

Technical Appendix to FORCED TO BE RICH? RETURNS TO COMPULSORY SCHOOLING IN BRITAIN

Paul J. Devereux and Robert A. Hart

ECONOMIC JOURNAL, doi: 10.1111/j.1468-0297.2010.02365.x

Data Appendix

2001 New Earnings Survey Panel Data-set (NESPD)

This dataset contains information from 1975 to 2001. We include part-time and full-time earners aged 18 to 64 as on January 1st of the survey year. We exclude cases where earnings are affected by absence. We also exclude observations if their ‘Hourly Earnings excluding overtime’ is missing or if their normal basic hours are missing. Because age is as at January 1st, year of birth is constructed as being equal to survey year – age – 1 and we keep the 1921–51 cohorts. We deflate wages and earnings using the British Retail Price Index and drop cases for which hourly wage observations are less than £1 or more than £150 (in December 2001 pounds). These exclusions are similar to those used by Card (1999). They imply the exclusion of 136 male observations (682 female) that have wages less than £1 and 91 male observations (0 female) that have wages greater than £150. To put these numbers in context, the minimum wage was £3.70 in 2001. We exclude the small number of cases where weekly hours are greater than 84 (99 cases), less than 1, or missing.

General Household Survey (GHS)

We use GHS files from 1979–98 (there is no 1997 survey) and include persons who report their age as being between 28 and 64.

Year of Birth. Year of birth is reported in the data from 1986–95 and in 1998. Year of birth is also available for women aged 16–49 in the 1983–5 surveys and we use this information where available. For cases where year of birth is unavailable, we impute year of birth as being (survey year – age) for persons who are interviewed between July and December, and as being (survey year – age – 1) for persons interviewed between January and June.

Earnings Measures. The usual weekly earnings measures in the GHS change in exact definition and name over time. We use PAYWEEK for the 1979–82 surveys, UGE for the 1984–91 surveys, GEIND for the 1992–96 surveys, and GREARN for the 1998 survey. Unlike the other weekly earnings measures, PAYWEEK excludes self-employment income so we add any positive self-employment income using the INCSELF variable. UGE is missing in the 1983 survey so we construct it in the same way that UGE is constructed by the GHS data team for the 1984 survey year. We construct an hourly earnings variable in the GHS using information on usual weekly earnings and usual weekly hours (WORKHRS). One drawback is that the earnings information includes overtime earnings but the hours variable excludes overtime hours. Despite this problem, we report estimates using the hourly wage variable in addition to weekly earnings. Manning (2000) also uses this variable and shows that it is highly correlated with the true hourly wage

(correlation = 0.98) because average overtime hours are relatively short (less than 3 per week) and because overtime hours are very weakly correlated with hourly earnings. As in the NESPD, we deflate wages and earnings using the British Retail Price Index. We set to missing cases for which hourly wage observations are less than £1 or more than £150 (in December 2001 pounds). We exclude cases where weekly hours are greater than 84, less than 1, or missing.

Schooling. Our measure of schooling is age left school. The relevant variables are AGELFTS from 1979–82 and AGELFTSC from 1983–98. We set schooling to missing for cases in which age left school is reported to be less than 10, greater than 24, or greater than the respondent's reported age.

Data for Specifications in Rows 3–5 of Table 2 and in Table 5

Variable construction and sample restrictions for these estimates are taken from Oreopoulos (2006) and from his STATA programs (with his kind assistance). All GHS variables are constructed from 1984–98 as this is the time period he used in estimation. Full details can be found in Oreopoulos (2006; 2008) and in his STATA programs on the AER website.

GHS: He uses Terminal Age of Education (TEA) as his schooling measure. For weekly earnings, he uses UGE for the 1984–91 surveys, GEIND for the 1992–96 surveys, and GREARN for the 1998 survey. He multiplies weekly earnings by 52 and refers to it as annual earnings. For year of birth, he uses that reported in the data in survey years 1987–95. He imputes year of birth as equal to survey year – age in other years. He drops cases with nominal earnings of £10,000 per week or more. His regressions always use nominal rather than real earnings.

Northern Ireland Data. The Continuous Household Survey (CHS) data from Northern Ireland used in Table 5 are exactly as defined by Oreopoulos (2006) and Oreopoulos (2008). The CHS surveys used are those from 1986–91, as well as 1996, 1998 and 1999. Earnings are reported in intervals in the CHS and the midpoint of each interval is used in estimation.

