WIOD Socio-Economic Accounts (SEA): Sources and Methods

Abdul Azeez Erumban

Reitze Gouma

Gaaitzen de Vries

Klaas de Vries

Marcel Timmer

April 2012

This project is funded by the European Commission, Research Directorate General as part of the 7th Framework Programme, Theme 8: Socio-Economic Sciences and Humanities.

Grant Agreement no: 225 281

1. Introduction

This document describes the sources and methods used for estimation of data on capital investment and stocks, and employment by skill type for the forty countries included in the WIOD database, called the Socio-Economic Accounts (SEAs). The SEAs contain annual data (1995-2009) for 35 industries on

- Industry output, value added, at current and constant price
- Capital stock and investment
- Wages and employment by skill type (low-, medium- and high-skilled).

The full set of variables included are given in Table 1

Table 1 Variables in the WIOD Socio-economic Accounts (SEA)

Values	***Description***
GO	Gross output by industry at current basic prices (in millions of national currency)
II	Intermediate inputs at current purchasers' prices (in millions of national currency)
VA	Gross value added at current basic prices (in millions of national currency)
COMP	Compensation of employees (in millions of national currency)
LAB	Labour compensation (in millions of national currency)
CAP	Capital compensation (in millions of national currency)
GFCF	Nominal gross fixed capital formation (in millions of national currency)
EMP	Number of persons engaged (thousands)
EMPE	Number of employees (thousands)
H_EMP	Total hours worked by persons engaged (millions)
H_EMPE	Total hours worked by employees (millions)
Prices	
GO_P	Price levels gross output, 1995=100
II_P	Price levels of intermediate inputs, 1995=100
VA_P	Price levels of gross value added, 1995=100
GFCF_P	Price levels of gross fixed capital formation, 1995=100
Volumes	
GO_QI	Gross output, volume indices, 1995 = 100
II_QI	Intermediate inputs, volume indices, 1995 = 100
VA_QI	Gross value added, volume indices, 1995 = 100
K_GFCF	Real fixed capital stock, 1995 prices
Additional variables	
LABHS	High-skilled labour compensation (share in total labour compensation)
LABMS	Medium-skilled labour compensation (share in total labour compensation)
LABLS	Low-skilled labour compensation (share in total labour compensation)
H_HS	Hours worked by high-skilled persons engaged (share in total hours)
H_MS	Hours worked by medium-skilled persons engaged (share in total hours)
H_LS	Hours worked by low-skilled persons engaged (share in total hours)

The sources used for Industry output, value added, at current and constant price are described in more detail in "*Sources for National Supply and Use Table Input files*" , Abdul Azeez Erumban, Reitze Gouma, Gaaitzen de Vries, Klaas de Vries and Marcel Timmer, April 2012. The current document describes the sources and methods for estimating labour and capital input.

For factor input requirements we collected country-specific data on detailed labour and capital inputs for all 35 industries. This includes data on hours worked and compensation for three labour types (low-, medium- and high-skilled labour) and data on capital stocks and compensation. These series are not part of the core set of national accounts statistics reported by NSIs. The database builds upon the data collected in the EU KLEMS project (see www.euklems.net described in O'Mahony and Timmer 2009) by updating it and extending it to a larger set of countries. Within EU KLEMS this type of data is available for about 15 OECD countries up to the year 2007. We extend this data to include also a large set of less developed countries and update to 2009. This extensive coverage of the SEAs in WIOD makes it a unique database compared to what is currently available.

Skills in the WIOD SEAs are defined on the basis of educational attainment levels. Data on number of workers by educational attainment are available for a large set of countries (e.g. Barro and Lee, 2010), but WIOD provides an extension in two directions. First, the WIOD SEAs provide industry level data, reflecting the large heterogeneity in the skill levels used in various industries (compare e.g. agriculture and financial and business services). This has been documented in e.g. Jorgenson and Timmer (2011) for the OECD countries, and this heterogeneity is even stronger in less developed countries. Moreover, the WIOD SEAs also provide relative wages by skill type that reflect the differences in remuneration of workers with different levels of education. The wage data is made consistent with the quantity data and can be used in conjunction to analyse distributional issues such as relative income shares.

The capital data in the WIOD SEAs include investment and capital stocks at current and constant prices. While this type of data is available for the total economy (see e.g. Total economy Database The Conference Board) there is no large-scale database that provides industry level detail. Heterogeneity of capital and investment flows across industries is even bigger than for labour, and taking account of this is crucial in any analysis of the role of capital in structural change and economic growth.

In this paper, we start out with the general approach used for labour (section 2) and capital input (section 3), followed by a discussion of sources and specific methods used on a country-by-country basis (section 4).

2. Wages and employment by skill types

Data on wages and employment by skill types are not part of the core set of national accounts statistics reported by NSIs; at best only total hours worked and wages by industry are available from the National Accounts. Additional material has been collected from employment and labour force statistics. For each country covered, a choice was made of the best statistical source for consistent wage and employment data at the industry level. In most countries this was the labour force survey (LFS). In most cases this needed to be combined with an earnings surveys as information wages are often not included in the LFS. In other instances, an establishment survey, or social-security database was used. Care has been taken to arrive at series which are time consistent, as most employment surveys are not designed to track developments over time, and breaks in methodology or coverage frequently occur. For most OECD countries labour data was taken from the EU KLEMS database (www.euklems.org, described in O'Mahony and Timmer 2009), revised and updated. For countries not in EU KLEMS new sources have been used.

Labour compensation of self-employed is not registered in the National Accounts, which as emphasised by Krueger (1999) leads to an understatement of labour's share. This is particularly important for less advanced economies that typically feature a large share of self-employed workers in industries like agriculture, trade, business and personal services. We make an imputation. For advanced countries, we assume that the compensation per hour of self-employed is equal to the compensation per hour of employees. For emerging countries this assumption is not plausible as a large part of informal workers earns much less than the average wage of low-skilled workers. Instead, we used additional information which differs by country. This is described in the country notes below.

In WIOD three skill types of labour are being distinguished. Skill type is defined on the basis of the level of educational attainment of the worker. Educational systems and attainment levels are not always comparable across countries in a straightforward manner. We use the 1997 International Standard Classification of Education (ISCED) classification to define low, medium and high skilled labour. The definition of skills is given in Table 1. For more information on ISCED, see http://www.uis.unesco.org/Education/Pages/international-standard-classification-of-education.aspx

Table 2 Definition of skills in WIOD SEA

WIOD skill-type	1997 ISCED level	1997 ISCED level description
Low	1	Primary education or first stage of basic education
Low	2	Lower secondary or second stage of basic education
Medium	3	(Upper) secondary education
Medium	4	Post-secondary non-tertiary education
High	5	First stage of tertiary education
High	6	Second stage of tertiary education

Data has been collected for both the number of workers, and their wages. If available the data refers to all workers including self-employed and family workers, but mostly they refer to employees only. This is indicated for each country in the country source notes below. Numbers refer to numbers of workers and do not adjust for differences in hours worked. The latter is preferable, but based on the available evidence there is no clear relationship between hours worked and skill-level.

In principle, data is constructed for the period from 1995 to 2009. Annual availability of data differs across countries. Also, data on wages is scarcer than for number of workers and more imputations have to be made. For each country we have at least one observation of wages in this period, to ensure that country-specific skill-premia are reflected in the data, and match with the definition for quantities. For most countries we have at least three observations across the period to reflect the changes in skill premia that take place over time.

The level of industry detail also varies across countries and depends crucially on the sample sizes of the surveys on which our estimates are based. Quantity data by skill type is available for at least 14 industries in all countries, up to 35 in some (e.g. India and Mexico). In order to derive shares for all 35 WIOD industries we assume that the skill distribution of workers in sub-industries is similar as the shares for more aggregate industries.

Relative wage data is scarcer. Sometimes it is available at a more aggregate level than data for quantities. e.g. for China only for 3 broad sectors. We assume that relative wages of both high-skilled and medium-skilled workers relative to low-skilled in sub-industries is the same as in the more aggregate industry. As relative wages will differ much less across industries than quantities this assumption can more easily be made at higher levels of aggregation. For example, while in say China the relative number of high to low-skilled workers is much higher in chemicals than in textile manufacturing, the relative wages of high-to-low skilled workers in both industries will be closer together. Also, relative wage data is not always available on an annual basis. In those cases we interpolate relative wages for years in-between. As relative wages develop only slowly, this assumption is relatively harmless.

For the European countries labour shares and labour compensation shares are estimated based on shares from the EUROSTAT Labour Force Surveys (LFS). The LFS provides data on the fraction of High, Medium and Low skilled labour in total workers (and sometimes hours worked) and labour compensation for 14 separate sectors of the economy, which sum to total economy. The shares of these 14 sectors are used as data for the more detailed underlying industries. LFS provides data up to 2009 for all countries, however, the starting year differs from country to country. Relative wage data is from the EU Structural Earnings Survey (SES) and EU Survey on Income and Living Standards (SILC). In order to complement the data such that it includes the skill shares from 1995 up to 2009, the LFS data needs to be extrapolated backwards to 1995. This is done by using the skill-shares from the March 2008 release of the EU KLEMS database insofar available. For non-EU countries various country-specific data sources are used. Details can be found in the section on country sources and methods.

3 Investment and capital stocks

The WIOD SEAs contain investment and capital stock series by industry at both current and constant prices. The series cover all fixed assets as defined in the SNA 1993. Data on capital stocks is only available up to 2007 unless otherwise indicated. The EU KLEMS database provides investment, stocks and capital services data cross-classified by both industry and eight asset types. For those countries for which EU KLEMS data was available we took the investment and stock estimates directly. These are given in the so-called capital input files on the EU KLEMS website and the files can be used when more detailed information on stocks or services is needed.[1]

This type of data is available for a limited set of OECD countries in the EU KLEMS database, but not for the majority of the 40 WIOD countries. For the other countries, capital stocks have been constructed on the basis of the Perpetual Inventory Method (PIM) in which the capital stock (K) in year t is estimated as the sum of the depreciated capital stock in year t-1 plus real investment (I) in year t:

$$K_t = (1-d)K_{t-1} + I_t$$

with d the depreciation rate. The depreciation rates are taken to be geometric and industry-specific and given in Appendix Table 1. They take into account the differences in the composition of capital assets in various industries and vary from less than 4% in e.g. Education and Public Administration to more than 10% in financial and business services. This takes into account the larger share of long-lived assets as buildings and structures in the former, and the larger share of short-lived assets like ICT-equipment and software in the latter. The industry-specific rates are based on weighted asset-specific depreciation rates used in EU KLEMS. The 8 asset types are weighted by the asset distribution within industries for Spain and averaged over the period 1995-2006. The industry-specific rates are assumed to be the same for all countries due to lack of sufficient country-specific information.

