

SCHOOLING DATA, TECHNOLOGICAL DIFFUSION

AND THE NEOCLASSICAL MODEL

By ANGEL DE LA FUENTE AND RAFAEL DOMÉNECH

Growth economists have spent more than forty years slowly chipping away at the Solow residual, largely by attributing increasingly larger chunks of it to investment in human capital. A few years ago we were reasonably certain that this was the way to go. But an increasing number of studies seem to be telling us that the effect of schooling variables on productivity vanishes when we turn to what seem to be the appropriate econometric techniques for the purpose of estimating growth equations. Should we take these results at face value? Before we do so and abandon the only workable models we have, it seems sensible to search for ways to reconcile recent empirical findings with some kind of plausible theory. In this paper we argue that we can make a fair amount of progress in this direction by combining two ingredients: better data on human capital, and a further extension of the human capital-augmented neoclassical model that allows for cross-country productivity differentials and for technological diffusion.

I. Some New Data

Poor data quality is widely recognized as a prime suspect for the counterintuitive results on human capital and growth found in some recent studies. If human capital stocks have been measured with error (and we have every reason to believe this is the case), their first differences will be even less accurate than their levels, a fact that will bias the relevant coefficients towards zero in many sensible specifications. To assess the importance of the problem, we have constructed new educational attainment series for a sample of 21 OECD countries covering the period 1960-90 and checked their performance against some previously available data sets in a

number of standard growth accounting specifications.¹ As we will see, the results are consistent with our initial suspicion that data quality can make a big difference.

Our schooling series are essentially a revised version of (a subset of) Robert J. Barro and Jong-Wha Lee's (1996) data set. We have attempted to increase the signal to noise ratio in the data by exploiting a variety of sources not used by these authors, and by eliminating sharp breaks in the series that can only arise from changes in data collection criteria. Our approach has been to collect all the information we could find on educational attainment in OECD countries, both from international publications and from national sources, and use it to try to reconstruct a plausible pattern, reinterpreting when necessary some of the data from international compilations as referring to somewhat broader or narrower schooling categories than the reported one. This clearly involves a fair amount of guesswork, but as we argue in detail elsewhere, there seems to be no feasible alternative given the lack of homogeneity of the primary data.²

The series we construct differ significantly from Barro and Lee's both in their cross-section and in their time-series profiles. Although the correlation of average years of schooling across data sets is relatively high (0.88), there are significant differences in the relative positions of a number of countries. Perhaps more important for our purposes here is the fact that the time profiles of our schooling series are considerably smoother and more plausible than those of Barro and Lee's original data.

[Insert Figure 1]

Barro and Lee's series display a large number of sharp breaks that give a distorted image of the pattern of human capital accumulation and may obscure its relationship with productivity growth. This is clearly illustrated in Figure 1, where we have plotted the fitted distribution of the annualized growth rate of average years of schooling for all countries and years in each data set. The difference in the range of this variable across data sets is enormous: while our annual growth rates range between 0.15% and 2%, Barro and Lee's go from -1.35% to 7.80%; moreover, 15.9% of their observations are negative, and 19% of them exceed 2%. The elimination of these

implausible observations, moreover, yields a completely different picture of schooling growth over time. As shown in Figure 2, the correlation of the growth rate of average years of schooling across data sets is almost nil.

[Insert Figure 2]

We suspect that these features may help explain why the Barro and Lee data often generate implausible results in growth regressions, particularly when these are estimated using panel or first difference specifications. A first indication of this is that the coefficient of a univariate regression of the growth rate of productivity on the growth rate of schooling (with both variables measured as deviations from their contemporaneous sample averages) increases from 0.174 (with a t-ratio of 1.56) with the Barro and Lee data to 1.211 (with a t-ratio of 3.92) with our revised series. The results we report in the following section are also consistent with this hypothesis.

