Aarhus, January 24 – 2023

Net Appendix to
the transition of education

Martin Paldam

This appendix uses the same variables as the main paper – see Table A1. All sections, tables, and figures with an A before the number are in this Appendix, where the Figures are numbered as the section. Without A they are in the main paper. The variables are defined in Table 1 of main paper.

As is the main paper data for all former and present OPEC countries are deleted.

Content

Table A1. The three samples, for the non-OPEC countries

Panel representation		Dimension			Ideal	Unified		
Nr	Sample	Countries	Period	Unit	*N**	*N*	Gap	Correlation
1	(*E-share*, *income*)	141	1970-2017	annual	141 *x* 48 = 6,768	3,173	53.1%	0.42
2	(*EC-share*, *income*)	141	1970-2017	annual	141 *x* 48 = 6,768	3,173	53.1%	0.72
3a	(*School*, *income*)	122	1970-2010	5-year	122 *x* 9 =1,098	1,077	1.95%	Av 0.76
3b	(*School*, *income*)	122	All 13 age groups		1,098 *x* 13 = 14,274	14,001	1.95%	0.74

The *School* and *income* data are available for 122 countries all years except the first three. The missing data are for 10 new countries that came out of the USSR and Yugoslavia and Slovakia.
For 1970 and 1975 the *income* data are for 112 countries only, missing all 10 countries.
For 1980 the *income* data covers 121 countries, where only Slovakia is missing. Sample 3a is for one out of 13 age groups. Sample 3b all 13 age groups are merged.

A1 Variations of the bandwidth for the curves of Figures 2b, 3, 5b and 6b

The *bw*-experiments reports the four main kernels from the main paper in bold surrounded with three curves for smaller *bw*'s and three curves for larger *bw*'s.