For many countries long time-series of investments are available and there is no need to have information on an initial stock estimate. In cases where investment series were too short, and did not start at least 20 years prior to 1995, an initial capital stock for 1995 had to be estimated. Two alternatives were used: the ICVAR and Harberger method.

ICVAR method. For countries were no investment data before 1995 was available, the ICVAR method was used (see e.g. Timmer 1999 for an application). In the ICVAR method, industry specific ratios of value added to capital stocks were used of a country at a similar stage of development (often Spain). These industry-specific ratios (averaged over 5 years to smooth out business cycle fluctuations) were applied to the 1995 value added to derive the 1995 capital stock. For years after 1995 the PIM method was used based on this 1995 estimate.

Harberger method. For countries for which investment series were available for a number of years before 1995, but less than 20, an initial capital stock for the year in which investment series start is estimated

[1] For the European countries for which EU KLEMS capital stocks data is used the Textiles, Textile, Leather and Footwear sector (NACE Rev.1.1 sector 17t19) has been split into Textiles and Textile (industry 17t18) and Leather and Footwear (industry 19) separately using Value Added shares. The same was done to split the EU KLEMS aggregate sector Transport and Storage (industry 60t63) into the transport of water, land and air and other transport activities separately.

using the Harberger method (Harberger, 1978; Easterly and Levine, 2001). The Harberger method can be written as:

$$K_0 = (i/(g+d)) * GO,$$

Where K_0 is the initial capital stock in constant 1995 prices, GO is gross output by industry in constant 1995 prices, i is the investment rate (GFCF in 1995 prices divided by GO in 1995 prices by industry), g is the average growth rate of output, and d is the total depreciation rate by industry. For most countries, gross output series extend back to 1970 and we used the average growth rate for the period 1970-1994 as an estimate for *g*. For other countries were only gross output series from 1995 onwards.

Capital stocks are estimated at the level of 35 industries unless otherwise indicated.

Table 3 Information on Capital stock estimates

Country	Begin year of invest-ment series	Method to generate initial capital stock	Additional remarks
1. Australia			EU KLEMS
2. Austria			EU KLEMS
3. Belgium	1970	Harberger	
4. Bulgaria		ICVAR	Total Economy investment data from UN. This is split to 35 industries using VA shares. The total economy investment deflator is applied.
5. Brazil	1995	ICVAR	
6. Canada			EU KLEMS
7. China	1995	ICVAR	
8. Cyprus	1995	ICVAR	EUROSTAT investment for 29 industries is used. Split to 35 industries is based on VA shares.
9. Czech Republic			EU KLEMS
10. Germany			EU KLEMS
11. Denmark			EU KLEMS
12. Spain			EU KLEMS
13. Estonia	1995	ICVAR	STAN investment series.
14. Finland			EU KLEMS
15. France			EU KLEMS
16. United Kingdom			EU KLEMS
17. Greece	1980	PIM.	Initial stock used from data delivered by Statistics Greece in November 2006 for EU KLEMS. STAN

			investment series and Greek implicit depreciation rates are used.
18. Hungary			EU KLEMS
19. Indonesia	1960		Gross capital stock data from Bank of Indonesia
20. India	1950	Net capital stock	Based on 3 asset types
21. Ireland			EU KLEMS
22. Italy			EU KLEMS
23. Japan			EU KLEMS
24. Korea			EU KLEMS
25. Lithuania	1995	ICVAR	EUROSTAT investment for 29 industries is used. Split to 35 industries is based on VA shares.
26. Luxembourg	1995	ICVAR	EUROSTAT investment for 29 industries is used. Split to 35 industries is based on VA shares.
27. Latvia	1995	ICVAR	EUROSTAT investment for 29 industries is used. Split to 35 industries is based on VA shares. UN total economy investment deflator is used.
28. Mexico	1988	ICVAR	
29. Malta	1995	ICVAR	EUROSTAT investment for 29 industries is used. Split to 35 industries is based on VA shares. UN total economy investment deflator is used.
30. Netherlands			EU KLEMS
31. Poland	1995	ICVAR	STAN investment series.
32. Portugal		PIM	EU KLEMS up to 2005 updated with STAN investment and implicit PRT industry depreciation rates.
33. Romania		ICVAR	Total Economy investment data from UN. This is split to 35 industries using VA shares. The total economy investment deflator is applied
34. Russia	1995	Net capital stocks	
35. Slovak Republic	1995	ICVAR	STAN investment series.
36. Slovenia		PIM	EU KLEMS to 2006, updated using STAN investment and implicit SVN industry depreciation rates
37. Sweden			EU KLEMS
38. Turkey	1970	Harberger	
39. Taiwan	1981	Gross Capital stock	For splitting some sectors distribution of an initial stock based on Harberger approach is used
40. USA			EU KLEMS until 2007 and updated using aggregate trend

4 .Country-specific sources

In this section the detailed sources and methods are described on a country-by-country basis.

1. Australia

Capital stock data and employment data by skill type are taken from the EU KLEMS 2009 release. To split industry 17t19 and 60t63 into 17t18 and 19, and 60, 61, 62, and 63 we use the more detailed EU KLEMS 2008 release which had to be updated in the following way.

To update the employment series to 2008 and 2009, we assume a constant employment to value added ratio taken from 2007 and apply this to value added in 2008 and 2009. The latest year for data on employment by skill type is 2005. For 2006-2009, we assume a constant ratio of skill-type in overall employment from 2005.

Capital stock data runs until 2007. The update of the GFCF series to 2009 are based on the aggregate investment series from the National Accounts. The capital stock is subsequently updated combining constant GFCF series and the average depreciation ratios.

2. Austria

Employment data by industry are taken from OECD's STructural ANalysis (STAN) database.

The skill shares in hours worked and in labour compensation are calculated based on the LFS data for 2002-2009. Prior to 2002 the growth trend of the skill shares in EU KLEMS has been used to extrapolate the series backwards to 1995.

Data on investment, investment prices and capital stocks are taken directly from the November 2009 release of the EU KLEMS database in which data are available up to 2007.

3. Belgium

Employment data are taken from the STAN database. Data on hours worked are only available for employees, with data missing for the breakdown of industries G, I and K. Average hours per employee are calculated for the aggregate industries and used to calculate the split of these industries. The average hours worked by total persons engaged are assumed to be equal to that of employees.

For the skill shares the LFS data is used for the period 2002-2009. Prior to 2002 the series have been extrapolated backwards using the EU KLEMS skill shares from the March 2008 release.

Capital stock estimates by industry are based on the investment data from the STAN database. The initial capital stock is estimated for 1970 using the Harberger method. The PIM method is used to build up the capital stock up to 2009 as described above.

4. Bulgaria

Employment data by industry are taken from EUROSTAT. For industry aggregates G, I and K there is no industry detail available and is based on value added shares from EUROSTAT Structural Business Statistics (SBS). For 2007, 2008 and 2009 the shares are set equal to 2007.

No detailed information on educational attainment in the labour force is available for Bulgaria; therefore the Portuguese industry specific skill-shares have been used.

Capital stock estimates are based on the total economy investment data from the United Nations' National Accounts. The investment data in current prices has been disaggregated using VA shares. The total economy investment deflator is applied to all industries. The initial capital stock is estimated using the ICVAR approach described above. The capital stock to value added ratios and the industry depreciation rates from Spain are used.

5. Brazil

The national accounts provide annual series of employment by industry. The employment series refer to occupations. They are an integral part of the supply and use table framework used by the statistical office, and the series include informal and own-account workers (IBGE, 2008). The integration of value added and employment ensures internal consistency in the time series for Brazil. Similar to value added series, we split up distributive trade industries using employment shares from the *Pesquisa anual de comércio*, and separate transportation services from business services and personal and community services using shares from the *Pesquisa anual de serviços*.

Hours worked as well as wage and employment shares by skill type are derived by industry from the annual Brazilian household surveys (*Pesquisa Nacional por Amostra de Domicílios, PNAD*). PNAD is available from 1981 onwards, but the methodology changed substantially in 2002. PNAD collects data each year from a representative national sample of households, with a sample size ranging from 291,000 to 525,000 individuals.[2] The survey reports each year on a range of variables that form the basic data set. Questions are asked on subjects pertaining to the household and to individuals within the household. Using the PNAD micro data, hours worked and wage and employment shares by skill type are calculated at the broad (15-) sector level.[3] The relative broad sector level is to assure a representative sample. Individual sampling weights from IBGE to calculate total employment are applied to these variables. For more disaggregated sub-industries, hours worked, wage and employment by skill shares from the parent industry are used.

We use the ISCED mapping from the World Education Indicators Programme (UNESCO, 2007) to match educational attainment with ISCED. Low-skilled includes: Elementar (primário); Regular do ensino fundamental ou do primeiro grau; Educação de jovens e adultos ou supletivo do ensino fundamental ou do primeiro grau; alfabetização de jovens e adultos; crèche; classe de alfabetização; maternal, jardim de infância. Medium-skilled includes: Médio primeiro ciclo; Médio Segundo ciclo; Regular do ensino médio ou do segundo grau; educação de jovens e adultos ou supletivo de ensino médio ou do Segundo grau.

[2] During 1981-2004, PNAD covered Brazil, except for the rural areas of Ronônia, Acre, Amazonas, Roraima, Pará, and Amapá.
[3] The sectors distinguished are: Agriculture (AtB); Mining (C); 15t19; 20t29; 30t37; Public Utilities (E); Construction (F); 50t55; 60t64; Finance (65t67); 70t74; Government administration (L); Education (M); Health (N); and other services (O).

High-skilled includes: superior – graduação; Mestrado ou doutorado. Using the educational attainment and wage rates from PNAD, we find large differences in the wage rate of the high-skilled compared to the low-skilled. On average, the ratio of high- to low-skilled is about 4. The ratio is similar to those found in the Socio-Economic Database for Latin America and the Caribbean published by CEDLAS and The World Bank (SEDLAC, 2011).