II. A Simple Model

In this section we estimate a simple growth regression that extends a standard aggregate production function with human capital by allowing for technological diffusion and for permanent TFP differences across countries. Following de la Fuente (1996), we estimate an equation of the form

$$(1) \mathbf{D}q_{it} = \mathbf{G}_0 + \mathbf{g}_i + \mathbf{h}_t + \mathbf{aD}k_{it} + \mathbf{bD}h_{it} + \mathbf{1}b_{it} + \mathbf{e}_{it}$$

where \mathbf{D} denotes annual growth rates (over the subperiod starting at time t), q_{it} is the log of output per employed worker in country i at time t , k the log of the stock of physical capital per worker, h the log of the average number of years of schooling of the adult population and \mathbf{h}_t and \mathbf{g}_i are fixed time and country effects. The only non-standard term, b_{it} , is a technological gap measure which enters the equation as a determinant of the rate of technical progress in order to allow for a catch-up effect. This term is the Hicks-neutral TFP gap between each country and the US at the beginning of each subperiod, given by

$$(2) b_{it} = (q_{US,t} - \mathbf{a}k_{US,t} - \mathbf{b}h_{US,t}) - (q_{it} - \mathbf{a}k_{it} - \mathbf{b}h_{it})$$

To estimate the model we substitute (2) into (1) and use NLS on the resulting equation with data on both factor stocks and their growth rates. Notice that in this specification the country dummies will pick up permanent cross-country differences in relative TFP levels that will presumably reflect differences in R&D investment and other omitted variables. The parameter λ measures the rate of (conditional) technological convergence. The productivity data are taken from an updated version of Teresa Dabán, Rafael Doménech and César Molinas (1997), who replicate Robert Summers and Alan Heston's (1991) data set for the OECD using a set of purchasing power parities specific to this sample. We use pooled data at five-year intervals starting in 1960 and ending in 1990 for Barro and Lee's and our own data set, and in 1985 for the one constructed by Vikram Nehru, Eric Swanson and Ashutosh Dubey (1995).

[Insert Table 1]

The pattern of results that emerges in Table 1 as we change the source of the human capital data is consistent with our hypothesis about the importance of educational data quality for growth estimates. The human capital variable is significant and displays a reasonable coefficient with our revised data (D&D), but not with the Barro and Lee (B&L) or Nehru et al (NSD) series, which actually produce a negative human capital coefficient. Moreover, the coefficients of the stocks of physical and human capital estimated with our data are quite plausible, with α only slightly above capital's share in national income (which is 0.35 in this sample) and β only slightly below N. Gregory Mankiw, David Romer and David N. Weil's (1992) preferred estimate of 1/3.

III. How Full Is the Neoclassical Glass?

As Mankiw (1995) has argued, most of the results found in the early convergence literature are consistent with an extended neoclassical model built around an aggregate production function that includes human capital as a productive input but assumes that all countries have access to a common technology.³ Mankiw's conclusion that such a model provides a satisfactory account of the growth process and of the determinants of income levels, however, has been challenged by

Nazrul Islam (1995), Lant Pritchett (1995) and Francesco Caselli, Gerardo Esquivel and Fernando Lefort (1996) among others. These studies have produced rather discouraging results that suggest, in particular, that educational investment is not productive and that the bulk of income differences across countries has little to do with differences in stocks of productive factors.

The validity of the augmented neoclassical model depends on the extent to which cross-country productivity differentials can be attributed to factor endowments rather than to total factor productivity (TFP). In this section we will attempt to gauge the relative importance of these two factors using the model and the data described above. The exercise is similar in spirit to the one performed by Peter Klenow and Andrés Rodríguez-Clare (1997), but it is conducted using a refined data set that should help improve the quality of TFP estimates and an empirically-based set of production function parameters.

We recover the Hicks-neutral technological gap between each country and a fictional average economy to which we attribute the observed sample averages of log productivity (q) and log factor stocks per employed worker (k and h). Thus, we define relative TFP ($tfprel$) by

$$(3) \quad tfprel_{it} = (q_{it} - \mathbf{a}k_{it} - \mathbf{b}h_{it}) - (q_{av_{it}} - \mathbf{a}k_{av_{it}} - \mathbf{b}h_{av_{it}}) = qrel_{it} - (\mathbf{a}krel_{it} + \mathbf{b}hrel_{it})$$

where av denotes sample averages and rel deviations from them. To obtain a summary measure of the importance of TFP as a source of productivity differentials, we regress relative TFP on relative productivity. (Notice that the regression constant will vanish because both variables are measured in deviations from sample means). The estimated coefficient gives the fraction of the productivity differential with the sample average explained by the TFP gap in a typical country.

