

WORKING PAPERS SES

Testing the Validity of the
Compulsory Schooling Law
Instrument

Benjamin Bolzern
and
Martin Huber

**N. 480
III.2017**

Testing the Validity of the Compulsory Schooling Law Instrument

Benjamin Bolzern* and Martin Huber**

*bolzern@gmail.com, University of Bern

**martin.huber@unifr.ch, University of Fribourg

Abstract: Changes in compulsory schooling laws have been proposed as an instrument for the endogenous choice of schooling. It has been argued that raising minimum schooling exogenously increases the educational attainment of a subset of pupils without directly affecting later life outcomes such as income or health. Using the method of Huber and Mellace (2015) and data from the Survey of Health, Ageing and Retirement in Europe, we jointly test random instrument assignment, weak monotonicity of education in the instrument, and the instrument exclusion restriction. The satisfaction of these restrictions permits identifying the local average treatment effect of education on those choosing more schooling as a reaction to the law change. Our results do not point to the invalidity of the schooling law instrument, though we acknowledge that even asymptotically, testing cannot detect all possible violations of instrument validity.

Keywords: instrumental variable, schooling laws, schooling reforms, treatment effects, LATE, tests.

JEL classification: C26, I12, I21, I28, J24.

Corresponding author: Martin Huber, University of Fribourg, Bd. de Pérolles 90, 1700 Fribourg, Switzerland;
martin.huber@unifr.ch.

1 Introduction

To evaluate the effects of education on later life outcomes, a vast literature aims at tackling the potential endogeneity of schooling decisions by instrumental variables (IV), see for example Angrist and Krueger (1991). A valid IV causes exogenous variation in education without influencing the outcome other than through education. Harmon and Walker (1995), Oreopoulos (2006), and others exploit changes in compulsory schooling laws (CSL) as an arguably valid IV, as they imply a shift in minimum years of school attendance that seems orthogonal to the characteristics of the individuals targeted by the policy. Analysis typically compares the schooling decisions and outcomes of adjacent cohorts that are either affected or unaffected by changes in CSL. However, cohort specific effects on the outcomes of interest may jeopardize IV validity, in particular when the number of cohorts used in the IV approach is non-negligible.

To provide statistical insights on the IV validity of CSL, we jointly test random instrument assignment, monotonicity of education in the instrument, and the instrument exclusion restriction by applying the method of Huber and Mellace (2015) to data from the Survey of Health, Ageing and Retirement in Europe (SHARE).¹ These assumptions are necessary for identifying the local average treatment effect (LATE) on those who change their schooling decision in reaction to a change in CSL, the so-called compliers in the denomination of Angrist, Imbens, and Rubin (1996). Considering health- and income-related outcomes as well as various numbers of cohorts, the test does not provide evidence for the invalidity of the instrument. As a word of caution, we acknowledge that even asymptotically, the approach of Huber and Mellace (2015) cannot detect all possible violations of IV validity when the complier share is larger than zero.²

The remainder is organized as follows. Section 2 briefly reviews the testing approach of Huber and Mellace (2015). Section 3 describes our data. Section 4 presents the testing results.

2 Testing approach

We assume a binary treatment D ($D = 1$: attaining or exceeding a particular level of education vs. $D = 0$: not attaining this level), whose impact on outcome Y (health or income) is of policy

¹See Brunello, Fort, and Weber (2009) for a study exploiting changes in CSL based on the SHARE data.

²This caveat also applies to the other IV tests suggested in the context of the just identified heterogeneous treatment effect model by Kitagawa (2015) and Mourifie and Wan (2016).