Labour shares in value added are estimated by adding the wage income of self-employed workers to labour compensation. We combine numbers of self-employed workers from an earlier vintage of the input-output tables (annual data from 1990 to 2003) with the average wage rate of self-employed workers in SEDLAC (2011). The wage income of own-account workers is added to the compensation data by sector, except for government administration, education, and health services were we do not make iimputations. For agriculture (were 99 per cent of persons engaged are employees) and leather and footwear manufacturing, the imputation resulted in a labour share larger than value added. In these instances we used the compensation data.

Gross Fixed Capital Formation (GFCF) for tangible assets (construction, transport equipment, computing equipment, communication equipment, and other machinery), both in current and previous year prices, is obtained from IBGE/National accounts and the SUTs for 1995-2008. GFCF in residential construction and other construction is available from the historical national accounts for 1901-2000. GFCF by industry is available in an investment matrix for the year 2005. The investment matrix (*Matriz de Absorção de Investimento*) is developed as part of a large investment project, see Dweck et al. (2009) for documentation.

GFCF by asset types is combined with GFCF by industry. We take the GFCF by asset from the national accounts as the row total border, the GFCF shares by industry as the column total border, and the distribution of asset type by industry for 2005 as initial starting values before applying RAS. This way, we obtain longitudinal investment series by industry and asset type. An initial capital stock by industry in 1995 prices is estimated using the average (1995-2007) capital to value added ratio for Spain from the EU KLEMS database (Timmer et al., 2010), and this capital stock is subsequently updated combining constant GFCF series and average depreciation rates for Spain in the perpetual inventory method.

6. Canada

For Canada we take the employment numbers and skill distribution of employment and compensation from EU KLEMS. The data is taken from the latest update of the EU KLEMS data, which provides data until 2010. See EU KLEMS source descriptions for more details. This data, however, did not had sub-sectors 17t18 & 19 and 60, 61, 62 & 63. In order to split employment and labor compensation in these sectors, we use compensation shares obtained from IO tables. For splitting capital stock, we assume the same capital/output ratio as in the parent industry.

7. China

Employment series by main sectors for China are from various issues of the *China Statistical Yearbook* (CSY) (NBS, various issues) and detailed industrial employment series are from various issues of the

China Industrial Economic Statistics Yearbook (*CIESY*) by the Department of Industrial and Transportation Statistics, which is part of the National Bureau of Statistics (NBS), and the *China Labour Statistical Yearbook* (*CLSY*). The three broad-sector series based on population censuses are considered the most consistent estimates and used by other scholars such as Bosworth and Collins (2008). At detailed industry levels in manufacturing, the *CIESY* data provide employment data for enterprises at or above the "designated size" and the *CLSY* data provide less detailed data for total employment (including all below the "designated size"). The industry data construction follows Wu and Yue (2010) in principle but using the above broad-sector estimates in Wu (2011) as control totals. Wu and Yue (2010) use all industry level census data, namely, China's 1985 and 1995 Industrial Censuses and the 2004 and 2008 Economic Censuses, to make industry level estimates consistent and then allocate the additional employment (below the "size" and outside the "system" or self-employed) into labour-intensive industries only. The industry level estimates are reconciled with the national totals.

For 2004, employment is split by educational attainment using the economic census 2004 (summary table 1-13). Educational attainment in the economic census 2004 distinguishes: Junior school and below; High school; Vocational college; Undergraduates; and Graduates and above. We consider junior school and below as "low-skilled". High school and vocational college are classified as "medium-skilled", whereas undergraduates and above are considered "high-skilled"(Wu and Yue, 2010; UNESCO, 2007). For agriculture, the educational attainment is from the labour force survey. Educational attainment by industry is extrapolated using trends from the labour force survey for the period from 2002 to 2008. 2009 shares are assumed similar to 2008. For years before 2002, we used the growth rate in primary (for low-skilled), secondary (for medium-skilled), and tertiary (for high-skilled) educational attainment from Barro and Lee (2010). Subsequently, these shares are normalized to sum to 1.

The main data source for relative wages by educational attainment and broad sectors of the economy are imputed using micro data from the *China Household Income Project* (CHIP) surveys for 1995, 2002, and 2007. Individuals observed in the CHIP surveys were drawn from larger samples of the National Bureau of Statistics (NBS) using a multistage stratified probability sampling method. The CHIP study is considered the best available data source on household income and expenditures and the only available source for wage data by educational attainment. The CHIP survey is split into an urban and a rural survey. The definition of wage income varies considerable between urban and rural questionnaires preventing the combination of both surveys. We use the urban surveys for imputing relative wages. However, while income levels differ substantially between the wage income definition in the urban and rural surveys, ratios of high to low-skilled (and medium to low-skilled) are comparable across the surveys.

The broad sectors distinguished are agriculture, other industries, manufacturing, transport, storage and communication, distributive trade, other market services, and government services. To correct for outliers, we drop the 1st and 99th percentile of wage income. We distinguish between three educational groups (matching with the China economic census 2004):

- LOW: Never schooled; Classes for eliminating illiteracy; Elementary school; and Junior middle school
- MEDIUM: Senior middle school (including professional middle school) and Technical secondary school
- HIGH: Junior college; College/University; Graduate

Relative wage rates are obtained for 1995, 2002, and 2007. Ratios between these survey years are interpolated. The wage ratio of high-skilled to low-skilled workers increases over time in the survey, rising from 1.41 in 1995 to 1.71 in 2007 for the total economy.

Labour shares in value added are from labour compensation provided in the input-output tables. Before the first Economic Census in 2004, the income of self-employed and their employees are included in labour compensation (NBS, 2003). While profits related to owners (informal entrepreneurs) should be part of gross operating surplus, we consider the labour compensation in the input-output tables before 2004 closest to the definition of labour compensation in value added. After the economic census, two changes in the income GDP accounting method introduce a break in the labour share time series by industry (Bai and Qian, 2010). First, profits of state-owned and collective-owned farms are included in labour compensation, introducing an upward break in the agricultural labour shares. Second, income of self-employed owners is subsequently included in gross operating surplus. We use the adjustment factors for both changes at the sector level in Bai and Qian (2010) to arrive at consistent time series following the definition of labour shares before the 2004 Economic Census.

An initial estimate of current gross fixed capital formation series is based on the aggregate investment series and the distribution of gross fixed investment by sectors from the CSY. Gross fixed investment by sector shows no detail for manufacturing and runs from 2003-2009. For earlier years, we assumed a constant ration of industry investment to total investment. For manufacturing detail, we used the distribution in Gross Operating Surplus from the input-output tables. An initial capital stock by industry in 1995 prices is estimated using the average (1995-2007) capital to value added ratio for Spain from the EU KLEMS database (Timmer et al., 2010), and this capital stock is subsequently updated combining constant GFCF series and average depreciation rates for Spain in the perpetual inventory method. In collaboration with the China Industry Productivity team, we are currently improving the capital estimates. A new version of the capital series will become available by mid-2012.

8. Cyprus

Employment data by industry are taken from EUROSTAT. For the period prior to 2000 only aggregate employment data is available for industry aggregates D, G, I and K. Data at the detailed industry level therefore is estimated by using Value Added growth rates and normalizing on the aggregate totals.

Data from the LFS are taken for the skill shares for 2002-2009. Portuguese skill-shares from EU KLEMS have been used to extrapolate the series backwards to 1995. Relative wage data is from the EU Structural Earnings Survey (SES) and EU Survey on Income and Living Standards (SILC) and are available for 2002-2008. For 2009 relative wages are assumed to be the same as the 2008 level.

Capital stock estimates are based on the investment data from EUROSTAT. The initial capital stock is estimated using the ICVAR approach, employing Spanish capital stock to value added ratios. The PIM method is used to build up the capital stock up to 2009. The Spanish industry depreciation rates are used.

9. Czech Republic

Employment data by industry are taken from the STAN database. Data on hours worked for the year 2009 are estimated by keeping average hours worked by an employee or person engaged equal to its 2008 level.

The skill shares in hours worked and labour compensation are calculated based on the LFS data for 2002-2009. Prior to 2002 the growth trend of the skill shares in EU KLEMS has been used to extrapolate the series backwards to 1995. Relative wage data is from the EU Structural Earnings Survey (SES) and EU Survey on Income and Living Standards (SILC) and are available for 2002-2008. For 2009 relative wages are assumed to be the same as the 2008 level.

Data on investment, investment prices and capital stocks are taken directly from the November 2009 release of the EU KLEMS database in which data are available up to 2007.

10. Germany

Employment data by industry are taken from the STAN database. The average hours worked by an employee or person engaged of the aggregate industry are used to compute missing data on hours worked at the detailed industry level for industries G, I and K. Detailed information on hours worked in the manufacturing sector is only available from 2002 onwards. In order to estimate hours worked at the detailed industry level prior to 2002, average hours worked have been estimated by using the growth of average hours worked for the total manufacturing sector and applying it to the level of average hours worked in 2002 for the detailed industries and extrapolating backwards. Data for 2009 are available for industry aggregates only, so the growth rates of these aggregate industries are used to estimate data for detailed industries for 2009.

The skill shares in hours worked and labour compensation are calculated based on the LFS data for 2002-2009. Prior to 2002 the growth trend of the skill shares in EU KLEMS is used to extrapolate the series backwards to 1995. Relative wage data is from the EU Structural Earnings Survey (SES) and EU Survey on Income and Living Standards (SILC) and are available for 2002-2009.

Data on investment, investment prices and capital stocks are taken directly from the November 2009 release of the EU KLEMS database in which data are available up to 2007.

11. Denmark

Employment data by industry are taken from the STAN database. In some industries (17t19, 23t25, 60t63, 70t74) industry detail is missing for the underlying industries for 2008 and 2009. In these cases the growth of the aggregate industry is used to extrapolate the series. For data on hours worked growth trends of average hours worked are used multiplied by employment in persons.

The skill shares in hours worked and labour compensation are calculated based on the LFS data for 2002-2009. Prior to 2002 the growth trend of the skill shares in EU KLEMS is used to extrapolate the series backwards to 1995. Relative wage data is from the EU Structural Earnings Survey (SES) and EU Survey

on Income and Living Standards (SILC) and are available for 2004-2008. For 2009 relative wages are assumed to be the same as the 2008 level. For 2002-2003 the relative wages are assumed to be equal to the 2004 level.