The average TFP share in relative productivity rises consistently over the sample period, from 0.353 in 1960 to 0.472 in 1990. That is, TFP differences seem to have become relatively more important over time in explaining productivity disparities. Towards the end of the sample period, one half of the productivity differential with the sample average can be traced back to differences in technical efficiency, with the other half being attributable to differences in factor stocks. The message is similar if we use Klenow and Rodríguez-Clare's estimates of the TFP gap, as the TFP

share estimated with these data is 0.495 in 1985.⁴ These figures stand approximately half way between the conclusions of Mankiw (1995), who attributes the bulk of observed income differentials to factor endowments, and those of Caselli et al (1996) and some other recent panel studies, where fixed effects that presumably capture TFP differences account for most of the observed cross-country income disparities.⁵ We view our results as an indication that, while the augmented neoclassical model prevalent in the literature does indeed capture some of the key determinants of productivity, there is a clear need for additional work on the dynamics and determinants of the level of technical efficiency.

IV. Conclusion

A number of authors have recently called attention to the crucial role of technical efficiency in understanding productivity disparities across economies and questioned the capacity of the human capital-augmented neoclassical model with a common technology to explain the international distribution of income. In this paper we have assessed the quantitative importance of this factor using a simple growth specification that can be seen as a further extension of (the technological components of) an augmented neoclassical model that allows for cross-country differences in TFP levels and for technological diffusion. We have estimated this specification using a revised data set on schooling for a sample of OECD countries and found that it explains 80% of the variation in the growth rate of productivity and that it yields sensible technological parameters. We have then used the model and the underlying data to quantify the contributions of factor stocks and levels of technical efficiency to observed productivity differentials. Our results show that the relative importance of TFP differences is considerable and that it has increased over time to account for about one half of the productivity differentials observed at the end of the sample period. These findings reinforce recent calls by Edward C. Prescott (1998) and other authors for better models of technical progress as a key ingredient for understanding international income dynamics while preserving an important role for factor stocks as a source of cross-country income disparities.

REFERENCES

- Barro, Robert J. and Sala i Martin, Xavier.** "Convergence." Journal of Political Economy, April 1992, 100(2), pp. 223-51.
- Barro, Robert J. and Lee, Jong-Wha.** "International Measures of Schooling Years and Schooling Quality." American Economic Review, May 1996 (Papers and Proceedings), 86(2), pp. 218-23.
- Caselli, Francesco; Esquivel, Gerardo and Lefort, Fernando.** "Reopening the Convergence Debate: a New Look at Cross-country Growth Empirics." Journal of Economic Growth, September 1996, 1(3), pp. 363-89.
- de la Fuente, Angel.** "On the Sources of Convergence: A Close Look at the Spanish Regions." CEPR Discussion Paper no. 1543, 1996.
- de la Fuente, Angel and Doménech, Rafael.** "Human Capital in Growth Regressions: How Much Difference Does Data Quality Make?" CEPR Discussion Paper no. 2466, 2000.
- Dabán, Teresa; Doménech, Rafael and Molinas, César.** "International and Intertemporal Comparisons of Real Product in OECD Countries: a Growth Sensitivity Approach." Review of Income and Wealth, March 1997, 43(1), pp. 33-48.
- Islam, Nazrul.** "Growth Empirics: a Panel Data Approach." Quarterly Journal of Economics, November 1995, 110(4), pp. 1127-70.
- Klenow, Peter and Rodríguez-Clare, Andrés.** "The Neoclassical Revival in Growth Economics: Has it Gone Too Far?" in Ben Bernanke and Julio Rotemberg, eds., Macroeconomics Annual 1997. Cambridge, MA: MIT Press, 1997, pp. 73-102.

Mankiw, N. Gregory. "The Growth of Nations." Brookings Papers on Economic Activity, 1995, I, pp. 275-326.

Mankiw, N. Gregory; Romer, David and Weil, David N. "A Contribution to the Empirics of Economic Growth." Quarterly Journal of Economics, May 1992, 107(2), pp. 407-37.

Nehru, Vikram; Swanson, Eric and Dubey, Ashutosh. "A New Database on Human Capital Stocks in Developing and Industrial Countries: Sources, Methodology and Results." Journal of Development Economics, April 1995, 46(2), pp. 379-401.

Prescott, Edward C. "Needed: A Theory of Total Factor Productivity." International Economic Review, August 1998, 39(3), pp. 525-51.

Pritchett, Lant. "Where Has All the Education Gone?". Mimeo, World Bank, 1995.

