5 New estimates of the formation and stock of human capital

1. INTRODUCTION

After defining human capital (chapter 2), collecting data (chapter 3), and looking at the development of the education systems (chapter 4), we will combine this quantitative and qualitative evidence into a set of estimates of the stock of human capital for the period 1890 to 2000 that conforms to the definition we use. Such a set of estimates is important as we need to construct a stock that reflects, as far as possible, all aspects of human capital. That is, not only must it reflect the quantity of human capital, but also the change in its quality.

To that end, we start in section 2 with a brief outline of the data and some of the problems we encountered. In section 3, we turn to the estimation procedure of the stock of human capital in the 1980s and 1990s based on household surveys. We then move on to the construction of time series of the stock of human capital in section 4. This includes both the estimation of the appreciation/depreciation of the stock and the construction of a Perpetual Inventory Method (PIM) to bring the human capital series back to 1890. To this end we use the human capital estimates for age cohorts from the household surveys together with private and public expenditure on education and foregone wages. Following this construction of the human capital stock, in section 5 we turn to the analysis of this series. After estimating the subjective margins of error and taking a closer look at the components, we claim that this stock has a considerable degree of plausibility in view of the definition of human capital used. Therefore, it is more suitable in growth regressions than most of the current human capital proxies. We end in section 6 with a brief conclusion.

2. DATA AND MEASUREMENT ISSUES

There are three methods to estimate human capital.[91] First, there are proxies of human capital. These became especially popular with the Penn World Tables (Summers and Heston 1988; 1991) as it became possible to perform cross-country analyses which required a large human capital database. The most famous example of such proxies is the dataset by Barro and Lee (1993; 2001) consisting of five-yearly estimates of ‘average years of education’ in the population aged 15 (or 25) and over. However,

[91] For excellent overviews of the available human capital estimates see Le, Gibson, and Oxley (2003) and Wöβmann (2003).

besides the low signal (they only to a limited extent reflect 'true' human capital), the main problem is that these data only capture the quantity of human capital.

There exist two alternative estimation methods. The first one is the prospective method, which looks at future earnings to calculate human capital. The second one is the retrospective method, which focuses on the investments in human capital. The advantage of these two methods is that they express human capital in monetary terms and keep into account the heterogeneity of labour. However, the definition of human capital may be problematic. For example, should it include the costs of rearing a child? In addition, these methods often require much data which are mostly only available for recent years, so only for a few years estimates can be made.

Some methods have been developed in order to combine the strengths and to avoid the weaknesses of the pro-and retrospective methods (Tao and Stinson 1997; Dagum and Slottje 2000). Yet, these methods are also data-intensive. Therefore, up till today, only a handful of studies use these techniques.

Since the combined approach offers the highest accuracy, our objective is to construct a human capital stock with a much more limited demand for data. In addition, we aim to adapt our estimates in such a way that they can easily be used to empirically test the new growth theories. We will use a slightly modified definition of the OECD (2001, 18), defining human capital as "the knowledge, skills, and competencies embodied in individuals that facilitate the creation of personal, social and economic well-being."[92] This definition excludes factors which relate to either physical aspects, like the costs of raising a child, or that relate to non-physical factors, which are nevertheless inherent to a physical person, such as innate ability. In other words, human capital consists of all forms of knowledge acquired with the exclusion of both innate abilities and the costs of reproducing labour.

The factors affecting the acquiring of knowledge, which together make up human capital, are generally unobservable. This is also the reason why most studies that tried to calculate the monetary value of human capital either used the input (costs) or the output (future earnings), or a combination of them. Although applying it in a different way, we use a latent variable approach[93] similar to Dagum and Slottje (2000).

[92] Please note that we excluded 'human attributes' from this definition as innate characteristics do not have an investment component.

[93] A latent, or unobservable, variable approach makes use of the relation between the unobservable variable and other available variables. It is likely that human capital is strongly related to social variables, e.g. age, sex, social status.

We estimate the probability of an individual having a 'relatively high educational level' and construct an index reflecting 'educational capital', that is, the total per capita stock of human capital minus 'experience' and 'on the job training' (see table 5.1). Next, we use the total spending on the education of low-educated young individuals in order to estimate a benchmark level of 'educational capital' for the entire population aged 16-65. This is based on Tao and Stinson (1997), just like the use of an earnings equation to estimate experience and on the job training.

Table 5.1: The division of the estimated stock of human capital on the basis of surveys by data, method of estimation.

Share of human capital	Data	Method
Educational Capital	a) Public expenditure on education (local and national)	Cost-based
	b) Private direct expenditure (school fees, books)	Cost-based
	c) Indirect expenditure (foregone wages)	cost-based
	d) Non-government, non-private effect**	Residual
Experience*	Experience part of earnings	Income-based

*Experience does also include 'on-the-job training' as it can also cause a rise in wage by age-class.
** While estimating the 'educational capital' index, we condition these index values on variables such as age, sex, and occupation. However, it is likely that, especially for older persons, the 'educational capital' (and as a consequence the index values) is higher than one might expect purely based on public and private expenditure and on foregone wages. Therefore, subtracting the latter from 'educational capital' gives the 'non-government, non-private' part of educational capital as can be inferred from this table. As a consequence, the 'non-government, non-private' effect consists of the share of educational capital that cannot be attributed to public or private expenditure on education or to foregone wages. This effect therefore mainly consists of factors such as home education.

In short, our method has a cost- and an income-based component. The estimates of the human capital stock we obtain from the household surveys in the next section reflect the effect of public and private expenditure on education, the foregone wages (the cost-based component), 'experience', and 'on the job training' (the income based components). While the latter components can only directly be estimated using household surveys, the former can also be estimated using alternative sources, such as:

a) Skilled and unskilled yearly wages.
b) Per student public expenditure on education per level of education.
c) Per student private expenditure on education.

These latter sources will be used in section 4 to bring the estimated human capital stock further back in time.

We will not go into detail regarding the alternative sources under point a-c because these have been extensively discussed in chapter 3 and appendix A.1 and A.8.

The only point we want to make here is that for Japan for the period 1954-2002 we use the more detailed wage data by education level and by sex which are available from the *Japan Statistical Yearbook* and the *Historical Statistics of Japan*. For the period before 1954 a simple distinction in skilled and unskilled wages was used as is given in appendix A.1.

The household surveys require more attention. In order to estimate the human capital stock we need surveys with individual level data for Japan, India, and Indonesia (see table 5.2). For Indonesia we used the Sakernas (Indonesian Labour Force Survey) and for India the National Sample Survey. The latter surveys report for about every fifth

Table 5.2: Overview of the surveys used for the estimation of the human capital stock for India, Indonesia, and Japan in the 1990s

Country	Name	Sort survey	Issuer	Remarks
India	National Sample Survey	National sample	India, National Sample Survey Organisation	The subject Employment & unemployment covered in the surveys used in this text are issued once every five years.
Indonesia	Sakernas (Indonesia Labour Force Survey)	Labour market	Indonesia, Badan Pusat Statistik	Started in 1976, but the data for the early years are unreliable. Therefore we used the surveys from the early 1990s.
Japan	International Social Survey Programme	Social survey aimed at the entire population	ISSP, Japan	Covers yearly changing social aspects. Has a common core with data on age, earnings, education.

year data on the labour force. For Japan we used the International Social Survey Programme (ISSP).

At this point two problems arise. First, for some individuals in a survey no information on education was given. This occurs in each survey a few times. Yet, it is possible that omitting these observations would create a bias, since these individuals were likely to have a lower number of years of education on average. Therefore, we imputed the missing education variables using a regression on other variables such as age, occupation, sex, marital status. An example of this exercise is presented in table 5.3 for Japan although the same also applies to India and Indonesia. It is clear that the number of imputations in the variable 'years of education' is only marginal, never exceeding 1% of the sample. Equally, on average the imputed 'years of education' are lower than the average of the sample. Although the share of missing observations in each sample is small, their effect can be considerable when one looks at the results per age. As we need the results per age to calculate possible appreciation or depreciation of human capital, we need to correct for these missing observations.

The second point of concern in using household surveys is that of weighting. Although, for example, the surveys for Japan have no weights since they are constructed such that they are representative of the Japanese society, some

Table 5.3: Number of imputations of missing education variables and the average years of education of the entire sample and of the imputed individuals in the household surveys in Japan 1993-2002

	Sample*		Imputations		
Survey	Size of the sample	average years education in the sample	Number of imputations in each sample	% in sample	average years of education of the imputed values
1993	1.138	12.1	4	0.35%	9
1994	1.130	12.0	2	0.18%	9
1995	1.064	12.4	2	0.19%	12
1996	1.045	12.3	6	0.57%	12.2
1998	1.131	12.6	3	0.27%	12
2000	937	12.7	5	0.53%	12
2002	899	12.8	1	0.11%	10

*Including the imputed values

modifications need to be implemented. Even in surveys where such weights are present, it may be necessary to calculate a set of alternative weights in order to let the age structure of the sample perfectly reflect the actual age structure in the society. This is important for two reasons. First, we need the human capital per age later on to construct historical estimates of the human capital stock using a PIM. Second, if age-classes are not perfectly reflecting the national situation, the estimate of the stock of human capital for the survey used might be too high or too low. For example, assume that the estimated human capital for persons aged 16 is much higher than that of persons aged 65 while the number of persons aged 65 in the sample is overrepresented. As we will see in the next section, we use a regression analysis to estimate educational capital, which is a share of human capital. If the number of persons aged 65 is overrepresented, the estimates of average per capita educational capital will be underestimated.