Table A1
Reduced Form and 2SLS Estimates (GHS)

	(1)	(2)	(3)	(4)	(5)
	Schooling	Log Wage	Log Earnings	Log Wage	Log Earnings
<i>Men</i>					
Including immigrants, excluding self-employed					
Law	0.458 (0.036)**	0.020* (0.009)	0.031 (0.010)*		
Schooling N = 44,256				0.043 (0.019)*	0.067 (0.024)*
Excluding immigrants and self-employed					
Law	0.501 (0.025)**	0.017 (0.009)	0.031 (0.011)*		
Schooling N = 41,039				0.034 (0.017)	0.062 (0.022)*
Excluding immigrants, including self-employed					
Law	0.471 (0.024)**	0.023 (0.012)	0.029 (0.012)*		
Schooling N = 46,995				0.049 (0.025)	0.061 (0.026)*
<i>Women</i>					
Including immigrants, excluding self-employed					
Law	0.529 (0.032)**	0.006 (0.010)	0.002 (0.018)		
Schooling N = 39,886				0.011 (0.018)	0.003 (0.034)
Excluding immigrants and self-employed					
Law	0.560 (0.037)**	0.003 (0.011)	−0.014 (0.017)		
Schooling N = 36,872				0.005 (0.019)	−0.025 (0.031)
Excluding immigrants, including self-employed					
Law	0.549 (0.044)**	0.009 (0.012)	−0.016 (0.018)		
Schooling N = 38,771				0.017 (0.022)	−0.029 (0.033)

* significant at 5%; ** significant at 1% level. All specifications include a quartic in year of birth and a quartic in age.

Robust standard errors in parenthesis allow for clustering by year-of-birth.

Table A2
Reduced Form and 2SLS Effects of 1947 Law on Schooling and Weekly Earnings (Pooled Sample)
 (Persons aged 28–60)

	1 st Stage: Schooling			Reduced Form: Log Weekly Earnings			2SLS: Log Weekly Earnings		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Row 1 N = 82,381	0.502 (0.034)**	0.500 (0.034)**	0.499 (0.033)**	0.012 (0.011)	0.015 (0.010)	0.018 (0.012)	0.023 (0.022)	0.030 (0.020)	0.036 (0.023)
Row 2 N = 50,301	0.435 (0.045)**	0.437 (0.045)**	0.442 (0.049)**	0.011 (0.023)	0.017 (0.020)	0.016 (0.020)	0.025 (0.054)	0.038 (0.046)	0.035 (0.046)
Row 3 N = 49,930	0.344 (0.053)**	0.343 (0.052)**	0.373 (0.061)**	0.011 (0.023)	0.014 (0.020)	0.014 (0.020)	0.031 (0.067)	0.040 (0.058)	0.039 (0.056)
Row 4 N = 51,643	0.373 (0.054)**	0.373 (0.054)**	0.408 (0.063)**	0.031 (0.027)	0.029 (0.029)	0.035 (0.028)	0.083 (0.076)	0.078 (0.081)	0.086 (0.075)
Row 5 N = 51,643	0.376 (0.054)**	0.376 (0.053)**	0.411 (0.063)**	0.047 (0.020)*	0.044 (0.022)	0.052 (0.023)*	0.125 (0.061)*	0.118 (0.069)	0.126 (0.067)
Age Controls	None	Quartic	Age Dummies	None	Quartic	Age Dummies	None	Quartic	Age Dummies

Row 1: Preferred Sample and Specification (female dummy included).

Row 2: Oreopoulos Sample Period (1984–1998) (female dummy included).

Row 3: Oreopoulos Sample Period, Law variable and Schooling Variable (female dummy included).

Row 4: Oreopoulos Sample Period, Law variable, Schooling Variable, Earnings variable, and Sample Restrictions (female dummy included).

Row 5: Oreopoulos Specification and Sample.

Estimates from the 1979–98 GHS for persons aged 28–60. All specifications include a quartic function of year-of-birth. Robust standard errors in parenthesis allow for clustering by year-of-birth.

* significant at 5%; ** significant at 1%.

Table A3
2SLS and Two Sample 2SLS Estimates of the Return to Schooling
 (Persons aged 28–60)

	Hourly Wage			Weekly Earnings		
<i>Men</i>						
GHS 2SLS Estimates	0.049 (0.027)	0.052 (0.028)	0.052 (0.030)	0.058 (0.027)*	0.065 (0.027)*	0.072 (0.031)*
N = 44,736						
NESPD	0.035 (0.015)*	0.031 (0.016)	0.025 (0.017)	0.036 (0.015)*	0.032 (0.017)	0.022 (0.018)
Two Sample 2SLS Estimates N = 992,625						
Age Controls	None	Quartic	Age Dummies	None	Quartic	Age Dummies
<i>Women</i>						
GHS 2SLS Estimates	0.029 (0.018)	0.030 (0.018)	0.028 (0.021)	−0.017 (0.032)	−0.010 (0.030)	−0.001 (0.034)
N = 37,645						
NESPD	−0.005 (0.017)	−0.004 (0.017)	−0.007 (0.017)	−0.006 (0.029)	−0.003 (0.028)	−0.007 (0.029)
Two Sample 2SLS Estimates N = 693,366						
Age Controls	None	Quartic	Age Dummies	None	Quartic	Age Dummies

All specifications include a quartic function of year-of-birth.

Two Sample Two Stage Least Squares estimation: First Stage comes from the GHS. Reduced Form estimates are from the NESPD.

Standard errors in parenthesis. Two Sample 2SLS standard errors are calculated using the delta method.

* significant at 5%; ** significant at 1%.