Data on investment, investment prices and capital stocks are taken directly from the November 2009 release of the EU KLEMS database in which data are available up to 2007.

12. Spain

Employment data by industry are taken from the STAN database. For data on hours worked prior to 2000 data is missing for detailed industries. These values are estimated by applying the growth rate of average hours of the aggregate industries to detailed industries. The resulting average is multiplied by employment in persons which is available. The values of hours worked for detailed industries are normalized to sum to the available industry aggregates.

The skill shares in hours worked and labour compensation are calculated based on the LFS data for 2002-2009. Prior to 2002 the growth trend of the skill shares in EU KLEMS is used to extrapolate the series backwards to 1995. Relative wage data is from the EU Structural Earnings Survey (SES) and EU Survey on Income and Living Standards (SILC) and are available for 2002-2008. For 2009 relative wages are assumed to be the same as the 2008 level.

Data on investment, investment prices and capital stocks are taken directly from the November 2009 release of the EU KLEMS database in which data are available up to 2007.

13. Estonia

Employment data by industry are taken from the STAN database. Prior to 2000 there is no data available on hours worked. Estimations are calculated by keeping average hours worked by an employee or person engaged equal to the 2000 level. The disaggregation of industry 60t63 is based on the shares of the compensation of employees (COMP).

Data from the LFS is taken for the skill shares for 2002-2009. However, no EU KLEMS data on educational attainment in the labour force is available for Estonia. Therefore the Portuguese skill-shares from EU KLEMS have been used to extrapolate the series backwards to 1995. Relative wage data is from the EU Structural Earnings Survey (SES) and EU Survey on Income and Living Standards (SILC) and are available for 2002-2008. For 2009 relative wages are assumed to be the same as the 2008 level.

Capital stock estimates are based on the investment data from the STAN database. The initial capital stock is estimated using the ICVAR approach, employing Spanish capital stock to value added ratios. The PIM method is used to build up the capital stock up to 2009. The Spanish industry depreciation rates are used.

14. Finland

Employment data by industry are taken from the STAN database.

The skill shares in hours worked and labour compensation are calculated based on the LFS data for 2002-2009. Prior to 2002 the growth trend of the skill shares in EU KLEMS has been used to extrapolate the series backwards to 1995. Relative wage data is from the EU Structural Earnings Survey (SES) and EU Survey on Income and Living Standards (SILC) and are available for 2002-2008. For 2009 relative wages are assumed to be the same as the 2008 level.

Data on investment, investment prices and capital stocks are taken directly from the November 2009 release of the EU KLEMS database in which data are available up to 2007.

15. France

Employment data by industry are taken from EUROSTAT rather than STAN due to more complete data availability in EUROSTAT. For data on hours worked the average hours worked from EU KLEMS are taken and multiplied by the EUROSTAT employment data. Average hours worked data are updated to 2009 using the growth trend of average hours worked from the aggregates at the 6 branch level from EUROSTAT. Growth trends of lower level industries are updated with the growth trend of the parent industry.

The skill shares in hours worked and labour compensation are calculated based on the LFS data for 2002-2009. Prior to 2002 the growth trend of the skill shares in EU KLEMS has been used to extrapolate the series backwards to 1995. Relative wage data is from the EU Structural Earnings Survey (SES) and EU Survey on Income and Living Standards (SILC) and are available for 2002-2008. For 2009 relative wages are assumed to be the same as the 2008 level.

Data on investment, investment prices and capital stocks are taken directly from the November 2009 release of the EU KLEMS database in which data are available up to 2007.

16. United Kingdom

Employment data by industry are taken from EUROSTAT. Data for employment is only available at the 6 branch level. Compensation of employees' data is used to estimate the decomposition of the aggregate industries for the employment variables. No breakdown of the aggregate industries 50t52, 60t64 and 70t74 is available in EUROSTAT for any of the NA variables. Therefore EU KLEMS Value Added data from the March 2011 update are used as shares to estimate values at the detailed industry level for all variables. For data on hours worked only total economy aggregates for total persons engaged are available. Employment shares (from persons engaged and employees) are used to distribute total hours worked over the detailed industries.

The skill shares in hours worked and labour compensation are calculated based on the LFS data for 2005-2009. Prior to 2005 the growth trend of the skill shares in EU KLEMS has been used to extrapolate the

series backwards to 1995. Relative wage data is from the EU Structural Earnings Survey (SES) and EU Survey on Income and Living Standards (SILC) and are available for 2005-2009.

Data on investment, investment prices and capital stocks are taken directly from the November 2009 release of the EU KLEMS database in which data are available up to 2007.

17. Greece

Employment data by industry are taken from the STAN database. Hours worked by employees are estimated for the period 1995-2000 by assuming a fixed ratio of employees to self-employed.

The skill shares in hours worked and labour compensation are calculated based on the LFS data for 2002-2009. Prior to 2002 the growth trend of the skill shares in EU KLEMS has been used to extrapolate the series backwards to 1995. Relative wage data is from the EU Structural Earnings Survey (SES) and EU Survey on Income and Living Standards (SILC) and are available for 2002-2008. For 2009 relative wages are assumed to be the same as the 2008 level.

Capital stock estimates are based on the investment data from the STAN database. The initial capital stock is based on detailed investment data for 7 asset types delivered to us by Statistics Greece in November 2006. These investment series start in 1980 and also provide an initial capital stock. These stocks were aggregated over the asset types. The 1995 total asset capital stock is taken and the PIM method is used to build up the capital stock up to 2009. The Spanish industry depreciation rates are used.

18. Hungary

Employment data by industry are taken from the STAN database. Data on hours worked by employees is incomplete; however, data for aggregate industries is available for the whole period. Hours worked at the detailed industry level are estimated by assuming that the average number of hours worked is the same as for the aggregate industry. Data for hours worked by total persons engaged is missing detail prior to 2000 and after 2007. The distribution is kept constant at the 2000 or 2007 level.

The skill shares in hours worked and labour compensation are calculated based on the LFS data for 2002-2009. Prior to 2002 the growth trend of the skill shares in EU KLEMS has been used to extrapolate the series backwards to 1995. Relative wage data is from the EU Structural Earnings Survey (SES) and EU Survey on Income and Living Standards (SILC) and are available for 2002-2008. For 2009 relative wages are assumed to be the same as the 2008 level.

Data on investment, investment prices and capital stocks are taken directly from the November 2009 release of the EU KLEMS database in which data are available up to 2007.

19. Indonesia

Indonesian employment data are taken from input-output tables. For the non-benchmark years, they are interpolated using the growth rate of employment in the parent industry, obtained from GGDC 10 sector database. For the years after the last benchmark IO year, 2005, the employment numbers are extrapolated using growth rates from ILO until 2008 and from BPO, National labor Force Survey data for 2008 and 2009. Both these data are comparable, and are from the August version of the Survey. These data were, however, available only for broad sectors (9 sectors for other years, and 18 sectors for 2007 and 2008). Therefore, we assumed the same growth rate as in the parent industries. Also industries 60 to 63 and 70 to 74 are split using compensation shares, obtained from IO tables.

In order to calculate the number of self-employed workers we use ILO data. We obtained the share of self-employed in total persons from ILO for 1992, 1997, 1998 and 2000-2008 for 9 broad sectors. For missing years, these ratios are interpolated, and for sub-sectors the parent industry share is used. These ratios are applied to the time-series employment generated using IO table employment. The estimated self-employed number of workers is subtracted from total persons (EMP) to obtain total employees (EMPE).

Total compensation is taken from NAS and for industry distribution; we use shares from IO tables. For non-benchmark years, we interpolate industry shares, and for years after 2005, we assume a constant 2005 industry distribution. Compensation for self-employed is computed using India's relative wage rate of self-employed to employees.

Skill shares of employees and compensation are taken from two publications of Badan Pusat Statistik (BPS): *'Laborer/Employee situation in Indonesia'*, and '*Labor force situation in Indonesia',* which are based on National Labor Force Survey. Educational attainment levels have been defined as follows:

Indonesian Educational categories	**WIOD categories**
No schooling	LOW
Did not complete primary school	LOW
Primary school	LOW
Junior High School: General	LOW
Junior High School: Vocational	LOW
Senior High School: General	MEDIUM
Senior High School: Vocational	MEDIUM
Diploma I/II	HIGH
Diploma I/II	HIGH
Diploma III	HIGH
University	HIGH

Estimates of capital stock are constructed using perpetual inventory method. An initial stock for the aggregate economy has been estimated for 1978 using Harberger approach. Then we construct a series of aggregate capital stock in 1995 prices using investment data obtained from the United Nations National Accounts database. For this purpose, we use the depreciation rates obtained from detailed Spanish data.

This aggregate capital stock has been subsequently distributed across industries, using the capital stock distribution of Gross Capital Stock estimates prepared by Directorate of Economic and Monetary Statistics, Bank Indonesia. For those sectors for which this data was not separately available, we use the capital output ratio for the parent industry. This is done for subsectors of 17t19; 23t25; 29t37; GtH; JtK and MtQ. Since no GFCF data was separately obtained, we have imputed investment from the published capital stock data as $I_t = K_t - K_{t-1}(1- \delta)$. Then the aggregate GFCF data obtained from National Accounts is redistributed across sectors using the industry distribution of this imputed investment series.

20. India

Comprehensive statistics on employment in India are relatively less frequent compared to other economic variables such as output or value added. In addition, the quality of available employment data in India is widely discussed among Indian researchers (see Himanshu, 2011; Unni and Raveendran 2007; Sundaram and Tendulkar 2004 among others) and consequently there has been an improvement in the quality of employment statistics over time. Nevertheless, by now, it is widely acknowledged that the quality of the data is still inadequate (Srinivasan, 2010).[4] We try to estimate employment by industries in a more consistent way, making use of several available sources on employment in India (For a detailed discussion on the construction of employment series, see de Vries et al, 2011). Two major sources of employment data by industries covering the entire economy are the decennial population censuses and the Employment and Unemployment Surveys of the National Sample Survey Organization (NSSO). Recently, India also brings out an economic census which also provides employment data during successive economic censuses. Other sources, which cover only selected segments of the economy, includes the Directorate General of Employment and Training (DGET), NSSO surveys on unregistered manufacturing, and the Annual Survey of Industries (ASI). The existence of multiple sources of data, nevertheless, hardly helps make a meaningful comparison across sources or over time, as they differ in coverage, sectors and more importantly worker definitions and frequency of availability (see Srinivasan, 2010).

