Economics Program Working Paper Series

Constructing a Data Set on Labour Composition Change

Boele Bonthuis

The Conference Board

December 2011

EPWP #11 – 04

THE CONFERENCE BOARD

Economics Program
845 Third Avenue
New York, NY 10022-6600
Tel. 212-339-0420
http://www.conference-board.org/publications

Constructing a data set on labour composition change*

Boele Bonthuis†

December 14, 2011

Abstract

In this paper the construction of a data set on the change in labour composition is described, thereby creating an insight into this construction process and consequently providing a guideline for using the data. A labour service function as a translog function of all labour input is defined. Using data on share in compensation and hours worked by the different labour inputs we can estimate the change in labour composition. For share of compensation we use data provided by EUKLEMS and Eurostat and use these data to construct an estimate for non-EUKLEMS countries and non-EUKLEMS years. For the construction process of the share of hours worked by different labour types we relied largely on the existing data construction methodology of The Conference Board Total Economy Database 2010, however minor changes to these data set with respect to sources and the interpolation process are made. Ultimately these data are consolidated into one measure of the change in labour composition.

---

*The author is grateful to comments and suggestions by Vivian Chen and Ataman Ozyildirim. Any remaining errors are the responsibility of the author.

†At the time of the data set construction Boele Bonthuis was research assistant at The Conference Board. E-mail: boele@boelebonthuis.com, website: www.boelebonthuis.com

1 Introduction

In this paper the construction of a data set on labour composition is described. This data set is constructed to replace the data set on labour quality in The Conference Board Total Economy Database 2010. That data set was constructed using data on educational attainment and two sets of constant productivity differences among skill groups; one set for each economic group.[1] The reason to replace this data set is that although educational attainment gives a reasonably good impression of the skill level of the labour force it does not necessarily reflect productivity differences. Moreover, a constant productivity difference is not a reasonable assumption; the interaction between different types of labour is not incorporated and the change over time in productivity is not accounted for.

In order to improve the labour service data we tried to separate the change in raw labour (total hours or total employment) and the change in composition of labour input. In order to do this we made the assumption that change in compensation of employees gives a reasonably good insight into the change of productivity of a certain type of labour input. However, it is impossible to assert that compensation levels actually accurately reflect productivity differences between different countries. The fact that compensation levels of western countries even corrected for prices are much higher than the levels in non-western countries, reflects not only productivity differences but also; among other things; institutional factors, unemployment rates and labour specific technology. However, it could be argued that changes in compensation reflect similar changes in productivity across countries. Therefore, we stick to changes in labour composition instead of an actual composition level. First, we will look at the theoretical background of our methodology. Second, we will look at the data availability and estimation technique and finally we will present the main results and conclude.

2 Methodology

The process of constructing this data set can be divided into two steps. First we build the theoretical framework for our data, using and slightly altering existing growth accounting techniques. Then we investigate the possible data sources which we can use to finally construct this data set.

[1] One economic group is composed of advanced countries, the other is composed of developing countries.

2.1 Theoretical framework

In order to determine what kind of labour service function (raw labour and labour composition) is feasible to estimate we have to determine the characteristics of the production function. Let production be some function of technology (A), capital (K) and labour services (L):[2]

$$F_t(A_t, K_t, L_t(H)), \tag{1}$$

in which $L_t(H) = L(H_{i,t}, ..., H_{j,t})$ is a function of labour input; hours worked by different labour types. We assume that both $F_t(.)$ and $L_t(.)$ are differentiable for all feasible values of input, labour markets work perfectly and firms maximize profits. We can therefore write the wage equation for labour type i as:

$$\begin{aligned} w_{i,t} &= \frac{\delta F_t(.)}{\delta H_{i,t}} \\ &= \frac{\delta F_t(.)}{\delta L_t(.)} \frac{\delta L_t(.)}{\delta H_{i,t}}. \end{aligned} \tag{2}$$

We see that wage levels depend on the values of all inputs. However, using equation (2) we can show that the shares of total earnings ($v_{i,t}$) depend only on labour input:

$$\begin{aligned} v_{i,t} &= \frac{w_{i,t} H_{i,t}}{\sum_i w_{i,t} H_{i,t}} \\ &= \frac{\frac{\delta L_t(.)}{\delta H_{i,t}} H_{i,t}}{\sum_i \frac{\delta L_t(.)}{\delta H_{i,t}} H_{i,t}} \end{aligned} \tag{3}$$

Following Jorgenson et al. (1987) we define the labour service function as a trans-log function of all labour inputs:

$$ln(L_t) = \sum_i \alpha_i ln(H_i) + \frac{1}{2} \sum_i \sum_j \beta_{i,j} ln(H_i) ln(H_j), \tag{4}$$

with the following restrictions:[3]

$$\sum_i \alpha_i = 1, \tag{5}$$

$$\beta_{i,j} = \beta_{j,i} \tag{6}$$

$$\sum_i \beta_{i,j} = 0. \tag{7}$$

[2] For now we focus on the single country case.

[3] Therefore equation (4) is homogenous of degree 1.

It can be shown that equation (3) is equal to:

$$\begin{aligned} v_{i,t} &= \frac{\delta ln(L)}{\delta ln(H_{i,t})}, \\ &= \alpha_i + \sum_j \beta_{i,j} ln(H_{j,t}). \end{aligned} \tag{8}$$

Therefore, equation 4 can be rewritten as:

$$ln(L_t) = \frac{1}{2}\left[\sum_i (\alpha_i ln(H_{i,t}) + v_i ln(H_{i,t})\right] \tag{9}$$

And total hours worked can be separated from labour composition because of homogeneity of degree 1:

$$ln(L_t) = ln(H_t) + ln(L_t^c) = ln(H_t) + \frac{1}{2}\sum_i [\alpha_i + v_i] lnh_{i,t} \tag{10}$$

In which $h_{i,t}$ is the share of total hours worked. Using the restrictions imposed on equation (4) we can therefore write the logarithmic change in labour composition as:

$$\Delta ln L_t^c = \frac{1}{2}\sum_i [[v_{i,t} + v_{i,t-1}][lnh_{i,t} - lnh_{i,t-1}], \tag{11}$$

which is a convenient equation to calculate given that data on share in compensation and share in total hours worked are available for a reasonably large set of countries.[4]

2.2 Estimation method

For countries and years in the EUKLEMS data set (see Timmer et al. (2007) for details) we can calculate equation (11) using data on shares in compensation and hours worked for different types of labour.

However, for non EUKLEMS countries and years this is more difficult. We need to generate a data set on share in compensation and share in total hours worked. In the January 2010 version of the Total Economy Database we have already estimated shares in total hours worked for 104 countries for the period 1960-2050 for low, medium and high skilled workers, using data sets of EUKLEMS (see Timmer et al. (2007)), Barro and Lee (2000) and IIASA (see KC et al. (2010)). The same technique is used in this version of the database. However,

[4]Data on the share in compensation of a specific labour type are more common than wage level data, for instance.

instead of using Barro and Lee (2000) we use Cohen and Soto (2001) where possible. The reason for this is that Cohen and Soto (2001) use a more detailed technique for calculating the educational attainment ratios; they fully exploit the demographic nature of a population, dividing the population into cohorts of 5 years, each with an individual educational attainment structure (for details see appendix 5.1). For now, we will briefly explain how all four databases are consolidated into one data set on educational attainment. From here on the data sets by EUKLEMS, Barro and Lee (2000), Cohen and Soto (2001) and the IIASA will be referred to as EUK, BL, CS and IIA respectively.

