ILO Estimates and projections of the economically active population: 1980-2020 (Fifth edition)

Methodological description

Update August 2008

Contents

Preface

The 5th Edition of the estimates and projections of the economically active population EAPEP Database is the result of a joint collaboration between the ILO Bureau of Statistics and the ILO Employment Trends Unit.

In this project, the Employment Trends Unit had primary responsibility for developing the historical estimates portion of the database (1980-2003), whilst the Bureau of Statistics had primary responsibility for developing the projections.

This collaborative project utilised new and enhanced methodologies to improve the EAPEP labour force estimates and projections, while also establishing a system to guarantee more frequent and reliable data updates. The resulting models and methodologies will be the basis for subsequent updates of the EAPEP Database by the Bureau of Statistics and the Employment Trends Unit.

This paper was prepared by Steven Kapsos (ILO Employment Trends UNIT) for the estimates, James Brown and Fiifi Amako Johnson (University of Southampton) for the basic model of the projections and Farhad Mehran, Ferdinand Lepper and Christophe Vittorelli (ILO Bureau of Statistics) for the projections.

1. Introduction

The ILO programme on estimates and projections of the economically active population is part of a larger international effort on demographic estimates and projections in which several UN agencies contribute. Estimates and projections of the total population and its components by sex and age group are produced by the UN Population Division, the economically active population by the ILO, the agricultural population by FAO and the school attending population by UNESCO.

The main objective of the ILO programme is to provide member states, international agencies and the public at large with the most comprehensive, detailed and comparable estimates and projections of the economically active population in the world and its main geographical regions. The first edition was published by the ILO Bureau of statistics in 1971 (covering 168 countries and territories, with reference period 1950-1985)[1]; the second edition in 1977 (with154 countries and territories and reference period 1975-2000)[2]; the third edition in 1986 (with 156 countries and territories and reference period 1985-2025)[3]; and the fourth edition in 1996 (with 178 countries and territories and reference period 1950-2010)[4].

The present fifth edition covers 191 countries and territories and 29 economic and geographical groupings. The reference period for the estimates is 1980-2006 and for the projections, 2007-2020. The basic data are single-year labour force participation rates by sex and eleven age groups in five-year age intervals, the last age group being 65 years and above. The data are available at the ILO main website on labour statistics: http://laborsta.ilo.org. Due to updated information (new UN population projections and estimates on activity rates up to 2003) and a further developed methodology, the results of the 5th edition may be different from those of the 4th edition for the same year/country for both the total estimates and the projections. A comparison of the results of the two editions is thus not useful.

The purpose of the present note is to describe the main elements of the estimation and projection methodologies adopted for the fifth edition. Both the estimation and projection parts deviate substantially from the procedures adopted in the previous editions: they have been designed to a much greater extent on consistent models with minimum parameter adjustments. It simplifies the methodological descriptions and makes the numerical results essentially reproducible.

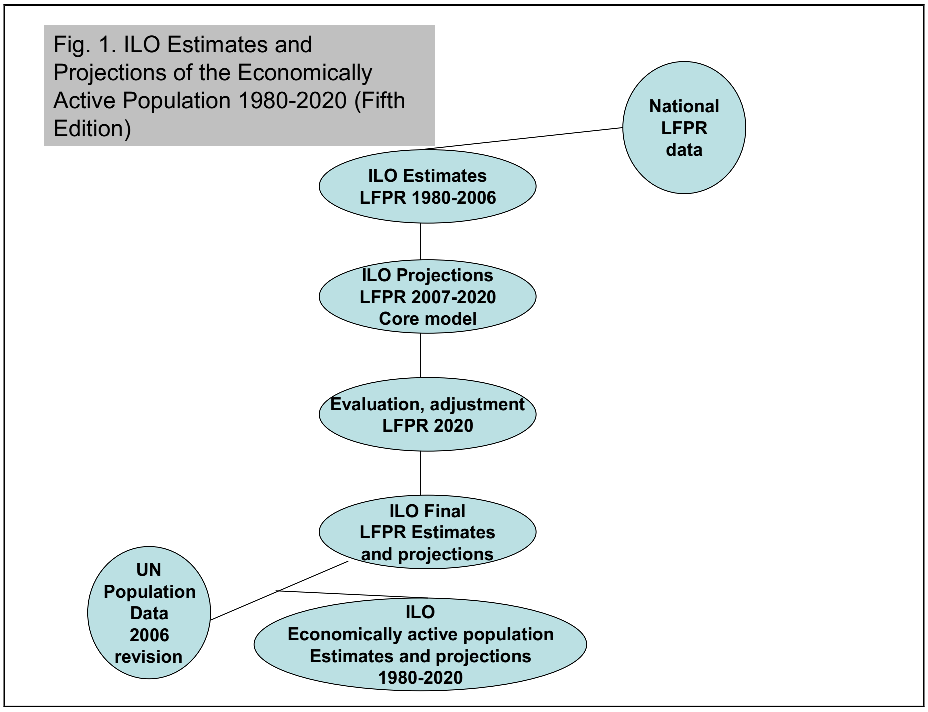
[1] ILO, Labour force projections, 1965-85 (1st edition, Geneva 1971).

[2] ILO, Labour force projections, 1950-2010 (2nd edition, Geneva, 1976).

[3] ILO, Economically Active Population: Estimates and projections, 1950-2025 (3rd edition, Geneva, 1986).

[4] ILO, Economically Active Population : Estimates and projections, 1950-2010 (4th edition, Geneva, 1996).

The following chart (fig. 1) depicts the main steps involved:



The underlying national labour force data used for producing harmonised single-year ILO country estimates of labour force participation rates (LFPR) by sex and standard age groups are described in Section 2. Also described in that section is the statistical treatment of missing values and the estimation models for countries for which no or limited data were available. The projection methodology is described in Section 3. It begins by describing the core model based on scenarios concerning the pattern of convergence or divergence of male and female labour force participation rates over the projection horizon. The procedures used for evaluating the results and, where necessary, adjusting the parameters of the core model are then outlined with a few numerical examples. Finally, Section 4 describes the procedures used for transforming the labour force participation rates into counts of the economically active population and those used for summing the country-level data into the main geographical and economic aggregates. Annex 1 lists the countries and territories, and the geographical and economic groupings covered by the project. Annex 2 present standardized sex and age profiles of estimated and projected labour force participation rates for each country.

2. Estimation model 1980-2006: data and methodology

Introduction

This section contains two main parts. The first part provides an overview of the criteria used to select the baseline national labour force participation rate (LFPR) data that serve as the key input into the ILO's Economically Active Population Estimates and Projections (EAPEP) 5th Edition database. This section includes a discussion of non-comparability issues that exist in the available national LFPR data and concludes with a description of the LFPR data coverage, after taking into account the various selection criteria. The second part describes the econometric model developed for the treatment of missing LFPR values, both in countries that report in some but not all of the years in question, as well as for those countries for which no data are currently available.

Data selection criteria and coverage

Overview

The EAPEP database is a collection of country-reported and ILO estimated labour force participation rates. The database is a complete panel, that is, it is a cross-sectional time series database with no missing values. A key objective in the construction of the database was to generate a set of *comparable* labour force participation rates across both countries and time. With this in mind, the first step in the production of the historical portion of the 5th Edition of the EAPEP database was to carefully scrutinize existing country-reported labour force participation rates and to select only those observations deemed sufficiently comparable. In the second step, a weighted least squares econometric model was developed to produce estimates of labour force participation rates for those countries and years in which no country-reported, cross-country comparable data currently exist. The following describes the sources of data non-comparability, the process through which data were either selected or eliminated and the resulting data coverage and database structure.

Non-comparability issues

In order to generate a set of sufficiently comparable labour force participation rates across both countries and time, it is necessary to identify and address the various sources of non-comparability. The main sources of non-comparability of labour force participation rates are as follows:[5]

Survey type – country-reported labour force participation rates are derived from several types of survey data including labour force surveys, population censuses, establishment surveys, insurance records or official government estimates. Data taken from different survey types are often not comparable.

Age group coverage – non-comparability also arises from differences in the age groupings used in measuring the labour force. While the standard age-groupings used in the EAPEP Database are 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64 and 65+, some countries report non-standard age groupings, which can adversely affect comparability.

Geographic coverage – some country-reported labour force participation rates correspond to a specific geographic region, area or territory. Geographically-limited data are not comparable across countries.

Others – Non-comparability can also arise from the inclusion or non-inclusion of military conscripts; variations in national definitions of the economically active population, particularly with regard to the statistical treatment of "contributing family workers" and "unemployed, not looking for work"; and differences in survey reference periods.

[5] This section draws heavily on the labour force participation data comparability discussion in the Key Indicators of the Labour Market (KILM), 4th Edition, Geneva, ILO.