Therefore we constructed a simple weight to counter these problems. We simply divided the percentage of age j in the total population aged 16-65 in the national census by the percentage of age j in the total population aged 16-65 in the sample. In other words:

$$Weight_j = \frac{\%inCensus_j}{\%inSample_j} \tag{5.1}$$

Here, j is the respective age of the individual. The weight is thus the same for all individuals, i, at the same age.

3. A NEW METHOD OF ESTIMATING THE HUMAN CAPITAL STOCK: USING HOUSEHOLD SURVEYS

Using above mentioned definition and data, in this section we start with outlining the method used to estimate the stock of human capital for the years for which we have household surveys. The historical estimates, based on a perpetual inventory method, are presented in section 4.

As we have seen in table 5.1, we can define the human capital stock as follows:

$$\textit{Human capital stock} \equiv \textit{total investments} + \textit{non-government, non-private effect} + \textit{experience} \tag{5.2}$$

These three components of the stock of human capital we estimate in the following three steps. In step 1 we start with total investments (which consists of the sum of all public and private expenditure on formal education plus foregone wages) and with the non-government, non-private effect. The result, which we call educational capital, is an index with a value indicating each individual's relative human capital. In the second step we convert this index in a monetary value by calculating the total public and private expenditure plus foregone wages for a young individual who is likely to have no, or only a low, non-government, non-private effect. Using this benchmark we can use the index to estimate the educational capital for all individuals. In the third step we estimate a Mincer-equation with the value of educational capital per individual to estimate experience. These three steps are outlined in more detail below:

Step 1: We need to estimate a latent variable ec_i (educational capital per individual).

Following Kendrick (1976), educational capital (which we define as the sum of total investments plus the non-government, non-private effect, see equation 5.2) is basically nothing more than the sum of all expenditure on education in the population in year t. However, summing all expenditure over time is extremely data-intensive. For example, a person aged 65 in 1990 probably started his education around 1932. To estimate all educational expenditure in 1990, we therefore need time series on public and private expenditure on education and on foregone wages going back to 1932. In addition, investments in education can depreciate (or appreciate) which we cannot directly observe. Finally, it is likely that especially older persons also acquired skills through non-formal education such as home education. Thus, directly taking 'years of education' from the household survey and use those to calculate all investments would likely underestimate total educational capital as it ignores the non-government, non-private effect. Therefore, we must correct the number of 'years of education' per person in the sample in such a way that it also reflects the non-government, non-private effect, i.e. factors such as 'home education'. This we can do by regressing a dummy variable indicating whether a person has more (1) or less (0) years of formal education than the 'median of 'years of education in an occupational category'[94] on variables indicating 'home education' such as age and sex (older people and women are more likely to have been subject to home education). In this way we say how much the 'real' years of education per person (corrected for the absence of home education) is likely to be.

The first step is thus to estimate a latent (or unobservable) variable by using a probit model. Now assume that the probability of having a relatively high educational level compared to the occupational group (that is: having an education level higher than the median of a certain occupational group)[95] depends on the unobserved educational capital (ec_i), (see table 5.1). This latent variable is determined by explanatory variables in such a way that the larger the value of ec_i, the greater the probability of having a relatively high educational level. This ec_i is expressed as:

$$ec_i = \beta_1 + \beta_2 Sex_i + \beta_3 Province_i + \beta_4 Age_i + \beta_5 RealForgoneWage_i + \beta_6 RealGovExp_i$$
$$+\beta_7 RealPrivExp_i + \beta_8 Sex_i * Age_i + \varepsilon_i \qquad \varepsilon \sim N\left(0,\sigma^2\right) \qquad (5.3)$$

[94] The main advantage of taking a dummy variable is 1) it removes outliers and 2) it allows for a stronger correction of unobserved factors of educational capital such as 'home education'.
[95] It is preferable to estimate the median per occupational group because it avoids placing occupations with totally different education levels in one group. For example, persons working in agriculture generally have less years of education than in some other occupational classes. Therefore, using more occupational classes is more precise. It seems, however, that the results do not significantly change even if we estimate the median educational level for the entire sample. This is one reason why we prefer the median. In a normal distribution, the mean, median, and mode are equivalent. When the distribution deviates from normality the median is preferable.

, where *i* indicates an individual, *Sex* is a dummy indicating whether a person is a male (1) or female (0), *Province* is a dummy indicating province or region[96], and *Age* is the age in years. Further we have *RealForegoneWage*, which is the wage foregone during the years one followed education.[97] We calculated only the foregone wage from the end of compulsory education.[98] If no compulsory education was enforced by law, we start at age 10.[99] Because children tend to get a lower wage than adults, we opted to attach the following weight to the adult wages:

Age	10	11	12	13	14	15	16	17	18	19	20
%wage	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%

So there is no overestimation of human capital in time periods when children start to work earlier. Also, this approach can uniformly be applied to both developed and developing countries. As a consequence, we avoid the problem of having to use different values and so enforce differences between developed and developing countries. Further variables in equation (5.2) are *Real Government expenditure* per student per level of education, and *Real private Expenditure*, which is the average household expenditure on education in year *t*.[100] Finally we have inserted a cross term of *sex* and *age*. We inserted this cross term in order to capture the effect of 'education at home', which cannot be seen directly from government expenditure or private expenditure. For example, it is likely that girls were more susceptible to home education than boys because they, for example, wove or took care of younger children. In addition this happened more often in the early decades of the twentieth century than later.

[96] For Japan these are the call areas: Hokkaido, Kanto, Chubu, Kinki, Chugoku, and Kyushu.
[97] It might be possible that *foregone wages* is correlated with physical capital, *K*. However, since we use *foregone wages* as an **indicator** for a relative high level of education, instead of estimating a structural model, this does not poses a problem. In addition, as mentioned before, foregone wages is often correlated with other monetary variables on human capital. Hence, it is often left out of the regression.
[98] This is difficult in itself because, for example in India, some states may have different ages for compulsory education. However, for Japan, we took 10 years for the period before 1908. Between 1908 and 1946 it was 12 years, and 15 years for all persons following education thereafter. For India we took compulsory education until age 14 since the 1950s and for Indonesia before 1994 until 12 years and until 15 years thereafter.
[99] As we attach 0% to children aged 10, we *de facto* start at age 11.
[100] All monetary independent variables, i.e. government expenditure, private expenditure, and foregone wages, have a large chance of causing strong collinearity. Therefore, it might sometimes be better to leave them out of the equation. This we also did in most cases.

Therefore, inserting the cross term might capture this effect.[101] In the same way, the inclusion of 'age' as an independent variable might pick up the effect that elderly persons are more likely to have had a greater share of home education compared to formal education.[102]

In the previous paragraph we described equation (5.3). But, like we mentioned, ec_i is unobservable. Therefore we estimate a probit regression.[103] Assume that Y=1 means that an individual has a relatively high educational level and Y=0 that he has not. Now assume that there is a critical level of the unobservable variable, ec_i^*, when this is the case. So, if $ec_i > ec_i^*$ a person will have a relatively high educational level (higher than the median). This critical level is also unobservable, but if we assume that it is normally distributed with the same mean and variance we can estimate both the critical level and the index of the entire variable ec_i. Now, given the assumption of normality, the probability that ec_i^* is less than or equal to ec_i can be computed from a standardized normal CDF as:

$$P_i = P(Y=1 \mid X) = P(ec_i^* \leq ec_i) = P(Z_i \leq \beta_1 + \gamma' X_i) = F(\beta_1 + \gamma' X_i) \qquad (5.4)$$

Here $P(Y=1 \mid X)$ is the probability that a person has a number of years of education higher than the median given the values of X (the vector of the explanatory variables).

[101] We have to be aware that it might be necessary to insert quadratic functions of the monetary variables to capture any possible non-linearities. Ignoring this possibility might bias the estimation result as it is by no means certain that persons with 20% chance of having a relatively high educational level also have 2 times more educational capital as those persons with 10% chance of such a relatively high educational level. Inserting these quadratic variables may counter this problem. For each of these monetary variables, we decide whether to include the quadratic variable if the optimum is within the range of the survey. For example, if we take government expenditure, we might get:

$$aLog\left(RealGovExp\right)_i + b\left(Log\left(RealGovExp\right)_i\right)^2$$

For the optimum, we obtain:

$$Log\,RealGovExp_i^* = -\frac{a}{2b}$$

Now if $Log\,RealGovExp_{\min} < Log\,RealGovExp_i^* < Log\,RealGovExp_{\max}$, we decide to include the non-linearity in the regression. In other words, as the estimated optimum falls within the range of the variable in the dataset, we include the non-linearity in the regression. Otherwise, we will exclude it. However, in actual calculation most often one will find that there is no need to include quadratic terms, partly because they are strongly correlated with the other variables.

[102] Part of the reason why this effect is picked up is because we use the human capital of persons around 16 years as a basis to calculate the human capital for all other persons (see step 2 in the text). Consequently, the upward effect of variables such as home education for, mostly older, persons is retained.

[103] This part is largely based on Gujarati (2003, 608-610).