As with the previous database, we deem EUK to be most accurate. Therefore, we have to transform the other databases into EUK format. EUK classifies the total hours worked into low, medium and high skilled levels, male and female and three different age groups. For the comparison with the three other databases we will focus on the division of skill levels.[5] The biggest difference between the three data sets that have to be transformed and the EUK data set is the skill categorisation and the participation ratio in hours worked (see Bonthuis (2009) for details). Since there is a lack of data on these issues for most countries and time periods, we estimate a statistical relationship to determine which transformation is needed. However, always bear in mind that the estimated coefficients contain implicit assumptions about categorisation and participation. Since CS, BL and IIA all categorise the same way we first consolidate these three data sets, using CS as the base data set. Then, we transform the consolidated data set into EUK format. Therefore, we estimate three sets of statistical relationships: between CS and BL, between CS and IIA and between EUK and CS. The overlapping years and countries on which these relationships are based are shown in table 1.

[5] However, for the construction of the final database the gender and age division will also be used where possible.

Table 1				
Time coverage	CS	BL	IIA	EUK
Start year	1960	1960	2000	1970
End year	2010	2000	2050	2005
Intervals	10	5	5	1
Overlapping countries	CS	BL	IIA	EUK
CS	95			
BL	87	140		
IIA	83	95	123	
EUK	19	24	24	25

To determine the statistical relationship between the data sets we estimate the following equations, for CS and BL:[6]

$$s_{CS,0,c,t} = \beta_1 s_{BL,0,c,t} + e_{0,c,t}, \tag{12}$$

$$s_{CS,1,c,t} = \beta_2 s_{BL,0,c,t} + \beta_3 s_{BL,1,c,t} + e_{1,c,t}, \tag{13}$$

$$s_{CS,2,c,t} = \beta_4 s_{BL,1,c,t} + \beta_5 s_{BL,2,c,t} + \beta_6 s_{BL,3,c,t} + e_{2,c,t}, \tag{14}$$

$$s_{CS,3,c,t} = \beta_7 s_{BL,2,c,t} + \beta_8 s_{BL,3,c,t} + e_{3,c,t}, \tag{15}$$

in which subscripts *0, 1, 2* and *3* stand for *no schooling, primary schooling, secondary schooling* and *tertiary schooling* respectively, subscripts *c, t* are *country* and *time.* For CS and IIA we estimate:

$$s_{CS,0,c,t} = \beta_1 s_{IIA,0,c,t} + e_{0,c,t}, \tag{16}$$

$$s_{CS,1,c,t} = \beta_2 s_{IIA,0,c,t} + \beta_3 s_{IIA,1,c,t} + e_{1,c,t}, \tag{17}$$

$$s_{CS,2,c,t} = \beta_4 s_{IIA,2,c,t} + \beta_5 s_{IIA,3,c,t} + e_{2,c,t}, \tag{18}$$

$$s_{CS,3,c,t} = \beta_6 s_{IIA,2,c,t} + \beta_7 s_{IIA,3,c,t} + e_{3,c,t}, \tag{19}$$

[6]The variables included in the equations are determined by sequentially eliminating all insignificant variables.

Finally for EUK and CS we estimate:[7]

$$h_{l,c,t} = \beta_1 s_{CS,0,c,t} + \beta_2 s_{CS,1,c,t} + \beta_3 s_{CS,2,c,t} + e_{l,c,t}, \quad (20)$$

$$h_{m,c,t} = \beta_4 s_{CS,2,c,t} + e_{m,c,t}, \quad (21)$$

$$h_{h,c,t} = \beta_5 s_{CS,3,c,t} + e_{h,c,t}, \quad (22)$$

Since including a constant would mean that countries only deviate slightly from some world constant in educational attainment constants are not included. Especially since we are transforming one of the data sets, to connect to the other, this would even suggest that the constant is invariable over time. Therefore, we choose to implement just a scalar transformation. Within each of the three systems there is a clear relation between the separate equations. Since all dependent variables add up to one, we cannot simply assume that all equations can be estimated separately. Using the Seemingly Unrelated Regression approach we get a more efficient estimator than under OLS.[8] The results for CS and BL, CS and IIA, and EUK and CS are shown in tables 2, 3 and 4 respectively.

Table 2

CS BL	$s_{CS,0,c,t}$	$s_{CS,1,c,t}$	$s_{CS,2,c,t}$	$s_{CS,3,c,t}$
$s_{BL,0,c,t}$	.968	.033	.	.
	(.008)*	(.008)*		
$s_{BL,1,c,t}$	.	.917	.083	.
		(.011)*	(.011)*	
$s_{BL,2,c,t}$	.	.	.937	.061
			(.007)*	(.006)*
$s_{BL,3,c,t}$	.	.	.377	.635
			(.020)*	(.017)*
R^2	.968	.928	.922	.920
N	423	423	423	423

Standard errors in parenthesis

* Significant on a 1% level

[7]Italy is excluded from the sample because of doubtful accuracy of the data. The low skilled category in Italy comprises of only people without formal education. Therefore, according to the data set, more than 80% of the population is medium skilled since 1964.

[8]SUR still assumes $E(e_i) = 0$ but $E(e_i e_j) \neq 0$

Table 3

CS IIA	$s_{CS,0,c,t}$	$s_{CS,1,c,t}$	$s_{CS,2,c,t}$	$s_{CS,3,c,t}$
$s_{IIA,0,c,t}$	1.030	.	.	.
	(.012)*			
$s_{IIA,1,c,t}$	.	1.000	.	.
		(.014)*		
$s_{IIA,2,c,t}$	.	.138	.787	.055
		(.018)*	(.023)*	(.016)*
$s_{IIA,3,c,t}$	.	.	.293	.674
			(.064)*	(.053)*
R^2	.956	.923	.941	.889
N	166	166	166	166

Standard errors in parenthesis

* Significant on a 1% level

Table 4

EUK CS	$h_{l,c,t}$	$h_{m,c,t}$	$h_{h,c,t}$
$s_{CS,0,c,t}$	.867	.	.
	(.075)*		
$s_{CS,1,c,t}$	.997	.	.
	(.016)*		
$s_{CS,2,c,t}$	.173	.839	.
	(.029)*	(.030)*	
$s_{CS,3,c,t}$	.	.	.972
			(.043)*
R^2	.903	.942	.791
N	48	48	48

Standard errors in parenthesis

* Significant on a 1% level

In table 2 and 3 we expect the coefficients along the diagonal to be close to one, after all we do not expect the different data sets to be too far apart when it comes to categorisation. Indeed, we see that for most coefficients in table 2 this is true. Only the coefficient for tertiary schooling is quite different from one. Apparently BL overestimate (or CS underestimate) the share of the population which has attained tertiary schooling. This can have two explanations. First, the sample size of persons who have obtained tertiary schooling is relatively small, therefore, a small difference in absolute figures can have a significant difference in shares. Second, as explained in the appendix, the assumption of CS that the oldest cohort at time t has the same educational attainment as the second oldest cohort at time t-1 amounts to the effect that the information about the survivors of the oldest cohort is discarded. Therefore, assuming an increasing educational attainment trend, forward (backward) extrapolating would overestimate (underestimate) the overall educational attainment. It is quite likely that backward extrapolation is used more often because actual educational attainment data (the starting point of the extrapolation) are probably more recent because of the improvement in data collection. Therefore, it is likely that the share of the population with tertiary schooling is estimated to be bigger in the BL data set than in the CS data set. Which of the estimates is closer to reality can only be determined through examination of the the respective methodologies. As explained before, we deem CS to be more accurate.

Furthermore, we expect the sum of coefficients of a certain category (sum of a particular row of the tables) across equations to be close to one, this should hold for all three systems. The reason for this is that three data sets (CS, BL and IIA) observe the entire population of 15+, hence an individual observed in CS also has to be included in BL and IIA, even though this individual could be categorised differently. Therefore, to ensure that all persons in CS are represented in BL and IIA the coefficients should (by approximation) add up to one. Indeed we see that this restriction is satisfied for most coefficients. For the relationship between CS and EUK this restriction is less stringent since the coefficients also represent the participation ratio, which means that not all persons in CS (i.e. the inactive) are represented in EUK.

