY2}^2$: <9	10.79%
More	91.20%	85.22%	84.45%	88.38%	$\mathcal{L}_{SY2}^3$: <10	14.73%
N	116,797	179,705	36,489	366,381	N (% of total)	33,390 (9.11%)
1940-1949						
Equal	1.79%	3.00%	3.16%	2.53%	$\mathcal{L}_{SY2}^1$: <8	2.53%
Less	1.38%	5.64%	6.09%	4.30%	$\mathcal{L}_{SY2}^2$: <9	5.41%
More	96.83%	91.36%	90.74%	93.17%	$\mathcal{L}_{SY2}^3$: <10	8.37%
N	131,875	297,498	82,481	548,870	N (% of total)	37,016 (6.74%)

Notes: 'N' is sample size, 'equal' is share of those with years of education equal to school law, and 'less' years of education and 'more' years of education respectively for those treated by the respective law. For the untreated group, share of those with less than 7, 8, and 9 years of education among untreated is given. 'Total' compares 'Untreated' against total of the sample in the equal to the law, less than the law and more than the law of schooling categories.

TABLE 2A—ESTIMATED ATE OF THE CSL $\beta_{y\mathcal{L}}$ via (7')

	SY								AK			
	1930-39				1940-49				1930-39		1940-49	
	Full		School		Full		School		Full	School	Full	School
	$\mathcal{L}_{SY1}$	$\mathcal{L}_{SY2}$	$\mathcal{L}_{SY1}$	$\mathcal{L}_{SY2}$	$\mathcal{L}_{SY1}$	$\mathcal{L}_{SY2}$	$\mathcal{L}_{SY1}$	$\mathcal{L}_{SY2}$	$\mathcal{L}_{AK}$		$\mathcal{L}_{AK}$	
$\boldsymbol{\beta_{y\mathcal{L}}}$												
$\mathcal{L}1$	.033	.037**	.014	.053**	.127**	.077**	.121**	.095**	-.014**	-.012**	.008**	.005
$\mathcal{L}2$	-.010	.022	-.0004	.024	.114**	.050*	.119**	.060**	-.009**	-.016**	.009**	.006
$\mathcal{L}3$	.014	.071**	.006	.106**	.110**	.095**	.105**	.125**	.0005	-.005	.010**	.016**

Notes: 1980 census, data for SY white male with positive weekly earnings, data for AK male with positive weekly earnings; ** denotes significance at the 1% level and * at the 5% level; robust cluster adjusted (for SY) standard errors are used; estimation of (7').

TABLE 3A—TEST FOR $\mathcal{L}_{AK}$ INDUCED PARAMETRIC SHIFT

	1930-1939			1940-1949		
	Full	School(7-12)	Higher(13-15)	Full	School(7-12)	Higher(13-15)
Full (t-stat)	.0576** (104.)	.0571** (63.2)	.0405** (7.19)	.0648** (116.)	.0723** (76.2)	.0265** (6.73)
N	329,311	188,486	53,883	486,707	228,087	106,941
Treated[1] (t-stat)	.0571** (73.3)	.0579** (45.4)	.0381** (4.87)	.0641** (81.9)	.0727** (54.1)	.0250** (4.60)
N	161,711	95,303	28,075	254,384	118,162	56,884
Untreated[2] (t-stat)	.0581** (74.0)	.0563** (44.0)	.0430** (5.32)	.0655** (82.7)	.0717** (53.6)	.0283** (4.94)
N	167,600	93,183	25,808	232,323	109,925	50,057
Chow Test (Treated-Full)	0.65 [0.4184]	0.72 [0.3949]	0.19 [0.6652]	0.76 [0.3842]	0.25 [0.6137]	0.18 [0.6673]
Chow Test (Untreated-Full)	0.56 [0.4549]	0.63 [0.4286]	0.19 [0.6617]	1.05 [0.3053]	0.17 [0.6769]	0.18 [0.6673]

Notes: 1980 census, male with positive weekly earnings, p-values in (.). [1] The treated sub-sample is defined as those born in the first and second quarters. [2] Untreated sub-sample is defined as those born in the third and fourth quarters.