Here Z_i is the standard normal variable, i.e. $Z \sim N(0, \sigma^2)$. F is the standard normal CDF. Since *P* presents the probability of having a relatively high educational level, it is measured by the standard normal curve from $-\infty$ to ec_i. Now we can obtain information on the stock of educational capital, ec_i. We take the inverse of equation (5.4) to obtain:

$$ec_i = F^{-1}(ec_i) = F^{-1}(P_i) = \beta_1 + \gamma' X_i \qquad (5.5)$$

In other words, after having estimated *P*, the chance of having a relatively high educational level from equation (5.3) (or, more formally written, equation (5.4)), we can take its inverse to estimate an index of educational capital.[104] This index thus represents the number of years of schooling per individual corrected for factors such as home education.

Step 2: Estimating the monetary value of the educational capital.

We now have an index of educational capital, i.e. of the sum of total investments and the non-government, non-private effect (see equation (5.2)). If we want to turn this into a monetary value, we run into some distortions:
1) the depreciation of money (even though we used real 1990 monetary units, problems with unaccounted inflation may seriously hamper the estimates),
2) depreciation/appreciation of human capital,
3) the 'non-government, non-private' effect which is unobserved even using the income- or cost-based approaches.
These three problems may cause biases in the value of the estimated educational capital.

To convert the educational capital index into a monetary value for each individual, we have to estimate the educational capital stock for some sort of baseline entrant similar to Tao and Stinson (1997). Using the index we can extend the value of the educational capital stock of the baseline entrant to the other individuals in the sample.

[104] What we are actually doing is to condition the Y on the X-values. This means that we first construct a variable indicating a relatively high level of education based on years of education. Then we calculate the probability of having a relatively high number of years of education. The final step is to condition the y-variable to the x-variables to construct the educational capital index. This index thus strongly resembles the original input (years of education) but is not completely the same. There are two main advantages in first transforming 'years of education' into a dummy variable. First, it removes outliers. Second, using a probit allows us to estimate home education as well (see table 5.1 and the text). If we would simply take 'years of education' we would miss home education. However, using a probit, we might condition it on variables such as age*sex in order to capture the effect of home education.

Yet, because the estimation of the educational capital stock faces the three problems mentioned above, we have to choose a baseline entrant to minimize these three problems.[105] As mentioned before, our human capital stock includes people between the ages 16 to 65 (in the labour force). Therefore, we opt for a person aged 16 with solely primary education, for whom we calculate his foregone wage, and his private and government expenditure incurred in 1990 prices. Summing these yields the educational capital of the baseline entrant in 1990 prices. Now, we can estimate the value of educational capital of other individuals in the sample using the educational capital index from step 1. If we calculate the value of average educational capital per age in the sample and we multiply it with the number of persons at that age in the total population we arrive at the value of the total stock of educational capital.

Step 3: Estimating the income-based part of human capital: 'experience' and 'on-the-job training'.

In the previous two steps, we estimated the educational capital stock, which was the total investment (public, private, and foregone wages) in human capital, added with the 'non-private, non-government' effect. Yet, as we can see in equation 5.2, human capital also has an income-based part, i.e. 'experience' and 'on-the-job training'. To estimate this part, we start by estimating an earnings equation, which we can do in a cross-section for each household survey separately:

$$LnE_i = \alpha + \beta_1 ec_i + \beta_2 t_i + \beta_3 t_i^2 + \varepsilon_i \qquad (5.6)$$

E_i and ec_i denote per capita wage and the value of the educational capital stock (from step 2) respectively, t_i and t_i^2 are variables included to capture the effect of experience and 'on the job training', and ε is the error term, assumed to be independent and identically distributed. The variable t is calculated as age minus 'school duration' minus 6 years. For those with lower secondary education or primary education levels, t is estimated as age minus 15 years. This is necessary to avoid overestimating the experience of persons with only lower education as children generally gain less experience than adults (Dougherty and Jimenez 1991). As a consequence, $\beta_2 t_i - \beta_3 t_i^2$ is the interest on experience and on-the-job training from individual i with age (m_i). To

[105] We can of course also calculate the costs of education for different persons and compare the results.

arrive at an estimate of experience, we simply take the mean interest per age *j*, $\overline{m}_j$. Then, for each age *j*, we sum the mean interest of the previous ages, so the experience at age *j* is $Exper_j = \sum_{16}^{j} \overline{m}_j$.

In other words, we calculate the interest on 'experience' for each age (from age 16 to age 65). As it is unlikely that persons aged 16 have a different innate ability than persons aged 65, this avoids an ability bias. Next, we sum the interest for each age to obtain the estimates of value of the relevant experience. For example, if we want to calculate the experience at age 20, we sum the interest on experience from age 16 to 20. The same we do for all other ages from 16 to 65.

However, this method can only be performed for persons with wage>0. It is, however, unlikely that persons outside of the labour force (with wage=0) have no experience at all. In order to correct for this sample selection bias we apply a method based on Heckman (1979).

First, we assume that the estimated experience from regression (5.6) also holds for persons with *wage=0*. Hence, the experience estimated by equation (5.6) for persons aged 30 is not only used for those persons aged 30 with *wage>0* but also for those persons aged 30 with *wage=0*. This creates a strong overestimation of experience as it is unlikely that the persons outside the labour force have the same level of experience as those in the labour force.

Second, we run a probit regression where we estimate the chance that someone is in the labour force, i.e. *wage>0*, Hence, we run a regression of whether someone is in the labour force (*Y=1*, i.e. *wage>0*) or outside the labour force (*Y=0*, i.e. *wage=0*) on a vector of explanatory variables such as age, sex, educational capital, and marital status. This results in an estimate for each person of the chance he or she is in the labour force.

Third, we multiply this chance for each person with the calculated experience of a person of that age. This means that persons with a high probability of being in the labour force, even if they are at present outside the labour force, have, given their educational capital and age, relatively much experience. The basic idea is that those persons with a higher chance of being in the labour force also have more chance of having had a job earlier and, hence, of having more experience.

4. BRINGING THE HUMAN CAPITAL ESTIMATES BACK IN TIME: A PEREPETUAL INVENTORY METHOD

Above three steps resulted in estimates of the monetary value of the human capital stock in some years in the 1980s and 1990s. The results are given in table 5.4. There are three important aspects to note. First, the per capita human capital stock was fractionally higher in Indonesia compared to India. This difference was around 3.5%. Of course, the total human capital stock was much higher in India because of its far larger population. Second, the per capita human capital stock in Japan is much higher than in India and Indonesia. The per capita stock in India is around 3.3% and in Indonesia 3.4% of that of Japan. This is remarkable because, if we had taken 'average years of education', the per capita stocks of India and Indonesia compared with Japan would have been 53.8% and 65.3% respectively. If assume, somewhat unrealistically, that 'average years of education' only consists of the quantity, and the stock of the newly estimated human capital consists of both the quality and quantity of human capital, then the quality of human capital would amount to 1-0.033-(1-0.538) = 50.5% of the difference of the per capita stock of human capital between India and Japan and 1-0.034-(1-0.65)= 61.6% of the difference between Indonesia and Japan. It thus seems that the quality of human capital is lower in Indonesia than it is in India. Indeed, comparing India to Indonesia, we find that the quality of human capital

Table 5.4: Total and per capita stock of human capital and total stock of gross fixed non-residential physical capital in Indonesia, India, and Japan in 1990 international USD, converted at purchasing power parity (PPP)

	Indonesia			**India**			**Japan**		
	Physical capital	Human Capital		Physical capital	Human Capital		Physical capital	Human Capital	
	Total	Total	Per capita	Total	Total	Per capita	Total	Total	Per capita
	Billion USD	Billion USD	USD	Billion USD	Billion USD	USD	Billion USD	Billion USD	USD
1993	649.0	348.7	1,851	2,530.1	1,589.6	1,787	4,458.1	6,721	53,907
1994							4,610.3	6,782	54,248
1995							4,768.6	7,256	57,888
1996	870.8	397.7	2,008				4,943.9	7,868	62,622
1997							5,112.7		
1998								8,407	66,590
1999	1,053.6			3,768.0	2,001.7	2,028			
2000								8,642	68,206
2001									
2002		534.8	2,378					9,408	74,036

Source: Human capital: this chapter; physical capital: appendix A.2.

would explain 1-0.944-(1-0.853) = -9.1% of the difference. In other words, the gap in per capita human capital would be 9.1% larger if India had the same quality of human capital as Indonesia.[106] Third, the amount of the estimated human capital is lower than the physical capital stock in India and Indonesia. The exception is Japan, where the estimated human capital stock is significantly larger than the stock of physical capital. This corresponds with the finding of Judson (2002, 217) that the human capital stocks in rich countries are significantly larger relative to both physical capital and GDP than in poorer countries.

These estimates of the stock of human capital also contain information of the human capital per age for persons aged 16 to 65 years old. We assumed that persons of these ages were eligible for the active labour force and their human capital was thus included in the stock of human capital. If we know the appreciation/depreciation of human capital by age, we can use this information to estimate how much human capital is added to the stock between circa 1950 and 2000. This information allows us to bring the human capital stock from the 1980s and 1990s back to 1890 which we need to cover the entire twentieth century.