While questions are raised about the methodology and often the observed trends in employment in the data, most researchers implicitly acknowledge the fact that the NSSO "quinquennial surveys provide perhaps the only comprehensive database on employment in the country" (Rangarajan et al., 2011). Given the absence of a better alternative, we also largely depend upon the NSSO employment surveys in the employment series for India. Fortunately, the concepts used in the consecutive NSSO surveys since the 1970s has remained almost the same, making inter-temporal comparison relatively feasible. NSSO provides the share of workers in different segments of the economy in total population. This information, along with population figures from decennial population census is used to derive the number of workers in any given industry.

NSSO defines work as any activity perused for income (pay, profit or family gain), thus including any economic activity that results in the production of goods and services. It provides employment data under

[4] Srinivasan (2010) argues that it is the complexity of Indian labor market such as shifting from self-employment to wage employment in off seasons and frequent entry to and exit from the work force, that causes much of the conceptual, measurement and data gathering problems.

three major definitions; 1) Usual Principal Status (UPS); 2) Current Weekly Status (CWS); and 3) Current Daily Status (CDS). While the UPS considers a person as employed depending on the activity pursued by him/her for the major part of the previous year, the CWS considers her employed if she has worked even for one hour during the previous week Under the CDS approach each day of the seven days preceding date of survey is considered as the reference period. A person is considered to be working for the entire day, if she had worked at least for four hours during the day. Among these concepts the Usual Status is the most liberal and widely used concepts (Aggarwal, 2004). We take the employment under the 'usual principal and subsidiary status' (UPSS) definition. This includes all persons who worked for at least 30 days during the past year. The employment definition is to a large extent comparable over the various rounds of the survey, and has a wider acceptance as a measure of employment (Bosworth and Collins, 2008). In addition, this employment definition is used in the national account statistics for India. We use the various quinquennial major rounds of the EUS from the 38th (1983) to the 61st (2004/05) to generate employment for these years. This series of employment is comprehensive in that it is inclusive of casual workers, regular and salaried workers and self-employed workers.

As is evident from the discussion above, there is no time-series data on employment in India, except for the organized segments of the economy. ASI provides annual employment data for registered manufacturing, and DGET provides the same for all registered segments of the economy, in particular at a very detailed level for the recent years. In addition NSSO also provides annual employment data derived from a smaller sample of households (see Srivivasan, 2010 for a discussion of these surveys). We have explored the possibility of using all these information, whenever appropriate, to derive a time-series of employment. For instance, we observe that the thin round based annual employment numbers are quite volatile, particularly at industry level. It has been often suggested that the use of thin rounds may pose little problem if one uses for national aggregate, despite a small sample size resulting in an increase in the variance. But this may not be advised at the detailed industry level.

We obtain time-series of employment for the organized segment of the economy from DGET, and calculate the unorganized employment for survey years as a residual (see Sakthivel and Joddar, 2006). Subsequently, we linearly interpolate the ratio of unorganized to organized employment, which is used to generate time-series of employment in the unorganized sector. The sum of the two for the years in between the survey years is the total employment. Here we make an assumption that the ratio of unorganized to organized grows linearly between two consecutive survey years. In order to examine how sensitive our final results to this assumption, we also generated time-series employment using alternative approaches. NSSO also provides wage rates by workers in its employment surveys. Assuming that wage rate grows linearly between two consecutive survey years, which is a sensible assumption, we imputed annual wage rate. National Accounts provide information on total labour compensation, which is divided by the imputed wage rate to obtain implicit employment. Growth rate of these are used to interpolate time series of employment for non-survey years. The results are quite comparable. [5]We also explored the possibility of using growth rates of annual thin round surveys of NSSO to generate a time-series of employment at the aggregate level, and for broad sectors. To accommodate the annual fluctuations from the thin sample, but at the same time retain the levels in the major round years, we employ a procedure which uses the movement of the thin sample minus the average annual growth rate of the major rounds

[5] A caveat may be added however. The wage rate based calculation excludes self-employed workers, as NAS compensation data does not provide self-employed wages separately. Self-employed workers constitute a major chunk of India's labour force (Srinivasan, 2010).

over the five year periods. We observed that the annual fluctuations in the thin round data is remarkably high, even at aggregate level, posing doubts about the reliability of this data. We also tried linear interpolation of employment to arrive at time-series for years in between successive NSSO rounds. From National Accounts, one can obtain industry output, and assuming that labour productivity grows linearly, one can impute employment series. Nevertheless, we do not follow this route, as it harms the purpose of conducting a proper productivity analysis.

For the skill distribution of workers, we consider three types of educational categories from NSSO, which will correspond with WIOD classification. These are:

- low skilled: up to primary education,
- medium skill : primary to higher secondary education
- high skilled: higher secondary and above.

The share of each of the above categories in total number of persons in each industry is computed from NSSO data for the major rounds. NSSO also provides statistics on average daily wages of workers of particular characteristics. For instance, it is possible to estimate the wage earnings of a high skilled worker from the NSSO unit level data. However, such information is available only for regular and casual workers, while no information on earnings is available for self-employed persons. Therefore, we impute the wages for self-employed using a Mincer wage equation, where we regress wage rates of casual and regular workers on their characteristics (age, gender, education, location, marital status, social exclusion and industry). The estimated coefficients are then used to impute earnings of self-employed.

Employee compensation (COMP) for broad sectors of the economy is taken directly from national accounts statistics (NAS). NAS provides three components of Net Domestic Product (NDP); compensation to employees; gross operating surplus; and mixed income. For the sub-sectors of the economy we had to make additional assumptions. For the manufacturing sector NAS provides data separately for registered and unregistered segments. We use detailed information from Annual Survey of Industries (ASI) to obtain sectoral distribution of NAS data for the organized sector, and unit level data obtained from NSSO survey on unorganized manufacturing for the unorganized sector. While the ASI data is available annually, the NSSO data for unorganized sector was available only for the survey years (1989-90, 1994-95, 2000-01 and 2005-06) and therefore, we interpolate the distribution linearly for the years in between. For two non-manufacturing sectors trade and other services we use value added distribution to split aggregate data into subsectors.

Typically mixed income consists of income of self-employed, which can be considered as the implicit labor compensation of the self-employed and the gross operating surplus. Therefore, in order to obtain the total labor compensation (LAB), we had to separate that part of mixed income which can be attributed to labor compensation. In order to split mixed income into self-employed compensation and operating surplus, we use the imputed wage rate for self-employed from the NSSO data (see discussion above). The relative wage rate of self-employed to employees is computed from NSSO data, which is applied to the wage rate of employees consistent with NAS to obtain the wage rate of self-employed consistent with NAS. Self-employed compensation is then estimated using the calculated wage rate. If the estimated

compensation is higher than the reported mixed income the entire mixed income is considered as self-employed compensation. The sum of compensation from NAS and the estimated self-employed compensation is taken as total compensation (LAB).
Hours worked is computed using information on average number of days worked per week from NSSO. Assuming 8 hours of work per day and 52 working weeks in a year, we arrive at total number of hours worked. The same number of hours worked per employee is assumed for both total persons and total employees.

Capital stock is constructed since 1950 using Perpetual Inventory Method, where the initial stock for 1950 is taken from published net capital stock data provided by the Central Statistical Organization (CSO). Then the GFCF data by industries, obtained from CSO for the period 1950-2009 are used to construct a series of capital stock by sectors. Since there was no GFCF separately available for sub-sectors of manufacturing from CSO, the total manufacturing investment has been distributed across industries using shares from Annual Survey of Industries for organized manufacturing, and National Sample Survey Organization's Survey on Unorganized Manufacturing. Three asset types are included: they are construction, machinery, transport equipment. The corresponding depreciation rates used are respectively 2.5 per cent, 8.00 per cent and 10 per cent.

21. Ireland

Employment data by industry are taken from the STAN database. The STAN database does not provide data on hours worked by industry. Therefore data on average hours worked by an employee or person engaged from the EU KLEMS database, March 2011 update, are used to estimate hours worked. Average hours worked data are updated to 2009 using the growth trend of average hours worked from the aggregates at the 6 branch level from EUROSTAT. Growth trends of lower level industries are updated with the growth trend of the parent industry. In STAN there is no decomposition available for the transport and storage sector, industry 60t63, for any of the variables. Therefore Value Added shares from EU KLEMS are used for all variables. Data on the total number of employees prior to 1998 is unavailable. The ratio of the total number of persons engaged over employees in 1998 is used to estimate values prior to 1998 for employees. The distribution of the compensation of employees is used to estimate data for the breakdown of industries G and K.

The skill shares in hours worked and labour compensation are calculated based on the LFS data for 2002-2009. Prior to 2002 the growth trend of the skill shares in EU KLEMS has been used to extrapolate the series backwards to 1995. Relative wage data is from the EU Structural Earnings Survey (SES) and EU Survey on Income and Living Standards (SILC) and are available for 2004-2008. For 2009 relative wages are assumed to be the same as the 2008 level. Similarly the values for 2002-2003 have been set equal to 2004 values.

Data on investment, investment prices and capital stocks are taken directly from the November 2009 release of the EU KLEMS database in which data are available up to 2007.

22. Italy

Employment data by industry are taken from the STAN database. Average hours worked of the aggregate industry are used to estimate hours data for the breakdown of industries G, I and K.

The skill shares in hours worked and labour compensation are calculated based on the LFS data for 1998-2009. Prior to 1998 the growth trend of the skill shares in EU KLEMS has been used to extrapolate the series backwards to 1995. Relative wage data is from the EU Structural Earnings Survey (SES) and EU Survey on Income and Living Standards (SILC) and are available for 1998-2008. For 2009 relative wages are assumed to be the same as the 2008 level.

Data on investment, investment prices and capital stocks are taken directly from the November 2009 release of the EU KLEMS database in which data are available up to 2007.

23. Japan

Capital stock data and employment data by skill type are taken from the EU KLEMS 2009 release. To split industry 17t19 and 60t63 into 17t18 and 19, and 60, 61, 62, and 63 we use the more detailed EU KLEMS 2008 release. Data run until 2007.

To update the employment series to 2009, we assume a constant employment to value added ratio for 2008 and 2009. For employment and labour compensation by skill type, we assume a constant ratio from 2005 for the period from 2006-2009.