Once we have established the three systems we first consolidate all data categorised like CS. We take CS as basis, and fill in the gaps using the transformed data from BL and IIA. If there are no observations for a country in CS we simply use the transformed BL and/or IIA data, if CS is missing years we extrapolate using the growth rate of the transformed data.[9] Once we have established a combined data set of CS, BL and IIA we transform this data set into EUK categorisation using the relationship between EUK and CS. For countries in EUK we simply start with original EUK data. For missing years (before 1970 and after 2005) we extrapolate using growth rates of the transformed BL-EUK-IIA data set. For countries not in the EUK data set we simply use the transformed data. The missing data point in between observations are interpolated using constant yearly growth rates.[10] It can happen that not all three categories at a certain point in time add up to one, which is why normalisation is needed for all extrapolated, interpolated and transformed data. Normalisation means that each observation is divided by the sum off all three categories. However, this adjustment is minor since the sum of most data at a certain point in time is close to one. Now we have a data set on share in total hours worked by low, medium and high skilled labour, containing 140 countries with a maximum time range of 1960-2050.

Needless to say that to construct a data set on labour composition, we still need data on the share in compensation. For few countries there exist data sets on wages or labour compensation.[11] Therefore, we need to construct a reasonable estimate. Using EUKLEMS data we can re-create relative wages using:

$$\frac{w_{i,c,t}}{\bar{w}_{c,t}} = \frac{w_{i,c,t}H_{i,c,t}}{\bar{w}_{c,t}H_{c,t}}\frac{H_{c,t}}{H_{i,c,t}} = \frac{v_{i,c,t}}{h_{i,c,t}}, \tag{23}$$

in which $\bar{w}_{c,t}$ is the average wage and $H_{c,t}$ is total hours worked in country c at time t. For non-EUKLEMS countries we average over every EUKLEMS country (excluding Italy) for every year.[12] We then get: $\bar{w}_{i,t}/\bar{\bar{w}}_t$. Multiplying this with the share of hours worked, we create an estimate for $v_{i,c,t}$ for every country. For non-EUKLEMS years we estimate the trend

[9] This is of course always the case for years included in IIA.

[10] This is one of the minor changes made compared to the old data set on share of hours worked.

[11] This exists merely for Western countries, for which we already have original data on the share of total compensation

[12] Eurostat data on average wages $\bar{\bar{w}}_t$ in combination with the hours share and the compensation share is used to calculate the average of every type of wage.

of $w_{i,c,t}/\bar{w_{c,t}}$ and using the projected values we calculate $v_{i,c,t}$ for every country and year:

$$\frac{\bar{w}_{l,t}}{\bar{\bar{w}}_t} = \beta_{0,l} + \beta_{1,l}\frac{\bar{w}_{l,t-1}}{\bar{\bar{w}}_{t-1}} + e_{l,t} \tag{24}$$

$$\frac{\bar{w}_{m,t}}{\bar{\bar{w}}_t} = \beta_{0,m} + \beta_{1,m}\frac{\bar{w}_{m,t-1}}{\bar{\bar{w}}_{t-1}} + e_{m,t} \tag{25}$$

$$\frac{\bar{w}_{h,t}}{\bar{\bar{w}}_t} = \beta_{0,h} + \beta_{1,h}\frac{\bar{w}_{h,t-1}}{\bar{\bar{w}}_{t-1}} + e_{h,t} \tag{26}$$

Because the normal fixed effects panel estimator would return inconsistent results (the lagged dependent variable is by definition dependent on $\beta_{0,i}$) we use an Arellano and Bond dynamic panel estimation; further lags of the dependent variable are used as instruments for the first lag, with maximum lag length 7 to avoid over-identification. The trend results are as follows:

Table 5 Trend relative wages

	$\frac{\bar{w}_{l,t}}{\bar{\bar{w}}_t}$	$\frac{\bar{w}_{m,t}}{\bar{\bar{w}}_t}$	$\frac{\bar{w}_{h,t}}{\bar{\bar{w}}_t}$
Constant	.008	.046	.099
	(.011)	(.008)*	(023)*
$\frac{\bar{w}_{l,t-1}}{\bar{\bar{w}}_{t-1}}$	.983	.	.
	(.015)*		
$\frac{\bar{w}_{m,t-1}}{\bar{\bar{w}}_{t-1}}$	.	.948	.
		(.008)*	
$\frac{\bar{w}_{h,t-1}}{\bar{\bar{w}}_{t-1}}$	.	.	.935
			(.013)*
P(Wald X^2)	.000	.000	.000
N	487	487	487

Standard errors in parenthesis

* Significant on a 1% level

The trend is relatively stable over time since the coefficient of the lagged variable is close to one. Eventually the system described here converges to .047, .88 and 1.52 for low, medium and high relative wages respectively. Technically speaking this could be a reasonable outcome since relative wages need to be centered around 1.

We now have a complete coverage of either original data or estimates on shares in hours worked and shares in compensation of different labour inputs. For non-EUKLEMS countries this is restricted to a breakdown in low, medium and high skilled labour. However, for

EUKLEMS countries we have a breakdown in skill levels (three types), gender (two) and age (three cohorts), creating 18 types of labour inputs. Using this information we can estimate the change of labour composition using equation 11.

3 Results

In this section we present the results for a selection of countries. Given the size of the final data set we cannot present all results. The chosen selection consists of all G8 countries, BRIC countries and South Korea, giving a total of twelve countries (since Russia is both in G8 and BRIC).[13]

Looking at the results we can see that the labour composition variable can be quite volatile; more volatile than the overall labour service function. The reason for this is that the interaction between the variables in the labour composition function is much stronger. Suppose that hours worked by high skilled labour increases. In the labour service function this only has an effect on the part of the function containing hours worked by high skilled labour. However, if the hours worked by high skilled labour increase, the share in hours worked by high skilled labour also increases. Meanwhile, the combined shares in hours worked by other labour types will fall by the same amount. Even though the share in compensation by each labour type is deemed to be implicit in the labour composition function (see equation (4) from which equation (8) is derived) it can be that this share differs in reality from its long term equilibrium. Therefore, a labour composition figure, constructed from 18 types of labour input, can change rapidly over time, even though the data used has smooth transitions. We can see this is indeed the case for shares of hours worked and shares in compensation, see figure 1 (for a more technical discussion see appendix 5.2 and 5.3).

Since the labour composition variable is quite volatile and difficult to interpret from year to year, we have calculated the rolling ten year average growth rates for the selected countries, which are shown in figure 2. As we can see the pattern differs significantly from country to country.

For most North-American and European countries (excluding the UK) we see steady or declining labour composition growth rates. We see that South-Korea has had stunning growth rates with only the 70's as a slowing period of less than a half percent (ignoring projections

[13]South Korea is chosen because of its interesting development of labour composition