There are two ways in which we can utilise this information to bring the human capital stock back to 1890. Both methods are strongly related to a Perpetual Inventory Method (PIM). A PIM requires two data series. First, we need the annual flow of human capital that is added to the stock. This is the gross fixed human capital stock. Second, we need the yearly appreciation (a percentage increase in value) or depreciation (a percentage decrease in value) of the human capital. Now, there are two ways in which can generate a series of human capital. First, we sum all flows that are added to the stock over 50 years (for persons aged 16 to 65), minus their depreciation (or plus their appreciation). The definition of human capital than becomes:

Human capital stock in year t ≡ sum of all additions to the stock of human capital in year t-1 to t-50 minus the depreciation (or plus the appreciation) (5.7)

[106] Assuming 'average years of education' is the quantity of human capital and 'Total HC' is the quality and quantity of human capital. We want to know what difference between the stocks of human capital can be explained by the quality. To arrive at this, we first estimate 1-%'average years of education'. In other words, 1-the percentage of 'average years of education' compared to the country with which you want to compare it (0.538 if you want to compare India to Japan). Now you have solely the quantity. Now simply subtract this from 1-%'Total HC', which is the total difference minus quantity and quality.

The disadvantage of this method is that it sums human capital accumulation over 50 years which means that the first estimate of the human capital stock is only available 50 years after the first estimate of human capital accumulation. Yet, the main advantage is that we can directly apply the appreciation/depreciation figures by age without reverting to constructing an average for the complete stock.[107] Therefore, we prefer this method. To extend this series back to the year in which we have the first observation of human capital accumulation we use the following equation:

Human capital stock in year t-1 ≡ (human capital stock in year t minus the human capital accumulation in year t-1)/(1-depreciation in year t-1) (5.8)

The main disadvantage of this method is that we have to assume a depreciation (or appreciation) of the total stock.

In the next five steps (step 4-8) we will thus use equation (5.7) and (5.8) to bring the human capital stock based on the household surveys back to 1890. Step 4 to 7 focus solely on equation (5.7). In step 4 we estimate the appreciation/depreciation by age and in step 5 and 6 we estimate the human capital accumulation in the second and first half of the twentieth century respectively. As we now have all the necessary variables, we turn in step 7 to the estimation of the human capital stock based on equation (5.7). In step 8 we bring the human capital stock back in time for the last 50 years using equation (5.8). These five steps are outlined in more detail below:

<u>Step 4: Estimating the appreciation/depreciation of human capital.</u>

To estimate equation (5.7), we need the appreciation/depreciation of the stock of human capital by age class. Therefore, we estimate the percentage change by age class from the human capital estimates based on the household surveys in step 1-3 above. For example, we estimate how much the per capita human capital for persons aged 40 in year *t* increases or decreases compared to the per capita human capital of persons aged 41 in year *t+1*.[108]

[107] In addition, it can be considered a cross-check for our human capital estimates based on household surveys in step 1-3. If the sum of the human capital accumulation over 50 years minus their depreciation (or plus appreciation) equals the human capital estimates based on the household surveys, this suggests that the appreciation/depreciation figures do not significantly change over time.

[108] This is no appreciation/depreciation because a) in the first years a person enters the human capital stock, he or she might still form human capital at secondary or higher schools, b) as a correction for mortality is excluded, this might increase the per capita human capital because mortality is higher among

The results are presented in table 5.5. In Japan and India we notice an increase in appreciation by age, followed by a decrease. This suggests that educational capital is relatively constant over time and that the pattern is dominated by experience, where appreciation declines at later ages. In Indonesia, however, the pattern of appreciation fluctuates over time. We might explain this development by the situation that Japan was already more advanced in education than was India and Indonesia at the start of the century. As secondary and higher education are more likely to retain their value (no chance of relapsing in illiteracy, better chances of retraining skills when one's job becomes economically obsolete), the appreciation/depreciation pattern of experience (first increasing and later decreasing) is likely to dominate. That this pattern also dominates in India might be attributed to the focus on secondary and higher education at the start of the century.

Table 5.5: Yearly appreciation of the per capita stock of human capital by age class in Japan, India, and Indonesia based on surveys in the 1990s

age class	Japan	Indonesia	India
16-20	3.7%	7.0%	5.1%
21-25	4.9%	4.9%	6.1%
26-30	5.8%	3.5%	9.4%
31-35	5.3%	1.9%	10.2%
36-40	4.9%	1.5%	7.7%
41-45	4.8%	3.4%	6.1%
46-50	4.6%	4.9%	3.7%
51-55	4.4%	4.9%	2.1%
56-60	3.7%	2.7%	-1.4%
61-65	2.8%	7.7%	-5.5%

Note: A negative value indicates a depreciation.

Step 5: Estimating the gross fixed human capital formation in the second half of the twentieth century

To calculate the stock of human capital, equation (5.7) indicates that we also need the human capital that is added to the stock by people who enter the stock of human capital (i.e. who are 16 years old). We use the appreciation/depreciation figures from step 4 to estimate the gross fixed human capital formation (GFHCF, gross increase of the human capital stock in year *t*) back to around 1945 for persons aged 16. We will call this the

persons with lower levels of human capital, and c) such parts of the human capital stock as 'experience' and 'on the job training' are formed later in life causing an increase of the appreciation (or decrease the depreciation) of the per capita human capital stock if it is not corrected for.

$GFHCF_{16}$ to distinguish it from the total GFHCF ($GFHCF_{tot}$) which consists of all investments (also of persons older than 16) added to the stock in year *t*.

The starting points are the per age class estimates of human capital based on the household surveys (step 1-3). Say, we take the survey for 1993. A person aged 17 in the 1993 survey was 16 in 1992. Therefore, we took the per capita human capital from a person aged 17 in the 1993 survey and corrected it for the percentage change in per capita human capital from age 16 to age 17 obtained in step 1. In this way we arrived at the per capita human capital of a person aged 16 in 1992.We did the same for a person aged 18 in the 1993 survey (aged 16 in 1991), etc. In this way we brought the per capita human capital of a person aged 16 back to 1945.[109]

Since we had surveys available for several years, we carried out this exercise for all these years. This yielded multiple, rather homogenous, estimates of the per capita human capital stock for a person aged 16. We then took the average and multiplied it with the total population aged 16. In this way we arrived at the $GFHCF_{16}$.

Step 6: Estimating the gross fixed human capital formation in the first half of the twentieth century

In step 5 we brought the $GFHCF_{16}$ back in time to 1945. Yet, given that equation (5.7) indicates that we have to sum this variable over 50 years, this means that we can only estimate the first human capital stock in 1995. However, our final objective is to bring it back in time to 1895. To this end, we need to estimate the $GFHCF_{16}$ for the period 1895-1945 as well. As pointed out in equation (5.2), the human capital stock (and thus also the GFHCF) consists of an observed component (government and private expenditure on education plus foregone wages), an unobserved component (the 'non-government, non-private' effect), and experience. However, it is not necessary to calculate experience as we are solely estimating the human capital accumulation for a person aged 16 (who just enters the human capital stock) and thus has no experience yet. We will thus only have to estimate the accumulation of what we called in step 1-3 'educational capital' for a person aged 16.

First we estimate the observable component of the $GFHCF_{16}$. This can be calculated (assuming 6 years of primary education from age 6 to age 11, 3 years of

[109] The implicit assumption is thus that the percentage change in per capita human capital from age x to age x+1 remains the same between 1945 and 2002.

lower secondary education from age 12 to 14, and three years of higher secondary education, age 15-17)[110] as:

$$GovInvest_{t,16} = \sum_{i=6}^{11}\left(\tfrac{1}{6} * PrimEnrolled_{t-i} * GovExpPrim_{t-i}\right) + \sum_{i=2}^{4}\left(\tfrac{1}{3} * LowSecEnrolled_{t-i} * GovExpLowSec_{t-i}\right)$$
$$+ \tfrac{1}{3} * UpSecEnrolled_{t-1} * GovExpHighSec_{t-1}$$

$$PrivInvest_{t,16} = \sum_{i=6}^{11}\left(\tfrac{1}{6} * PrimEnrolled_{t-i} * PrivExp_{t-i}\right) + \sum_{i=2}^{4}\left(\tfrac{1}{3} * LowSecEnrolled_{t-i} * PrivExp_{t-i}\right)$$
$$+ \tfrac{1}{3} * UpSecEnrolled_{t-1} * PrivExp_{t-1}$$

Here, $GovInvest_{t,16}$ and $PrivInvest_{t,16}$ are the cumulated government and private investment in persons aged 16 in year *t* respectively. Further *Prim*, *LowSec* and *UpSec* enrolled give the number of students enrolled at that level of education in year *t*. The fraction before those variables indicates the number of years of education at each level. The sum signs indicate how many years have to be summed before age 16 is reached. For example, 6 years of primary education, 3 years of lower secondary, and 1 year of upper secondary education as the final two years of the latter education level are intended for persons above age 16. We did not include foregone wages in some cases because compulsory education may last up to age 16 and we only started to include foregone wages after the end of compulsory education. However, we did include foregone wages for earlier decades when compulsory education did not last until age 16. If we sum the above-mentioned investments, we arrive at the $GFHCF_{16}$ without the 'non-government, non-private' effect.