The update of the GFCF deflator is based on the aggregate investment deflator. For the growth rate of the capital stock for 2008 and 2009, we assume a constant capital to value added ratio.

Capital stock data runs until 2007. The update of the GFCF series to 2009 are based on the aggregate investment series from the National Accounts. The capital stock is subsequently updated combining constant GFCF series and the average depreciation ratios.

24. Korea

For Korea the employment numbers and skill distribution of employment and compensation are taken from EU KLEMS. This data was available only until 2007, and therefore, for years after 2007, we apply 2007 sectoral shares to total employment imputed using growth rates obtained from OECD.

Skill shares for 2008 and 2009 are assumed to be the same as in 2007. See EU KLEMS source descriptions for more details.

The GFCF and capital stock data are also directly taken from EU KLEMS. The data is taken from March 2008 version of EU KLEMS which provides capital stock data until 2005. For the later years we construct aggregate capital stock using PIM and aggregate GFCF data, and redistribute across sectors using the last available year in EU KLEMS data.

25. Lithuania

Employment data by industry are taken from EUROSTAT. For data on hours worked, detail for industries for G, I and K is missing. Average hours worked per employee or person engaged are assumed to be equal to that of the aggregate in order to estimate data at the detailed industry level.

Data from the LFS is taken for the skill shares for 2002-2009. However, no EU KLEMS data on educational attainment in the labour force is available for Lithuania. Therefore the Portuguese skill-shares from EU KLEMS have been used to extrapolate the series backwards to 1995. Relative wage data is from the EU Structural Earnings Survey (SES) and EU Survey on Income and Living Standards (SILC) and are available for 2002-2008. For 2009 relative wages are assumed to be the same as the 2008 level.

Capital stock estimates are based on the investment data from EUROSTAT. The initial capital stock is estimated using the ICVAR approach, employing Spanish capital stock to value added ratios. The PIM method is used to build up the capital stock up to 2009. The Spanish industry depreciation rates are used.

26. Luxembourg

Employment data by industry are taken from the STAN database. In STAN no data are available on hours worked, therefore average hours worked by an employee or person engaged has been taken from the EU KLEMS March 2011 update to estimate these data. Average hours for 2008 and 2009 are kept constant at the 2007 level.

The skill shares in hours worked and labour compensation are calculated based on the LFS data for 2002-2009. Prior to 2002 the growth trend of the skill shares in EU KLEMS has been used to extrapolate the series backwards to 1995. Relative wage data is from the EU Structural Earnings Survey (SES) and EU Survey on Income and Living Standards (SILC) and are available for 2002-2008. For 2009 relative wages are assumed to be the same as the 2008 level.

Capital stock estimates are based on the investment data from EUROSTAT. The initial capital stock is estimated using the ICVAR approach, employing Spanish capital stock to value added ratios. The PIM method is used to build up the capital stock up to 2009. The Spanish industry depreciation rates are used. No investment deflator is available in EUROSTAT. Therefore the total economy investment deflator is used from the UN.

27. Latvia

Employment data by industry are taken from EUROSTAT. For the split of industries 61 and 62 Value Added shares from the EU KLEMS March 2011 update are used for all variables. For data on hours worked the average hours worked from EU KLEMS are taken and multiplied by the EUROSTAT employment data. Average hours worked data are updated to 2009 using the growth trend of average hours worked from the aggregates at the 6 branch level from EUROSTAT. Growth trends of lower level industries are updated with the growth trend of the parent industry.

Data from the LFS is taken for the skill shares for 2002-2009. However, no EU KLEMS data on educational attainment in the labour force is available for Latvia. Therefore the Portuguese skill-shares from EU KLEMS have been used to extrapolate the series backwards to 1995. Relative wage data is from the EU Structural Earnings Survey (SES) and EU Survey on Income and Living Standards (SILC) and are available for 2002-2008. For 2009 relative wages are assumed to be the same as the 2008 level.

Capital stock estimates are based on the investment data from EUROSTAT. The initial capital stock is estimated using the ICVAR approach, employing Spanish capital stock to value added ratios. The PIM method is used to build up the capital stock up to 2009. The Spanish industry depreciation rates are used. No investment deflator is available in EUROSTAT. Therefore the total economy investment deflator is used from the UN.

28. Mexico

Employment series by industry are obtained from the system of national accounts, which is based on economic censuses and annual surveys of industries. INEGI, the statistical office of Mexico, considers the number of jobs as a measure of employment taking additional jobs by a person into account. Data published in the Cuenta de Bienes y Servicios (CBS) by detailed industry are used from 2003 onwards. Series before 2003 are obtained extrapolating growth rates from the previous series in the CBS.

The split of employment by educational attainment is based on the Encuesta Nacional de Empleo (ENE) from 1995 to 2004 and the Encuesta Nacional de Ocupación y Empleo (ENOE) from 2005 onwards. Both surveys provide the number of years studied. We use the ISCED mapping from the World Education Indicators Programme (UNESCO, 2007) to derive the number of years of education for low-skilled (up to 9 years), medium skilled (9-17 years) and high skilled (over 17 years). Relative wages are obtained from the same sources.

Labour shares in value added are estimated by adding the wage income of self-employed workers to labour compensation. We combine self-employed workers by industry with the economy-wide average wage rate of self-employed workers in SEDLAC (2011). The wage income of own-account workers is added to the compensation data by sector, except for government administration, education, and health services were we do not make imputations.

GFCF for tangible assets (residential structures, other construction, transport equipment, and other assets) in current and constant prices is available from INEGI/National accounts for the period from 1988-2006. This data is combined with the Economic census for 1999 and 2004 for industry shares in total GFCF for tangible assets. Investment to value added ratios for industries AtB, L, M, N, and O for Spain (from the EU KLEMS database) were used to obtain the full set of industry shares. . An initial capital stock by industry in 1995 prices is estimated using the average (1995-2007) capital to value added ratio for Spain from the EU KLEMS database (Timmer et al., 2010), and this capital stock is subsequently updated combining constant GFCF series and average depreciation rates for Spain in the perpetual inventory method.

29. Malta

Employment data by industry are taken from EUROSTAT. Very limited information is available on employment. Data for total persons engaged is only available at the 6 branch level in EUROSTAT for 2000-2009. The breakdown for detailed industries is estimated using Value Added shares. For employees' data only total economy values are available for 2000-2009. The data for detailed industries has been estimated using the distribution of total persons engaged. EUROSTAT does provide data on total hours worked for the full period 1995-2000. The growth rate of total hours worked has been used to extrapolate data for employees and persons engaged backwards to 1995. The breakdown of hours worked is based on the distribution of total persons engaged. Hours worked by employees are estimated by multiplying the average hours worked by persons engaged by the total number of employees in each industry.

No detailed information on educational attainment in the labour force is available for Malta; therefore the Portuguese skill-shares have been used.

Capital stock estimates are based on the investment data from EUROSTAT. The initial capital stock is estimated using the ICVAR approach, employing Spanish capital stock to value added ratios. The PIM method is used to build up the capital stock up to 2009. The Spanish industry depreciation rates are used. No investment deflator is available in EUROSTAT. Therefore the total economy investment deflator is used from the UN.

30. Netherlands

Employment data by industry are taken from the STAN database. Industry detail is missing in the STAN data on hours worked by persons engaged prior to 2001 for D, G, I, K and LtP. Data are estimated based on the growth rates of total persons engaged in the respective industry and normalized to match totals.

The skill shares in hours worked and labour compensation are calculated based on the LFS data for 2002-2009. Prior to 2002 the growth trend of the skill shares in EU KLEMS is used to extrapolate the series backwards to 1995. Relative wage data is from the EU Structural Earnings Survey (SES) and EU Survey on Income and Living Standards (SILC) and are available for 2002-2008. For 2009 relative wages are assumed to be the same as the 2008 level.

Data on investment, investment prices and capital stocks are taken directly from the November 2009 release of the EU KLEMS database in which data are available up to 2007.

31. Poland

Employment data by industry are taken from EUROSTAT. There is no decomposition available of industry aggregates G and K for any of the variables, therefore Value Added shares from the EU KLEMS

March 2011 update are used to estimate data for detailed industries. From 2006 onwards (and from 1999 backwards for employment) industry detail is missing. Growth patterns of the aggregate industries are used to estimate data for the detailed industries. For data on hours worked the average hours worked from EU KLEMS are taken and multiplied by the EUROSTAT employment data. Average hours worked data are updated to 2009 using the growth trend of average hours worked from the aggregates at the 30 branch level from EUROSTAT. Growth trends of lower level industries are updated with the growth trend of the parent industry.

The skill shares in hours worked and labour compensation are calculated based on the LFS data for 2002-2009. Prior to 2002 the growth trend of the skill shares in EU KLEMS is used to extrapolate the series backwards to 1995. Relative wage data is from the EU Structural Earnings Survey (SES) and EU Survey on Income and Living Standards (SILC) and are available for 2002-2008. For 2009 relative wages are assumed to be the same as the 2008 level.

Capital stock estimates are based on the investment data from the STAN database up to 2007. For 2008 and 2009 the STAN investment data is updated with data from EUROSTAT. The initial capital stock is estimated using the ICVAR approach, employing Spanish capital stock to value added ratios. The PIM method is used to build up the capital stock up to 2009. The Spanish industry depreciation rates are used.

32. Portugal

Employment data by industry are taken from EUROSTAT. Industry detail is missing for industries G, I and K. Value Added shares are used from the EU KLEMS database, March 2011 update, and shares are assumed constant after 2006 for these industries. After 2006 employment data is only available at the 6 branch level from EUROSTAT. Growth rates of the aggregate industries are used to estimate values of the underlying industries. For data on hours worked the average hours worked from EU KLEMS are taken and multiplied by the EUROSTAT employment data. Average hours worked data are updated to 2009 using the growth trend of average hours worked from the aggregates at the 6 branch level from EUROSTAT. Growth trends of lower level industries are updated with the growth trend of the parent industry.

The skill shares in hours worked and labour compensation are calculated based on the LFS data for 2002-2009. Prior to 2002 the growth trend of the skill shares in EU KLEMS is used to extrapolate the series backwards to 1995. Relative wage data is from the EU Structural Earnings Survey (SES) and EU Survey on Income and Living Standards (SILC) and are available for 2002-2008. For 2009 relative wages are assumed to be the same as the 2008 level.