Data selection criteria

Taking these issues into account, a set of criteria was established upon which nationally-reported labour force participation rates would be selected for or eliminated from the input file for the EAPEP dataset.[6] The selection criteria include the following:

Selection criterion 1. Data must be derived from either a labour force survey or population census and population census data are included only if no labour force survey data exist for a given country. Labour force surveys are the most comprehensive source for internationally comparable labour force data. National labour force surveys are typically very similar across countries, and the data derived from these surveys are generally much more comparable than data obtained from other sources. Consequently, a strict preference was given to labour force survey data in the selection process. Yet, many developing countries without adequate resources to carry out a labour force survey do report LFPR estimates based on population censuses. Due the need to balance the competing goals of data comparability and data coverage, some population census-based labour force participation rates were included. However, a strict preference was given to labour force survey-based data, with population census-derived estimates only included for countries in which no labour force survey-based participation data exist. Data derived from official government estimates were not included in the dataset as the methodology for producing official estimates can differ significantly across countries and over time, leading to non-comparability.

Selection criterion 2. Only data corresponding to the 11 standardized age-groups (15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64 and 65+) are included. The inclusion of data corresponding to age-groups other than those listed above could result in a less comparable dataset. Therefore only data from these 11 standard age groupings were included.

Selection criterion 3. Only fully national (i.e. not geographically limited) labour force participation rates are included. Labour force participation rates corresponding to only urban or only rural areas were not included. This criterion was necessary due to the large differences that often exist between rural and urban labour markets.

Resulting input data file

Together, these criteria determined the data content of the final input file, which was utilized in the subsequent econometric estimation process (described below). Table 1 provides response rates and total observations by age-group and year. These rates represent the share of total potential observations for which real, cross-country comparable data exist.

[6] All labour force participation data in the EAPEP input file were selected from the ILO Key Indicators of the Labour Market (KILM) 5th Edition Database (Geneva, 2007), http://www.ilo.org/kilm. The main sources of data in the KILM include the ILO Yearbook of Labour Statistics (Laborsta) (Geneva, 2007), http://laborsta.ilo.org; the Organization for Economic Development and Cooperation (OECD) Labour Force Statistics Database, http://www.oecd.org; and the ILO Labour Market Indicators Library (LMIL), http://www.ilo.org/trends.

Table 1. Response rates by year, both sexes combined

Year	Proportion of potential observations	Number of observations
1980	0.17	732
1981	0.18	746
1982	0.17	720
1983	0.23	956
1984	0.19	778
1985	0.20	838
1986	0.22	942
1987	0.19	804
1988	0.22	942
1989	0.29	1212
1990	0.31	1298
1991	0.32	1324
1992	0.26	1072
1993	0.28	1180
1994	0.28	1182
1995	0.31	1304
1996	0.32	1364
1997	0.31	1316
1998	0.31	1320
1999	0.33	1398
2000	0.32	1352
2001	0.31	1304
2002	0.33	1384
2003	0.33	1382
2004	0.36	1514
2005	0.31	1322
2006	0.26	1100
Total	0.27	30786

The input file is also broken down by sex, however the number of both male and female observations is the same (15,393), thus only total figures are provided in the table. In total, comparable data are available for 30,786 out of a possible 113,434 observations, or approximately 27 per cent of the total. The total number of potential observations in the panel is determined by multiplying 191 countries * 11 age-groups * 2 sexes * 27 years = 113,454. It is important to note that while the percentage of real observations is rather low, 168 out of 191 countries (88 per cent) reported labour force participation rates in at least one year during the 1980 to 2006 reference period. Thus, some information on LFPR is known about the vast majority of the countries in the sample.

There is very little difference among the 11 age-groups with respect to data availability. This is primarily due to the fact that countries that report LFPR in a given year tend to report for all age groups. The main exception to this occurs in cases in which some reported age-groups do not conform to selection criterion 2. On the other hand, there is clear variation in response by year. In particular, coverage has tended to improve over time, as the lowest coverage occurred in the early 1980s. While the overall response rate is approximately 27 per cent, as will be shown in the next section, response rates vary substantially among the different regions of the world.

Missing value estimation procedure

Overview

This section describes the basic missing value estimation model developed to produce the EAPEP historical database. The model was developed by the ILO Employment Trends Unit as part of its ongoing responsibility for the development and analysis of world and regional aggregates of key labour market indicators including labour force, employment, unemployment, employment status, employment by sector and working poverty, among others.[7] The present methodology contains four steps. First, in order to ensure realistic estimates of labour force participation rates, a logistic transformation is applied to the input data file. Second, a simple interpolation technique is utilized to expand the baseline data in countries that report labour force participation rates in some years. Next, the problem of non-response bias (systematic differences between countries that report data in some years and countries that do not report data in any year) is addressed and a solution is developed to correct for this bias. Finally, the weighted least squares estimation model, which produces the actual country-level LFPR estimates, is explained in detail. Each of these steps is described below.

Step 1: Logistic transformation

The first step in the estimation process is to transform all labour force participation rates included in the input file. This step is necessary since using simple linear estimation techniques to estimate labour force participation rates can yield implausible results (for instance labour force participation rates of more than 100 per cent). Therefore, in order to avoid out of range predictions, the final input set of labour force participation rates is transformed logistically in the following manner prior to the estimation procedure:

$$y_{it}^{T} = \ln\left(\frac{y_{it}}{1-y_{it}}\right) \qquad (1)$$

where y_{it} is the observed labour force participation rate in country i and year t. This transformation ensures within-range predictions, and applying the inverse transformation produces the original labour force participation rates. The specific choice of a logistic function in the present context was chosen following Crespi (2004).

Step 2: Country-level interpolation

The second step in the estimation model is to fill in, through linear interpolation, the set of available information from countries that report in some but not all of the years in question. In many reporting countries, some gaps in the data do exist. For instance, a country will report labour force participation rates in 1990 and 1992, but not in 1991. In these cases, a simple linear interpolation routine is applied, in which "smoothed" LFPR estimates are produced using equation 2.

[7] A series of background papers on the Trends Unit's work related to world and regional aggregates including methodological documents on the relevant econometric models can be found at http://www.ilo.org/trends.

$$y_i^T = \frac{y_{i1}^T - y_{i0}^T}{t_1 - t_0}(t - t_0) + y_{i0}^T \qquad (2)$$

In this equation, y_i^T is the interpolated logistically transformed labour force participation rate in country *i*, and *t* is the year in which y_i^T is linearly interpolated. y_{i1}^T is the logistically transformed labour force participation rate in year t_{i1}, which corresponds to the closest reporting year in country *i* following year *t*. y_{i0}^T is the logistically transformed labour force participation rate in year t_{i0}, which is the closest reporting year in country *i* preceding year *t*. Accordingly, y_{i1}^T is bounded at the most recent overall reporting year for country *i*, while y_{i0}^T is bounded at the earliest reporting year for country *i*.

This procedure increases the number of observations upon which the econometric estimation of labour force participation rates in reporting and non-reporting countries is based and also provides a somewhat smoother, more stable series for use in the estimation.

Table 2. Response rates by estimation group

Estimation group	% of potential obs.	% of potential obs., post-interpolation	Obs.	Obs., post-interpolation
Developed Europe	80.1	85.2	10474	11130
Developed Non-Europe	87.0	93.2	4652	4984
CEE and CIS	24.1	42.2	3872	6766
East and South-East Asia	18.2	31.9	2384	4166
South Asia	19.6	52.9	1046	2830
Central America and the Caribbean	23.1	54.4	3424	8078
South America	34.0	58.4	2022	3470
Middle East and North Africa	9.6	34.9	1028	3736
Sub-Saharan Africa	6.5	18.6	1884	5414
Total	27.1	44.6	30786	50574

The increase in observations resulting from the linear interpolation procedure is provided in table 2. This table also provides a picture of the large variation in data availability among the different geographic/economic estimation groups. In total, the number of observations increased from 30,786 to 50,574 – that is, from 27 per cent to 44.6 per cent of the total potential observations. The lowest data coverage is in sub-Saharan Africa, in which the post-interpolation coverage is 18.6 per cent. East and South-East Asia, the Middle East and North Africa and CEE & CIS regions also have relatively low post-interpolation coverage, at 31.9 per cent, 34.9 per cent and 42.2, respectively. Post-interpolation coverage in all other regions is over 50 per cent, reaching 85.2 per cent in the Developed Europe region and 93.2 per cent in the Developed Non-Europe region. This resulting database represents the final set of harmonized real and estimated labour force participation rates upon which the multivariate weighted estimation model was carried out as described below.