Figure A1.1. *Income* explains *E-share* (Figure 2b), *bw* = 0.3 to 0.9

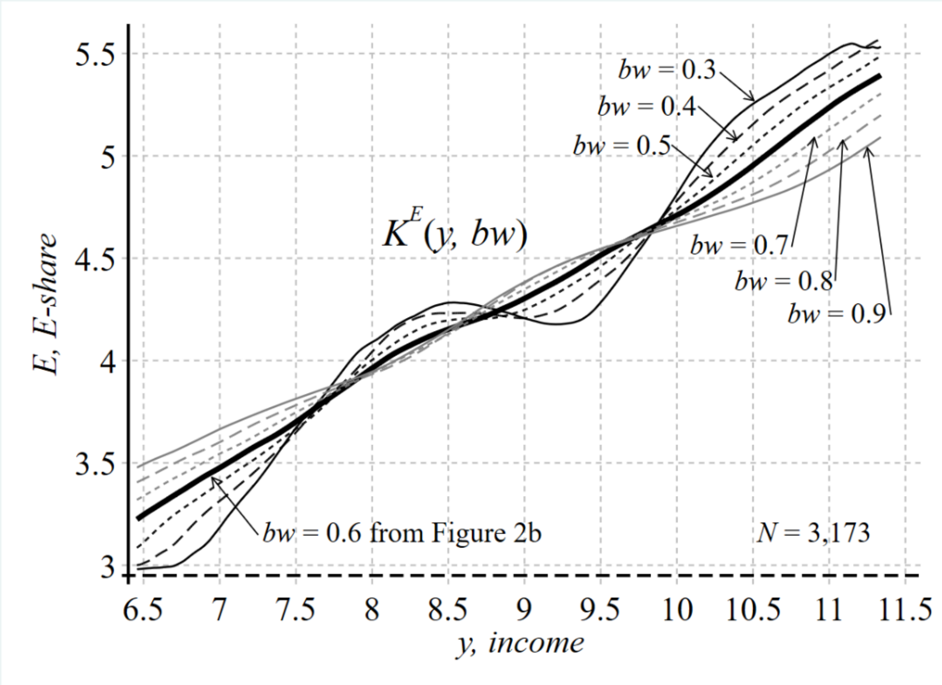


Figure A1.2. *E-share* explains *income* (Figure 3), *bw* = 0.4 to 1.6

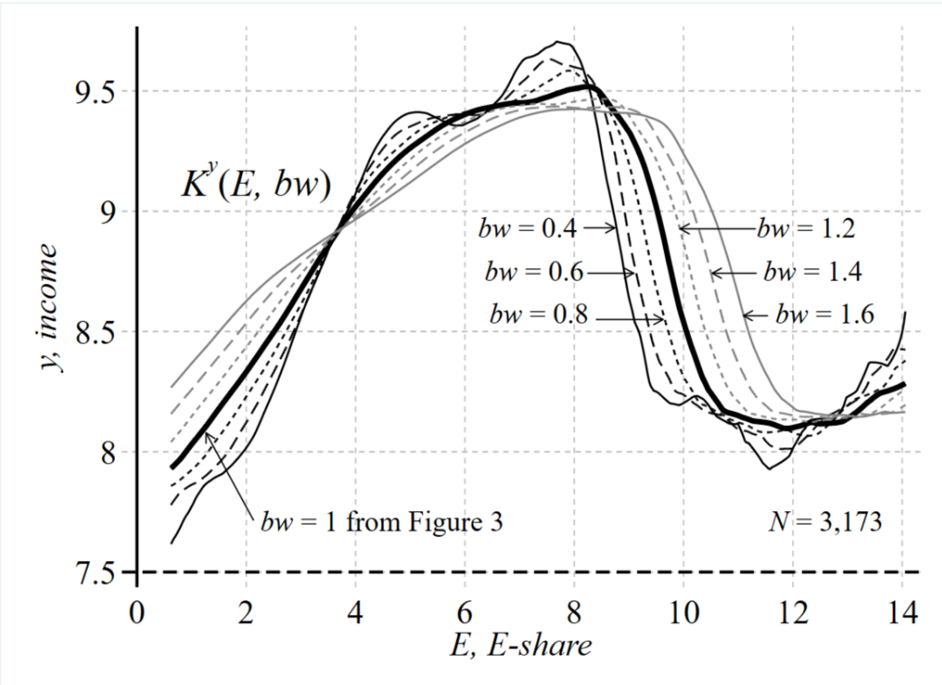


Figure A1.1 looks at the robustness of Figure 2b, i.e., the relation from *income* to the *E-share*. For small values of the *bw* the kernel curve has some cyclicality that seems to lack an explanation. However, for bw = 0.5 and up the curve is close to a straight line. Thus, the curve on Figure 2b is rather robust.

Figure A1.2 analyzes the strange kernel of Figure 3, where the *E-share* explains *income*. It moves systematically upward as *bw* does. This indicate that the curve is of dubious stability. However, the shape with a peak in the middle remains. As argued in the main paper Figure 2 is both simpler and more robust than Figure 3. Figure A1.1 and A1.2 confirm this. Thus, the demand explanation is 'better' than the supply explanation, as in the main paper.

Figure A1.3. *Income* explains *school*, age group 45-49 (Figure 6), *bw* = 0.1 to 0.6

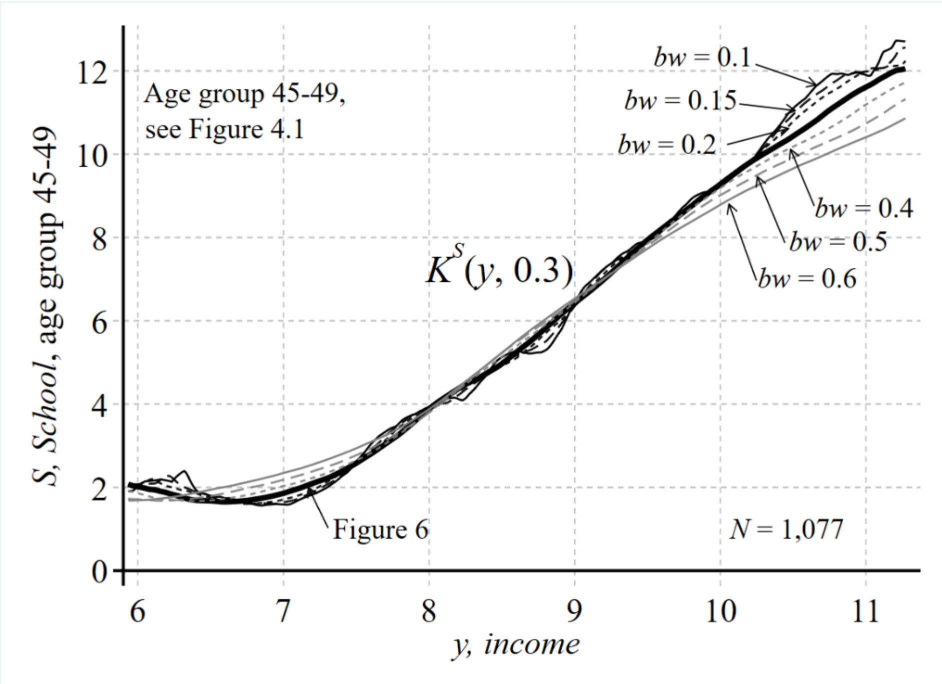
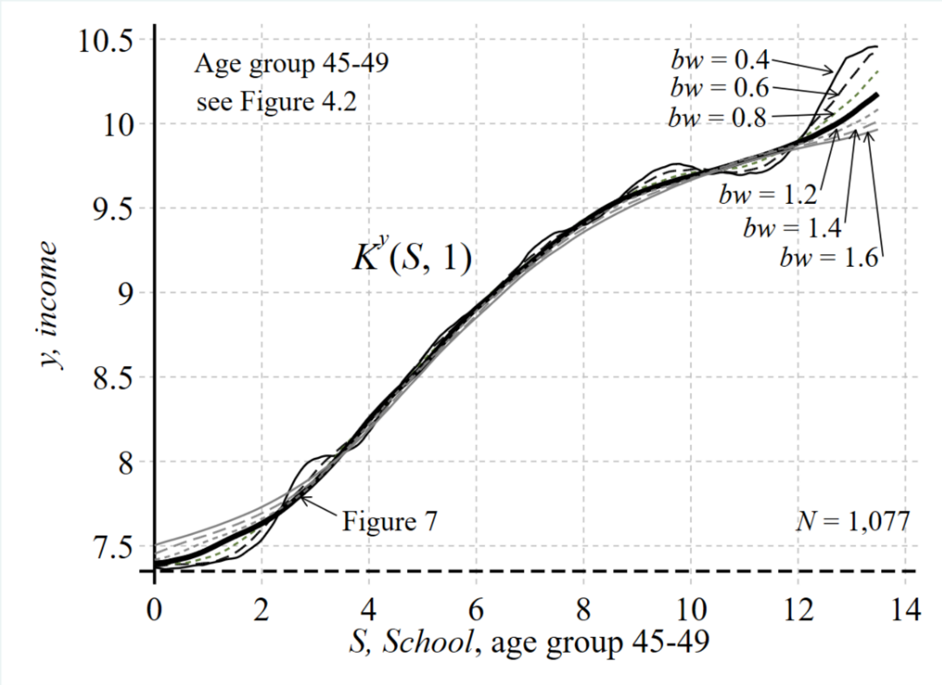


