



OESTERREICHISCHE NATIONALBANK

EUROSYSTEM

WORKSHOPS

Proceedings of OeNB Workshops

Current Issues of Economic Growth

March 5, 2004

No. 2

Opinions expressed by the authors of studies do not necessarily reflect the official viewpoint of the OeNB.

The presented articles were prepared for an OeNB workshop and therefore a revised version may be published in other journals.

Convergence of Educational Attainment Levels in the OECD: More Data, More Problems?

Jesús Crespo-Cuaresma

University of Vienna

1. Introduction

Finding a robust positive partial correlation between long-run economic growth and measures of educational attainment across world economies (after conditioning upon some other determinants of growth) has proved to be an extraordinarily difficult task for growth empiricists. In their influential survey on the empirics of economic growth, Durlauf and Quah (1999) report results for 16 empirical studies that used different educational variables in the specification. Only eight of the parameter estimates reported in Durlauf and Quah (1999) are positive, and not always significant. More recent studies (not included in Durlauf and Quah, 1999) find evidence of a negative partial correlation between human capital variables and economic growth (Pritchett, 1996, Benhabib and Spiegel, 1994).

One of the usual explanations for the failure in finding robust positive partial correlations between educational attainment and growth is related to the quality of the data usually employed as a proxy of human capital. Krueger and Lindahl (2001) and de la Fuente and Doménech (2002) present comparisons of different datasets of educational attainment (including the three datasets that will be used in this study). Using simple estimates of the signal-to-noise ratio, both contributions reach the conclusion that the widely used Barro-Lee dataset (Barro and Lee, 1993, 2001) performs poorly if the data is used in first differences, and that more recent datasets improve considerably the informational content of human capital proxies.

This contribution will focus on the comparative dynamics of the dispersion of educational attainment across OECD countries for the period 1960-1990. The question that is being tackled is: did educational attainment levels converge across OECD economies, or do we observe divergence in human capital accumulation?

This issue is of major relevance, since convergence in educational levels has been often claimed to be one of the motors for productivity convergence among industrialized countries (see for example Wolff, 2000). The aim of this note is to investigate the patterns of σ -convergence and σ -divergence (in the terminology introduced by Barro and Sala-i-Martin, 1992, for income levels) in educational attainment levels across OECD countries. To the knowledge of the author, there exists no comparative research tackling such an issue for different data sources. Depending on the dataset used in order to study the problem, it will be shown that the answer to the question concerning whether convergence or divergence in schooling years took place in the OECD between 1960 and 1990 can be very different.

This note presents results for the datasets by Barro and Lee (2001), Cohen and Soto (2001) and de la Fuente and Doménech (2000), which lead to contradictory answers to the question posed above. The Barro-Lee dataset is probably the most widely used reference for educational attainment in the economic literature and the Cohen-Soto dataset ranks very well in terms of signal-to-noise ratio in the comparison carried out by de la Fuente and Doménech (2002). De la Fuente and Doménech (2000, 2001, 2002) present a new database of educational attainment for 21 OECD countries. Relying on the primary sources, they correct, among other things, for changes classification criteria that may have led to implausible developments in earlier datasets. It should be noticed that, while the Barro-Lee and de la Fuente-Doménech datasets report data at a five-year periodicity, the Cohen-Soto dataset has a single observation for each decade.

This note is structured as follows. In section two, a general picture of the dynamics of the distribution of educational attainment across OECD countries is given based on each dataset. Section three presents the results of convergence/divergence tests based on the test statistic proposed by Carree and Klomp (1997) and section four concludes.

2. The Distribution of Educational Attainment: A Comparison of Datasets

The variable whose distributional dynamics we are interested in is “Average years of schooling of the adult population (over 25 years)”. This variable is reported by the three datasets being studied, and its basic descriptive statistics are given in table 1 for the sample of 21 OECD countries (Australia, Austria, Belgium, Canada, Switzerland, Germany, Denmark, Spain, Finland, France, United Kingdom, Greece, Ireland, Italy, Japan, Netherlands, Norway, New Zealand, Portugal, Sweden and United States) for each period.