Second, we need to estimate the unobservable part, the 'non-government, non-private' effect. As we have the total $GFHCF_{16}$ for the period 1945-2002, we simply have to subtract the $GFHCF_{16}$ without the 'non-government, non-private' effect from these figures to obtain the 'non-government, non-private' effect. As we had both the $GFHCF_{16}$ without the 'non-government, non-private' effect for 1895-2002 and the 'non-government, non-private' effect for the period 1945-2002, we used the following single equation regressions to backcast the latter:

Japan:

$$\text{'non-government, non-private effect'}_t = 0.043*trend - 0.711*Other_t + 0.003*Other_t^2 + \varepsilon_t$$

(SE)	(0.044)	(0.084)	(0.0004)

110 These ages may differ by country or time period.

India:

$$\text{'non-government, non-private effect'}_t = -0.018*trend - 0.056*Other_t - 0.001*Other_t^2 + \varepsilon_t$$

(SE) (0.009) (0.024) (0.0001)

Indonesia:

$$\text{'non-government, non-private effect'}_t = 0.023*trend - 0.349*Other_t - 0.001*Other_t^2 + \varepsilon_t$$

(SE) (0.003) (0.033) (0.0006)

Here $Other_t$ is the $GFHCF_{16}$ without the 'non-government, non-private effect in year *t*. We included $Other_t^2$ because it is likely that there are non-linearities. For example, it is likely that the 'non-government, non-private' effect is much larger relative to the formal education expenditures at the start of the century because it includes such measures as 'home education' and 'on the job training'. Indeed, we find this squared variable to be significant. Adding our backcasts to the estimates of the $GFHCF_{16}$ without the 'non-government, non-private' effect, gives the total $GFHCF_{16}$ for the late nineteenth century to 2000.

Step 7: Estimating the human capital stock in the second half of the twentieth century

We can now estimate the human capital stock based on equation (5.7). Since we are interested in the human capital stock of individuals aged 16-65, we sum the $GFHCF_{16}$ which we calculated in step 5 and 6 over 50 years while we correct for the percentage change in per capita human capital estimated in step 4.[111] As we sum over 50 years (age 16-65), summation can be done as:

$$\sum_{i=0}^{50} GFHCF_{16,t-i} \qquad \text{where } i = \text{age-15}. \qquad (5.9)^{112}$$

[111] As we link these human capital stock estimates with the estimates of the human capital stock from the household surveys (thus adapting the level of these human capital stock estimates to those of the household surveys), the implicit assumption is a constant mortality of the working age population (age 16 to 65) over time.

[112] This equation is unlikely to be biased by the breaks in human capital developing institution as sketched in chapter 4. Indeed, we do need to pick up these breaks in order to test for the presence of educational institutions in chapter 7. However, first, it is only estimated for circa 50 years (1950-2000) which period knows very little breakpoints as we pointed out in chapter 4. But, second, even if there were breakpoints, these would most likely take place in human capital accumulation. However, this is relatively independent of the breaks in the institutional development. Third, of course, depreciation could

However, the $GFHCF_{16}$ has to be corrected for appreciation/ depreciation (δ) for each age. As we allow for both an appreciation and depreciation, this means that it is possible that $\delta > 0$ or $\delta < 0$. This means that for a person aged 65 (*j*=50) his gross fixed human capital formation when he was 16 has to be multiplied with $(1+\delta_0)$, where δ_0 is the appreciation/depreciation from age 15 to 16 (which is non-existent as we assume that appreciation/depreciation only starts when a person enters the human capital stock at age 16), with $(1+\delta_1)$ for age 16 to 17, with $(1+\delta_2)$ for age 17 to 18,…., with $(1+\delta_{50})$ for age 64 to 65. This means that, as the appreciation/depreciation is only dependent on age (thus independent of time), the total appreciation/depreciation for persons aged θ can be estimated as:

$$\delta_\theta = \prod_{j=0}^{i} \left(1+\delta_j\right) \tag{5.10}$$

Combining equation (5.9) and (5.10), the total human capital stock (H) of a society can be estimated as:

$$H_t = \sum_{i=0}^{50} \left[GFHCF_{16,t-1} \cdot \prod_{j=0}^{i} \left(1+\delta_j\right) \right] \tag{5.11}$$

In other words, equation (5.11) is the more formal version of equation (5.7).

For example, we would like to estimate the human capital stock in 1950. As we defined the human capital stock as consisting of the human capital of all individuals between the ages of 16 and 65 (50 years), this means that the first persons that entered the 1950 human capital stock were 16 in 1900. Next, we apply the appreciation/depreciation figures (from step 4) to their gross fixed human capital formation (from step 6). This means that, if we take Japan, their gross fixed human capital formation appreciated with 3.6% yearly between age 16-20, with 4.9% between age 21-25, etc. In this way we arrive at how much their human capital would be in 1950 (when they were 65). We do the same for persons who are 16 in 1901 (who were 64 in 1950). This we do for all the years up to 1950. Finally, we sum all values to obtain the human capital stock in 1950.

---

be influenced by breaks. However, this is much less likely to be the case than for capital accumulation. After all, it is easier to increase the number of schools than to change the human capital depletion (for example because people are retiring) in a factory. Fourth, even if breakpoints affect depreciation, for example at time *t*, than still the possibility exists that only the depreciation of persons entering the labour force after year *t* is affected. Fifth, because we are talking about a stock variable, any minor changes will be smoothed.

Step 8: Estimating the human capital stock in the first half of the twentieth century

Unfortunately, we can only provide estimates for the human capital stock 50 years after the first observation of the $GFHCF_{16}$. In our case the first year in which the human capital stock can be estimated is around 1945. As pointed out at the start of this section, in order to go further back in time we need both the human capital accumulation and the yearly depreciation of the total stock of human capital (see equation (5.8)). For the first we only have the $GFHCF_{16}$, excluding any human capital obtained at age 17-22, when a person may still attend upper secondary school, college, or university. Estimates of the appreciation or depreciation can only be made if we have both the stock of human capital and the total GFHCF ($GFHCF_{tot}$). As we have the stock of human capital in 1945-2002, we first have to estimate $GFHCF_{tot}$.

Therefore, we have to add the human capital formed by persons aged 17-22 in year *t* to the $GFHCF_{16}$ which we estimated in step 5 and 6. Now, if we are confronted with a school system where we have three years of higher secondary education (age 15-17) and with either two years of college education (18-19) or four years of higher education (age 18-21), we can estimate the investment of persons of age 17-21 in year *t* as:

$$HighSecInvest_t = \left(\frac{HighSecEnrolled_t}{3}\right) * \left(\left[(0.7) * ForegoneWage_t\right] + \left[GovExpHighSec_t\right] + \left[PrivInvest_t\right]\right)$$

and

$$CollegeInvest_t = \left(\frac{CollegeEnrolled_t}{2}\right) * \left(\left[(0.8 + 0.9) * ForegoneWage_t\right] + \left[2 * GovExpCollege_t\right] + \left[2 * PrivInvest_t\right]\right)$$

and

$$UniversityInvest_t = \left(\frac{UniversityEnrolled_t}{4}\right) * \left(\left[(0.8 + 0.9 + 2) * ForegoneWage_t\right] + \left[4 * GovExpUniversit_t\right] + \left[4 * PrivInvest_t\right]\right)$$

The three equations above state that the average number of students per year and level of education has to be multiplied by the per capita foregone wages, corrected for the fact that younger children earn less, the per capita government expenditure for that education level, and the per capita private investment. This results in:

$$InvestAge16_t + HighSecInvest_t + CollegeInvest_t + UniversityInvest_t = GFHCF_t$$

, which means that the sum of $GFHCF_{16}$ plus investments of persons older than 16 in higher secondary education, college, and university results in the $GFHCF_{tot}$.

As we now know both the total gross fixed capital formation (including the gross fixed capital formation in year *t* of persons older than 16) and the stock of human capital for the period 1945-2002, we can calculate a yearly changing appreciation/depreciation of the estimated total stock of human capital for 1945-2002 as[113]

$$appreciation / depreciation = \frac{HC_t - HC_{t-1} - GFHCI_t}{HC_{t-1}} \quad (5.12)$$

This has the important advantage that the appreciation of the estimated stock mirrors the changes in the duration of compulsory education. This resulted in an average yearly appreciation of around 0.8% in the 1990s.[114]

Now we have the $GFHCF_{tot}$ for the period 1895-2002 and the yearly appreciation/depreciaton of the stock of human capital for the period 1945-2002. If we assume that the yearly appreciation depreciation remains the same as the average appreciation/depreciation of the period 1950-60 (thus excluding the immediate post-War period), we can estimate the stock of human capital before 1945 as

[113] An alternative method for directly estimating the appreciation/depreciation of the total stock of human capital, using solely the surveys, would be to estimate, using the total, per age, stock of human capital:

$$\frac{\left(\sum_{j=16}^{65} HC_{j,t+1}\right) - \left(\sum_{j=16}^{65} HC_{j,t}\right) - GFHCF_{t+1}}{\sum_{j=16}^{65} HC_{j,t}}$$

, where HC is the total stock of human capital per age *j* and *t* is the year of the survey. In other words, we estimate the percentage difference between the total human capital stock in year *t+1* and the total human capital stock in year t minus the human capital inserted in the stock of human capital in year *t+1*, the Gross Fixed Human Capital Formation (GFHCF). However, this approach gives two problems. First, an obvious drawback is that we can solely calculate the appreciation/depreciation of the stock of human capital for those years for which we have surveys. It is not clear at all, however, that these appreciation/depreciation figures remain the same over time. Second, in section 3 we already mentioned that foregone wages were calculated starting from the age at which compulsory education ended or, when there is no compulsory education, from age 10. This changed in all three countries in the middle of the twentieth century. In turns, this might have changed the rates of appreciation/depreciation. However, in the period before this change in the duration of compulsory education, we do not have surveys from which we can calculate a new figure for the appreciation/depreciation.