Figure A1.3 considers the nice transition curve of Figure 5b, where *income* explains *school*. It preserves its form all the way from bw = 0.1 to 0.6. This is a very robust kernel curve, as also confirmed by Figure A4.2 below.

Figure A1.4 looks at the curve of Figure 6b, where *school* explains *income.* As expected, it is practically equally nice as is Figure A1.3. It is clear *school* and *income* has much simultaneity.

Figure A1.4. *School* explains *income*, age group 45-49 (Figure 6b), *bw* = 0.4 to 1.6



A2 *E-share* and *School* over time

Both Figures 2 show kernels over time *t*. Both the *E-share* and *School* grow over time at a rate that may be constant, as a straight line can be drawn within the confidence intervals.

Figure A2.1 shows that the *E-share* is rising by about 0.15% per year. This is so little that there is no need to correct the D^E-variable for time.

Figure A2.1. The *E-share* over *t,* time

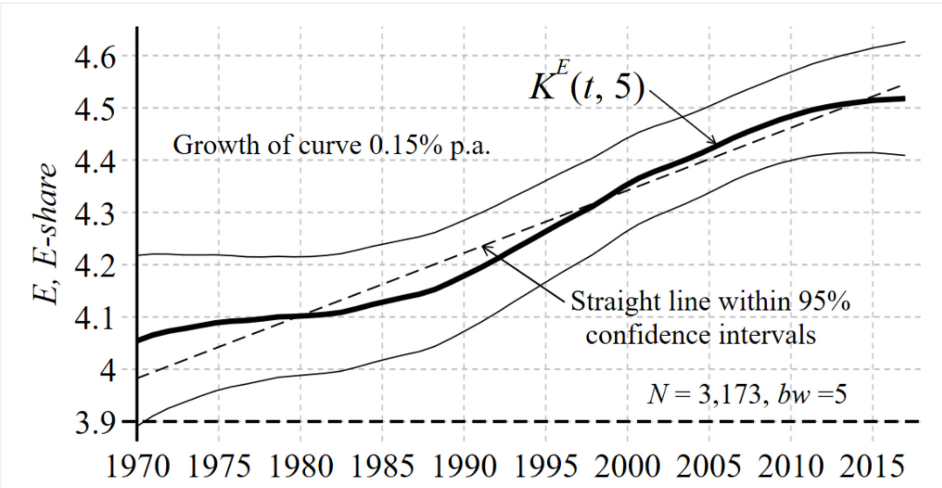
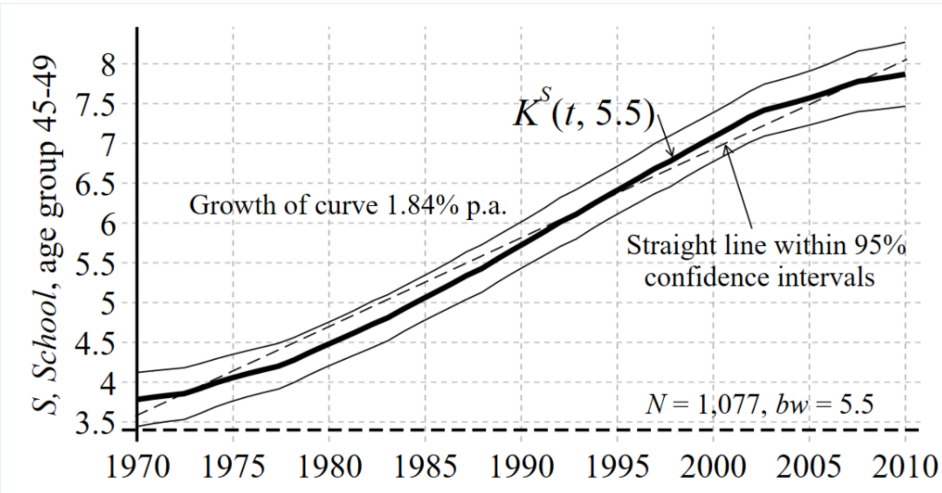


Figure A2.2 shows the development over time for *School*. It grows along a straight line with a rate of 1.84% per year. This is so much that the D^S variable has to be corrected for time giving the D^{ST}-variable.