Step 3: Calculation of response-probabilistic weights

Out of 191 countries in the EAPEP database, 23 do not have any reported comparable labour force participation rates over the 1980-2006 period. This raises the potential problem of non-response bias. That is, if labour force participation rates in countries that do not report data tend to differ significantly from participation rates in countries that do report, basic econometric estimation techniques will result in biased estimates of labour force participation rates for the non-reporting countries, as the sample upon which the estimates are based does not sufficiently represent the underlying heterogeneity of the population.[8]

The identification problem at hand is essentially whether data in the EAPEP database are missing completely at random (MCAR), missing at random (MAR) or not missing at random (NMAR).[9] If the data are MCAR, non-response is ignorable and multiple imputation techniques such as those inspired by Heckman (1979) should be sufficient for dealing with missing data. This is the special case in which the probability of reporting depends neither on observed nor unobserved variables – in the present context this would mean that reporting and non-reporting countries are essentially "similar" in both their observable and unobservable characteristics as they relate to labour force participation rates. If the data are MAR, the probability of sample selection depends only on observable characteristics. That is, it is known that reporting countries are different from non-reporting countries, but the factors that determine whether countries report data are identifiable. In this case, econometric methods incorporating a weighting scheme, in which weights are set as the inverse probability of selection (or inverse propensity score), is one common solution for correcting for sample selection bias. Finally, if the data are NMAR, there is a selection problem related to unobservable differences in characteristics among reporters and non-reporters, and methodological options are limited. In cases where data are NMAR, it is desirable to render the MAR assumption plausible by identifying covariates that impact on response probability (Little and Hyonggin, 2003).

Given the important methodological implications of non-response type, it is instructive to examine characteristics of reporting and non-reporting countries in order to determine the type of non-response present in the EAPEP database. Table 3 confirms significant differences between reporting and non-reporting countries in the sample.

[8] For more information, see Crespi (2004) and Horowitz and Manski (1998).

[9] See Little and Hyonggin (2003) and Nicoletti (2002).

Table 3. Per-capita GDP and population size of reporting and non-reporting countries

	Reporters	Non-reporters
Mean per-capita GDP, 2005 (2000 International $)	10568	4031
Median per-capita GDP, 2005 (2000 International $)	6376	1812
Mean population, 2005 (millions)	37.6	8.8
Median population, 2005 (millions)	8.1	4.0
Total countries	168	23

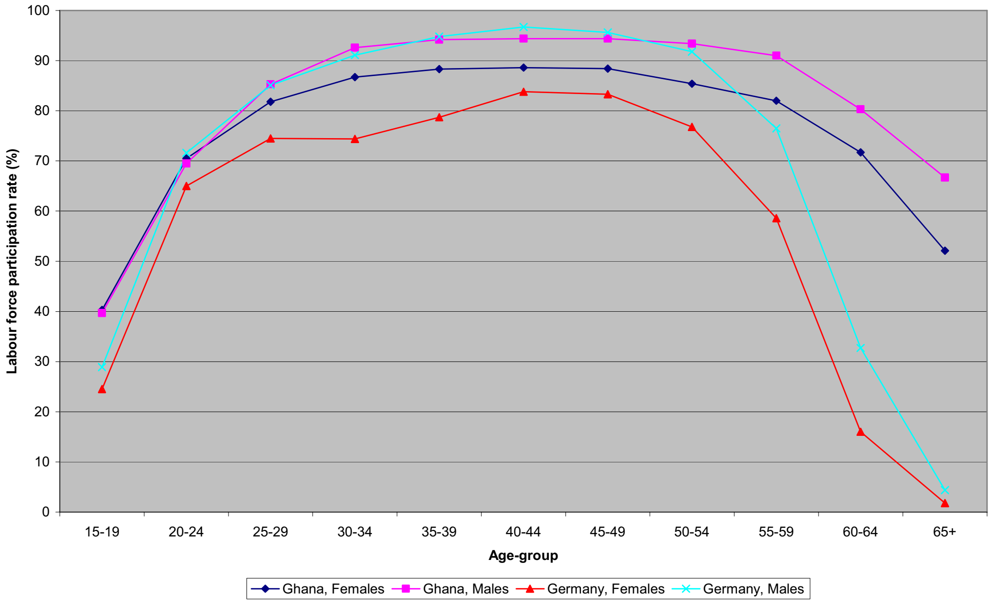
Sources: World Bank, World Development Indicators Database 2007; UN, World Population Prospects 2006 Revision Database.

The table shows that reporting countries have considerably higher per capita GDP and larger populations than non-reporting countries. In the context of the EAPEP database, it is important to note that countries with low per-capita GDP also tend to exhibit higher than average labour force participation rates, particularly among women, youth and older individuals. This outcome is borne mainly due to the fact that the poor often have few assets other than their labour upon which to survive. Thus, basic economic necessity often drives the poor to work in higher proportions than the non-poor. As economies develop, many individuals (particularly women) can afford to work less, youth can attend schooling for longer periods and, consequently, overall participation rates in developing economies moving into the middle stages of development tend to decline.[10]

This is demonstrated in figure 2, which depicts actual country-reported labour force participation rates by 5-year age-group in Germany and Ghana. Germany's per-capita GDP in 2005 stood at around $26,200, while Ghana's was approximately $2,150. While there is little difference with regard to male prime working-age labour force participation, female participation is considerably higher in Ghana, including during prime child-rearing years. In addition, the LFPR curves corresponding to women and men in Ghana are considerably flatter than the curves corresponding to their German counterparts. This reflects the considerably higher participation rates of youth and older workers in Ghana.

[10] See ILO, KILM 4th Edition, (Geneva, ILO, 2005) and Standing, G. *Labour Force Participation and Development* (Geneva, ILO, 1978).

Figure 2. Labour force participation rates by age-group in Ghana and Germany, most recent year



It appears that factors exist that co-determine the likelihood for countries to report labour force participation rates in the EAPEP input dataset and the actual labour force participation rates themselves. The missing data do not appear to be MCAR. Due to the existence of data (such as per-capita GDP and population size) that exist for both responding and non-responding countries and that are related to response likelihood, it should be possible to render the MAR assumption plausible and thus to correct for the problem of non-response bias.[11] This correction can be made while using the fixed-effects panel estimation methods described below, by applying "balancing weights" to the sample of reporting countries. The remainder of the present discussion describes this weighting routine in greater detail.

The basic methodology utilized to render the data MAR and to correct for sample selection bias contains two steps. The first step is to estimate each country's probability of reporting labour force participation rates. In the EAPEP input dataset, per-capita GDP, population size, year dummy variables and membership in the Highly Indebted Poor Country (HIPC) Initiative represent the set of independent variables used to estimate response probability.[12]

Following Crespi (2004) and Horowitz and Manski (1998), we characterize each country in the EAPEP input dataset by a vector (y_{it}, x_{it}, w_{it}, r_{it}), where y is the outcome of interest (the logistically transformed labour force participation rate), x is a set of covariates that determine the value of the outcome and w is a set of covariates that determine the probability of the outcome being reported. Finally, r is a binary variable indicating response or non-response as follows:

[11] Indeed, according to Little and Hyonggin (2003), the most useful variables in this process are those that are predictive of both the missing values (in this case labour force participation rates) and of the missing data indicator. Per-capita GDP is therefore a particularly attractive indicator in the present context.

[12] HIPC membership is utilized as an explanatory variable for response probability due to the fact that HIPC member countries are required to report certain statistics needed to measure progress toward national goals related to the program. As a result, taking all else equal, HIPC countries may be more likely to report labour force participation rates.

$$r_{it} = \begin{cases} 1 & \text{if } i \text{ reports} \\ 0 & \text{if } i \text{ does not report} \end{cases} \quad (3)$$

Equation 4 indicates that there is a linear function whereby the likelihood of reporting labour force participation rates is a function of the set of covariates:

$$r_{it}^{*} = w_{it}^{'}\gamma + \varepsilon_{it} \quad (4)$$

where a country reports if this index value is positive ($r_{it}^{*} > 0$). γ is the set of regression coefficients and ε_{it} is the error term. Assuming a symmetric cumulative distribution function, the probability of reporting labour force participation rates can be written as in equation 5.

$$P_i = F\left(w_{it}^{'}\gamma\right) \quad (5)$$

The functional form of *F* depends on the assumption made about the error term ε_{it}. As in Crespi (2004), we assume that the cumulative distribution is logistic, as shown in equation 6:

$$F\left(w_{it}^{'}\gamma\right) = \frac{\exp\left(w_{it}^{'}\gamma\right)}{1 + \exp\left(w_{it}^{'}\gamma\right)} \quad (6)$$

It is necessary to estimate equation 6 through logistic regression, which is carried out by placing each country into one of the 9 estimation groups listed in table 2. The regressions are carried out for each of the 11 standardized age-groups. The results of this procedure provide the predicted response probabilities for each age-group within each country in the EAPEP dataset.

The second step is to calculate country weights based on these regression results and to use the weights to "balance" the sample during the estimation process. The predicted response probabilities calculated in equation 6 are used to compute weights defined as:

$$s_{it}(w) = \frac{P(r_{it} = 1)}{P(r_{it} = 1 \mid w_{it}, \hat{\gamma})} \quad (7)$$

The weights given by equation 7 are calculated as the ratio of the proportion of non-missing observations in the sample (for each age-group and each year) and the reporting probability estimated in equation 6 of each age-group in each country in each year. By calculating the weights in this way, reporting countries that are more similar to the non-reporting countries (based on characteristics including per-capita GDP, population size and HIPC membership) are given greater weight and thus have a greater influence in estimating labour force participation rates in the non-reporting countries, while reporting countries that are less similar to non-reporting countries are given less weight in the estimation process. As a result, the weighted sample looks more similar to the theoretical population framework than does the simple un-weighted sample of reporting countries.