[114] This sounds somewhat as a circle argument. We could have estimated the appreciation/depreciation for the human capital from the years we have household surveys and applied it to the preceding years. Instead, we used the per capita appreciation/depreciation to calculate the stock of human capital as indicated in the text. As these per capita appreciation/depreciation figures are also obtained from the household surveys, this is a circle argument if the relation between the gross fixed human capital formation and the estimated human capital stock remains constant over time. Yet, this is unlikely as, for example, it may be argued that the depreciation increases (or appreciation declines) during the War-periods. Therefore, having a stock and a series of human capital accumulation, the decline in the stock in the War-periods is largely attributed to the depreciation. If we use the depreciation figures from the 1990s surveys for the entire period, and given the fact that we had the human capital accumulation data, the decreases (or increases) in depreciation/appreciation during certain periods would be completely reflected in the stock which is quite unlikely as a stock variable by definition is less susceptible to hectic fluctuations.

$HC_t = \frac{HC_{t+1} - GFHCF_{t+1}}{(1+appreciation)or(1-depreciation)}$. The final results are presented in appendix A.9-A.12.

5. PLAUSIBILITY AND DEVELOPMENT OF HUMAN CAPITAL

5.1 Introduction

In the previous sections we introduced our new method for estimating the stock of human capital. This stock, in constant per capita terms, is presented in figure 5.1. Three things are worth noticing. First, all three countries have more or less a straight rising curve. It seems that, before the 1930s, the growth of human capital is larger in Indonesia than it is in India. Between 1930 and 1950 (Indonesia) or 1970 (India) the per capita human capital grows strongly. Afterwards its growth again decreases because the period between 1930 and 1970 was the period of large increases in educational enrolment and attainment. The pattern of first constant, than rising, and finally again decreasing growth of human capital is called logistic. This is a pattern that is commonly

Figure 5.1

Per capita human capital in India, Indonesia, and Japan in log 1990 International USD, converted at PPP.

Source: Appendix A.12.

found for variables in growth studies (see for example Cameron 1989, 16; Astorga, Berges, and Fitzgerald 2005, 770).

The same pattern we also find in Japan except that, based on the observations in chapter 4, the growth of per capita human capital already started long before 1890 (for which we have no data). Consequently, the second important point is that the per capita human capital in Japan was much higher than in India and Indonesia, even around 1900. Whereas in Japan the estimated per capita human capital in 1900 in constant 1990 International dollars was around 2.682 USD, in India it was 361 USD and in Indonesia 130 USD. Interesting is also to note that the per capita human capital in Japan in 1895 is about as high as that of India and Indonesia in 2000. The final point to note is that Indonesia around 1890 had a lower per capita human capital than India. Yet, although they moved around the same patterns, Indonesia's growth in per capita human capital set in earlier than in India and was much stronger. As a result, after surpassing India around 1915, the gap in per capita human capital widened in favour of Indonesia until the 1940s. Afterwards the difference began to shrink until the two countries converged after 1970. A simple explanation would be that mass education set in earlier and was stronger, both qualitative and quantitative, in Indonesia compared to India. This corresponds well with our conclusions in the previous chapter on the Indian focus on secondary and higher education prior to the 1920s.

This development of the estimated per capita human capital stock seems plausible. Indeed, because its development seems to conform to the expectations, and because it is designed to include both the qualitative and quantitive aspects of education, we expect that this stock is a better indicator of human capital than most alternative measures. In this section, we will try to give an impression of its plausibility, both quantitatively and qualitatively.

We will look at the plausibility of the stock of human capital in three different ways. First, in section 5.2 we will look at the subjective margin of error. This shows how reliable the data and estimation methods are **given the definition we use for human capital**. It thus cannot give a definite answer on whether the definition is correct or not. In section 5.3, we will look at the composition of the human capital stock. What is its structure? How do these shares relate to the government expenditure on education or private expenditure? This already gives a gauge of how the stock of human capital relates to the basic inputs, which are also often used in alternative measures of human capital. Finally, in section 5.4, we will look at how the estimated human capital stock compares to GDP and the physical capital stock.[115]

5.2 Subjective margins of error

A first step would be to estimate how reliable the estimated stock is given the definition of human capital employed. Even the estimates of the reliability are necessary flawed but, as Feinstein and Thomas (2001, 14) already pointed out '[h]owever problematical such subjective assessments of unknown errors may be, they are much more informative than general statements formed from some favoured permutation of stock phrases (these estimates are very: 'approximate', 'imperfect', 'unreliable', 'tentative', 'uncertain', 'fragile'; they are: 'a best guess', 'a rough guide', 'an order of magnitude', 'a crude indication'; or, very occasionally, they are: 'reasonably reliable', 'broadly acceptable'; and so on).' We will therefore give some indications of the margins of error in this section.

Indeed, there are a lot of options to estimate such margins. Here, we will estimate for each component the margins of error and than aggregate them. The first step is to attribute margins of error to the rough data we collected in first instance, sometimes with a few modifications: government and private expenditure on education (see chapter 3 and appendix A.8.) and wages (see appendix A.1.). Following Chapman (1953, 231), we attached margins of error of 2.5% to 'firm figures', 7.5% to 'good estimates', 17.5% to 'rough estimates', and 40% to 'conjectures'. However, rather than the 95% confidence interval she used, we chose a 90% confidence interval. As a result, instead of dividing the margins of error by 2 to get the standard error, we divided the figure by 1.645.[116] This gave the errors for the different components of the stock of human capital. However, if the errors are derived independently, some errors will offset each other. As a consequence, the formula for the standard error of the whole, from the combined standard errors of the parts becomes: $\sigma_v = \sqrt{\sigma_x^2 + \sigma_y^2 + 2r_{xy}\sigma_x\sigma_y}$. In other words, if there is no relation between the two components x and y, the value of the correlation coefficient is 0 and the standard error is the average of the errors of the two

[115] A fourth test would be to place the stock of human capital in a growth regression. However, this will be the topic of the next two chapters.
[116] Given the relatively small samples, we might have to divide by a figure higher than 1.645. However, given the rough nature of these estimates and given that we already use a confidence interval of 90%, we decided not to correct for the small sample sizes.

Table 5.6: Subjective margins of error in the stock of human capital and some components for India, Indonesia and Japan for 1900-1950 and 1950-2000*

	Private and public expenditure on education	Gross fixed human capital formation	Stock
		Japan	
1900-1950	10.6%	19.9%	26.7%
1950-2000	7.2%	8.1%	14.7%
		India	
1900-1950	11.9%	29.3%	37.5%
1950-2000	10.6%	14.4%	43.4%
		Indonesia	
1900-1950	22.6%	22.4%	40.5%
1950-2000	12.6%	13.5%	32.5%

*90% probability.

components. But, if the correlation is positive, the standard error is larger, and if the correlation is negative the standard errors of the components offset each other. Clearly, it can be made far more complicated. However, it would be useless to make an elaborate estimation of the errors because these are also subject to error. The results of the errors of private and public expenditure are given in table 5.6. All fluctuate around 10%, except Indonesia in the 1900-1950 period. This is largely caused by the situation that we attached 40% unreliability to the estimates of private expenditure on education.

The second step is to estimate the margins of error for the gross fixed capital formation. We simply assumed that the surveys we used were 'firm estimates'. For the period without surveys (the period before circa 1950) we used the extrapolated standard error of the single equation model to estimate the 'non-private, non-government' part of the GFHCF back to 1890. Again, summing the expenditure and the 'non-government, non-private' effect, and using the correlation coefficients, gave the subjective margins of error as reported in table 5.6 for the GFHCF. These errors are on average about 6 %. (percentage points) higher than those of the private and public expenditure.

For the stock we had to make some more modifications. Indeed, we took 'firm estimates' for the stock estimated on the basis of the household surveys. However, these are only for accumulating the several parts of human capital and not for the reliability of these parts themselves. For the underlying data (the gross fixed human capital formation) we took the estimated margins of error and multiplied them with 50 (the years a person remains in the stock of human capital). Again, we used the correlations to arrive at around 20% margin of error for Japan, and around 38% for India and Indonesia. This means for example for Japan that in 1900, when we estimated the stock of human capital to be around 162.4 billion in 1990 International USD, given our definition of human capital the actual stock must be with 90% probability around ± 26.7%, or between 119 and 205.8 billion USD.

Admittedly, these margins of error are large at a first glance. However, compared to most other calculations of historical series our series seem to perform quite well. This has three reasons. First, we set the initial margin of error for the components relatively high. One could easily argue that the components are more reliable, which would decrease the margins of error. Second, even large projects that have more data available may have large margins of error. For example Kuznets (1941, 501-537), for his GDP estimates, arrived at a margin of error of around 20%. Although probably an overestimation, because he did not take into account that the errors of the individual components might be partly uncorrelated which would reduce his margins of error, this figure is still high (Feinstein and Thomas 2001, 7). However, the point is that even the present day estimates suffer from large margins of error. Finally, we have to note that human capital is, contrary to many other time series, unobservable. This is likely to increase the margin of error as well. Given these developments, our series do not perform badly at all.