Summers, Robert and Heston, Alan. "The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-88." Quarterly Journal of Economics, May 1991, 106(2), pp. 327-368.

* De la Fuente: Instituto de Análisis Económico (CSIC), Campus UAB, 08193 Bellaterra, Barcelona, Spain; Doménech: Departamento de Análisis Económico, Universidad de Valencia, Campus de los Naranjos, 46022 Valencia, Spain. This research was supported by the Spanish Ministry of Science and Technology through CICYT grants SEC99-1189 and SEC99-0820 and by the Spanish Ministry of Finance and Fundación Caixa Galicia. We would like to thank José E. Boscá, Jonathan Temple and Antonio Ciccone for their helpful comments and suggestions.

¹ These data are available at <http://iei.uv.es/~rdomenec/human.html>. They refer to all the countries included in Mankiw, Romer and Weil's (1992) OECD sample with the exception of Turkey.

² See de la Fuente and Doménech (2000) for a review of other data sets, a discussion of the procedure used to construct these series and detailed empirical results.

³ See for instance Barro and Sala-i-Martin (1992) and Mankiw, Romer and Weil (1992).

⁴ These authors actually report a number close to $2/3$ because they attribute to TFP differences an estimate of their indirect effects through induced factor accumulation. We consider only the direct contribution of the TFP gap.

⁵ Using our 1990 data and Caselli et al's most "plausible" parameter estimates ($\alpha = 0.107$ and $\beta = 0.00$), the share of TFP in relative productivity is 0.90.

Table 1: Results of the estimation with different human capital data

	[1]	[2]	[3]
<i>human capital data:</i>	<i>NSD</i>	<i>B&L</i>	<i>D&D</i>
a	0.510 (8.30)	0.409 (6.12)	0.373 (7.15)
b	-0.148 (2.62)	-0.057 (0.88)	0.271 (2.53)
I	0.100 (6.98)	0.063 (8.27)	0.068 (6.34)
<i>adj. R²</i>	0.840	0.811	0.809
<i>std. error reg.</i>	0.0074	0.0079	0.0079

- *Notes:* White's heteroscedasticity-consistent t ratios in parentheses. Only significant country dummies are left in the reported equation.

**Figure 1: Distribution of the annual growth rate of average years of schooling:
Barro and Lee vs. de la Fuente and Doménech.**

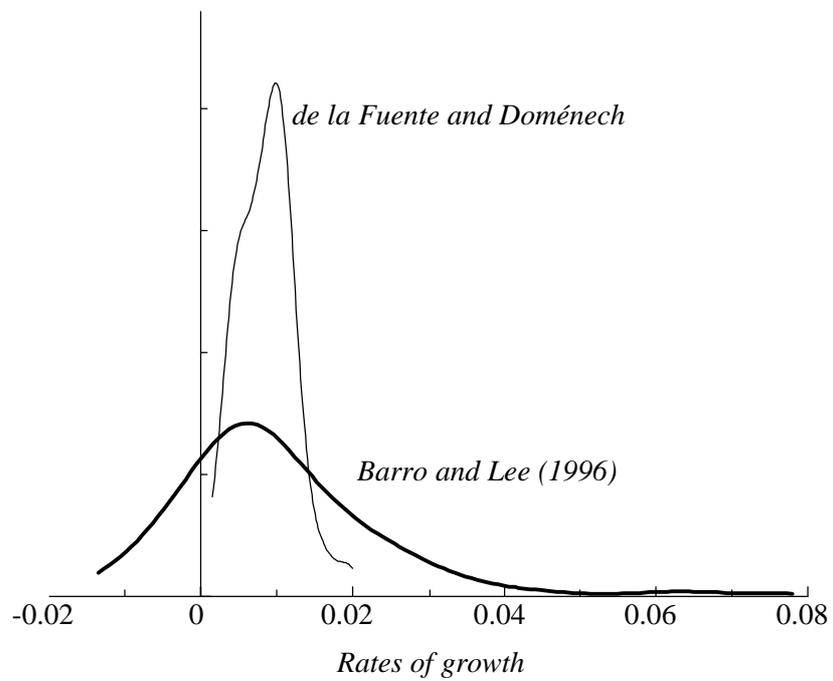


Figure 2: Correlation across data sets of the growth rate of average years of schooling.