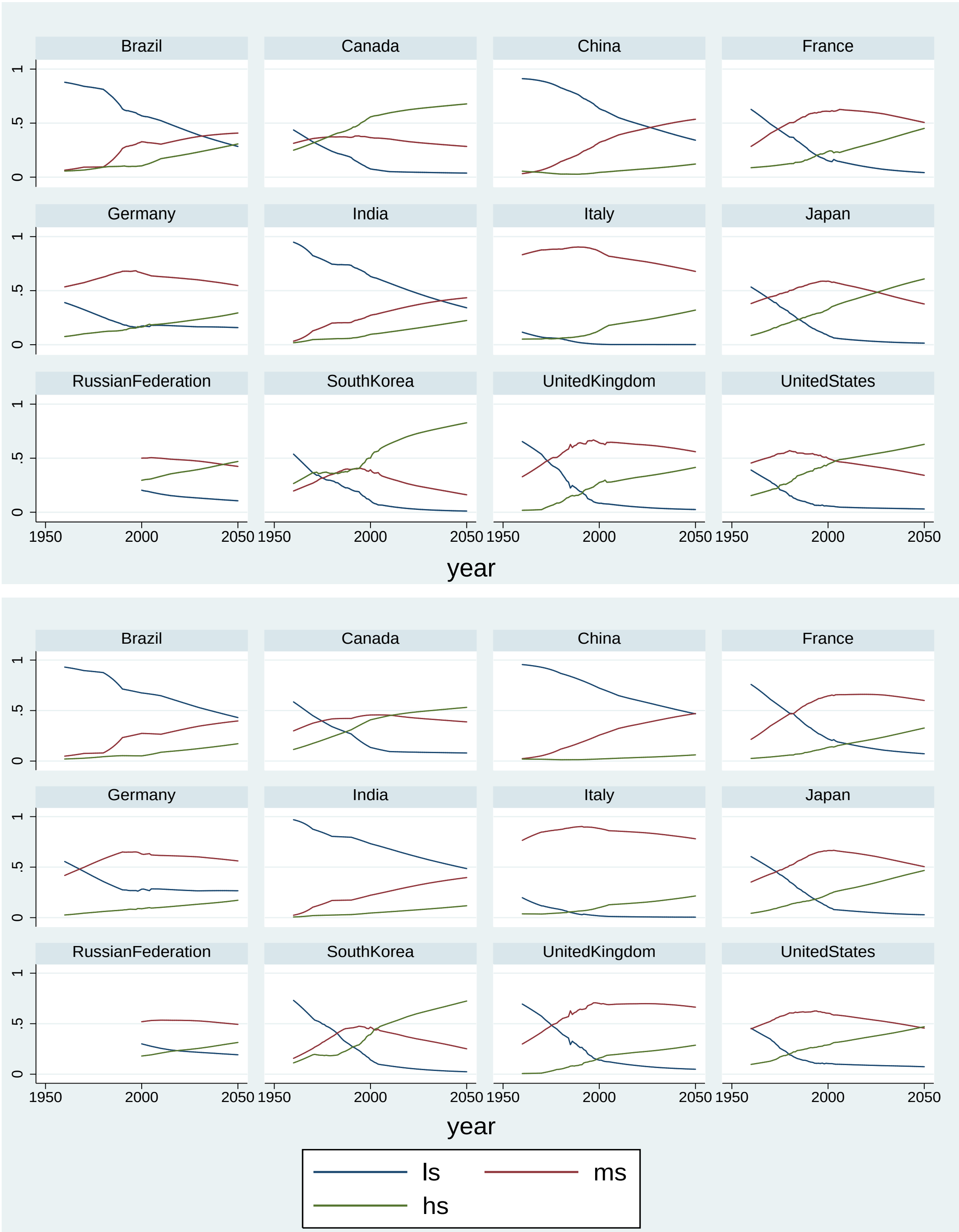

Figure 1: Share of compensation $(v_{j,t})$(upper panel) and share of hours worked $(h_{j,t})$ (lower panel), ls = low skilled, ms = medium skilled, hs = high skilled

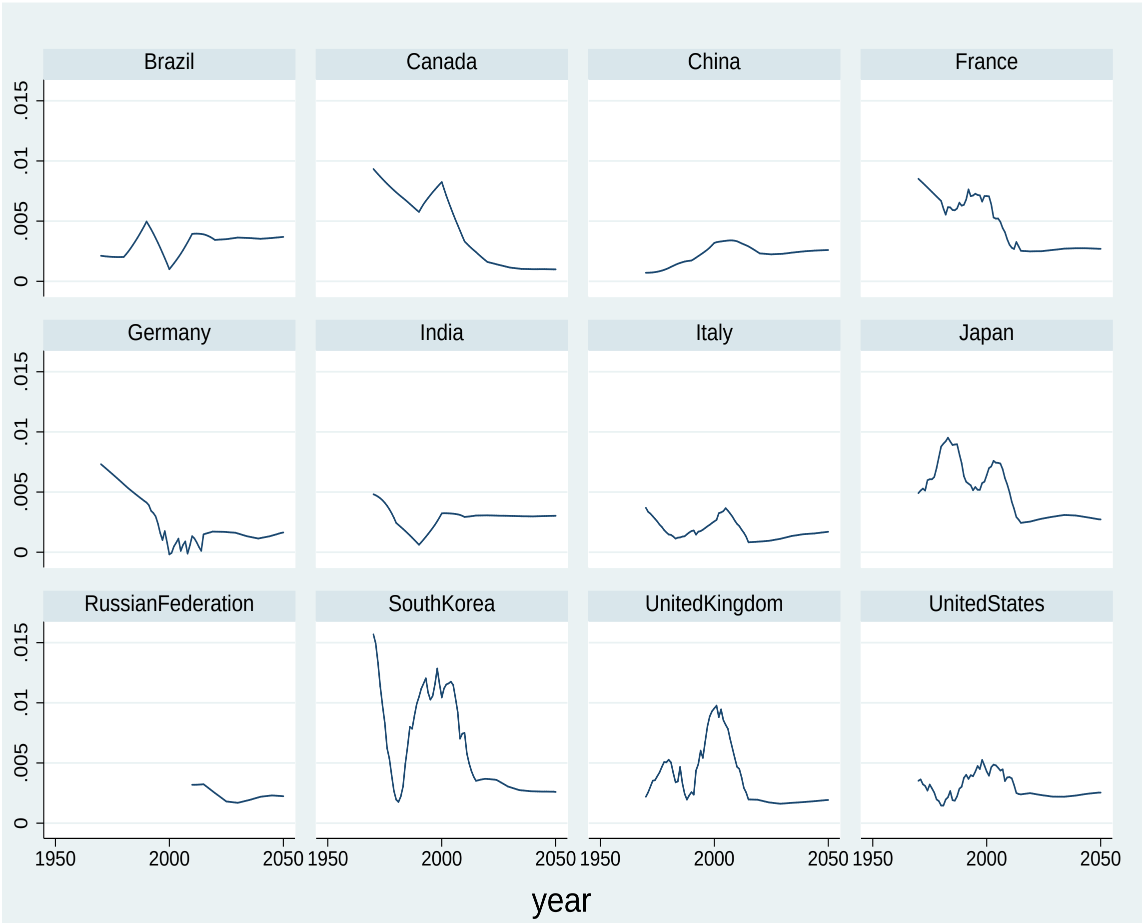

Figure 2: Average 10 year labour composition growth ($L_{c,t}$)

for a moment). This is due to the rapid decline in the share of total hours worked by low skilled labour in the previous century and the increase in; first medium skilled labour and then high skilled labour (South-Korea is currently the country with the highest share of high skilled labour). The UK exhibits relatively robust growth during the 90's and early 2000's. Interestingly enough it seems that this growth cannot be attributed to rapid growth in the share of hours worked by high skilled labour like the South-Korean case, but rather to a more than proportional rise in the share of compensation by high skilled labour, indicating a rise in productivity by this labour type.[14] Japan exhibits a combination of effects seen in the UK and South-Korea, a relatively modest increase in share of high skilled labour and a slightly more than proportional increase in share of compensation by high skilled labour, creating an average labour composition growth of above .5% for the entire previous century. Even though

[14]However, this can also indicate that high skilled labour improved their wage negotiation skills.