Figure A2.2. *School* for age group 45-49 over *t*, time

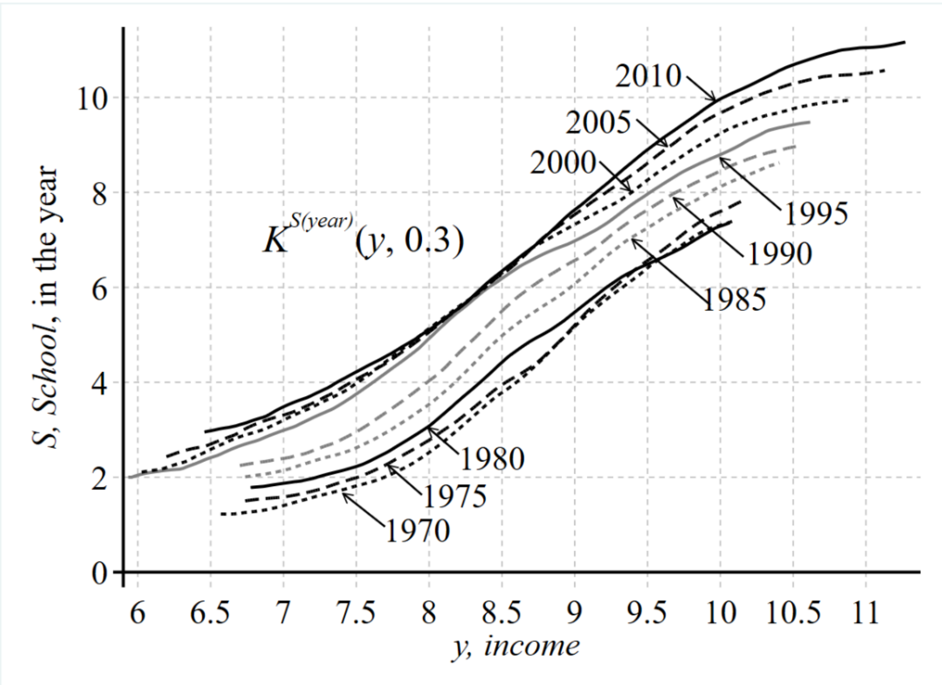


A3 *School* kernels for each of the 9 years

The kernels for the transition curve on Figure 2 are calculated for each age-group and all years. Figure A3 is estimated for all age groups and each year. The picture is much the same. Thanks to the stable growth in the HICs (High Income Countries) covered all years, the lines extend longer and longer each year. It is also easy to see that the number of School years at each income increases by 2½ over the full 40 years period covered.

Countries thus increases the *School* variable for two reason. (i) By 2.5 years over time even if income would be the same, and (ii) 1.1 years because income grows by 2-2½% per year. This is 3.6 years in the average country.

Figure A3. One Kernel for each of the 9 years covered by the Barro-Lee data



N is 1,456 for 1970 and 1975, 1,573 for 1980 and 1,586 the remaining six years.

A4 Scatter for *School* and *income* for the age group in the middle

The two curves are the middle curves from Figure 6a and 7a, with the scatter and the 95% confidence intervals included. They should be compared with Figures 6b and 7b where all age groups are merged. They look the same. The correlations are higher for the individual age groups than for the merged ones, but the confidence intervals ae narrower for the merged ones.