In sum, the results seem acceptable. For the private and public expenditure on education, we see that the margins of error decrease over time. The same is true for the gross fixed capital formation. In the latter variable, the margins of error are generally somewhat higher. This may be attributed to the situation that the 'non-government, non-private' effect has higher errors than the more visible private and public expenditure. Also for the estimated stock the margins of error decrease. The exception is India which shows a small increase of the error. Yet, this may be attributed largely to the increase in the 'non-government, non-private' effect which was especially strong in the period after 1950. Also we see that the margins of error are lower for Japan than for India and Indonesia. This is caused by the situation that the data for Japan are more reliable than for the other two countries.

5.3 A comparison of human capital and its components

A second step in evaluating the newly estimated stock of human capital is to look at its constituent parts. We chose to divide the stock of human capital into its main components, i.e. public expenditure, private expenditure, and the sum of the foregone wages and 'non-government, non-private effect'. The latter two are strongly interrelated as most of the 'home education' is caused by implicit foregone wages. Also, we decided not to split the 'non-government, non-private' effect into 'experience' and 'home education'. The main reason is that experience is difficult to separate from 'on the job training' and is small compared to 'home education'. Consequently, the bias for experience will be so large that it would not add much to the analysis.[117]

We start by assuming that the per capita depreciation figures are valid for 'government expenditure', 'private expenditure', and 'foregone wages'. The 'non-government, non-private' stock was calculated as the estimated stock of human capital minus the shares of the other parts of the stock. The main reason is that the informal component of human capital, which is what the 'non-government, non-private' effect mostly entails, may be subject to a higher depreciation (or a lower appreciation) than formal education because the latter is generally more intensive. The results are presented in table 5.7. We estimated a figure before and after 1950, because after 1950 the 'non-government, non-private effect' is more directly based on the estimations of the household surveys which is not the case before 1950. The findings below show a strongly different structure of human capital among the three countries.

We first notice that in Japan the percentage of government expenditure in the stock of human capital rises from 11% to 32%. Equally, there is a rise in the share of private expenditure. As a consequence, the final column, 'foregone wages and the "non-government, non-private part"' must decrease. However, if we look at these figures in absolute terms, we see that all categories increased, except the 'non-government, non-private part'. In other words, the decrease in 'non-government, non-private' effect was matched by a rise in formal education, reflected by a comparable increase in public and (to a lesser extent) private expenditure. Compared to Japan, India was on the other extreme. Although, just as in Japan, the share of government expenditure in the stock of

[117] However, for the interested reader: 'experience' plus 'on the job training' is calculated to be on average around 15% in the 1990s. There are also many other options to compare the estimated human capital. For example if we compare human capital by sex, we find that in 1993 for Japan these were 53,549 USD and 55,855 USD per capita for females and males respectively. In Indonesia and India these figures were much more skewed in favour of males. For example in 1993 in Indonesia these figures were 1,561 USD and 2,142 USD for females and males respectively.

Table 5.7: Percentage breakdown of the human capital stock for India, Indonesia and Japan before and after 1950*

	Government expenditure	Private expenditure	Foregone wages and 'non-government, non-private part'
		Japan	
1900-1950	10.8%	3.7%	85.5%
1950-2000	32.4%	14.2%	53.4%
		India	
1900-1950	24.6%	29.9%	45.5%
1950-2000	53.4%	23.4%	23.2%
		Indonesia	
1900-1950	23.7%	19.7%	56.6%
1950-2000	21.9%	11.6%	66.5%

* Under the assumption that the depreciation per age is the same for all forms of human capital. The ‘non-government, non-private part’ is calculated as 100% minus the rest. The reason is that it is likely that the others have about the same depreciation. However, depreciation is likely to be higher in the ‘non-government, non-private part’ meaning that if one would take the same depreciation for all, one would overestimate the share of ‘non-government, non-private part’.

human capital increased strongly, this was accompanied by a decrease in private spending. The development of Indonesia is between that of India and Japan. Just as in India, we find that there is a reduction of the share of private spending in the stock of human capital. However, just as in Japan, we find a relatively high share of ‘foregone wages and non-government, non-private effect’. This is largely caused by the increase in ‘foregone wages’.

This pattern suggests, as we found in chapter 4, that most countries go through the same cycle. They start from a less developed state of formal education. In this phase the share of home education, on the job training and private education is large. This corresponds to the situation in India and Indonesia in the first half of the twentieth century. In the second phase, countries move towards the development of mass public education. This public education competes in first instance with private education. As most public education is free, households which used to spend money on private education can now use this money for other purposes. Hence, expenditure on private education declines relative to expenditure on public education (see also Tilak 1984) which is the case in India and Indonesia in the post-World War II period and in Japan in pre-War period. In the third phase, due to economic growth, there is an increasing demand for persons with secondary and higher education. As public expenditure on these levels of education only covers part of their total costs, households must contribute more to the costs of formal education. This corresponds to the post-War period in Japan.

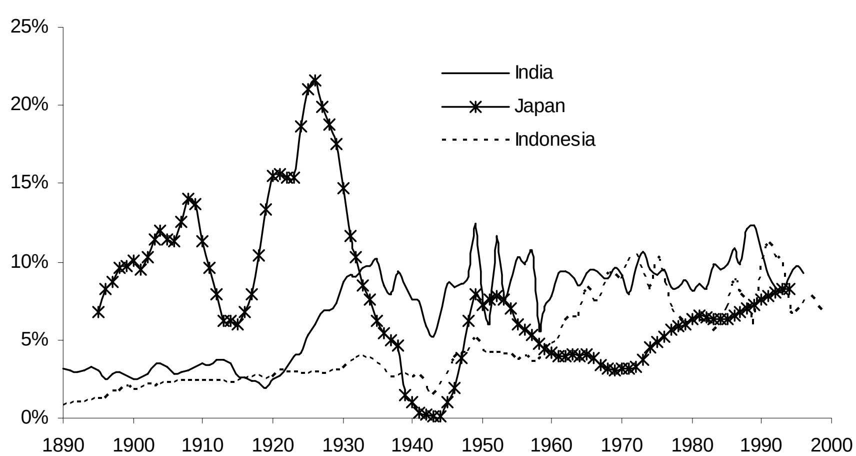
5.4 Some comparisons with GDP and physical capital

In this section we make some observations on the relation between our estimated human capital stock, GDP and the stock of physical capital. The latter two variables offer a consistent framework which we can use to test the plausibility of our estimates over time.

We presented the share of gross human capital formation (in current prices) in GDP in figure 5.4 below. However, there was a problem with the GDP we used. Normally, GDP consists of all forms of capital formation. However, current measures of GDP only include some parts of our human capital formation, most notably public and private expenditure on education. In other words, part of the estimated gross human capital formation is included in GDP, but a part is not. Therefore, we also included the remaining parts of human capital formation in GDP (see appendix A.13). Using this corrected GDP, we notice from figure 5.2 that the share of GHCF in GDP, with one

Figure 5.2

Share of gross human capital formation in human capital corrected GDP for India, Indonesia, and Japan, 1890 -2000 (based on current local prices)



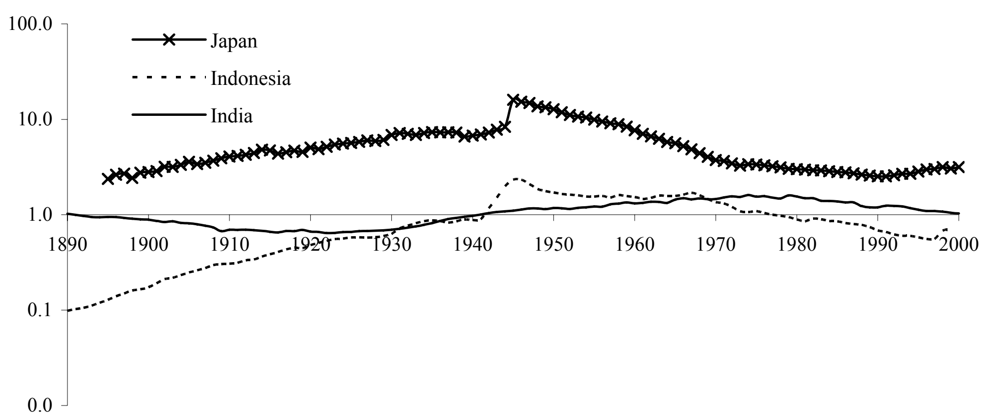
Source: Based on appendices A.9-11 and A.13.

small exception, never exceeds 20% of GDP, which is a likely figure. In addition, Japan clearly exceeds India and Indonesia, at least until the 1930s, in the share of human capital formation in GDP. But afterwards Japan's human capital formation as part of GDP is at, or below, the level of India and Indonesia. Given that Japan's stock of human capital remained well above that of India and Indonesia, this suggest that after gaining an initial lead in human capital, investments as part of GDP may decline while still retaining the lead.