interest. Z is the supposed instrument ($Z = 1$: exposed to CSL reform vs. $Z = 0$: not exposed). Denote by $D(z)$ the potential treatment state for instrument $Z = z$, and by $Y(d)$ the potential outcome for treatment $D = d$ (see Rubin, 1974, for a discussion of the potential outcome notation). The population can be categorized into four types (denoted by $T \in \{a, c, d, n\}$) as function of the potential treatments, see Angrist, Imbens, and Rubin (1996): compliers ($c : D(1) = 1, D(0) = 0$), always-takers ($a : D(1) = 1, D(0) = 1$), never-takers ($n : D(1) = 0, D(0) = 0$), and defiers ($d : D(1) = 0, D(0) = 1$). Under the following assumptions the LATE $E[Y(1) - Y(0)|T = c]$ is obtained by the probability limit of two stage least squares: (1) $E(Y^{d,1}|T = t, Z = 1) = E(Y^{d,0}|T = t, Z = 0) = E(Y^d|T = t)$ for $d \in \{0, 1\}$ and $t \in \{a, c, n\}$ (mean exclusion restriction); (2) $\Pr(T = d) = 0$ (monotonicity); (3) $\Pr(T = t|Z = 1) = \Pr(T = t|Z = 0)$ for $t \in \{a, c, n\}$ (unconfounded type); (4) $\Pr(T = c) > 0$ (existence of compliers).

Huber and Mellace (2015) show that under Assumptions (1)-(3), it must hold that³

$$\begin{aligned}
 E(Y|D = 1, Z = 1, Y \leq y_q) &\leq E(Y|D = 1, Z = 0) \leq E(Y|D = 1, Z = 1, Y \geq y_{1-q}), \\
 E(Y|D = 0, Z = 0, Y \leq y_r) &\leq E(Y|D = 0, Z = 1) \leq E(Y|D = 0, Z = 0, Y \geq y_{1-r}). \quad (1)
 \end{aligned}$$

$q = \Pr(D = 1|Z = 0)/\Pr(D = 1|Z = 1)$ corresponds to the share of always-takers conditional on $D = 1$ and $Z = 1$, and y_q is the q th quantile of Y given $D = 1$ and $Z = 1$. $r = \Pr(D = 0|Z = 1)/\Pr(D = 0|Z = 0)$ corresponds to the share of never-takers conditional on $D = 0$ and $Z = 0$, and y_r is the r th quantile of Y given $D = 0$ and $Z = 0$. Considering the first line of (1), the intuition of the testable constraint is as follows: $E(Y|D = 1, Z = 0)$ point identifies the mean potential outcome of the always-takers under treatment, as any subject with $D = 1, Z = 0$ must be an always-taker in the absence of defiers. Furthermore, the mean potential outcomes of the always-takers are bounded by the averages in the upper and lower outcome proportions with $D = 1$ and $Z = 1$ that correspond to the share of the always-takers in the mixed population with compliers: $E(Y|D = 1, Z = 1, Y \leq y_q)$, $E(Y|D = 1, Z = 1, Y \geq y_{1-q})$. $E(Y|D = 1, Z = 0)$ must lie within the latter bounds, otherwise the assumptions are necessarily violated. An analogous result applies to the mean potential outcome of never-takers under non-treatment. Any procedure suitable for testing multiple moment inequalities could be used for verifying (1), for instance the method by Chen and Szroeter (2014).

³Under Assumptions (1) - (3) and a violation of Assumption (4), the weak inequalities in (1) become equalities.

3 Data

Our treatment and outcome variables come from the first wave of the Survey of Health, Ageing and Retirement in Europe (SHARE) conducted in 2004, which targets individuals who are 50 years or older and reside in Europe or Israel. Our binary CSL instrument is based on schooling reforms in seven countries in the 1960s or 1970s: Austria, Denmark, France, Greece, Italy, the Netherlands, and Spain.⁴ Table 1 is based on Brunello, Fort, and Weber (2009) and provides the year of the reform, the pivotal cohort first potentially affected by it, and the expected change in minimum educational attainment as a consequence of the reform, measured according to the International Standard Classification of Education (ISCED). The ISCED levels are the following: 1 - primary education or first stage of basic education; 2 - lower secondary education or second stage of basic education; 3 - upper secondary education; 4 - post-secondary non-tertiary education; 5 - first stage of tertiary education; 6 - second stage of tertiary education.