Table 1: Descriptive Statistics, Average Years of Schooling for Adult Population, OECD Countries

Barro-Lee Dataset

| Year | Mean | Median | Maximum | Minimum | Std. Dev. | Skewness | Kurtosis |
|-------------|-------------|---------------|----------------|----------------|------------------|-----------------|-----------------|
| 1960 | 6.70 | 6.87 | 9.56 | 1.94 | 1.98 | -0.57 | 2.81 |
| 1965 | 6.79 | 7.17 | 9.42 | 2.24 | 1.88 | -0.59 | 2.93 |
| 1970 | 7.25 | 7.47 | 10.09 | 2.44 | 1.85 | -0.71 | 3.37 |
| 1975 | 7.50 | 7.73 | 11.00 | 2.79 | 1.94 | -0.49 | 3.20 |
| 1980 | 8.22 | 8.28 | 11.91 | 3.27 | 2.06 | -0.45 | 3.22 |
| 1985 | 8.38 | 8.40 | 11.71 | 3.57 | 1.93 | -0.55 | 3.44 |
| 1990 | 8.87 | 9.06 | 12.00 | 4.33 | 1.83 | -0.67 | 3.30 |

Cohen-Soto Dataset

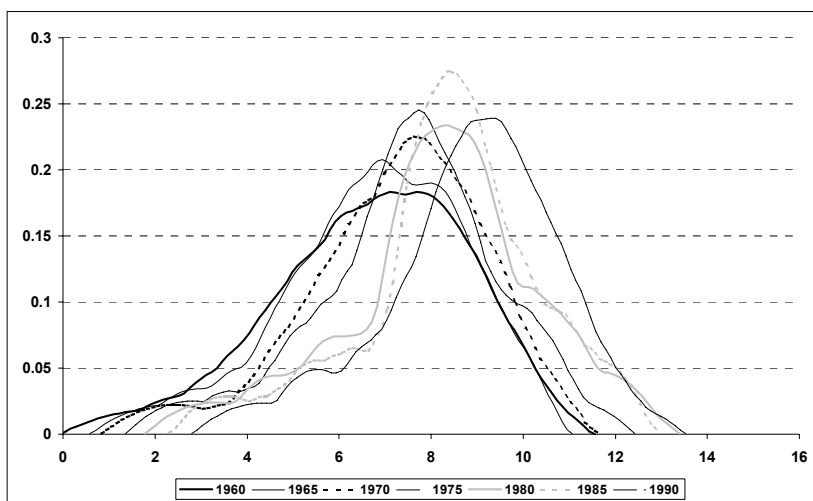
| Year | Mean | Median | Maximum | Minimum | Std. Dev. | Skewness | Kurtosis |
|-------------|-------------|---------------|----------------|----------------|------------------|-----------------|-----------------|
| 1960 | 8.07 | 8.68 | 10.96 | 3.15 | 1.86 | -0.86 | 3.45 |
| 1970 | 9.12 | 9.87 | 11.81 | 4.11 | 1.94 | -0.85 | 3.18 |
| 1980 | 10.23 | 10.72 | 12.65 | 5.57 | 1.90 | -0.77 | 2.86 |
| 1990 | 10.93 | 11.02 | 13.21 | 5.91 | 1.83 | -1.01 | 3.72 |