TABLE 4A—TEST FOR $\mathcal{L}_{SY}$ INDUCED PARAMETRIC SHIFT

TEST FOR $\mathcal{L}_{SY1}$ INDICATOR

	Full	School (7-12)	Higher (13-15)	School Binding[3]
1930-1939				
Full (t-stat)	.0607** (109.)	.0616** (68.2)	.0463** (12.9)	.0606** (67.2)
N	366,381	208,599	59,540	208,599
Treated[1] (t-stat)	.0607** (109.)	.0616** (68.1)	.0461** (12.9)	.0690** (8.89)
N	364,834	207,750	59,256	15,261
Untreated[2] (t-stat)	.0619** (8.87)	.0641** (5.10)	.0730 (1.63)	.0617** (61.0)
N	1,547	849	284	193,338
Chow Test (Treated-Full)	0.01 [0.9172]	0.01 [0.9236]	0.46 [0.4990]	1.26 [0.2607]
Chow Test (Untreated-Full)	0.02 [0.8848]	0.03 [0.8574]	0.38 [0.5395]	4.45 [0.0350]
1940-1949				
Full (t-stat)	.0643** (120.)	.0739** (80.5)	.0199** (8.40)	.0740** (80.7)
N	548,870	256,848	118,714	256,848
Treated[1] (t-stat)	.0644** (118.)	.0741** (79.4)	.0198** (8.28)	.0744** (9.26)
N	536,712	250,973	115,894	13,220
Untreated[2] (t-stat)	.0575** (19.9)	.0645** (12.6)	.0235 (1.51)	.0742** (73.2)
N	12,158	5,875	2,820	242,663
Chow Test (Treated-Full)	2.24 [0.1342]	1.49 [0.2220]	0.02 [0.8847]	0.00 [0.9580]
Chow Test (Untreated-Full)	5.04 [0.0248]	3.25 [0.0713]	0.05 [0.8195]	0.26 [0.6115]

TEST FOR $\mathcal{L}_{SY2}$ INDICATOR

	Full	School (7-12)	Higher (13-15)	School Binding[3]
1930-1939				
Full (t-stat)	.0607** (108.)	.0616** (68.2)	.0463** (12.9)	.0606** (64.2)
N	366,381	208,599	59,540	208,599
Treated[1] (t-stat)	.0604** (102.)	.0615** (64.5)	.0460** (12.2)	.0653** (21.0)
N	332,991	189,437	54,563	22,605
Untreated[2] (t-stat)	.0619** (35.4)	.0607** (22.0)	.0514** (4.38)	.0609** (60.2)
N	33,390	19,162	4,977	185,994
Chow Test (Treated-Full)	1.80 [0.1797]	0.06 [0.8056]	0.09 [0.7693]	2.59 [0.1076]
Chow Test (Untreated-Full)	0.37 [0.5417]	0.10 [0.7556]	0.20 [0.6569]	0.19 [0.6596]
1940-1949				
Full (t-stat)	.0650** (121.)	.0753** (82.0)	.0203** (8.56)	.0739** (80.5)
N	548,870	256,848	118,714	256,848
Treated[1] (t-stat)	.0653** (117.)	.0757** (79.2)	.0203** (8.26)	.0857** (27.5)
N	511,854	239,545	110,763	25,964
Untreated[2] (t-stat)	.0603** (30.9)	.0670** (20.3)	.0215* (2.36)	.0731** (71.0)
N	37,016	17,303	7,951	227,348
Chow Test (Treated-Full)	2.28 [0.1312]	1.96 [0.1617]	0.00 [0.9896]	14.05** [0.0002]
Chow Test (Untreated-Full)	4.23* [0.0396]	6.06* [0.0138]	0.02 [0.8966]	2.41 [0.1206]

Notes: 1980 census, white male with positive weekly earnings; p-values in [.]. [1] The treated sub-sample is defined as those born in a state with some school law in place, that is, minimum years of schooling unequal zero. [2] Untreated sub-sample is defined as those born in a state with no school law in place, that is, minimum years of schooling equal to zero. [3] Definition of treated and untreated change for this experiment. The treated sub-sample is defined as those who drop out after the minimum years of schooling and the untreated sub-sample comprises of the remaining observations.