Table 1: Schooling reforms

country	year of reform	pivotal cohort	expected change in minimum ISCED level
Austria	1962	1947	ISCED 1 to ISCED 2
Denmark	1971	1957	ISCED 2 to ISCED 3
France	1959*	1953	ISCED 2 to ISCED 3
Greece	1975	1963	ISCED 1 to ISCED 2
Italy	1963	1949	ISCED 1 to ISCED 2
Netherlands	1975	1959	ISCED 1 to ISCED 2
Spain	1970	1957	ISCED 1 to ISCED 2

Note: * The Berthoin Reform of 1959 was implemented in 1967.

We pool the data from the seven countries and define the binary treatment variable D to be one if an individual attains the expected minimum ISCED level after the respective CSL reform or a higher degree, and zero otherwise. For example, for an individual from Austria the treatment is one if she attained at least ISCED 2 and zero in the case of ISCED 1, while for Denmark, the treatment is one for individuals attaining at least ISCED 3 and zero otherwise.⁵

For the construction of the instrument Z , we make use of adjacent cohorts not yet exposed to

⁴Although present in the first panel wave, the following countries were excluded: 1) Belgium, because its reform of 1983 did not affect any individual in the first wave of SHARE; 2) Germany and Switzerland, since they did not experience a single national schooling reform; 3) Sweden, because it lacks variation in the treatment variable to be defined; 4) Israel, as its schooling reform of 1968 was only partially implemented.

⁵A striking feature in the data is that there exist untreated individuals who did receive the instrument, i.e. never-takers. In our CSL setting, such individuals should not exist if those laws were strictly enforced. One reason for their occurrence may be gradual implementation of the policy over time. For example, complete implementation of Italy's reform of 1963 was, according to Brunello, Fort, and Weber (2009), not immediate, but took 13 years. Our data confirms gradual compliance. Based on the sample for the outcome income, we find for Italy that among the pivotal and the following cohort there is a 18% share of never-takers. The share drops to 12% in the two following cohorts and 7% in the following four cohorts.

the respective CSL reform, for whom Z is zero, and exposed to the reform, for whom Z is one. In the analysis, we consider three different symmetric data windows around the point when the respective reform came into force. The smallest window contains four cohorts: the two cohorts prior to the pivotal cohort, the pivotal cohort, and the following cohort. Likewise, a medium-sized window with a total of eight cohorts and a large window with sixteen cohorts is chosen. We apply the test to six health- and income-related outcomes Y : frequency of alcoholic beverages consumed during the last 6 months measured on a scale from 0 (not at all) to 6 (almost every day), BMI, number of doctor visits in the last 12 months, self-reported health measured on an scale from 1 (poor) to 5 (excellent), labor earnings in euros in 2003, and pooled income in euros from earnings and pensions in 2003.

Table 2: Descriptive statistics

outcome	+/- 2 cohorts				+/- 4 cohorts				+/- 8 cohorts			
	obs	range Y	av D	av Z	obs	range Y	av D	av Z	obs	range Y	av D	av Z
alc. beverages	1124	0 - 6	0.46	0.67	2306	0 - 6	0.35	0.66	4766	0 - 6	0.25	0.57
BMI (rounded)	1116	15 - 51	0.46	0.67	2289	15 - 59	0.35	0.66	4711	15 - 60	0.25	0.57
doctor visits	1121	0 - 98	0.46	0.67	2295	0 - 98	0.35	0.66	4745	0 - 98	0.25	0.57
health	583	1 - 5	0.48	0.67	1157	1 - 5	0.37	0.65	2328	1 - 5	0.26	0.56
income in million	1130	0 - 0.33	0.46	0.67	2317	0 - 1.50	0.35	0.66	4785	0 - 100	0.25	0.57
pooled inc. in m.	1130	0 - 0.33	0.46	0.67	2317	0 - 1.50	0.35	0.66	4785	0 - 100	0.25	0.57