Figure A4.1. *Income* explains *school* – the age group in the middle

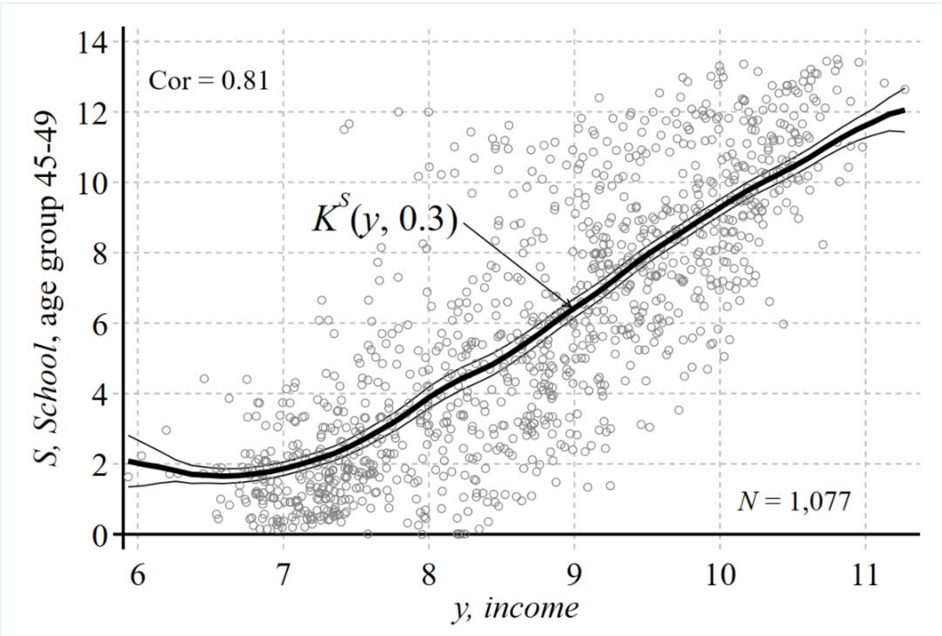
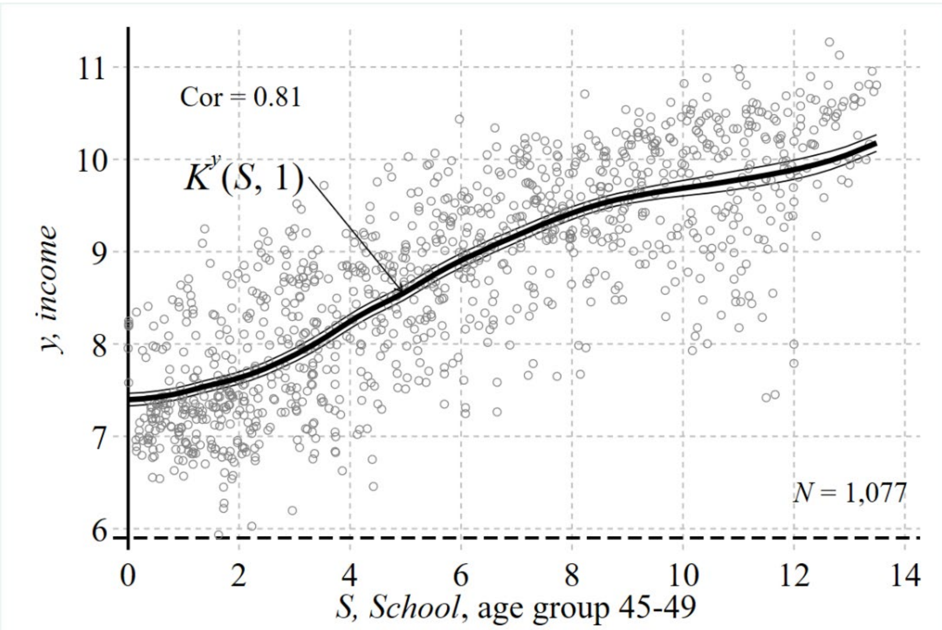


Figure A4.2. *School* explains *income* – the age group in the middle



A5 The straight transition line on Figure 2b

The straight line through $(E, y) = (3.25, 6.5)$ and $(5.25, 11)$, where

(1) $E = a + b\,y,$ where a and b are found by the standard method:

Two linear equations with two variables, a and b are:

(2) $3.25 = a + 6.5b$ and (3) $5.25 = a + 11b$

Or $a = 3.25 - 6.5b$ or $a = 5.25 - 11b$

Thus, b can be found as follows:

$3.25 - 6.5b = 5.25 - 11b$, so that $2 = 4.5\,b$

Or $b = 2/4.5 = 4/9$ so that $b \approx 0.44$

Thus, a can be found in two ways, giving the same result:

$a = 3.25 - 6.5 \cdot 4/9 = 3.25 - 2.89 = 0.36$

$a = 5.25 - 11 \cdot 4/9 = 5.25 - 4.89 = 0.36$

The equation is:

(2a) $E = 0.36 + 0.44\,y$ so that (2b) $y = 2.27\,E - 0.82$

Where (2a) is drawn on Figure 2a, and (2b) is dawn on Figure 3.

A6 The share of kids ages 6 to 19 in population

Figure 4 shows that the population in the primary school ages from 6-15 as a share of the population has a perfect transition form. Figure A6.1 show that the same applies to each of the 14 ages from 6 to 19. The curves for incomes above 9.5 are very close together. Figure A6.2 enlarge that part of the graph. The correlations between income and share falls rather smoothly from 0.846 to 0.636. The demographic transition is well known so Figure A6, is not surprising. However, it is still nice to see that the curve on Figure 4 is robust.

