# Average years of education in Hungary: annual estimates 1920-2006

## Introduction

These data and descriptions are obtained from the following paper:

Péter Földvári and Bas van Leeuwen, 'Average years of education in Hungary: annual estimates 1920-2006', 30 May 2008.

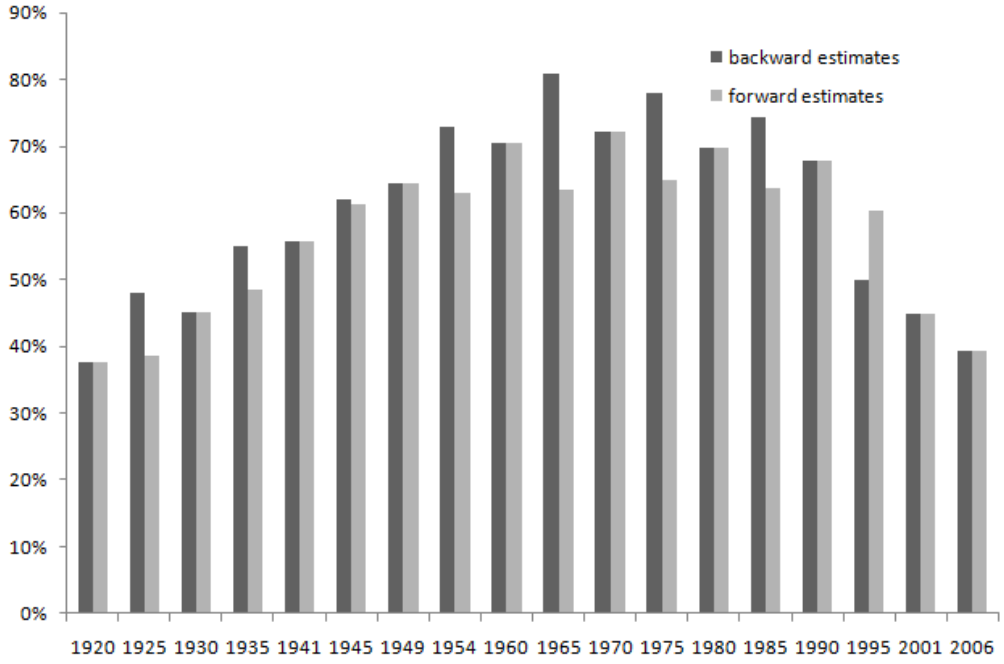
## Data construction

The estimates are based on a modified Barro and Lee (1993; 2003) methodology. This requires data on educational attainment from the censuses. The most reliable source of data is the population census available from the Hungarian Central Statistical Bureau (KSH, http://www.nepszamlalas.hu/eng/index.html). For the years of censuses (1920, 1930, 1941, 1949, 1960, 1970, 1980, 1990, 2001) we have good quality data available on the educational attainment of the population. We also used the estimates of the Government Commission on Population Policy (http://www.nepinfo.hu/) for 2006. For these benchmark years we can estimate the average years of education by assigning a length to each educational level attained.

In the next step, we used the Barro and Lee method to interpolate these benchmark years. Since the Barro-Lee method is designed to provide estimates for each fifth year, we need to modify it in order to arrive at annual estimates. An additional problem is that the Barro and Lee method has difficulties with dealing with repeaters and drop outs of the education system. Although in their 2001 estimates they try to solve this problem, the corrections must necessarily be limited, especially for the pre-WWII years when less and lower quality data are available. Their most important implicit assumption is that mortality is equal among different educational groups. A consequence of this approach is that when a population's educational attainment grows considerably, and

hence mortality decreases in the more educated groups, their method tend to underestimate the average years of education.

The mortality assumptions thus cause an underestimation of the growth of attainment (and the average years of education) when the attainment is estimated forward from the previous census. This is even more apparent in below Figure, where we present the




the share of the population with primary education estimated forward and backward (calculated back from the next census). As the mortality assumption leads to an underestimation of the growth of attainment, estimating backward results in an overestimation of average years of education.