Our evaluation samples based on the various cohorts only contain individuals without missing values in the treatment, the instrument, and the respective outcome. Furthermore, individuals who had immigrated to the respective countries at the age 11 or later are dropped, as they might not have been affected by the reform. Separately by outcomes and the number of cohorts, Table 2 reports the number of observations ('obs'), the empirical range of values in the outcome ('range Y '), the share of treated ('av D '), and the share of instrumented individuals ('av Z '). An implication of the age distribution in the first wave of SHARE is that our samples are skewed towards an increasingly smaller fraction receiving the instrument as more cohorts are included. The fraction receiving the instrument is approximately 46% when using 4 cohorts. This share drops to around 36% for 8 cohorts and to approximately 25% for 16 cohorts.⁶ We also note that the contribution of the seven countries to our samples is not uniform. Observations from Austria, France and Italy account for approximately 92% of observations in our samples, with each having a share of between 23% to 37% depending on the three sample sizes.⁷

⁶For this reason, we also test IV validity with asymmetric sampling windows, which keep the share of IV recipients roughly constant. The results are very similar to those reported in Section 4 and therefore omitted.

⁷A further point worth noting is that even though SHARE targets individuals who are 50 years old or older,

4 Empirical results

The testing results are presented in Table 3, in separate panels for different numbers of not affected (i.e. earlier) and affected (i.e. later) cohorts to be included in the analysis: ± 2 cohorts around the change in CSL, ± 4 cohorts, and ± 8 cohorts. The column ‘outcome’ contains the outcome variables for which (1) is tested, while ‘obs’ reports the number of observations. ‘first’ gives the first stage effect of Z on D , which is an unbiased estimate of the complier share if the LATE assumptions hold. We note that the first stage effect is statistically significant at the 5% level in any scenario considered. The last column ‘p-val’ provides the p-value when implementing the test based on the method by Chen and Szroeter (2014) and 999 bootstrap replications, see Huber and Mellace (2015) for details. We see that for any outcome and sample specification, the p-values are far from any conventional level of significance. This even holds for the smaller first stages obtained under a smaller number of cohorts, which ceteris paribus imply more narrow bounds in (1) and thus, a higher asymptotic power of the test. We therefore conclude that we find no statistical evidence for a violation of the CSL instrument in the SHARE data, but point out that even asymptotically, such tests cannot detect all potential violations of Assumptions (1)-(3) if the complier share is positive (i.e., Assumption (4) holds).

Table 3: Empirical results

outcome	± 2 cohorts			± 4 cohorts			± 8 cohorts		
	obs	first	p-val	obs	first	p-val	obs	first	p-val
alc. beverages	1124	7.6%	1.000	2306	10.6%	0.981	4766	32.1%	1.000
BMI	1116	7.4%	0.993	2289	10.5%	1.000	4711	32.0%	1.000
doctor visits	1121	7.6%	1.000	2295	10.3%	0.556	4745	31.8%	1.000
health	583	8.7%	1.000	1157	11.1%	1.000	2328	31.1%	1.000
income	1130	7.6%	0.979	2317	10.6%	0.886	4785	32.1%	0.715
pooled income	1130	7.6%	0.989	2317	10.6%	0.926	4785	32.1%	0.726

References

ANGRIST, J., G. W. IMBENS, AND D. RUBIN (1996): “Identification of Causal Effects Using Instrumental Variables,” *Journal of American Statistical Association*, 91, 444–472 (with discussion).

our samples also contain observations under the age of 50, as spouses of eligible respondents are included in the data set. However, their numbers are substantially lower and generally decrease in the year of birth. These spouses are predominantly female, causing an over-representation of women in particular in countries with relatively late reforms. For instance, 58% of our respondents in the BMI sample with 16 cohorts are female and 61% of the subsample receiving the instrument are female. However, countries with late reforms do not contribute importantly to our samples.