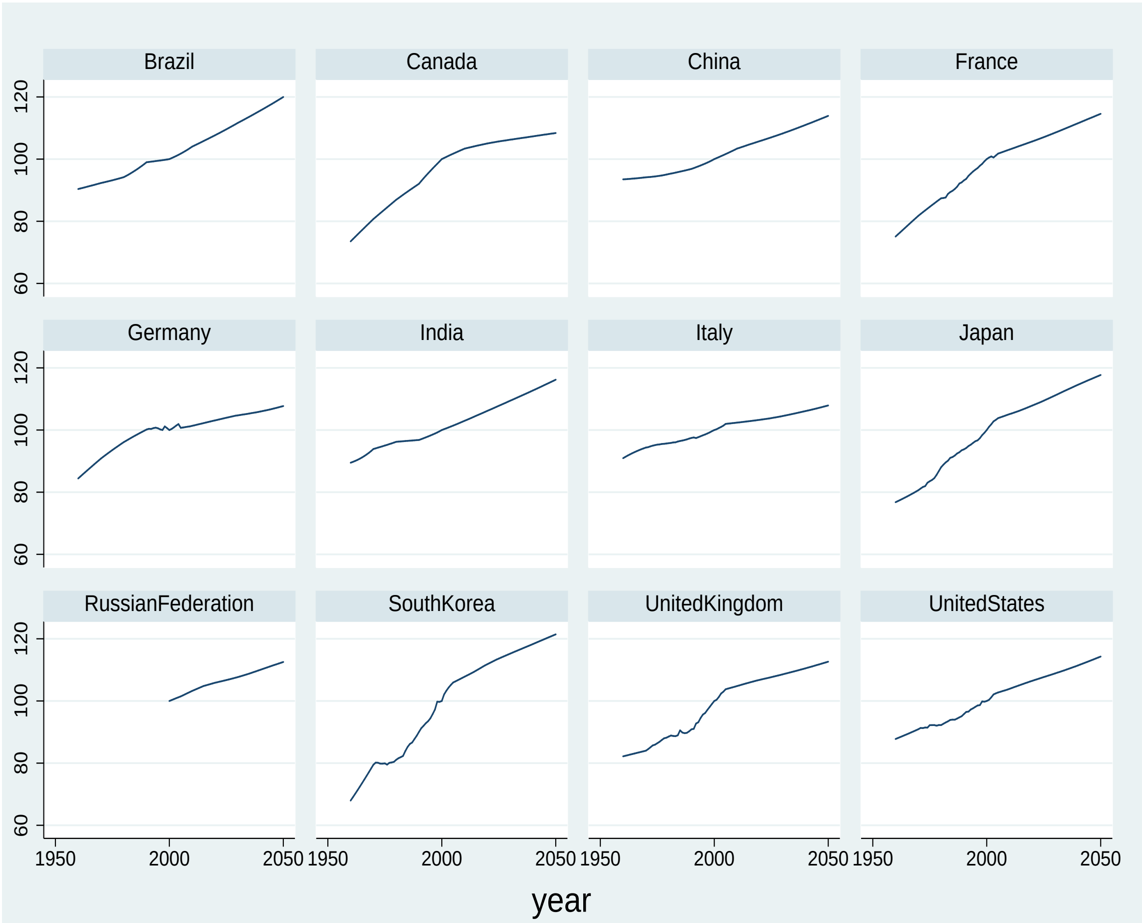

Figure 3: Labour composition growth (index, 2000=100)

most BRIC countries have experienced rapid economic growth during the last decade, most of this growth does not seem to be coming from labour composition growth. All four countries have growth rates of less than .5%.

In figure 3 we see the effect of indexing the labour composition growth rate. Again we see the same pattern as in the previous figure, with South Korea, Japan and the UK leading the way and the rest of the Western countries trailing with only Canada having significant increase early in the second half of the previous century. We see that even though all BRIC countries have relatively modest labour composition growth rates, they are expected to be picking up quite quickly during the coming decades.[15]

Overall we see that labour composition growth is not very large compared to other sources

[15] In the appendix 5.3 we have included a table showing for every decade the slowest 3 and fastest five labour composition growth rates and corresponding countries.

of growth. The relatively modest role of labour composition change in growth is what we would expect if we take into consideration the slow moving nature of educational attainment improvements.

The declining nature of labour composition growth could mean several things.[16] Classical growth theory would suggest that all countries eventually reach a steady state. Advanced countries would reach points close to this state (assuming similar transition paths) earlier than developing countries.[17] Slowing growth rates, in this case slowing labour composition growth rates, would reflect approaching the steady state. However, if we assume endogenous growth through human capital accumulation à la Uzawa (1965) and Lucas (1988), this would mean that these countries actually have stagnating human capital accumulation functions.[18]

4 Conclusion

We have successfully constructed a data set on the change in labour composition for a large set of countries (140) for a considerable length of time (1960-2050). The advantage over the previous labour quality data set by The Conference Board is that our current methodology is firmly rooted in existing growth accounting theory and reflects common assumptions made in this line of research, this data set is therefore deemed to more accurately reflect productivity differences between labour types.

[16] Germany's labour composition growth rate even turns negative for a short period of time in the 90's, but this is probably due to the re-unification of Germany at the beginning of that decade.

[17] Following Mankiw et al. (1992) it would mean that the capital accumulation function and human capital accumulation function (i.e. labour composition growth function) would take this form:

$$\Delta K = s_K Y - \delta K, \tag{27}$$

$$\Delta H = s_H Y - \delta H, \tag{28}$$

in which H is production, δ is the depreciation rate, ΔK is change in capital and ΔH is change in human capital, similar to our change in labour composition. ΔH declines when the steady state is approached (assuming for a moment a steady population size).

[18] In Uzawa (1965) the human capital accumulation function (in continuous time) takes the form of:

$$\dot{H} = \phi(1-u)H, \tag{29}$$

in which u is the part of labour used in production and ϕ is some constant denoting the efficiency of education. We see that this function is only declining if u declines (u should even decline at a faster rate to offset the initial rise in H the previous period).

However, as with the previous data set, caution is needed when using the data. Transformations used in constructing the data set are based on characteristics of a sample of all countries and therefore might not perfectly reflect real differences between all countries. We therefore stress the importance of judging the quality of the data when using it and interpreting the results.

At this point a couple of unresolved issues remain. First, an obvious improvement of the data set would be the inclusion of data for all countries on hours worked by skill types instead of shares of population by skill types. Second, data on shares of total compensation per skill type for all countries would also be an improvement of the data. Finally, we continue looking for a labour composition level at a certain point in time which is comparable across countries. Using this benchmark point we could apply the labour composition growth rates, creating comparable labour composition level data for all countries and all years.

Ultimately we can conclude that the creation of this data set is an important step forward compared to existing data sources, but that further improvement remains a considerable challenge.

5 Appendix

5.1 Barro and Lee vs. Cohen and Soto

In their paper CS focus primarily on average years of schooling. Since we are interested in the construction of the separate ratios of schooling levels we have to analyse their method for average years of schooling and infer their method with respect to the schooling level ratios. If we look at the definition of average years of schooling in CS's paper:

$$y_t = \sum_{g=1}^{G} I_t^g y_t^g \tag{30}$$

Where I_t^g is the population share of group g (five year age groups) in population 15+ and y_t^g is defined as:

$$y_t^g = \sum_{i=1}^{3} s_{i,t}^g D_{i,t} \tag{31}$$

In which $s_{i,t}^g$ is the fraction of group g having attained level i and $D_{i,t}$ is the duration of level i.[19] If we reverse the order of summation we can rewrite equations (30) and (31) in a more

[19]Note that instead of hours worked Barro and Lee as well as Cohen and Soto focus on shares of population.

convenient way for our purpose:

$$y_t = \sum_{i=1}^{3} D_{i,t} s_{i,t} \tag{32}$$

In which $s_{i,t}$ is:

$$s_{i,t} = \sum_{g=1}^{G} s_{i,t}^{g} I_t^g \tag{33}$$

(33) is the measure we are interested in; the fraction of population 15+ which attained schooling level i. For some years $s_{i,t}$ is observed for both CS and BL, then the only difference between BL and CS can be the use of different sources. However, for the years in which $s_{i,t}$ cannot be observed there exists a large extrapolation difference between CS and BL. CS use net intake rates for the different levels of schooling and they explain how these net intake rates are calculated, they then extrapolate forward according to:[20]

$$s_{i,t+5} = \sum_{g=1}^{g=3} I_{t+5}^{g} \hat{s}_{i,t+5}^{g} + \sum_{g=4}^{G} I_{t+5}^{g} s_{i,t}^{g-1} \tag{34}$$