To avoid these issues we use a simple correction method. We start with the standard Barro and Lee equations which we slightly modified to allow for the estimation of annual series:

$$h_{0,t} = H_{0,t}/L_t = h_{0,t-i}\left[1-\left(L15_t \cdot i/5 \cdot L_t\right)\right]+\left(L15_t \cdot i/5 \cdot L_t\right)\cdot\left(1-PRI_{t-i}\right)$$

$$h_{1,t} = H_{1,t}/L_t = h_{1,t-i}\left[1-\left(L15_t \cdot i/5 \cdot L_t\right)\right]+\left(L15_t \cdot i/5 \cdot L_t\right)\cdot\left(PRI_{t-i}-SEC_t\right)$$

$$h_{2,t} = H_{2,t}/L_t = h_{2,t-i}\left[1-\left(L15_t \cdot i/5 \cdot L_t\right)\right]+\left(L15_t \cdot i/5 \cdot L_t\right)\cdot SEC_t-\left(L20_t \cdot i/5 \cdot L_t\right)\cdot HIGH_t$$

$$h_{3,t} = H_{3,t} / L_t = h_{3,t-i}\left[1-\left(L15_t \cdot i/5 \cdot L_t\right)\right]+\left(L20_t \cdot i/5 \cdot L_t\right)\cdot HIGH_t$$

where *h* is attainment per level of education (0= no education, 1=prim, 2=sec, 3=high). *H* is the total number of people in the population with a certain education level, *i* is the length of the forward estimation (so if one wants to estimate attainment in year *t* based on attainment 5 years ago, *i* would be 5, yielding the standard Barro-Lee formula), *L* is the total population aged 15 years or older, *L15* is total population aged 15-19, *L20* is total population aged 20-24, *PRI, SEC, HIGH* are the enrolment ratios in primary, secondary, and higher education respectively.

As can be seen in above Figure, this method leads to an underestimation of the actual attainment (and hence, of the average years of education) because of the assumption that mortality is independent is of the education level. If we calculate backwards, the bias is exactly reversed (see Figure). Simply rewriting above equations we arrive at the formula to estimate the attainment backwards:

$$h_{0,t-i} = \left(h_{0,t} - \left(L15_t \cdot i/5 \cdot L_t\right)\cdot\left(1-PRI_{t-i}\right)\right)\Big/\left(1-L15_t \cdot i/5 \cdot L_t\right)$$

$$h_{1,t-i} = \left(h_{1,t} - \left(L15_t \cdot i/5 \cdot L_t\right)\cdot\left(PRI_{t-i} - SEC_t\right)\right)\Big/\left[1-\left(L15_t \cdot i/5 \cdot L_t\right)\right]$$

$$h_{2,t-i} = \left(h_{2,t} - \left(L15_t \cdot i/5 \cdot L_t\right)\cdot SEC_t + \left(L20_t \cdot i/5 \cdot L_t\right)\cdot HIGH_t\right)\Big/\left[1-\left(L15_t \cdot i/5 \cdot L_t\right)\right]$$

$$h_{3,t-i} = \left(h_{3,t} - \left(L20_t \cdot i/5 \cdot L_t\right)\cdot HIGH_t\right)\Big/\left[1-\left(L15_t \cdot i/5 \cdot L_t\right)\right]$$

The next step is that we estimate the attainments both backward and forward. We have census data for about every 10 years. So, if we have census data for 1960 and 1970, we calculate the figure for 1965 from both the 1960 and the 1970 benchmark and calculate their average. This should net out the effect of repeaters, mortality, and drop outs (which means removing the downward bias caused by using the forward method under the assumption that these three factors remain constant between two benchmarks). These assumptions are far less restrictive than assuming that mortality is uncorrelated with education or simply ignoring repeaters and dropouts.

As we now have attainment figures for 1965 and 1970, we again calculate attainment for the year between these benchmarks (1967). We do the same for 1969 (based on 1967 and 1970), etc. The main rule is that in each step one needs to estimate the year that is in the middle between two benchmarks. If the forward estimate

(underestimation) is made over a longer time period than the backward estimate (overestimation), the downward bias could not be offset by the upward bias. Using this method, all implausible fluctuations drop out.

**References**

Barro, R. W. and J.-W. Lee, “International Comparisons of Educational Attainment”, *Journal of Monetary Economics* 32, no. 3 (December 1993): 363-394.

Barro, R. W. and J.-W. Lee, “International Data on Educational Attainment: Updates and Implications”, *Oxford Economic Papers* 53, no.3 (July 2003): 541-563.