In absolute figures, gross human capital formation in Japan exceeds that of India and Indonesia. This corresponds with Judson (2002, 217) who finds that richer countries have relatively more human capital than poorer ones. Indeed, if we look at the estimated human capital output ratio (see figure 5.3), this clearly is the case. Over the entire century, the human capital-output ratio for Japan exceeds that of India and Indonesia. Except for a peak around 1942, caused by a decline in GDP because of the start of World War II, the Japanese ratio is between 2 and 8. Indeed, this peak is for over 95% caused by a decline in GDP. In India and Indonesia the ratio remains fairly

Figure 5.3

Logarithm of the estimated human capital-output ratio for India, Indonesia, and Japan, 1890 -2000 (based on constant 1990 international USD, converted at PPP)



Source: Appendices A.2 and A.12.

constant at around 1. Just as in Japan, we notice a peak in Indonesia around 1942 which is also caused by the War. India, which was much less hard hit, does not exhibit such a peak in the ratio.

The ratios seem to be in a reasonable order of magnitude. Indeed, for the gross fixed non residential physical capital stock we find capital-output ratios of on average 1 for all three countries.[118] Therefore, especially when keeping in mind that adding residential capital may strongly increase the ratio, the human capital-output ratios do not seem to be implausible. The same is confirmed in some alternative studies on human capital present for Japan and India. For example, Panchamuki (1965, 310) found that the annual average increase in human capital formation was smaller than that of physical capital formation in India in the 1950s. The same development was also found by Gounden (1967, 356). We found that in India the growth of physical capital formation was also higher than that of human capital formation. This is contrary to the observed trends in the USA where human capital formation between 1900 and 1956 rose much stronger than physical capital formation. This seems to be true for developed countries in general, as the Ministry of Education (1963) (see also Dore 1964, 68), just as we did, recorded the same development for Japan.

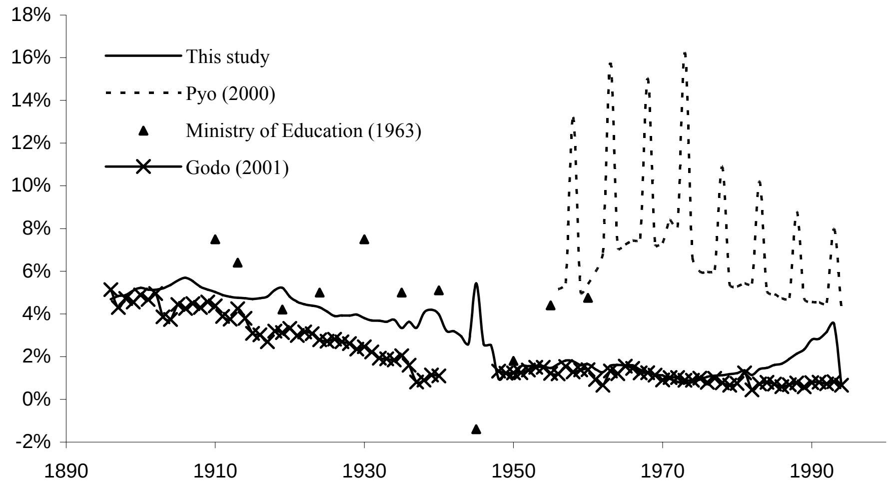
Thus in Japan before 1945, human capital was supposed to grow faster than physical capital while in India the opposite was true. Indeed, this is what we find in figure 5.5. Japan shows an increase in the ratio of human to physical capital up to 1945 while India experiences a small decrease. In Indonesia, however, the ratio remains about constant until the 1940s whereafter it starts to increase. The increase in the ratio of human to physical capital in the 1940-1950 period in Indonesia and Japan is largely caused by World War II which had a negative impact on the stock of physical capital.

However, in the second half of the twentieth century, both countries also experienced a fast decline in this ratio. An objection to this view is raised by Pyo and Jin (2000, 301) who argue that the ratio of human to physical capital in Japan increased in the period 1955-1996. However, their estimates of intangible human capital, based on the method proposed by Kendrick (1976), show implausible growth rates of close to 8 percent on average over the post-War period. The Kendrick method is cost- based, which means that it is largely based on educational expenditure. Yet, we expect that 'true' human capital growth rates must be between measures of the quality of education, which are largely based on educational expenditure (Ministry of Education 1963; Pyo and Jin 2000), and of the of the quantity of education which are largely based on

[118] India exceeds Indonesia and Japan slightly in this respect.

Figure 5.4

A comparison of the growth rates of per capita stock of human capital in Japan, 1895-2000, all monetary variables are based on constant 1990 international USD, converted at PPP



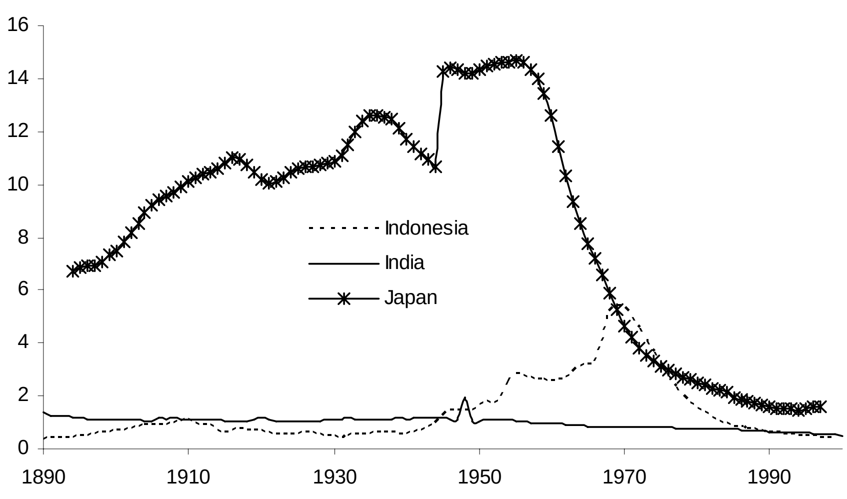
Note: The growth rates of the Ministry of Education (1963) are annualized growth rates based on five- yearly averages

'average years of education' (Godo 2001) (see figure 5.4).

Indeed, Indonesia and Japan started to experience after 1950/1960 a faster

Figure 5.5

Estimated human capital stock-gross fixed non-residential physical capital stock ratio for India, Indonesia, and Japan, 1890 -2000 (based on constant 1990 international USD, converted at PPP)



Source: Appendices A.2 and A.12.

growth of physical compared to human capital (see figure 5.5). This development was not present in India. This is not surprising as India had over the entire twentieth century the highest physical capital-output ratio. In addition, the War harmed the physical capital much more in the former two countries than it did in India. They thus had a relatively small stock of physical capital after World War II, the American occupationand the Dutch police actions respectively.

6. CONCLUSION

This chapter was intended to give a description and analysis of the estimated stock of human capital that we use for our analyses in the next chapters. Our objective was to correct some of the problems with which most human capital proxies are associated. However, we are aware that this alternative human capital stock is also an approximation. Although we hope that it follows the historical development of human capital, it still remains far from perfect.

With these limitations in mind, we constructed an alternative stock of human capital, a human capital stock that solely indicates direct expenditure on human capital (Expenditure HC), and 'average years of education', the latter as given in chapter 3. From our estimates of the alternative human capital stock we can draw two important conclusions. First, it seems that human capital appreciates. To be more specific, if one looks at the human capital of an individual (thus without taking account of mortality), it has an appreciation of around 4.6% for human capital in Japan. In Indonesia and India, this figure is even higher. If we look at the stock of human capital, we see that its appreciation/depreciation is low in all three countries.[119] We estimated that in the 1990s the yearly appreciation in Japan was 0.8%, in India 2.6%, and in Indonesia 3%. These low figures, compared to the much higher depreciation figures of physical capital, are not surprising given that some aspects of human capital appreciate while others depreciate. Second, the growth of the newly estimated per capita stock of human capital is relatively slow at the start and end of the century. The growth increased in the mid-twentieth century. This causes a perfectly logistic curve which is found in many growth studies for social and economic variables.

[119] The yearly appreciation/depreciation is also presented in appendix A.12.

This alone already shows that the newly estimated stock seems to be a reasonably good indicator of human capital. The same follows when we look at some alternative indications of its plausibility. There are three possible indications, namely the subjective margins of error of the data, the division of the stock of human capital in its constituent shares, and a comparison with GDP and physical capital.

First, the subjective margins of error are around 30% for the estimated stocks of human capital. Although this seems large, we have to keep in mind that we are estimating an unobserved variable, a situation which may strongly increase the error. Also, even variables that can be observed directly, like GDP, may suffer from large errors. Furthermore, we find that generally the reliability rises over time and that the reliability for Japan is higher than that for India and Indonesia, which is what we expected to find.

The second method of analysing the stock of human capital was to look at its constituent parts. We found the pattern we have also noted in chapter 4: Japan knew a strong increase in private education which, with a strong rise in public education, compensated for the decrease of the combined share of foregone wages and the 'non-government, non-private' effect in the newly estimated stock of human capital. In India and Indonesia, however, the share of private expenditure decreased. This shows that in developing economies, the rise in public expenditure was nullified by the decrease in private expenditure and 'home education'. Yet, in the latter aspect Indonesia differed from India. The combined share in the stock of human capital of foregone wages and the 'non-government, non-private' effect was in Indonesia much larger than in India.

The third method tries to relate the human capital stock to both GDP and the stock of physical capital. We found that the share of gross human capital formation in GDP was plausible for all three countries, never exceeding 20%. The human capital-output ratio was on average higher than the physical capital-output ratio. Yet, this is partly because we used the gross fixed non-residential physical capital stock, thus excluding residential capital. Including residential capital may lower the human-physical capital ratio considerably.