de la Fuente-Domenech Dataset

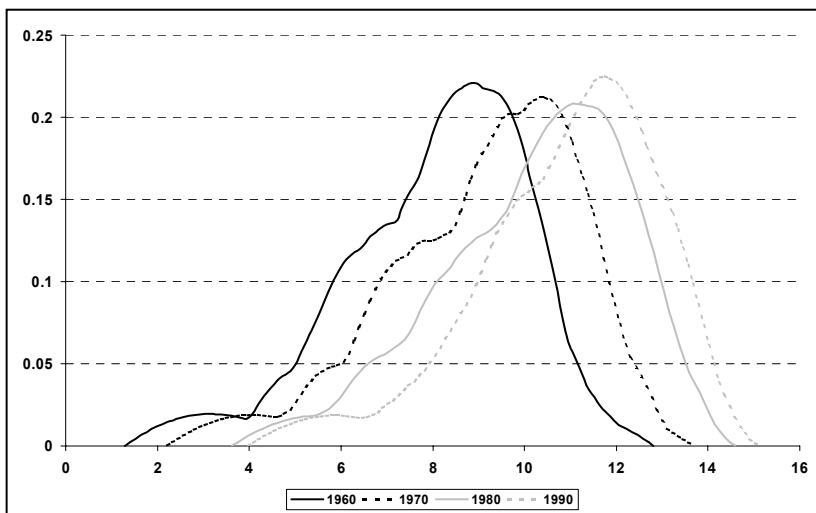
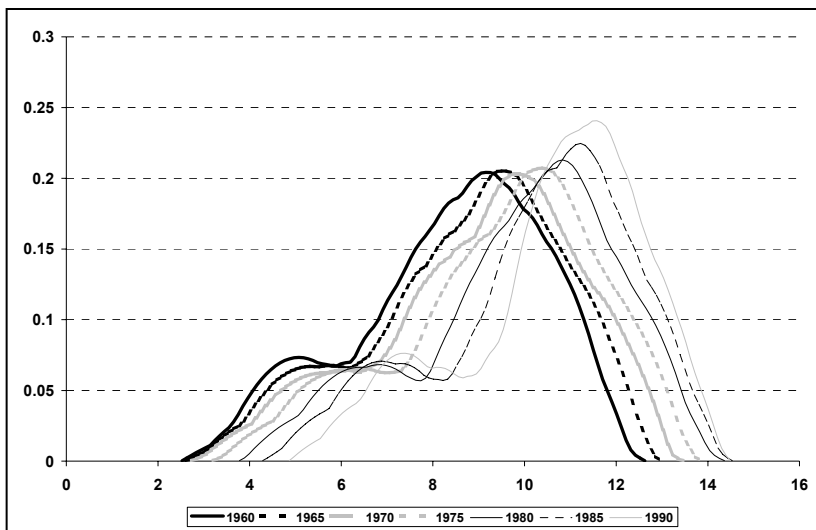
| Year | Mean | Median | Maximum | Minimum | Std. Dev. | Skewness | Kurtosis |
|-------------|-------------|---------------|----------------|----------------|------------------|-----------------|-----------------|
| 1960 | 8.36 | 8.57 | 10.78 | 4.37 | 1.95 | -0.63 | 2.34 |
| 1965 | 8.69 | 8.84 | 10.96 | 4.62 | 1.99 | -0.66 | 2.38 |
| 1970 | 9.02 | 9.10 | 11.32 | 4.87 | 2.02 | -0.69 | 2.46 |
| 1975 | 9.45 | 9.57 | 11.76 | 5.29 | 2.03 | -0.72 | 2.50 |
| 1980 | 9.87 | 9.94 | 12.41 | 5.73 | 2.03 | -0.72 | 2.54 |
| 1985 | 10.28 | 10.48 | 12.76 | 6.06 | 1.99 | -0.76 | 2.62 |
| 1990 | 10.64 | 10.97 | 12.95 | 6.41 | 1.90 | -0.81 | 2.73 |

While all datasets present a steady increase on the average level of educational attainment for the countries in the sample, the dynamics of the distribution of schooling across OECD countries implied by the three datasets are very different. Charts 1 to 3 present the density estimates (using an Epanechnikov kernel) for each dataset and each year reported. All datasets present some degree of left skewness, with the Cohen-Soto dataset presenting the highest asymmetry.¹ The de la Fuente-Doménech dataset presents a quasi-twin peaked distribution (with a relatively high concentration of mass for values in the interval 4-8 years) which is more relevant for the most recent observations and is not directly observable in the other data collections.

Chart 1: Kernel Density Estimates: Barro-Lee Dataset



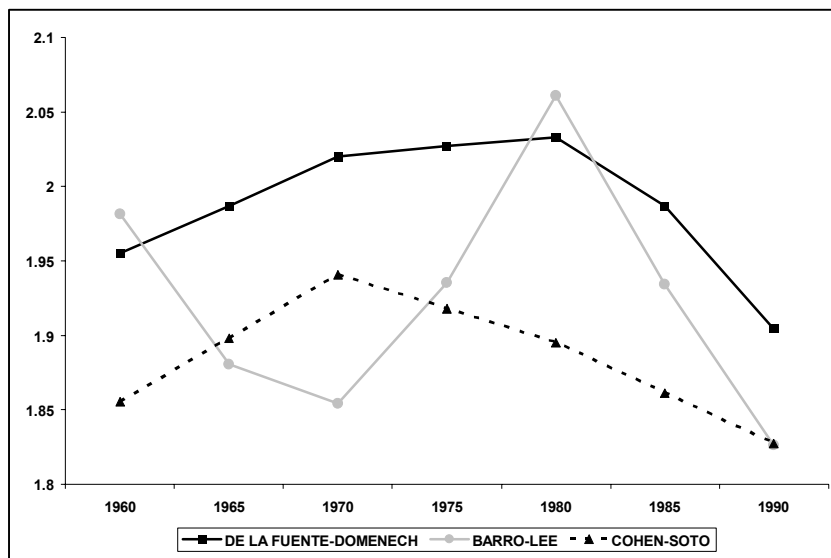
¹ It should be noticed, however, that, independently of the dataset used, the overall shape of the distribution of the variable in each period is not significantly different from that of a Gaussian distribution when tested using the Jarque-Bera test.