Data on investment, investment prices and capital stocks are taken directly from the March 2008 release of the EU KLEMS database in which data are available up to 2005. After 2005 the data on capital stocks are estimated using EUROSTAT investment series. The PIM method is applied to the available EU KLEMS capital stocks. Industry depreciation rates are calculated from the EU KLEMS capital input data using the same method that was applied to calculate the Spanish industry depreciation rates.

33. Romania

Employment data by industry are taken from EUROSTAT. Data for 2009 is only available for aggregates; estimates for detailed industries are made by using the growth rates of the aggregate industries. Prior to 1999 employment data for hours worked is missing and for persons engaged/employees only aggregates are available. When aggregates are available the decomposition is based on the growth rates of the aggregate industries. Estimations for the hours worked data prior to 1999 are based on data from ILO on (the growth of) average hours worked at the aggregate level, which are applied to the detailed industry level in combination with employment data in persons.

No detailed information on educational attainment in the labour force is available for Romania; therefore the Portuguese skill-shares have been used.

Capital stock estimates are based on the total economy investment data from the United Nations' National Accounts. The investment data in current prices has been disaggregated using VA shares. The total economy investment deflator is applied to all industries. The initial capital stock is estimated using the ICVAR approach described above. The capital stock to value added ratios and the industry depreciation rates from Spain are used.

34. Russia

The main source for employment series in Russia is the system of national accounts employment statistics, which provides full-time equivalent jobs by one-digit sectors for the period from 2003-2008. Importantly, this source includes households that produce goods and services for own consumption. In Russia, the share of hours worked from these activities by households is estimated at about 12-15 per cent of total hours worked, and 57.8 per cent of total hours worked in agriculture (Rosstat, 2009). For disaggregation and backward extrapolation of employment series to 1995, we used the *Balance of labour force*, the *Full circle employment survey*, and the *Labour force survey* for particular industries.

Shares of employment by skill type are currently assumed similar to that of the Czech Republic.

Net capital stock series in constant prices are calculated on the basis of nominal gross fixed capital formations by type of assets, investment deflators, and net capital stock values for the starting year. Data on gross fixed capital formation by industry matches with the official series for the total economy. GFCF consist of investments to fixed capital in organizations, nonmaterial assets and investments of households. Investments in organizations are available in industries, data of non-material assets is for the total economy only, and households' investments are calculated as the difference between GFCF and total investments to fixed capital and non-material assets.

Data on nominal investments were broken down by industry and types of assets. Imputations were done in two steps. First, official data on investments to fixed assets and non-material assets in organizations were broken down by types of assets with the data on new acquisitions for large and medium firms from the survey F11. Second, investments of households were allocated by industries and types of assets in proportion to new asset acquisitions from the official Balance of Households Property.

Net capital stock values in industries in the beginning of 1995 were obtained from the official Balance of fixed assets. The data were broken down by types of assets with data of large and medium firms from the survey F11. Investment series were deflated with official investment deflators for the total economy. Since official data before 2004 is available in the old industrial classification OKONKh, the nominal investment and capital stock series for these years were bridged with NACE 1.0 with the detailed official concordance table for investments.

35. Slovak Republic

Employment data by industry are taken from the STAN database.

The skill shares in hours worked and labour compensation are calculated based on the LFS data for 2002-2009. Prior to 2002 the growth trend of the skill shares in EU KLEMS is used to extrapolate the series backwards to 1995.

Capital stock estimates are based on the investment data from the STAN database. The initial capital stock is estimated for 1995 using the Harberger method. The PIM method is used to build up the capital stock up to 2009. The Spanish industry depreciation rates are used.

36. Slovenia

Employment data by industry are taken from the STAN database. The STAN database does not include data on hours worked; therefore average hours worked data from the EU KLEMS database, March 2011 update, are used. However, Slovenian data in EU KLEMS only runs up to 2006 so average hours worked are assumed constant at the 2006 level for 2007, 2008 and 2009.

The skill shares in hours worked and labour compensation are calculated based on the LFS data for 2002-2009. Prior to 2002 the growth trend of the skill shares in EU KLEMS is used to extrapolate the series backwards to 1995. Relative wage data is from the EU Structural Earnings Survey (SES) and EU Survey on Income and Living Standards (SILC) and are available for 2005-2008. For 2009 relative wages are assumed to be the same as the 2008 level. Similarly the relative wages for 2002-2004 have been set equal to the 2005 values.

Data on investment, investment prices and capital stocks are taken directly from the November 2009 release of the EU KLEMS database in which data are available up to 2006. After 2006 the data on capital stocks are estimated using STAN investment series. The PIM method is applied to the available EU KLEMS capital stocks. Industry depreciation rates are calculated from the EU KLEMS capital input data using the same method that was applied to calculate the Spanish industry depreciation rates.

37. Sweden

Employment data by industry are taken from the STAN database. Employment data for detailed industries in G is missing. These are estimated by applying the shares from COMP.

The skill shares in hours worked and labour compensation are calculated based on the LFS data for 2002-2009. Prior to 2002 the growth trend of the skill shares in EU KLEMS is used to extrapolate the series backwards to 1995. Relative wage data is from the EU Structural Earnings Survey (SES) and EU Survey on Income and Living Standards (SILC) and are available for 2002-2008. For 2009 relative wages are assumed to be the same as the 2008 level.

Data on investment, investment prices and capital stocks are taken directly from the November 2009 release of the EU KLEMS database in which data are available up to 2007.

38. Turkey

Turkish employment data inclusive of self-employed are taken from Turkstat *Labor Force Surveys*. These data are available for the broad sectors of the economy and therefore further information from *Annual business enterprise survey* is used to split aggregate sectors to subsectors. This has been done for splitting sectors manufacturing, trade, transport and storage and other services. We obtained detailed data at ISIC 3.1 2 digit level business survey data for 1992 till 2001 separately for firms employing 1 to 9 workers and 10+ workers for manufacturing sector only. Since 2003 the same data was obtained for all firms (1 to 9 and 10+) for manufacturing and services sectors. The industry distribution of total persons engaged from this survey is applied to the broad sector aggregates obtained from labor force survey to obtain sectoral employment data.

From Turkstat labor force survey, we also obtained total self-employed workers (employers + own account workers + unpaid family worker) in Agricultural sector and non-agricultural sector since 2002. Since the non-agricultural sector data was not available by sub-sectors, we use the information from Annual business survey statistics to split this into sub-sectors. Annual business statistics also provides data on total employees and total persons. We impute the self-employed by industries using the ratio of Self-employed to total persons (Total persons-Total employees/Total Persons) from business statistics multiplied with total persons from Labor force survey. The distribution of this imputed series has been applied to the total reported self-employed in non-agricultural sector to obtain the controlled sectoral self-employed workers.

Compensation (COMP) for each sector is computed using comp/GDP ratio from input output tables, controlled for the time-series of total compensation from NAS. In order to compute total compensation (LAB), we first impute self-employed compensation using the relative wage rate of self-employed/employees in India, applied to Turkey. LAB is then the total of the imputed compensation and COMP. Since this imputation produced very high compensation for agriculture, we set the LAB share in GDP for agriculture as 0.9.

Skill distribution of employment is also obtained from Turkstat labor force survey, for the broad sectors, separately. For the sub-sectors, we use the skill distribution of the sectoral aggregate (for instance all manufacturing sectors are assumed to have the same skill distribution). The educational categories used are:

- Low-skilled =Below high school (primary, secondary, illiterate, and others);

- Medium-skilled = High school and vocational high school; and
- High-skilled =University and above.

Compensation shares for these educational categories are arrived at using data on average earnings for educational categories. We obtained average earnings for educational categories for the aggregate economy for 2006 and 2010 from earning statistics of Turkstat. Since this information is obtained only for the aggregate economy, compensation shares by industries are imputed using relative wages of educational groups for the aggregate economy, as wage rate of a given skill category relative to wage rate of low skilled (e.g. wage rate of medium skill/low skill) multiplied with employment share of the given skill category (e.g. medium skilled in total). For years between 2006 and 2010, we linearly interpolated the relative wages, and for years before 2006, we used a fixed relative wage rate.

Total Hours worked is computed using Hours per employee for the aggregate economy obtained from the Total Economy Database provided by the Conference Board. We assume same hours worked per employee for all industries, in order to impute total hours in each industry.

Turkish capital stock is constructed using investment data obtained from Turkstat for the same broad sectors of the economy. As in the case of employment, we supplemented this information by detailed sectoral data from *Annual business statistics*, which is used to split aggregate sectors to sub-sectors. A time series since 1970 has been constructed, with the initial stock for 1970 is calculated using Harberger approach (see discussion for Taiwan).

39. Taiwan

Employment in Taiwan is taken from National Statistics of Taiwan. Most recent version (for 2000 to 2010) of *Employed Persons by industry* available from Directorate General of Budget, Accounting and Statistics (DGBAS) is used. We take total employment provided by this database, which is defined as "persons who, during the reference week, undertake a paid job or work 15 hours or more as unpaid family workers", as total persons engaged (EMP). For earlier years they are interpolated backwards using growth rates from previous versions. The most recent data available since 2001, however, had remediation services included in water supply. 2001 data was also available in the earlier series, where it was not included. We subtract remediation services from water supply for all years after 2001 using 2001 shares. Remediation is included in other services. However, this data was not available for all WIOD sectors and therefore, we had to split some sectors. For manufacturing sector, we use the distribution from detailed pay-roll statistics to split total manufacturing into sub-sectors. In order to split some of the service sectors, viz. trade into 3 sectors, transport into 4 sectors, and other services (OtP) into two sectors we use labor compensation shares.