- ANGRIST, J., AND A. KRUEGER (1991): “Does Compulsory School Attendance Affect Schooling and Earnings?,” *Quarterly Journal of Economics*, 106, 979–1014.
- BRUNELLO, G., M. FORT, AND G. WEBER (2009): “Changes in Compulsory Schooling, Education and the Distribution of Wages in Europe,” *Economic Journal*, 119, 516–539.
- CHEN, L.-Y., AND J. SZROETER (2014): “Testing Multiple Inequality Hypotheses: A Smoothed Indicator Approach,” *Journal of Econometrics*, 178, 678–693.
- HARMON, C., AND I. WALKER (1995): “Estimates of the Economic Return to Schooling for the United Kingdom,” *American Economic Review*, 85, 1278–1286.
- HUBER, M., AND G. MELLACE (2015): “Testing Instrument Validity for LATE Identification Based on Inequality Moment Constraints,” *Review of Economics and Statistics*, 97, 398–411.
- KITAGAWA, T. (2015): “A Test for Instrument Validity,” *Econometrica*, 83, 2043–2063.
- MOURIFIE, I., AND Y. WAN (2016): “Testing Local Average Treatment Effect Assumptions,” *forthcoming in the Review of Economics and Statistics*.
- OREOPOULOS, P. (2006): “Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter,” *The American Economic Review*, 96, 152–175.
- RUBIN, D. B. (1974): “Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies,” *Journal of Educational Psychology*, 66, 688–701.

Authors

Benjamin BOLZERN
University of Bern, Department of Economics, Email: bolzern@gmail.com

Martin HUBER
University of Fribourg, Department of Economics, Bd. de Pérolles 90, 1700 Fribourg, Switzerland.
Phone: +41 26 300 8274; Email: martin.huber@unifr.ch; Website: <http://www.unifr.ch/appecon/en/team/martin-huber>

Abstract

Changes in compulsory schooling laws have been proposed as an instrument for the endogenous choice of schooling. It has been argued that raising minimum schooling exogenously increases the educational attainment of a subset of pupils without directly affecting later life outcomes such as income or health. Using the method of Huber and Mellace (2015) and data from the Survey of Health, Ageing and Retirement in Europe, we jointly test random instrument assignment, weak monotonicity of education in the instrument, and the instrument exclusion restriction. The satisfaction of these restrictions permits identifying the local average treatment effect of education on those choosing more schooling as a reaction to the law change. Our results do not point to the invalidity of the schooling law instrument, though we acknowledge that even asymptotically, testing cannot detect all possible violations of instrument validity.

Citation proposal

Benjamin Bolzern, Martin Huber. 2017 «Testing the Validity of the Compulsory Schooling Law Instrument». Working Papers SES 480, Faculty of Economics and Social Sciences, University of Fribourg (Switzerland)

Jel Classification

C26, I12, I21, I28, J24

Keywords

instrumental variable, schooling laws, schooling reforms, treatment effects, LATE, tests

Working Papers SES collection

Last published

- 474 Grossmann V., Schäfer A., Steger T., Fuchs B.: Reversal of Migration Flows: A Fresh Look at the German Reunification; 2016
- 475 Pesenti A.: The Meaning of Monetary Stability; 2016
- 476 Furrer O., Sudharshan D., Tsiotsou Rodoula H., Liu Ben S. : A Framework for Innovative Service Design; 2016
- 477 Herz H., Taubinsky D.: What Makes a Price Fair? An Experimental Study of Transaction Experience and Endogenous Fairness Views; 2016
- 478 Zehnder C., Herz H., Bonardi J.-P.: A Productive Clash of Cultures: Injecting Economics into Leadership Research; 2016
- 479 Huber M., Wüthrich K.: Evaluating local average and quantile treatment effects under endogeneity based on instruments: a review; 2017

Catalogue and download links

<http://www.unifr.ch/ses/wp>
http://doc.rero.ch/collection/WORKING_PAPERS_SES

Publisher

Université de Fribourg, Suisse, Faculté des sciences économiques et sociales
Universität Freiburg, Schweiz, Wirtschafts- und sozialwissenschaftliche Fakultät
University of Fribourg, Switzerland, Faculty of Economics and Social Sciences

Bd de Pérolles 90, CH-1700 Fribourg
Tél.: +41 (0) 26 300 82 00
decanat-ses@unifr.ch www.unifr.ch/ses