Step 4: Weighted multivariate estimation

The final step is the estimation process itself. Countries are again divided into the 9 estimation groups listed above, which were chosen on the combined basis of broad economic similarity and geographic proximity.[13] Having generated response-probabalistic weights to correct for sample selection bias, the key issues at hand include 1) the precise model specification and 2) the choice of independent variables for estimating labour force participation.

In terms of model specification, taking into account the database structure and existence of unobserved heterogeneity among the various countries in the EAPEP input database, the choice was made to use panel data techniques with country fixed effects, with the sample of reporting countries weighted using the $s_{it}(w)$ to correct for non-response bias.[14] By using fixed effects in this way, the "level" of known labour force participation rates in each reporting country is taken into account when estimating missing values in the reporting country, while in non-reporting countries, the weighted average fixed effect among reporting countries in each estimation group is used to estimate these countries' labour force participation rates. More formally, the following linear model was constructed (and run on the logistically transformed labour force participation rates):

$$Y_{it}^{T} = \ln\left(\frac{y_{it}}{1-y_{it}}\right) = \alpha_i + x_{it}^{'}\beta + \mu_{it} \qquad (8)$$

where y_{it} is the observed labour force participation rate in country i and year t and x_{it} is a set of explanatory covariates of the labour force participation rate and e_{it} is the error term. The main set of covariates included is listed in table 4.[15]

[13] Schaible (2000) discusses the use of geographic proximity and socio-economic status to define estimation domains for data estimation including for ILO labour force participation rates. See also Schaible and Mahadevan-Vijaya (2002).

[14] Crespi (2004) provides a test comparing the bias resulting from different missing value estimation models and finds that the weighted least squares model using fixed-effects provides the smallest relative bias when estimating unemployment rates.

[15] Covariate selection was done separately for each of the estimation groups. Full regression results corresponding to the EAPEP Version 5 database are published in Kapsos (2007) "World and regional trends in labour force participation: Methodologies and key results", Economic and Labour Market Papers 2007/1, Geneva, ILO.

Table 4. Independent variables in fixed-effects panel regression

Variable	*Source*
Per-capita GDP	World Bank, World Development Indicators 2007; IMF, World Economic Outlook October 2007.
Per-capita GDP squared	World Bank, World Development Indicators 2007; IMF, World Economic Outlook October 2007.
Real GDP growth rate	World Bank, World Development Indicators 2007; IMF, World Economic Outlook October 2007.
Lagged real GDP growth rate	World Bank, World Development Indicators 2007; IMF, World Economic Outlook October 2007.
Share of population aged 0-14	United Nations, World Population Prospects 2006 Revision Database
Share of population aged 15-24	United Nations, World Population Prospects 2006 Revision Database
Share of population aged 25-64	United Nations, World Population Prospects 2006 Revision Database

In the context of the EAPEP database, there are two primary considerations in selecting independent variables for estimation purposes. First, the selected variables must be robust correlates of labour force participation, so that the resulting regressions have sufficient explanatory power. Second, in order to maximize the data coverage of the final EAPEP database, the selected independent variables must have sufficient data coverage.

In terms of variables related to economic growth and development, as mentioned above, per-capita GDP is often strongly associated with labour force participation.[16] This, together with the substantial coverage of the indicator made it a prime choice for estimation purposes. However, given that the direction of the relationship between economic development and labour force participation can vary depending on a country's stage of development, the square of this term was also utilized to allow for this type of non-linear relationship.[17] Annual GDP growth rates were used to incorporate the relationship between participation and the state of the macro-economy.[18] The lag of this term was also included in order to allow for delays between shifts in economic growth and changes in participation.

Changes in the age-structure of populations can also affect labour force participation rates over time. This happens at the country-wide level, since different age cohorts tend to have different labour force participation rates, and thus changes in the aggregate age structure of a population can affect the overall participation rate. More importantly for the present analysis, however, is the potential impact that demographic changes can have on intra age-group participation rates within countries. Changes in population age structure can affect the overall burden for caring for dependents at home, thus affecting individuals' decisions to participate in labour markets. This can have a particularly important effect on women's

[16] See also Nagai and Pissarides (2005), Mammen and Paxon (2000) and Clark et al. (1999).

[17] Whereas economic development in the poorest countries is associated with declining labour force participation (particularly among women and youth), in the middle- and upper- income economies, growth in GDP per capita can be associated with rising overall participation rates – often driven by rising participation among newly empowered women. This phenomenon is the so-called "U-shaped" relationship between economic development and participation. See ILO, KILM 4th Edition and Mammen and Paxon (2000).

[18] See Ngai, L. and Pissarides (2005), Fortin and Fortin (1998) and McMahon (1986).

decision to enter into work.[19] In order to incorporate these types of demographic effects, the share of population aged 0-14 (young age-dependent), 15-24 (working-age youth) and 25-64 (prime working age) were incorporated to varying degrees in regions in which an important relationship between participation and demographics was found. These variables are by definition correlated and thus increase the presence of multicolinearity in the regressions. However it was determined that this did not present a prohibitively significant problem in the context of the present estimation procedure.

In all estimation groups, a set of country dummy variables was used in each regression in order to capture country fixed effects. A dummy variable to indicate whether the observation was pre- or post-1990 was also included in regressions carried out for the Central and Eastern Europe (CEE) and Commonwealth of Independent States (CIS) estimation group, as was an interaction between this dummy variable and the per-capita GDP and per-capita GDP squared variables. These variables were used to capture the important effects of the dissolution of the Soviet Union on labour markets throughout the region. A preliminary examination of the input data revealed that countries in the South Asia estimation group exhibit a particularly large degree of heterogeneity in labour force participation rates, especially with regard to female participation. In order to estimate robust labour force participation rates in non-reporting countries in this estimation group, it was necessary to introduce a dummy variable to further subdivide economies in the region based on observed national labour market characteristics and prevailing cultural norms with regard to male and female labour market participation. This variable was significant in more than 70 per cent of the regressions carried out for the estimation group. Finally, the constant $\alpha_{i,}$ given in equation 8, is country-specific and captures all the persistent idiosyncratic factors determining the labour force participation rate in each country. The end result of this process is a balanced panel containing real and imputed cross-country comparable labour force participation rates for 191 countries over the period 1980-2003. In the final step, these labour force participation rates are multiplied by the total population figures given in the United Nations World Population Prospects 2004 Revision database, which gives the total labour force in each of the 191 countries, broken down by age-group and sex.

3. Projection: 2007-2020

a. Core model

Background

The fourth edition of projections was based on a whole set of parametric models (six in total) and within any one country usually at least two models were used to cover the range of age-sex projections although no information was given on how this choice was made. In addition, it was necessary to make ad-hoc adjustments to many of the projections to produce an acceptable set of projections across the range of ages and between males and females. It was recorded that these had taken place but not the nature of the adjustments. This meant it was virtually impossible to reproduce the projections given the estimates, even knowing the base model that had been used.

[19] Bloom and Canning (2005), Falcão and Soares (2005), O'Higgins (2003), Clark et al. (1999), Fullerton (1999) and McMahon (1986) provide some examples of the relationships between population structure (and demographic change) and labour force participation rates for different groups of the population.

Bearing this in mind, the approach taken with the fifth round core model has been to simplify the approach by using a single flexible parametric model that can then be adjusted in a transparent way to reflect the specialist knowledge of ILO statisticians and economists. The aim of the core model is to produce the initial set of estimates based on a set of automated procedures. It is expected that for most countries the estimates for one or two age groups will then have to be adjusted to produce a final set of projections that are consistent across the ages and between the sexes. The adjustments will be based on ILO specialist knowledge and the behaviour of the estimates prior to the projection window.

The basic model

The parametric form for the basic model is linear but fitted to the logit (logistic transformation) of the proportion participating, scaled to fit between the values y_{min} and y_{max} chosen for each age-sex group by the program (see the next section). Typically, one value will be historical and one will represent the extreme long-term assumption. This implies that the participation rate at time t is then given by

$$y_t = y_{min} + \frac{y_{max} - y_{min}}{1 + e^{a+bt}}$$

where the parameters a and b are based on fitting the model to the most recent estimates for an age-sex group within a country and the projections come from extrapolating t beyond the end of the estimates. Transparent adjustments (in the sense that they are easily recorded) can then be made to the default values for y_{min} and y_{max} to ensure a plausible profile across age groups and sensible relationships between participation rates for males and females.

Default choices for y_{min} and y_{max}

The program uses information from the most recent ten-year window of estimates to choose default values for y_{min} and y_{max}, with the flexibility for the user to make informed changes to those defaults. The basic premise is that if the male and female rates are converging for a particular age group, this will continue. The alternative is either non-convergence based on a constant difference between rates or divergence where the rate of divergence comes from the most recent estimates.