Figure A6.1. The share in the population for 14 ages from 6 to 19

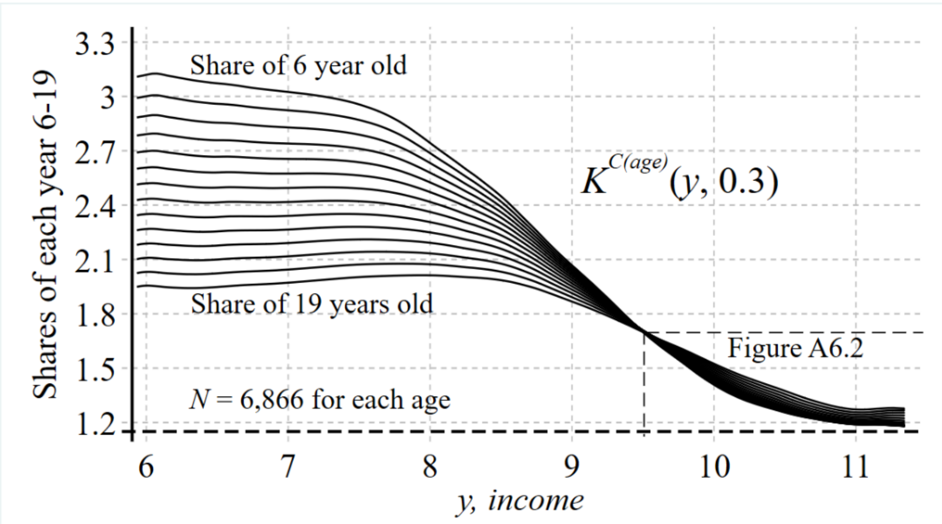
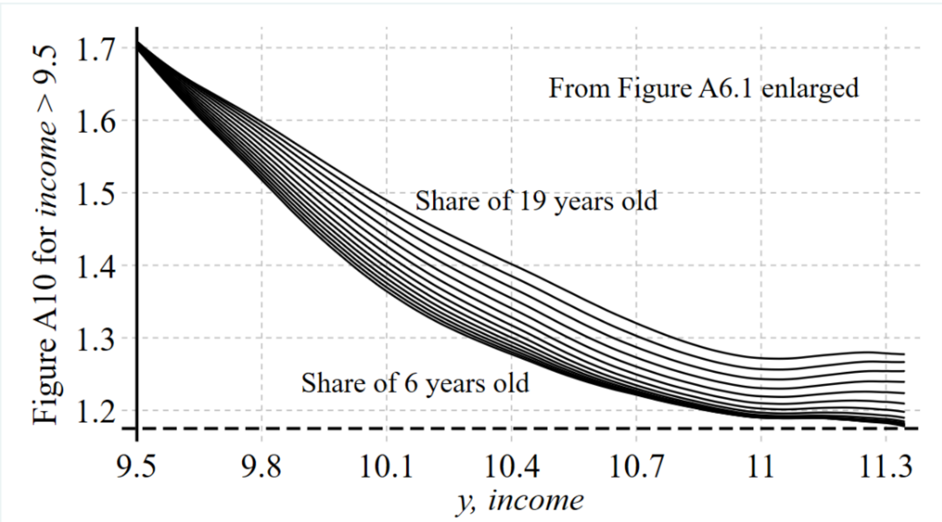


Figure A6.2. High income part of Figure A6.1



Estimated as kernel regressions, the lines fall systematically at the low-income level as the age goes up, but they increase systematically at the high end on Figure A6.2

A7 *Income-School* correlations, Fig. 3 without outliers, and D^E-histogram

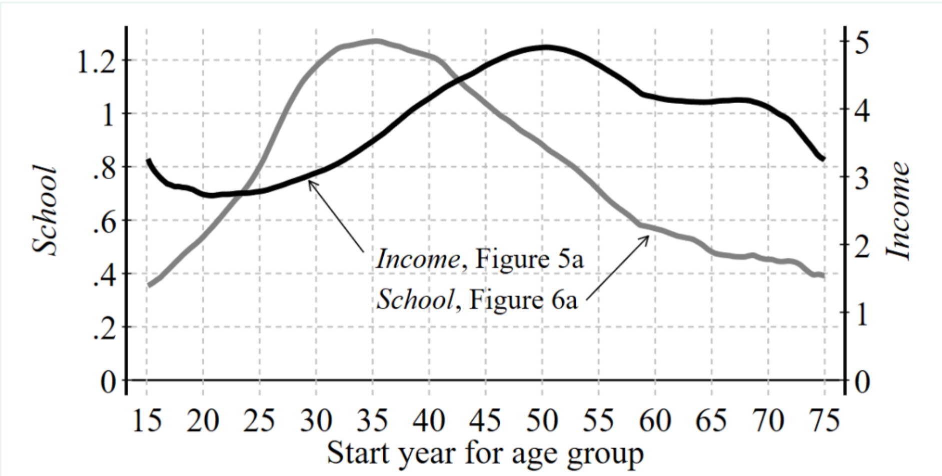
Figures 5a and 6a brings curves for the 13 age groups of the Barro-Lee data set. Each curve is calculated for 1,077 observations for all countries and years. Table A2 shows how *income* and *school* are correlated in each age group. The first two rows give the age group.

Table A7. Correlations of *income* and *school*, for each age group

	Age group													Av
From	15	20	25	30	35	40	45	50	55	60	65	70	75	15
To	19	24	29	34	39	44	49	54	50	64	69	74	Up	up
	Correlation between the 1,077 group observations in the age													
Cor	0.631	0.789	0.795	0.793	0.784	0.782	0.777	0.777	0.768	0.761	0.749	0.748	0.735	0.761

Figure A7.1 shows the vertical distance between the top and bottom curves on 5a and 6a. The black curve shows that *school* differ by (1.22 – 0.7)/1.22 = 0.42 when explained by *income.* The gray curve shows that *income* differ by (5.2 – 1.4)/5.2 = 0.73 when explained by *school.* Thus, Figure 5a gives more parallel curves than Figure 6a, as is also visible on the two Figures.

Figure A7.1. Two range curves, vertical distance between curves on Figures 5a and 6a



The *income* curve is for Figure 5a, where *income* explains *School*. The *school* curve is for Figure 6a, where *school* explains *income*. Each point is calculated as the vertical distance between the max and min observation.