Where in the first part of the right hand side the estimated attainment ratio is used:

$$\hat{s}_{i,t+5}^{g} = R_{i,t+5i-5g} - R_{i+1,t+5(i+1)-5g} \tag{35}$$

R_i is the net intake rate in level i which is equivalent to BL's enrollment ratios[21]. So written out we get:[22]

$$\begin{aligned}
s_{0,t+5} &= I_{t+5}^{1}(1 - R_{1,t}) + I_{t+5}^{2}(1 - R_{1,t-5}) + \\
&\quad I_{t+5}^{3}(1 - R_{1,t-10}) + \sum_{g=4}^{G} I_{t+5}^{g} s_{0,t}^{g-1}, && (36)\\
s_{1,t+5} &= I_{t+5}^{1}(R_{1,t} - R_{2,t+5}) + I_{t+5}^{2}(R_{1,t-5} - R_{2,t}) + \\
&\quad I_{t+5}^{3}(R_{1,t-10} - R_{2,t-5}) + \sum_{g=4}^{G} I_{t+5}^{g} s_{1,t}^{g-1}, && (37)\\
s_{2,t+5} &= I_{t+5}^{1}(R_{2,t+5}) + I_{t+5}^{2}(R_{2,t} - R_{3,t+5}) + \\
&\quad I_{t+5}^{3}(R_{2,t-5} - R_{3,t}) + \sum_{g=4}^{G} I_{t+5}^{g} s_{2,t}^{g-1}, && (38)\\
s_{3,t+5} &= I_{t+5}^{2} R_{3,t+5} + I_{t+5}^{3} R_{3,t} + \sum_{g=4}^{G} I_{t+5}^{g} s_{3,t}^{g-1} && (39)
\end{aligned}$$

[20]See section 2.1 of Cohen and Soto (2001).

[21]$R_0 = 1$ and if i exceeds 3 and/or the time subscripts exceeds $t + 5$ then $R = 0$.

[22]In which 0, 1, 2 and 3 stand for no schooling, primary schooling, secondary schooling and tertiary schooling respectively.

Next we look at the method BL use. First we rewrite sys.2 of Bonthuis (2009):

$$
\begin{aligned}
s_{0,t+5} &= I^1_{t+5}(1-R_{1,t}) + (1-I^1_{t+5})\sum_{g=1}^{G} I^g_t s^g_{0,t} && (40)\\
s_{1,t+5} &= I^1_{t+5}(R_{1,t}-R_{2,t+5}) + (1-I^1_{t+5})\sum_{g=1}^{G} I^g_t s^g_{1,t} && (41)\\
s_{2,t+5} &= I^1_{t+5}R_{2,t+5} - I^2_{t+5}R_{3,t+5} + (1-I^1_{t+5})\sum_{g=1}^{G} I^g_t s^g_{2,t} && (42)\\
s_{3,t+5} &= I^2_{t+5}R_{3,t+5} + (1-I^1_{t+5})\sum_{g=1}^{G} I^g_t s^g_{2,t} && (43)\\
& && (44)
\end{aligned}
$$

Now we have written BL in the same way as CS. If we subtract (36) from (40),(37) from (41),(38) from (42), and (39) from (43) we have the actual difference between BL and CS.

$$
\begin{aligned}
\Delta s_{0,t+5} &= (1-I^1_{t+5})I^1_t s^1_{0,t} - I^2_{t+5}(1-R_{1,t-5}) + \\
&\quad (1-I^1_{t+5})I^2_t s^2_{0,t} - I^3_{t+5}(1-R_{1,t-10}) + \\
&\quad \sum_{g=4}^{G}[(1-I^1_{t+5})I^{g-1}_t - I^g_{t+5}]s^{g-1}_{0,t} + (1-I^1_{t+5})I^G_t s^G_{0,t}, && (45)\\
\Delta s_{1,t+5} &= (1-I^1_{t+5})I^1_t s^1_{1,t} - I^2_{t+5}(R_{1,t-5}-R_{2,t}) + \\
&\quad (1-I^1_{t+5})I^2_t s^2_{1,t} - I^3_{t+5}(R_{1,t-10}-R_{2,t-5}) + \\
&\quad \sum_{g=4}^{G}[(1-I^1_{t+5})I^{g-1}_t - I^g_{t+5}]s^{g-1}_{1,t} + (1-I^1_{t+5})I^G_t s^G_{1,t}, && (46)\\
\Delta s_{2,t+5} &= (1-I^1_{t+5})I^1_t s^1_{2,t} - I^2_{t+5}R_{2,t} + (1-I^1_{t+5})I^2_t s^2_{2,t} - \\
&\quad I^3_{t+5}(R_{2,t-5}-R_{3,t}) + \sum_{g=4}^{G}[(1-I^1_{t+5})I^{g-1}_t - I^g_{t+5}]s^{g-1}_{2,t} + \\
&\quad (1-I^1_{t+5})I^G_t s^G_{2,t}, && (47)\\
\Delta s_{3,t+5} &= (1-I^1_{t+5})I^1_t s^1_{3,t} + (1-I^1_{t+5})I^2_t s^2_{3,t} - I^3_{t+5}R_{3,t} + \\
&\quad \sum_{g=4}^{G}[(1-I^1_{t+5})I^{g-1}_t - I^g_{t+5}]s^{g-1}_{3,t} + (1-I^1_{t+5})I^G_t s^G_{3,t}, && (48)
\end{aligned}
$$

Now let's see under what condition the difference is 0. In order to compare we have to use the following equations. We use for $g=1$:

$$
I^1_{t+5} = \frac{N^1_{t+5}}{N_{t+5}} = \frac{\beta_{t+5}N_t}{N_{t+5}} \qquad (49)
$$

For all $g > 1$ and $g < G$:

$$I_{t+5}^{g} = \frac{N_{t+5}^{g}}{N_{t+5}} = \frac{(1-\delta_{t+5}^{g-1})N_{t}^{g-1}}{N_{t+5}} \tag{50}$$

For $g = G$:

$$I_{t+5}^{G} = \frac{N_{t+5}^{G}}{N_{t+5}} = \frac{(1-\delta_{t+5}^{G-1})N_{t}^{G-1} + (1-\delta_{t+5}^{G})N_{t}^{G})}{N_{t+5}} \tag{51}$$

and:

$$N_{t+5} = (1-\bar{\delta}_{t+5} + \beta_{t+5})N_t \tag{52}$$

In which N is the number of people in a certain category, δ_{t+5} is the death rate between t and $t+5$ (subscript indicates cohort death rate and bar means average death rate) and β_{t+5} is the birth rate between t and $t+5$. Equation (51) holds because the group that was already in G at time t remains in group G at time $t+5$.