Choosing the historical value

For each age-sex group the program fits a linear regression to detect whether the most recent estimates are increasing or decreasing. If the rates are decreasing the program takes a maximum value from the historical estimates. Conversely, if the rates are increasing the program takes a minimum value from the historical estimates.

Detecting convergence / divergence

For each age group, a model is fitted to the difference in participation rates between males and females. This model can detect whether:

type one: the rates are diverging;

type two: the rates are crossing just before the projection window;

type three: the rates will converge quickly during the projection interval (*predict the difference goes to zero within ten years of the last estimate*);

type four: the rates will converge slowly during the projection interval;

type five: the rates are a constant distance apart (*the slope parameter is not significant*).

Once one of the five scenarios has been selected for the male and female rates within each age-group, it is then necessary to select appropriate future values for either y_{min} or y_{max}.

Type One: The average between the rates is projected forward based on the ratio change in that average prior to the projection window.

$$\text{average}_t = \frac{\text{males}_t + \text{females}_t}{2}, \qquad \text{ratio}_t = \frac{\text{average}_t}{\text{average}_{t-1}},$$

$$\text{projected average} = \text{average}_{2000} * \overline{\text{ratio}}^5,$$

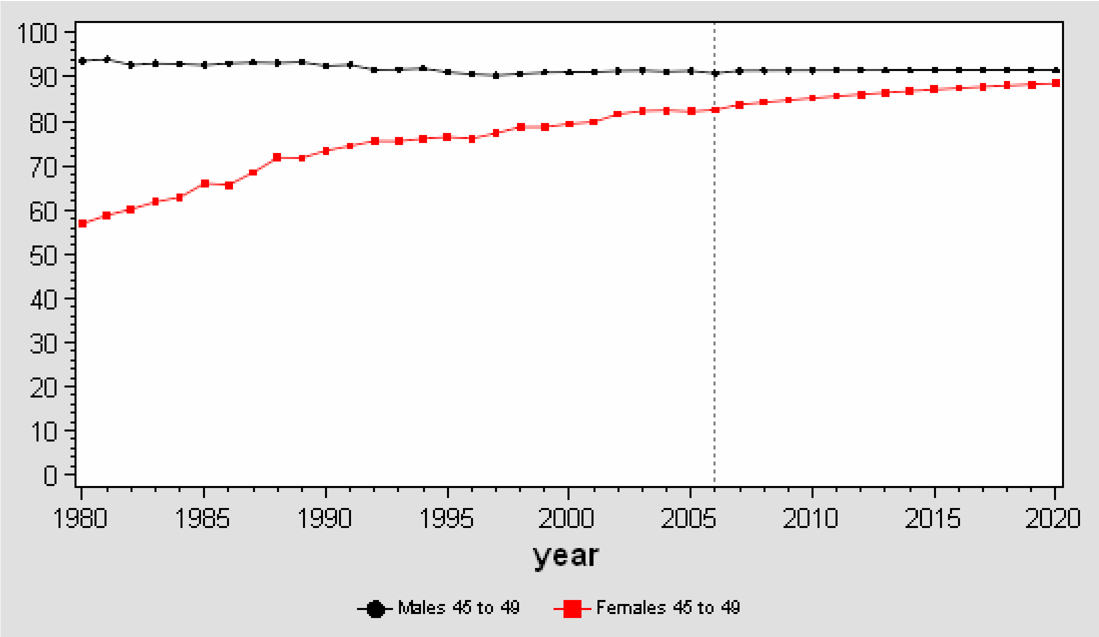
where male_t is the male LFPR at time t and female_t is the female LFPR at time t and $\overline{\text{ratio}}^5$ is the average ration for a period of 5 years.

As the rates are diverging, the difference is also projected forward based on ratio change in the difference prior to the projection window. This assumes that the mid-point between the rates given by the projected average will tend to continue along its historical trajectory and the divergence will also continue.

Type Two: This works in a similar way by projecting forward the average and the difference after the crossing point to get plausible values for the extreme future points in the logistic model.

Type Three: The average is again projected forward and this becomes the convergence point provided it is not inconsistent with the estimates for each sex. The graph below demonstrates why this is necessary as convergence is often achieved by the female rates rising much faster than the male rates are falling. Simply taking the average at the end would force male rates to fall much further in order to achieve long-term convergence.

Figure 2: Male and female LFPR for age-group45-49 (Canada)



For the age groups covering the 20s and 30s a gap is forced between males and females based on the assumption that child-bearing will always imply slightly lower rates for females than males, even if they are appearing to converge.

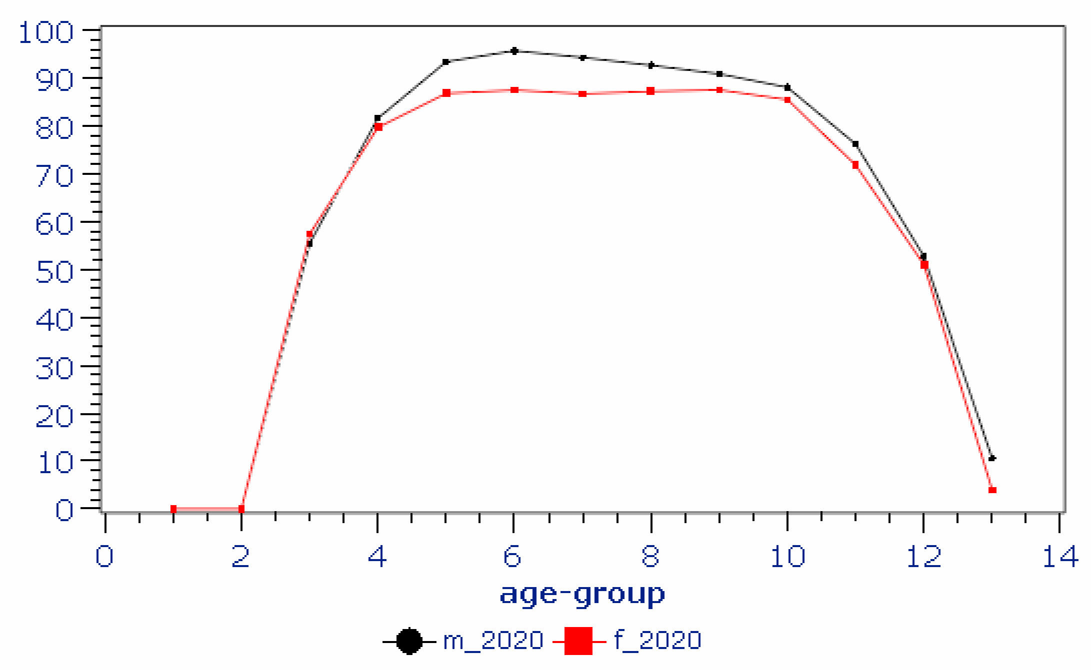
Type Four: The same as above but the average is projected by $\overline{\text{ratio}}^{10}$ to give the convergence point.

Type Five: As above the average is projected forward by $\overline{\text{ratio}}^{5}$ and then the common difference is applied to this point. This allows for the fact that both rates may be increasing or decreasing over time.

Output from the program

The program gives the users plots of the age profile for males and females at different times, the start of the projection window, the end of the published projections at 2020, and the 'long-term' profile based on the minimums and maximums in the logistic models. The program also shows the projected proportion of the total population in the labour market.

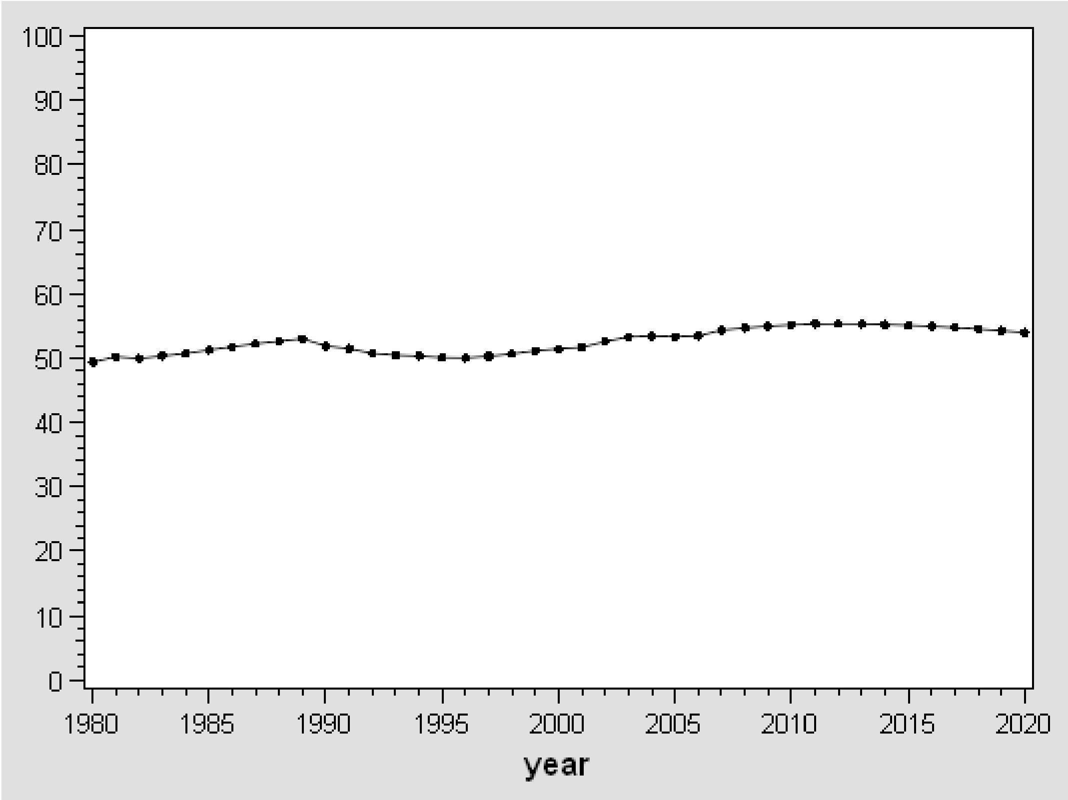
Figure 3: Default Output from Canada, male and female LFPR for 2020 [20]