Chart 2: Kernel Density Estimates: Cohen-Soto Dataset*Chart 3: Kernel Density Estimates: de la Fuente-Doménech Dataset*

The differences in terms of the dynamic behavior of the second moment of the distribution, and thus of the dispersion of educational levels across OECD countries, can be grasped from the charts in Table 1 and from chart 4, which plots

the evolution of the cross-country standard deviation of years of schooling for the three datasets used in the study.²

Chart 4: Evolution of the Standard Deviation of Schooling Years



The dynamics of the dispersion of educational attainment in the OECD are extremely different depending on the source employed. The Barro-Lee dataset shows a relatively strong reduction of the cross-country standard deviation of years of schooling for the 1960s and 1980s, interrupted by a decade of divergence in the 1970s. The Cohen-Soto dataset, on the other hand, presents a slow path of continuous reduction in the dispersion of educational attainment since 1970, which was preceded by a decade of increase in the standard deviation of the variable. Finally, the de la Fuente-Doménech dataset only shows a decline in dispersion across OECD countries in the 1980s, with steady increases in the standard deviation of educational attainment in the 1960s and 1970s. To sum up, the only period for which all three datasets report the same type of dynamics in the dispersion measure is 1980-90, where all of them report convergence in schooling across OECD countries (in the sense of a reduction of the dispersion of educational attainment in the OECD). Notice that the dispersion dynamics in the Barro-Lee dataset are much more volatile than those resulting from the Cohen-Soto and de la Fuente-Doménech datasets.

The overall dynamics of educational attainment dispersion deliver thus a

² For the Cohen-Soto dataset, the observations corresponding to 1965, 1975 and 1985 were interpolated linearly.

completely different message depending on the dataset used to assess convergence in schooling for OECD countries. A further issue that needs to be tackled concerns the actual statistical significance of the changes in dispersion observed in the different data.

3. Testing for Convergence of Schooling Levels

In order to assess the statistical significance of the reductions and increases in the dispersion of educational attainment presented in table 1 and chart 4, the test introduced by Carree and Klomp (1997) will be used. Correcting an earlier proposal from Lichtenberg (1994), Carree and Klomp (1997) propose two different statistics to test for convergence. We will use the test statistic T_2 , defined as

$$T_2 = (N - 2.5) \ln \left[1 + \frac{(\hat{\sigma}_1^2 - \hat{\sigma}_T^2)}{4(\hat{\sigma}_1^2 \hat{\sigma}_T^2 - \hat{\sigma}_{1T}^2)} \right], \quad (1)$$

where $\hat{\sigma}_1^2$ is the variance of the variable being investigated in the initial period, $\hat{\sigma}_T^2$ is the variance of the variable in the final period, $\hat{\sigma}_{1T}$ is the covariance between the variable in the initial and final period, and N is the number of observations. Under the null hypothesis of equal dispersion in the initial and final periods, T_2 has a limiting $\chi^2(1)$ distribution.³

Table 2: Tests for σ -Convergence/Divergence

| Period | Barro-Lee | Cohen-Soto | de la Fuente-Doménech |
|---------|-----------|------------|-----------------------|
| 1960-65 | 3.28* | | 1.02 |
| | | 3.05* | |
| 1965-70 | 0.04 | | 1.43 |
| 1970-75 | 0.50 | | 0.05 |
| 1975-80 | 1.03 | 0.70 | 0.05 |
| 1980-85 | 4.78** | | 2.73* |
| 1985-90 | 0.64 | 1.30 | 7.37*** |

*(**)[***] stands for 10% (5%) [1%] significant. The figures refer to the T_2 test statistic in Carree and Klomp (1997), $\chi^2(1)$ distributed under the null hypothesis of equal variance.

³ The choice of T_2 over the other alternative put forward in Carree and Klomp (1997), T_3 , can be easily justified as T_3 would require the estimation of an autoregressive parameter for the educational attainment data. Given the reduced size of the time dimension in the panel of OECD countries for the data available, inference based on this test statistic would render unreliable results.