In order to compute total employees (EMPE), we first construct a series of self-employed workers. We obtained number of self-employed workers along with total workers for years 1993, 2002, 2007 and 2009 from DGBAS, '*Yearbook of Manpower Survey Statistics'*. The number of self-employed by industry has been derived by applying ratios of self-employed (including employers, own-account workers and unpaid

family workers) to employees. This is done for 14 industry groups (Agriculture, Mining & Utilities, Manufacturing, Construction, Trade, Hotels, Transport & communication, Finance, Real estate, Business services, Education, Health, Other services, Public administration). For more detailed industries the ratio of the parent industry is used and the ratios for years in-between have been linearly interpolated. The ratios are derived from the same survey as used for total persons engaged to be consistent. We could also obtain the average labor compensation for self-employed from the same source. Employee compensation (COMP) for the aggregate economy has been directly taken from the National Accounts. For subsectors this has been computed using compensation/GDP ratio for each industry obtained from input-output tables. For non-benchmark years, we interpolate this ratio and apply to sectoral GDP to obtain compensation. These are then normalized to the NAS total. Total compensation (LAB) is computed using the relative wage rate of self-employed to employees also obtained from 'Yearbook of manpower survey statistics'. By multiplying relative wage rate of self-employed to employees with the ratio of self-employed to employees, we obtain the ratio of self-employed compensation to employee compensation, which is applied to COMP to obtain LAB. We use the relative wage rate for the broad sectors of the economy and assume the same rates for nearest sub-sectors. Also, we did not have the self-employed wage rate for agriculture, and therefore, we used the relative wage rate for the total economy. In sector P, the estimated compensation was higher than GDP, hence it is set equal to GDP.

The income of self-employed have been estimated on relative wages of employees and self-employed from DGBAS, *Census of Industry*, 2006, Table 46. The ratios are derived for 13 broad non-agricultural industry groups. The ratio for agriculture is taken to be the total economy average. On average self-employed income is about 65% of employee income. These income ratios appeared to be rather stable over time and we kept them constant over the period.

Hours worked by employees are derived by multiplying number of employees with average annual hours worked per employee. Hours worked per employee are based on average monthly working hours by employees (including regular and irregular working hours) times 12. Industry detail is available for all manufacturing industries and ten broad sectors (Mining & Utilities, Construction, Trade, Transport & communication, Hotels, Finance, Real estate, Business services, Other services, Health). Annual data is taken from DGBAS, *Earnings and productivity survey.* Agriculture and Public Administration are based on Employed persons' weekly working hours from DGBAS, *Reports on the Manpower Utilization Survey* 2004-2007. For these sectors years before 2004 and after 2007 data are assumed constant respectively at 2004 and 2007 level. For sector Education we take the same hours per worker as in Public administration. Hours worked by self-employed are assumed to be the same as for employees (for each industry)

The skill distribution of workers is obtained from various issues of DGBAS, Yearbook of Manpower Survey Statistics. This data has been obtained for 14 broad sectors of the economy for years 1993, 2002, 2004, 2005, 2006, 2007, 2008 and 2009. For the missing years, the data has been interpolated, and for sub-sectors of the economy, we assume the skill distribution of the nearest aggregate sector. We also obtained the relative income of these educational categories for 2004-2009 period for the aggregate economy from DGBAS, Report on the Manpower Utilization Survey. This has been applied to the share of each educational category in total employment to obtain the compensation shares. For years before 2004, we assume 2004 relative wages

Capital stock for Taiwan is generated using Perpetual Inventory Method. We obtained investment and investment deflators data since 1981 from National Statistics of Taiwan, which was available for 19 sectors of the economy. These 19 sectors, however, did not strictly concord with the WIOD industry classification. Therefore, we had to split many sectors: manufacturing, trade, transport and storage, other services (71t74) and personal and house hold services. For these sectors we use the investment output ratio for the parent industry. From Statistics Taiwan, we obtained a series on real gross fixed capital stock in 1996 constant prices. This was available for industries Mining & Quarrying, Manufacturing, Electricity, Gas & Water, Construction, Trade & Eating and drinking Place, Transport, Storage & Communication, Financing, Insurance and Real Estate, Business Services, and Community, Social & Personal Services for the period 1961-1999. We take our initial stock for 1981 from this series after converting it into 1995 prices. For the sub-sectors, we split the parent industry's initial stock using the distribution of initial stock constructed using Harberger approach. In this case, the stock for 1981 is calculated using Harberger approach, where the steady state capital output ratio is calculated as $\frac{[i/y]}{(g+\delta)}$, where i is the average investment to output ratio (for the first 10 years), g is the average growth rate of output (for first 10 years) and δ the rate of depreciation. Depreciation rates are assumed to be the same over time, and are taken from detailed Spanish data.

40. USA

The US employment (EMP and EMPE), Hours, compensation (COMP) and skill distribution are updates of EU KLEMS database. We have imputed LAB by assuming the employees' wage rate for self-employed. GFCF and capital stock data are directly taken from EU KLEMS November 2009 version. For skill shares, we did not have separate data for industries 17 to 19 and 60 to 63. Therefore, we assumed the same skill distribution as in the parent industry. The same applies for both employment and compensation shares. The US skill categories and their corresponding skill categories in the WIOD are as follows:

US educational categories	**WIOD categories**
Before 1992	
Less than HS	LOW
Some high school	LOW
High school grad	MEDIUM
Some college	MEDIUM
College grad	HIGH
More than college	HIGH
After 1992	
8th grade or less	LOW
grades 9-12 no diploma	LOW
high school grad	MEDIUM
some college no degree, associate degree	MEDIUM
BA, BS	HIGH
More than BA	HIGH

References

Bai, Chong-En, and Zhenjie Qian (2010). The Factor Income Distribution in China: 1978-2007. *China Economic Review*, vol. 21(4), p. 650-670.

Barro, Robert and Jong-Wha Lee, April 2010, "A New Data Set of Educational Attainment in the World, 1950-2010." *NBER Working Paper* No. 15902

Bosworth, B., and S. M. Collins (2008). "Accounting for Growth: Comparing China and India", *Journal of Economic Perspectives,* vol. 22(1), p. 45-66.

CEC. *China Economic Census* (2004). Zhongguo jingji pucha nianjian 2004 [China Economic Census Yearbook 2004]. Beijing: China Statistics Press. In 4 volumes.

China MOA (2005). China, Ministry of Agriculture, *China Township and Village Enterprises Yearbook* Editorial Committee, editors. Beijing: China Agriculture Publishing House.

Dweck et al. (2009). *Matriz de Absorção de Investimento e Análise de Impactos Econômicos*. Rio de Janeiro: UFRJ, Instituto de Economia.

Easterly, W. and R. Levine (2002). "It´s Not Factor Accumulation: Stylized Facts and Growth Models".

Harberger, A (1978), Perspectives on capital and technology in less developed countries: in Artis, M and A. Nobay (eds.), *Contemporary Economic Analysis*, Croom Helm, London.

Instituto Brasileiro de Geografia e Estatistica (2008). *Sistema de Contas Nacionais Brasil* (2nd edition). Series Relatorios Metodologicos, 24.

NBS SY. National Bureau of Statistics of China (various issues). *China Statistical Yearbook* (Zhongguo tongli nianjian). Beijing: China Statistical Publishing House.

O'Mahony, M. and M.P. Timmer (2009), "Output, Input and Productivity Measures at the Industry Level: the EU KLEMS Database", *Economic Journal*, 119(538), pp. F374-F403

Rosstat (2009). *Labor and Employment in Russia*, 2009. Moscow: Rosstat.

SEDLAC (2011). *Socio-Economic Database for Latin America and the Caribbean* (CEDLAS and The World Bank). April 2011 version.

Srinivasan, T.N., (2010). "Employment and India's development and reforms," *Journal of Comparative Economics*, vol. 38(1), pages 82-106, March.

Timmer, M.P. (1999), "Indonesia's Ascent on the Technology Ladder: Capital Stock and Total Factor Productivity in Indonesian Manufacturing, 1975-1995", *Bulletin of Indonesian Economic Studies* , vol 35(1), pp. 75-97.

Timmer, M.P., M. O'Mahony, B. van Ark and R. Inklaar (2010). *Economic Growth in Europe*. Cambridge University Press.

UNESCO Institute for Statistics (2007). *World Education Indicators Programme*.

Wu, H. X. (2000). "China's GDP level and growth performance: Alternate estimates and the implications", *Review of Income and Wealth*, vol. 46(4), pp. 475-499.

Wu, H. X. (2002). "How Fast has Chinese Industry Grown? Measuring the Real Output of Chinese Industry, 1949-97" *Review of Income and Wealth*, vol. 48(2), pp. 179-204.

Wu, H.X and X. Yue (2010). "Accounting for Labor Input in Chinese Industry, 1949-2005". *Paper Presented at the 31st General Conference of the International Association for Research in Income and Wealth*, August 2010.

Appendix Table 1. Industry depreciation rates used in capital stock estimates for WIOD

Industry	code	1995-2006 average depreciation rate
AGRICULTURE, HUNTING, FORESTRY AND FISHING	AtB	4.7%
MINING AND QUARRYING	C	5.3%
FOOD , BEVERAGES AND TOBACCO	15t16	5.9%
Textiles and textile	17t18	5.9%
Leather, leather and footwear	19	5.9%
WOOD AND OF WOOD AND CORK	20	6.1%
PULP, PAPER, PAPER , PRINTING AND PUBLISHING	21t22	6.6%
Coke, refined petroleum and nuclear fuel	23	6.4%
Chemicals and chemical	24	6.2%
Rubber and plastics	25	6.6%
OTHER NON-METALLIC MINERAL	26	6.6%
BASIC METALS AND FABRICATED METAL	27t28	5.7%
MACHINERY, NEC	29	7.1%
ELECTRICAL AND OPTICAL EQUIPMENT	30t33	7.6%
TRANSPORT EQUIPMENT	34t35	9.0%
MANUFACTURING NEC; RECYCLING	36t37	6.0%
ELECTRICITY, GAS AND WATER SUPPLY	E	4.9%
CONSTRUCTION	F	5.4%
Sale, maintenance and repair of motor vehicles and motorcycles; retail sale of fuel	50	4.1%
Wholesale trade and commission trade, except of motor vehicles and motorcycles	51	4.5%
Retail trade, except of motor vehicles and motorcycles; repair of household goods	52	4.3%
HOTELS AND RESTAURANTS	H	4.0%
Other Inland transport	60	6.0%
Other Water transport	61	6.0%
Other Air transport	62	6.0%
Other Supporting and auxiliary transport activities; activities of travel agencies	63	6.0%
POST AND TELECOMMUNICATIONS	64	11.5%
FINANCIAL INTERMEDIATION	J	12.1%
Real estate activities	70	1.3%
Renting of m&eq and other business activities	71t74	12.0%
PUBLIC ADMIN AND DEFENCE; COMPULSORY SOCIAL SECURITY	L	3.2%
EDUCATION	M	3.7%
HEALTH AND SOCIAL WORK	N	7.1%
OTHER COMMUNITY, SOCIAL AND PERSONAL SERVICES	O	7.7%