20

Age-group	3	4	5	6	7	8	9	10	11	12	13
Age	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65+

Proportion in the Labour Market



Based on these plots the user can then adjust the minimum and maximum values based on expert knowledge of the type of economy each country has as well as the projected profiles. It was expected that the default values for y_{min} and y_{max} would be plausible for most age-sex groups in a country and that for those where it was not, the output would help the ILO specialist to adjust the default values accordingly.

b. Comparison with national projections

Based on the results of the first version of the 5th edition in 2006, an initial assessment of the ILO core projection model has been made by comparing the projection results with corresponding national projections of labour force participation rates by sex and age group. Because of the limited number of available national projections, the assessment was based on a selected sample of four developed countries (Canada, France, UK, and USA) and three developing countries (Iran, Mexico, and the Philippines). Furthermore, due to the variability of the projection horizon of the national projections, it was decided to use national projections with ten-year projection horizons and a target year closest to 2010.

Table 5 below shows the summary results in terms of the percentage difference between the national and the ILO core model projections. The results indicate that the core model LFPR projections are on average 4.7 percentage points higher than the corresponding national projections. The gap is substantially narrower among the developed countries than among the developing countries.

Table 5. Percentage point differences between ILO core model and national LFPR projections

	Mean	Median	Inter-quartile
Total	4.7	1.9	15.9
Developed countries	0.1	-0.4	5.3
Developing countries	5.8	2.9	19.3

Essentially, the same pattern emerges when the difference in the LFPR projections is assessed on the basis of the median, a more robust measure, somewhat tampering extreme differences between the national and the core model LFPR projections.

The last column of the table shows that there is a great variation in the gap between the national projections and the corresponding core model projections. The variation is much narrower among the developed countries than among the developing countries. But in all cases, the inter-quartile range is about ten times the median difference.

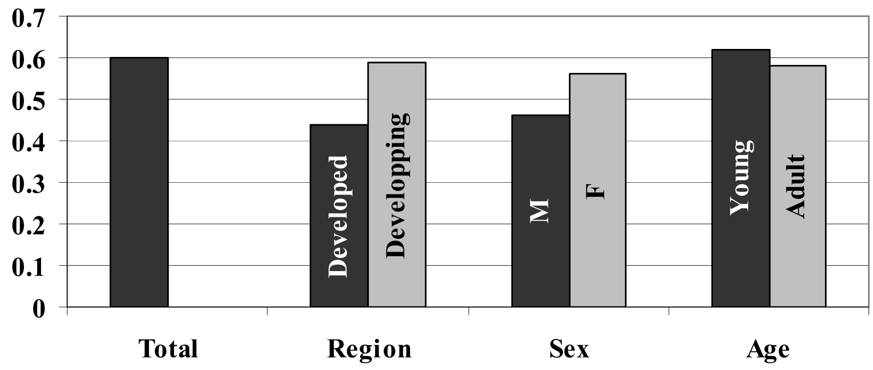
The next table (Table 6) records the differences between the core model and the national projections for men and women and for different age groups. It can be observed that the differences are virtually positive throughout the sex and age distribution, indicating that the core model over projects the labour force participation rate for both the male and female populations, and for all age groups except for the teenagers, 15 to 19 years old, and possibly for the youth population, 20 to 24 years old. The narrowest differences are for men and for the prime age groups, 25 to 44 years old.

Table 6. Percentage point differences between ILO core model and national LFPR projections by sex and age group

	Mean	Median	Inter-quartile
Male	-0.4	0.1	7.1
Female	9.7	8.8	23.5
15-19	-0.4	-2.1	17.0
20-24	-1.1	0.1	12.2
25-29	4.8	2.3	15.1
30-34	4.0	0.8	9.6
35-39	5.9	2.1	15.8
40-44	5.7	1.5	17.9
45-49	8.5	3.2	19.8
50-54	6.5	3.0	17.0
55-59	8.2	4.8	17.5
60-65	5.5	3.8	22.0
65+	5.2	2.1	16.8

A large part of the observed differences can be attributed to differences in the base-year data used for making the projections. The following figure (fig. 2) shows the values of the correlation coefficient between the base-year LFPR used in the ILO core-model against the corresponding rates used in the national projections. The correlations are statistically significant, and higher for developing countries than developed countries, for women and for men, and for adult than youth.

Figure 4: Correlation coefficient between base-year LFPR used in ILO core model and national projections



This analysis led to the decision that the projections should not be based on considerations derived from (only few available) country-specific results but on general and internationally comparable assumptions.

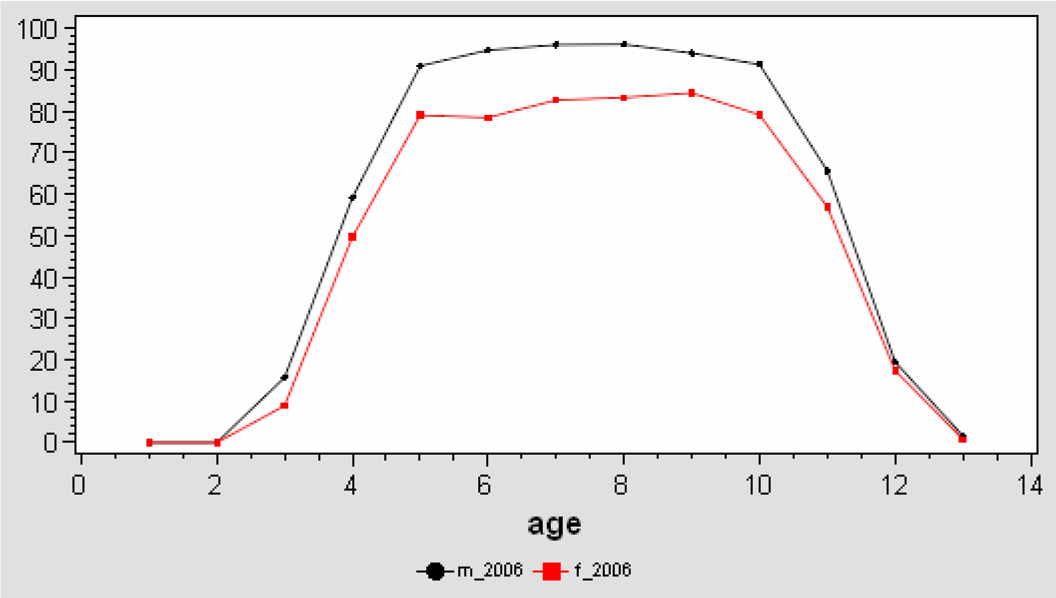
c. Adjustment of projection parameters

The consistency check of the labour force participation rates (LFPR) generated by the core model for each country has been based on 'profile' plots provided by the program. Each of these plots shows curves with the LFPR for men and women by age-group (3=15-19 years, 4=20-24 years, … , 12=60-64 years, 13=65 years and over) and has been checked for the years 2006 and 2020.