Using (49)-(52) in equation (45) gives:

$$\begin{aligned}\Delta s_{0,t+5} = & \left[\frac{(1-\bar{\delta}_{t+5})s_{0,t}^{1} - (1-\delta_{t+5}^{1})(1-R_{1,t-5})}{1-\bar{\delta}_{t+5}+\beta_{t+5}}\right] I_t^1 + \\ & \left[\frac{(1-\bar{\delta}_{t+5})s_{0,t}^{2} - (1-\delta_{t+5}^{2})(1-R_{1,t-10})}{1-\bar{\delta}_{t+5}+\beta_{t+5}}\right] I_t^2 + \\ & \sum_{g=3}^{G-1}\left[\frac{(\delta_{t+5}^{g} - \bar{\delta}_{t+5})s_{0,t}^{g}}{1-\bar{\delta}_{t+5}+\beta_{t+5}}\right] I_t^g + \left[\frac{(1-\bar{\delta}_{t+5})s_{0,t}^{G} - (1-\delta_{t+5}^{G})s_{0,t}^{G-1}}{1-\bar{\delta}_{t+5}+\beta_{t+5}}\right] I_t^G\end{aligned} \tag{53}$$

Equation (49) gives:

$$\begin{aligned}\Delta s_{1,t+5} = & \left[\frac{(1-\bar{\delta}_{t+5})s_{1,t}^{1} - (1-\delta_{t+5}^{1})(R_{1,t-5}-R_{2,t})}{1-\bar{\delta}_{t+5}+\beta_{t+5}}\right] I_t^1 + \\ & \left[\frac{(1-\bar{\delta}_{t+5})s_{1,t}^{2} - (1-\delta_{t+5}^{2})(R_{1,t-10}-R_{2,t-5})}{1-\bar{\delta}_{t+5}+\beta_{t+5}}\right] I_t^2 + \\ & \sum_{g=3}^{G-1}\left[\frac{(\delta_{t+5}^{g} - \bar{\delta}_{t+5})s_{1,t}^{g}}{1-\bar{\delta}_{t+5}+\beta_{t+5}}\right] I_t^g + \left[\frac{(1-\bar{\delta}_{t+5})s_{1,t}^{G} - (1-\delta_{t+5}^{G})s_{1,t}^{G-1}}{1-\bar{\delta}_{t+5}+\beta_{t+5}}\right] I_t^G\end{aligned} \tag{54}$$

Equation (50) gives:

$$\begin{aligned}\Delta s_{2,t+5} = & \left[\frac{(1-\bar{\delta}_{t+5})s_{2,t}^{1} - (1-\delta_{t+5}^{1})R_{2,t}}{1-\bar{\delta}_{t+5}+\beta_{t+5}}\right] I_t^1 + \\ & \left[\frac{(1-\bar{\delta}_{t+5})s_{2,t}^{2} - (1-\delta_{t+5}^{2})(R_{2,t-5}-R_{3,t})}{1-\bar{\delta}_{t+5}+\beta_{t+5}}\right] I_t^2 + \\ & \sum_{g=3}^{G-1}\left[\frac{(\delta_{t+5}^{g} - \bar{\delta}_{t+5})s_{2,t}^{g}}{1-\bar{\delta}_{t+5}+\beta_{t+5}}\right] I_t^g + \left[\frac{(1-\bar{\delta}_{t+5})s_{2,t}^{G} - (1-\delta_{t+5}^{G})s_{2,t}^{G-1}}{1-\bar{\delta}_{t+5}+\beta_{t+5}}\right] I_t^G\end{aligned} \tag{55}$$

and equation (51) gives:

$$\begin{aligned}\Delta s_{3,t+5} &= \left[\frac{(1-\bar{\delta}_{t+5})s^1_{3,t}}{1-\bar{\delta}_{t+5}+\beta_{t+5}}\right] I^1_t + \left[\frac{(1-\bar{\delta}_{t+5})s^2_{3,t}-(1-\delta^2_{t+5})R_{3,t}}{1-\bar{\delta}_{t+5}+\beta_{t+5}}\right] I^2_t + \\ &\sum_{g=3}^{G-1}\left[\frac{(\delta^g_{t+5}-\bar{\delta}_{t+5})s^g_{3,t}}{1-\bar{\delta}_{t+5}+\beta_{t+5}}\right] I^g_t + \left[\frac{(1-\bar{\delta}_{t+5})s^G_{3,t}-(1-\delta^G_{t+5})s^{G-1}_{3,t}}{1-\bar{\delta}_{t+5}+\beta_{t+5}}\right] I^G_t \qquad (56)\end{aligned}$$

Let's look for the conditions that return (53)-(56) are equal to 0. The only meaningful solution to this problem is if all elements in equation (53)-(56) are 0. Therefore, the following must hold for all four equations:

$$\bar{\delta}_{t+5} = \delta^g_{t+5} \text{ for all } g \qquad (57)$$

$$s^G_{i,t} = s^{G-1}_{i,t} \text{ for all } i \qquad (58)$$

For equation (53) in particular:

$$s^1_{0,t} = 1 - R_{1,t-5} \qquad (59)$$

$$s^2_{0,t} = 1 - R_{1,t-10} \qquad (60)$$

For equation (54):

$$s^1_{1,t} = R_{1,t-5} - R_{2,t} \qquad (61)$$

$$s^2_{1,t} = R_{1,t-10} - R_{2,t-5} \qquad (62)$$

For equation (55):

$$s^1_{2,t} = R_{2,t} \qquad (63)$$

$$s^2_{2,t} = R_{2,t-5} - R_{3,t} \qquad (64)$$

For equation (56):

$$s^1_{3,t} = 0 \qquad (65)$$

$$s^2_{3,t} = R_{3,t} \qquad (66)$$

Conditions (59), (60), (61), (62), (64), and (66) are likely to be satisfied since the enrollment rates at the particular dates reflect the educational attainment ratios for that particular cohort. Cohort 1 (aged 15-19) at time t, for instance, has finished primary schooling at time *t-5*. Information which is already incorporated in the education attainment ratios of period t. Since these ratios are observed by both BL and CS it is not very likely that the data of BL and CS diverge when such information is used a period later. Conditions (63) and (65) need some more explanation because they look different from the other conditions. Condition (63) is likely to hold since the enrollment rate in secondary schooling at time t is probably equal to the secondary schooling attainment ratio of cohort 1, because, cohort 1 is not yet enrolled in tertiary schooling. Condition (65) is likely to hold because a large part of cohort 1 is too young to be enrolled in tertiary schooling and therefore this cohort is also too young to have finished tertiary schooling.

Condition (57) is not very likely to hold, it is highly unlikely that the death rate of young cohorts is equal to the death rate of older cohorts. CS are making a valid point here and they argue correctly that their method fully exploits the demographic structure. Assuming that younger cohorts have lower death rates and higher educational attainment this results in an underestimation of educational attainment by BL in forward extrapolations and overestimation in backward extrapolations. However, CS methodology automatically adopts one assumption that is less likely to hold; the oldest cohort in time t is discarded in time $t+1$, this brings us to conditions (58). BL let a constant percentage of every cohort die off and therefore, still use the information about the oldest cohort (cohort G) in period t, CS on the contrary assume that the oldest cohort in time $t+5$ has the same educational attainment as cohort G-1 in time t, therefore, discarding the information about cohort G at time t. Again, assuming that younger cohorts have higher educational attainment than the old ones results in an overestimation on CS's side for forward extrapolation and an underestimation for backward extrapolation[23].

Overall, CS's method seems more viable because of the full use of demographic structure. However, there are three problems with their data. First of all, the data are only available for ten year time intervals. Second, they do not reveal much of their method, therefore, BL are still more transparent even in their flaws. Third, CS might actually overestimate (underestimate) educational attainment in forward (backward) extrapolation due to their assumption about the oldest cohort. If the oldest cohort is not very large compared to younger cohorts (like most Asian countries, except Japan) this is not a big problem, however,

[23]Unless of course cohort G at the end of time t dies off entirely.

in Western economies the oldest cohort is pretty large compared to young ones, therefore, creating a problem. Still, throughout this paper we assume that the CS data set is more accurate than BL.

5.2 Derivative of labour service and labour composition functions

In this section we show that the derivative of the labour service function and the labour composition function with respect to the same variable differs. We start with the labour service function (equation (4)), the derivative with respect to the amount of hours worked by labour type i can be expressed as follows:[24]

$$\frac{\delta ln(L)}{\delta H_i} = \frac{\alpha_i}{H_i} + \sum_j \beta_{i,j} \frac{ln(H_j)}{H_i} \tag{67}$$

$$\frac{\delta ln(L)}{\delta H_i} = \frac{1}{H_i}\left[\alpha_i + \sum_j \beta_{i,j} ln(H_j)\right] \tag{68}$$

$$\frac{\delta ln(L)}{\delta H_i} = \frac{1}{H_i}\left[\alpha_i + \sum_{j \neq i} \beta_{i,j}[ln(H_j) - ln(H_i)]\right] \tag{69}$$

We see that the term in brackets is equal to the share in labour compensation by type i (see equation (8)). Therefore, this derivative is always positive.