Table 2 presents the results of the test for all subperiods and datasets. For the Barro-Lee dataset, the 1960-65 and 1980-85 periods present significant σ -convergence in educational attainment levels for OECD economies. This last period of significant convergence in schooling is also found in the de la Fuente-Doménech dataset, together with further convergence in the subperiod 1985-90. The Cohen-Soto dataset paints a very different picture concerning the evidence of changes in the dispersion of educational attainment. In this case, the only statistically significant change in dispersion takes place in the subperiod 1960-70, and it is in the direction of σ -divergence in educational attainment.

The results of the Carree and Klomp (1997) test emphasize thus the contradictory results obtained by the visual analysis of the evolution of the dispersion of educational attainment. Not only do the overall dynamics of dispersion differ extremely across datasets, but also the statistically significant changes do not coincide for the different data. There is no single period for which all three dataset offer a unified picture of significant change in the standard deviation of schooling, and for some cases the answer to whether convergence took place is answered completely differently depending on the dataset used. The experience in the 1960s is a clear example: while using the Barro-Lee dataset one would conclude that there was (10%) significant convergence in the period 1960-65, using the Cohen-Soto dataset the conclusion would be that the 1960s were marked by (10%) significant divergence in educational attainment in OECD countries, and the de la Fuente-Doménech dataset does not find any significant change in the second moment of the distribution in the whole decade.

4. Conclusions

This note shows that the answer to the question whether convergence of educational attainment levels across OECD countries happened in the period 1960-1990 depends strongly on the dataset used for the analysis. Three datasets were studied, including two of the most recently developed data collections on human capital variables (Cohen and Soto, 2001, and de la Fuente and Doménech, 2000, 2002) and the most widely used dataset on educational attainment in the empirical economic growth literature (Barro and Lee, 1993, 2001). It was shown that the dynamics of dispersion in educational attainment for OECD countries differ enormously across datasets, as do the results of the Carree and Klomp (1997) test of significance in the change of standard deviation between periods. The three datasets provide contradictory conclusions on both the existence and evolution of convergence of educational attainment in industrialized countries.

References

- Barro, R. and J.W. Lee (1993). International measures of schooling years and schooling quality. *American Economic Review, Papers and Proceedings* 86, pp. 218-223.
- Barro, R. and J.W. Lee (2001). International data on educational attainment: Updates and implications. *Oxford Economic Papers* 53, pp. 541-563.
- Barro, R. and Sala-i-Martin, X. (1992). Convergence. *Journal of Political Economy* 100, pp. 223-251.
- Benhabib, J. and M. Spiegel (1994). The role of human capital in economic development: Evidence from aggregate cross-country data. *Journal of Monetary Economics* 34, pp. 143-174.
- Carree, M. A. and L. Klomp (1997). Testing the convergence hypothesis: A comment. *Review of Economics and Statistics* 79, pp. 683-686.
- Cohen, D. and M. Soto (2001). Growth and human capital: good data, good results. *CEPR Discussion Paper 3025*.
- De la Fuente, A. and R. Doménech (2000). Human capital in growth regressions: how much difference does data quality make? *CEPR Discussion Paper 2466* and *Economics Department Working Paper 262*, OECD, Paris.
- De la Fuente, A. and R. Doménech (2001). Schooling data, technological diffusion and the neoclassical model. *American Economic Review, Papers and Proceedings* 91, pp. 323-27.
- De la Fuente, A. and R. Doménech (2002). Human capital in growth regressions: How much difference does data quality make? An update and further results. *CEPR Discussion Paper 3587*.
- Durlauf, S. and D. Quah (1999). The new empirics of economic growth. Chapter 4 in *Handbook of Macroeconomics*, John B. Taylor and Michael Woodford (eds.), North-Holland, 1999.
- Krueger, A. and M. Lindahl (2001). Education for growth: why and for whom? *Journal of Economic Literature* XXXIX, pp. 1101-36.
- Lichtenberg, F. L. (1994). Testing the convergence hypothesis. *Review of Economics and Statistics* 76, pp. 576-579.
- Pritchett, L. (1996). Where has all the education gone? *World Bank Policy Research Working Paper 1581*, World Bank.
- Wolff, E.N. (2000). Human capital investment and economic growth: exploring the cross-country evidence. *Structural Change and Economic Dynamics* 11, pp. 433-472.