'Typical' LFPR distribution

The plot below (fig. 5) shows the 'typical' distribution for a developed country for the year 2006:

Figure 5: LFPR Profile for 2006, France [21]



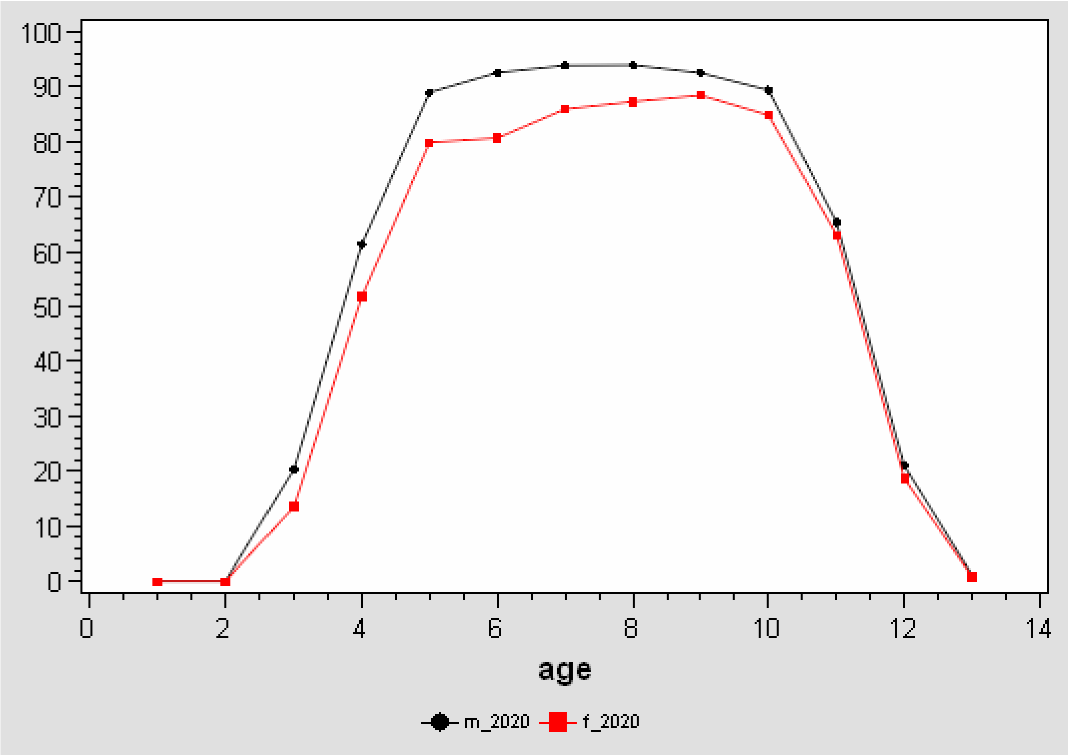
Male LFPR rise from the younger ages, have their maximum value between 30 and 55 years and then decline to a low value for those over 65 years. Female LFPR for younger ages have more or less the same value than for males. Because of motherhood they do not rise to the same extent than those for men and may even decline between 30 and 40 years. Up to 55 years the curve for female LFPR has its maximum value and then declines. Towards the end of the projection period convergence between male and female LFPR may happen fully or to a certain extent.

The plot for 2020 (fig. 6) has more or less the same characteristics. Except the fact that LFPR may have increased in general and female LFPR have – to a certain extent or fully – reached the level of LFPR for men.

21

Age-group	3	4	5	6	7	8	9	10	11	12	13
Age	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65+

Figure 6: LFPR Profile for 2020, France

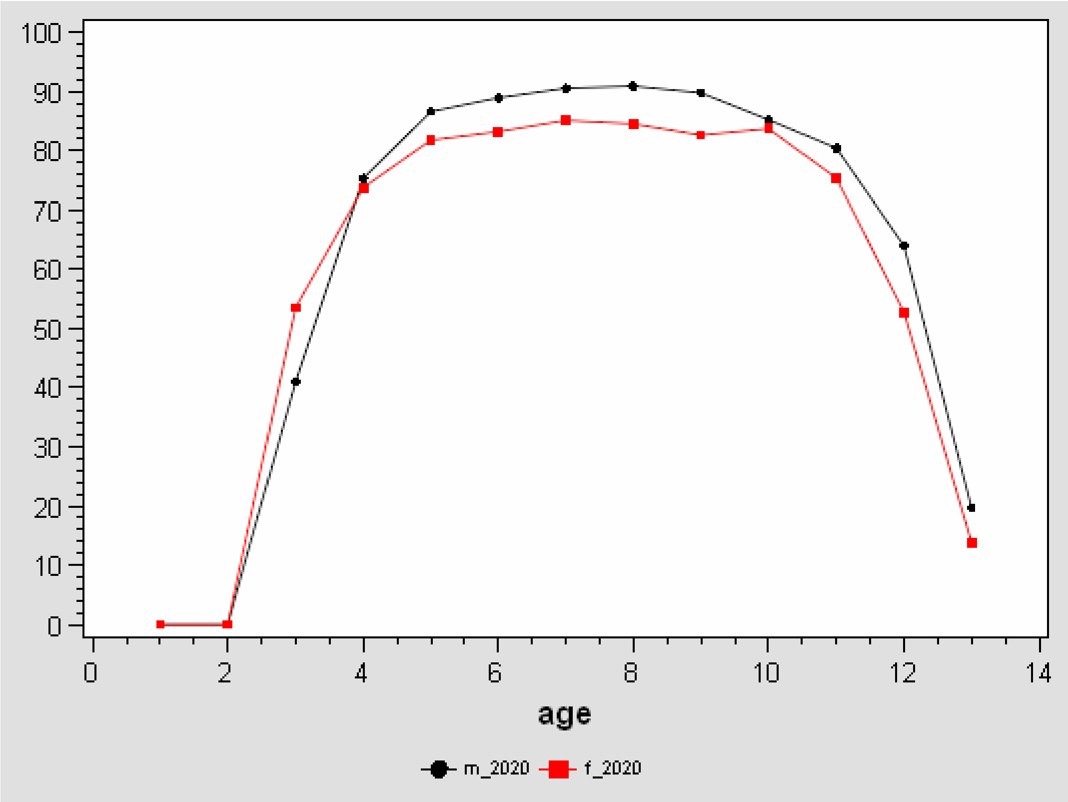


For developing countries the characteristic of the typical curves is more or less the same but some changes have to be made to the levels of LFPR. The gap between male and female LFPR is distinct in nearly all age-groups and the rates predominantly remain constant. Only for the age-group 15-19 years the model produces converging and sometimes declining values for male and female LFPR in several countries. In some cases an impact of motherhood on female LFPR in according ages is not evident because of the low level of these LFPR.

Adjusting the projections

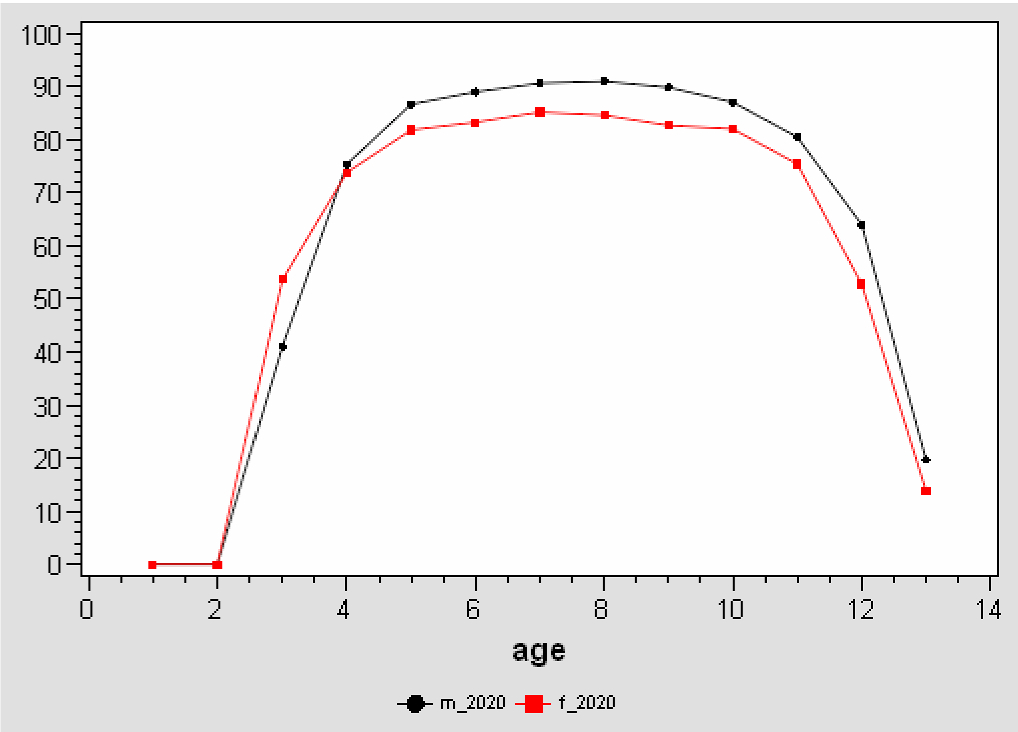
For nearly half of the 191 countries the model produced projections of LFPR that were close to this 'typical' distribution, so that no adjustments have been necessary. For the other half of the countries the plots for 2020 were checked for consistency and changes have been made to specific age groups. The plot below (fig. 7) shows for example an inconsistency in age-group 10 for both sexes, where LFPR for males are lower and for females are higher than would have been expected:

Figure 7: LFPR Profile for 2020, default output for Norway



By compelling the model to increase the value of the male LFPR and to decrease the value for the female LFPR in age-group 10 it could be achieved that the resulting plot came close to the 'ideal' distribution described above. The time series between the starting year of the projection and 2020 were automatically adjusted to these changes. The resulting plot is shown below (fig. 8).