Figure A7.2 is the same as Figure 3 but the extreme data are omitted by a deletion of all the 53 extreme observations. Not surprisingly, this makes the curve much nicer. Note that Figure A12 looks very much as Figure 6b and Figure A9. Thus, the *E-share* and the *School* data have a great deal in common, as they should.

Figure A7.2. Revised Figure 3. *E-share* explains *income*, 53 extreme values deleted

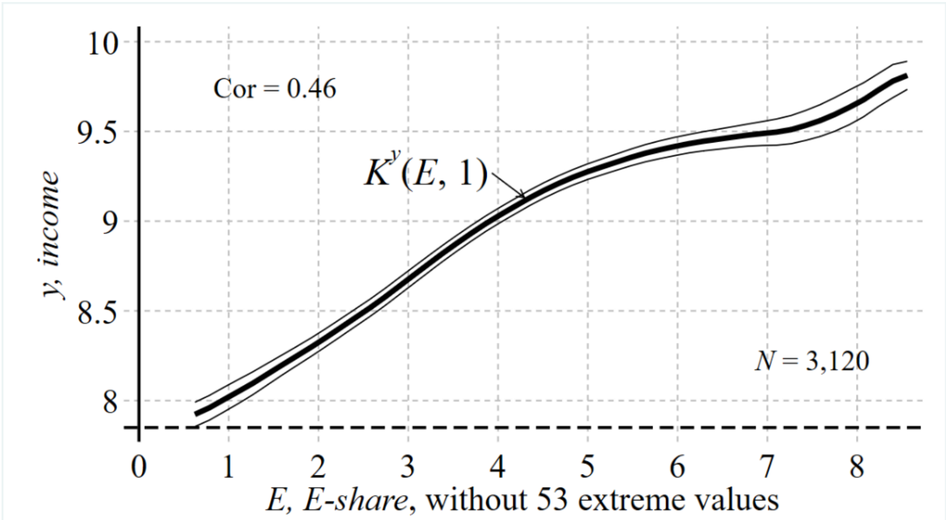


Figure A7.3. The frequency distribution of the D^E-variable, $N = 3{,}173$

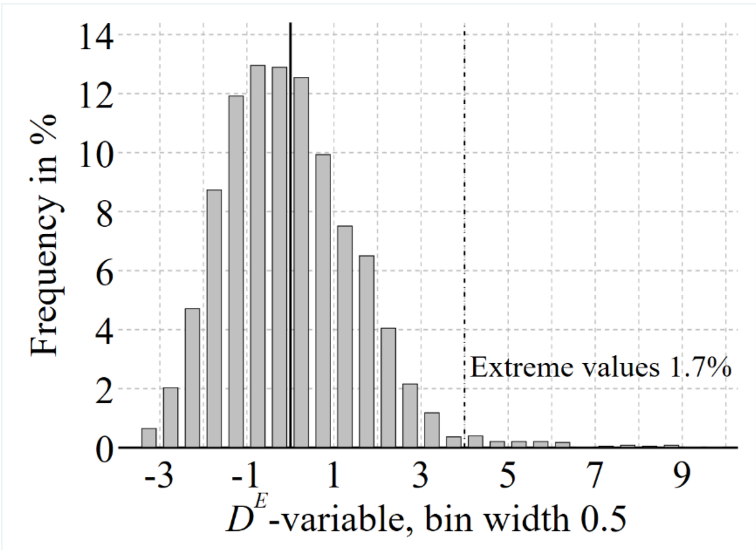


Figure A7.3 is the frequencies in % for the D^E-variable. The distribution is somewhat skew, and it has a long tail of extreme values to the right, as also suggested by Figure 3.

A8 Can D^E, the excess *E-share*, explain growth the same year?

The text analyzes an increase in the *E-share* and the growth rate much later. Figure A8.1 shows how the relation looks without lags. The 53 extreme observations are omitted. Figure A8.2 show that this does not affect the curve. Most of the curve do not deviate significantly from a horizontal line. But in the section where D^E is from about -2.3 to 2.6, where data are dense a significantly negative part of the curve appears. Clearly, in the short run it is an investment cost to expand education.