The derivative of the labour composition function is a bit more complex. First we start with deriving the labour composition function from the labour service function. Because of homogeneity of degree one of the labour service function, we can extract total hours worked and write the labour service function as follows:

$$ln(L^c) = \sum_i \alpha_i ln(h_i) + \frac{1}{2}\sum_i \sum_j \beta_{i,j} ln(h_i) ln(h_j), \tag{70}$$

The derivative is:

$$\frac{\delta ln(L^c)}{\delta H_i} = \sum_j \frac{\alpha_j}{h_j}\frac{\delta h_j}{\delta H_i} + \frac{1}{2}\sum_k \sum_j \beta_{k,j}\left[\frac{ln(h_j)}{h_k}\frac{\delta h_k}{\delta H_i} + \frac{ln(h_k)}{h_j}\frac{\delta h_j}{\delta H_i}\right] \tag{71}$$

The derivative of share of hours worked is:

$$\frac{\delta h_{j \neq i}}{\delta H_i} = -\frac{H_j}{H^2} \tag{72}$$

$$\frac{\delta h_{j=i}}{\delta H_i} = \frac{\sum_{j \neq i} H_j}{H^2} = \frac{H - H_i}{H^2} \tag{73}$$

[24] For notational convenience we leave out the time subscript

in which H is total hours worked. Plugging equation (72) and (73) in equation (71) and using restrictions (5)-(7) on the coefficients, results in:

$$\frac{\delta L^c}{\delta H_i} = \frac{1}{H_i}\left[\alpha_i - h_i + \sum_{j\neq i}\beta_{i,j}[ln(H_j) - ln(H_i)]\right] \tag{74}$$

$$\frac{\delta L^c}{\delta H_i} = \frac{\delta ln(L)}{\delta H_i} - \frac{\delta ln(H)}{\delta H_i} \tag{75}$$

$$\frac{\delta L^c}{\delta H_i} = \frac{\delta ln(L)}{\delta H_i} - \frac{1}{H} \tag{76}$$

As we can see this is indeed the derivative of the labour service function minus the derivative of total hours worked. We therefore see that the labour composition derivative is by definition always smaller than the labour service derivative. The sign of equation (74) depends on a couple of elements. First of all α_i is by definition positive, probably larger the more educated the labour type. Second, we see that total hours worked has a negative effect on labour composition. However, at a decreasing rate, the effect is smaller when H is larger.[25] Last but not least we look at the last term in the labour composition derivative. We first have to know the sign of $\beta_{i,j}$. If we look at equation (8) we would expect $0 < \beta_{i,i} < 1$, that is, we would expect that if the share of a certain labour type rises the share of income rises, however, by less than its proportion because the particular labour type becomes less scarce and therefore we would expect the wage of that labour type to decline. Since all betas have to add up to zero, at least some $\beta_{i,j}$ have to be negative, it is reasonable to assume that all are negative with differing sizes depending on the substitutability of the different labour types. Therefore, we expect the betas in equation (74) to be negative. The sign of the term within brackets depends on the relative sizes of the labour types. If $H_j > H_i$ the sign is positive, if $H_j < H_i$ the sign is negative, in this latter case the sum term is positive.

If we write the condition for which the derivative is positive and rearrange slightly using equation (8) we get:

$$\frac{\delta L^c}{\delta H_i} > 0 \tag{77}$$

$$\frac{H_i}{H} < \left[\alpha_i + \sum_{j\neq i}\beta_{i,j}[ln(H_j) - ln(H_i)]\right] \tag{78}$$

$$v_i > h_i \tag{79}$$

[25] This makes sense, since every marginal increase of a certain labour type has less effect the more labour is already in place.

Which means that the derivative of labour composition is positive as long as the share in compensation is larger than the share in hours worked.

5.3 Overview labour composition growth rates

In this section we present the slowest and fastest growers for six periods. The 2010 period is forecast.

Table A1 Top and bottom 3 growth rates (%)

Decade	Country	Average yearly growth rate
1960	Afghanistan	-0.17
	GermanyWest	-0.08
	PapuaNewGuinea	-0.07
	...	...
	Finland	0.97
	Jordan	1.26
	SouthKorea	1.57
1970	Poland	-0.15
	Uganda	0
	Singapore	0
	...	...
	Japan	0.88
	Jordan	0.88
	Portugal	1.13
1980	Argentina	-0.1
	Afghanistan	-0.02
	SouthAfrica	0
	...	...
	Spain	0.81
	Finland	0.98
	SouthKorea	1.05

Continued		
Decade	Country	Average yearly growth rate
1990	Tajikistan	-0.43
	USSR	-0.17
	Pakistan	-0.11
	...	...
	SouthKorea	1.05
	Greece	1.16
	Singapore	1.26
2000	Jordan	-0.09
	Zimbabwe	-0.07
	Ghana	0
	...	...
	Singapore	0.84
	Greece	1.01
	Portugal	1.16
2010	Italy	0.09
	Denmark	0.11
	Switzerland	0.11
	...	...
	DominicanRepublic	0.52
	Thailand	0.53
	Portugal	0.87

We see that average yearly growth rates hardly ever surpass the 1.5%. As we have seen in the previous chapter of this appendix this makes sense since we would expect labour composition growth to be relatively modest.

However, there are some interesting results in this section. For instance, we see a lot of countries in the top three that we would expect in top three, either through catching up or simply because of the high standards of their education system. For instance, we see Jordan in the top three (catching up) and we see South-Korea and Finland repeatedly in the top three (high standards in education). However, there are also some surprises, for instance, Western-Germany in the bottom three in the 60's with declining labour composition. Furthermore, we can doubt the (repeated) top three position of Portugal (70's, 10's) Spain (80's) and Greece (90's, 00's). In the ranking of these countries we can quite possibly see the effect of rapidly rising wages for certain labour types that are not necessarily related to rising productivity. This is partly confirmed by the data which shows more than proportional rises in share of total earnings by high skilled labour and to a lesser extend medium skilled labour. For these three European countries we see not only the effect of catching up but also the effect of joining the EU and subsequently adopting the Euro.

Therefore, as was stressed in the conclusion, these data should be treated with care and should always be cross checked with other sources of data, either qualitative accounts or quantitative data if available.

References

Barro, R. J. and Lee, J.-W. (2000). International data on educational attainment updates and implications. NBER Working Papers 7911, National Bureau of Economic Research, Inc.

Bonthuis, B. (2009). Measuring labour quality. Technical report, The Conference Board and University of Groningen. www.boelebonthuis.com.

Cohen, D. and Soto, M. (2001). Growth and human capital: Good data, good results. CEPR Discussion Papers 3025, C.E.P.R. Discussion Papers.

Jorgenson, D. W., Gollop, F. M., and Fraumeni, B. M. (1987). *Productivity and U.S. economic growth*. Harvard University Press, Cambridge, Mass.

KC, S., Barakat, B., Goujon, A., Skirbekk, V., Sanderson, W. C., and Lutz, W. (2010). Projection of populations by level of educational attainment, age, and sex for 120 countries for 2005-2050. *Demographic Research*, 22(15):383–472.

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, 22(1):3 – 42.

Mankiw, N. G., Romer, D., and Weil, D. N. (1992). A contribution to the empirics of economic growth. *The Quarterly Journal of Economics*, 107(2):407–37.

Timmer, M., van Moergastel, T., Stuivenwold, E., OMahony, M., and Kangasniemi, M. (2007). Eu klems growth and productivity accounts version 1.0, part i methodology. Technical report, EUKLEMS.

Uzawa, H. (1965). Optimum technical change in an aggregative model of economic growth. *International Economic Review*, 6(1).