Figure A8.1. *Growth* explained by D^E, the excess *E-share*, the same year

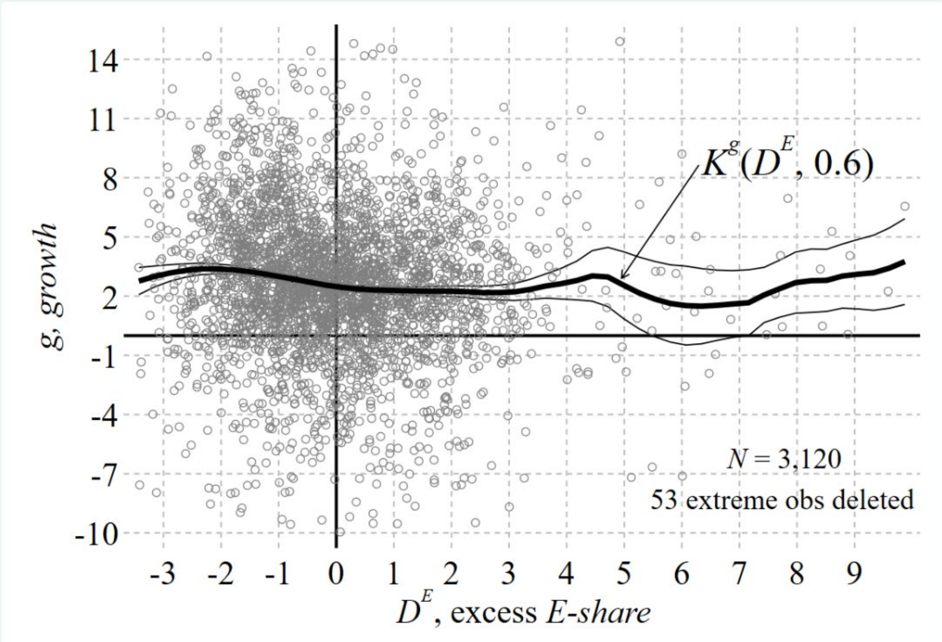
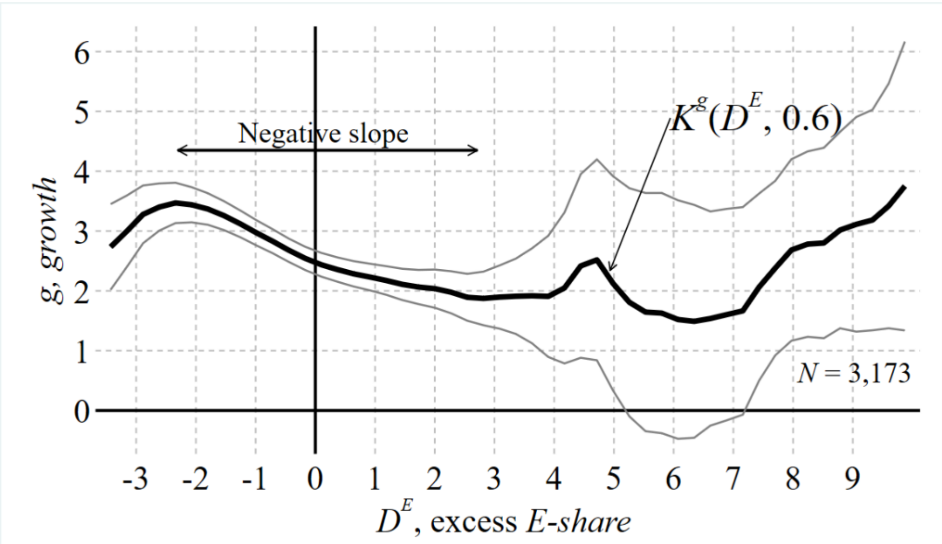


Figure A8.2. Same as Figure A8.1, for all observations and without scatter



A9 Tables for the School variable

Table A9.1 shows the average years of *school* for a member of the work force (the population from 35 to 64 for five country-groups every fifth year from 1970 to 2010.

Table A9.1. *School* for the country groups averages for all age groups in labor force

Group	Nr	1970	1975	1980	1985	1990	1995	2000	2005	2010	Change
Africa SS	28	1.14	1.34	1.54	1.79	2.19	2.64	3.01	3.47	3.99	2.85
Asia	16	2.33	2.65	3.03	3.58	4.17	4.87	5.59	6.21	6.98	4.65
Ex-soc	21	5.58	6.41	6.73	7.72	8.60	9.56	10.60	11.23	11.74	6.16
Lat Am	21	3.52	3.77	4.10	4.61	5.16	5.77	6.44	7.14	7.69	4.17
MENA+	11	0.82	0.97	1.33	1.82	2.35	2.91	3.56	4.24	4.78	3.96
West	25	6.89	7.43	7.97	8.32	8.95	9.60	10.27	10.95	11.63	4.74
Sum/av	122	3.44	3.80	4.35	4.87	5.46	6.11	6.78	7.40	7.99	4.55

Table A9.2 looks at the D^{ST}-variable that is adjusted for both income and time. The two trend scores are calculated as correlations to the 9 years. Column (4) shows that Sub-Saharan Africa has a normal School level when corrected for income and years. However, column (5) shows that the countries with the highest relative level also has the highest improvement and vice versa, so the counties are diverging in *School*.

Table A9.2. Some statistics for the D^{ST}-variable

	(1)	(2)	(3)	(4)	(5)
	Nr	Trend relative to years		Level	Correlation
Group		Kendall	Cor		(2) and (4)
Africa SS	28	-0.17	-0.19	0.00	0.71
Asia	16	-0.10	-0.04	0.47	0.52
Ex-soc	21	-0.23	-0.20	0.23	0.23
Lat Am	21	-0.19	-0.21	-0.10	0.64
MENA+	11	0.06	0.03	0.92	0.72
West	25	-0.23	-0.28	-0.56	0.61
Sum/av	122	-0.14	-0.15	0.16	0.58

The correlation analysis is done both by the Kendall rank correlation and the Pearson normal correlation (Cor). The results are so similar that that the correlation is not due to extreme observations.