Figure 8: LFPR Profile for 2020, corrected values for Norway



There were many countries with inconsistencies for more than one age-group and quite often more than one correction cycle was necessary to bring LFPR values into the ‘right’ dimension. Nevertheless, the method of changing target values for 2020 (and adjusting time series accordingly) was successful for nearly all countries. There were only 5 countries for which the changes had to be made manually to the time series because the model was not able to perform this automatically.

The trends of the participation rates produced by the basic model on the basis of these adjustments were consistent in general and no further adjustments have been necessary. The scenarios automatically determined for the relation between the development of male and female LFPR were mainly convergence (fully or incomplete) and common difference. There was no contradiction between the resulting projections. This was checked on the basis of plots showing the estimates and projections of the LFPR for each country by age-group and sex (Annex 2).

In developed countries male LFPR remain predominantly constant across the whole projection period and for almost all age-groups, whereas female LFPR tend to converge to these but do not always come to the same level within the projection period. Since the model did not consider information on future changes of the statutory pension systems it projected constant or slightly decreasing trends for LFPR of higher age-groups. The same applied to trends of LFPR for lower age-groups as a result of improving education and training. In a few countries female LFPR for lower age-groups are already higher than those for men. In these cases the model kept the difference constant.

For many developing countries the projections of LFPR were quite rough due to insufficient basic data. Nevertheless, the results are reasonable and indicate some common trends. In several of these countries there is a small increase for LFPR of younger ages and female LFPR have a tendency to rise a bit faster then those for male. Nevertheless, the gap between male and female LFPR remains quite large. For older ages the projections very often remain constant in many countries.

One special remark should be made in relation to the development of LFPR in countries with predominantly Muslim population. For the 4th edition of the ILO estimates and projections of the economically active population the projections of participation rates have been based upon national data until 1989. These data indicated a clear tendency of increase for many of these rates, especially for younger men and almost all age-groups of women. The development of female activity rates was analysed in Egypt in a study that has been published in the ILO Bulletin of Labour Statistics 1992-3. Its findings were also applied to the projections for other Muslim countries.

Projections of the 5th edition were now based upon estimates using national data until 2006. Beginning in 1990 data for age groups mentioned above started to decline or at least remained constant. Therefore, the computer model calculated a constant trend for future years and the resulting participation rate for women over all age groups remains at the same level whereas these rates increased in the 4th edition.

4. Aggregation

One of the goals of this project is to present estimates and projections not only of the participation rates (LFPR) but also of the total and economically active population by five-year age groups and by sex from 1980 to 2020, per country and area.

This section will present the methods employed to obtain the estimates and projections of:

- the total and economically active population by country;
- the total and economically active population and the LFPR by area.

All calculations, the participation rates were rounded to 1 decimal.

The total and economically active population by country

The estimates and projections of the total population were taken from the *2006 Revision* of the *World Population Prospects*, prepared by the United Nations Population Division[22]. They provide a comprehensive and consistent set of population data for the world's countries and their aggregates, including total population by five-year age group, major area, region and country for 1950-2050 for male, female and male plus female. Reorganization of the data in terms of format and country inclusion was carried out for consistency with the participation rates dataset.

Based on the LFPR as well as the total population, a calculation of the economically active population by country was possible with the following formula:

$$\text{active population} = \text{total population} \times \text{LFPR}$$

The total and economically active population and the LFPR by area

The country groupings are the same as those presented by the United Nations Population Division and in the last edition of *ILO Estimates and projections of the economically active population (fourth edition)*. These groupings are mainly geographical areas with sub-regional division[23].

To carry out these aggregations, following calculations were applied for each area:

- The total population of the area "j" is the sum of the total populations of each country composing this area:

$$A_j = \sum_{i=1}^{n_j} A_{ij}$$

with n_j : number of countries in area « j »

A_{ij} : total population for country « i » of area « j »

A_j : total population of area « j »

[22] United Nations, Department of Economic and Social Affairs, Population Division. World Population Prospects: The 2006 Revision.

[23] See annex 1, list of countries and regional division

- The economically active population of the area "j" is the sum of the active populations of each country composing this area:

$$B_j = \sum_{i=1}^{n_j} B_{ij}$$

with nj : number of countries in area « j »

Bij : active population for country « i » of area « j »

Bj : active population of area « j »

- The LFPR of the area "j" is the result of the division of the economically active population by the total population of this area:

$$C_j = \frac{B_j}{A_j} \times 100$$

with Aj : total population for area « j »

Bj : active population for area « j »

Cj : LFPR for area « j »

A complete dataset was built based on the total population, the economically active population and the LFPR by five-year age groups and by sex from 1980 to 2020, per country and area.

The following example (Table 7) shows the figures for the global area "WORLD" for both male and female in the period 2001 to 2010.

The output is presented in three tables: the first one shows the total population by five-year age groups in thousands; the second one, the economically active population by five-year age groups in thousands; and the third one, the LFPR by five-year age groups. It can be noted that for the economically active population, the minimum age selected is fifteen years.

The data can be viewed and downloaded freely on LABORSTA Internet (http://laborsta.ilo.org) of the ILO Bureau of Statistics.

Table 7.

WORLD (TOTAL)

*** Population by age (thousands)**
*** Economically active population by age (thousands)**
*** Economically active population by age (rates)**

Population by age (thousands)

Both sexes	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Total	6190776	6269062	6346965	6424783	6502736	6580890	6659199	6737619	6816060	6894431
Total (15+)	4342270	4421628	4501600	4581304	4660070	4737633	4814040	4889322	4963632	5037088
0-9	1233415	1231599	1230566	1230493	1231497	1233696	1237095	1241513	1246666	1252299
10-14	615092	615835	614799	612986	611169	609560	608064	606784	605762	605043
15-19	574047	584474	594197	602292	608129	611378	612265	611422	609796	608118
20-24	520574	528338	537665	547794	558082	568512	579019	588846	597055	603005
25-29	503223	503330	503378	504641	508015	513773	521583	530963	541151	551504
30-34	479044	484472	489068	492598	494980	496035	496056	496011	497195	500514
35-39	435277	443847	451786	459050	465604	471511	476762	481190	484583	486876
40-44	379173	388381	398390	408455	418006	426936	435360	443193	450380	456880
45-49	341232	346729	351146	355910	362040	369835	378924	388836	398825	408324
50-54	276012	288825	301558	313138	322843	330227	335595	339919	344609	350669
55-59	214590	221946	230867	240917	251700	263309	275653	287935	299128	308544
60-64	188551	189213	189861	191533	194968	200343	207403	215919	225496	235772
65+	430546	442073	453684	464976	475702	485775	495419	505089	515415	526883

Economically active population by age (thousands)

Both sexes	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Total (15+)	2843264	2890278	2936321	2986239	3036678	3085432	3133494	3181784	3229431	3276145
15-19	226477	227559	226623	227200	225840	222639	222556	220172	217775	215613
20-24	353030	356593	360736	366470	372321	378091	384633	390400	395005	397970
25-29	400738	399851	398675	398809	400842	404902	410929	418484	426850	435457
30-34	394213	398555	401626	404724	406167	406809	405824	404993	405272	407506
35-39	362321	370214	377537	384573	390348	395849	400318	404078	406812	408476
40-44	313806	321575	330852	340027	349392	357915	364829	371991	378488	384249
45-49	276711	281272	284822	288885	294535	301529	309272	317928	326685	335034
50-54	206702	216465	226788	235880	244091	250506	254222	257753	261576	266488
55-59	136564	141813	147973	154531	162321	170406	179292	187501	194969	201245
60-64	86465	87929	89445	91627	94070	97526	100950	105456	110482	115910
65+	86235	88451	91244	93512	96752	99261	100669	103029	105517	108198

Economically active population by age (rates)

Both sexes	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Total (15+)	65.5	65.4	65.2	65.2	65.2	65.1	65.1	65.1	65.1	65
15-19	39.5	38.9	38.1	37.7	37.1	36.4	36.3	36	35.7	35.5
20-24	67.8	67.5	67.1	66.9	66.7	66.5	66.4	66.3	66.2	66
25-29	79.6	79.4	79.2	79	78.9	78.8	78.8	78.8	78.9	79
30-34	82.3	82.3	82.1	82.2	82.1	82	81.8	81.6	81.5	81.4
35-39	83.2	83.4	83.6	83.8	83.8	84	84	84	84	83.9
40-44	82.8	82.8	83	83.2	83.6	83.8	83.8	83.9	84	84.1
45-49	81.1	81.1	81.1	81.2	81.4	81.5	81.6	81.8	81.9	82.1
50-54	74.9	74.9	75.2	75.3	75.6	75.9	75.8	75.8	75.9	76
55-59	63.6	63.9	64.1	64.1	64.5	64.7	65	65.1	65.2	65.2
60-64	45.9	46.5	47.1	47.8	48.2	48.7	48.7	48.8	49	49.2
65+	20	20	20.1	20.1	20.3	20.4	20.3	20.4	20.5	20.5

5. Bibliography

Bloom, D. and Canning, C. 2005. "Global demographic change: Dimensions and economic significance". Harvard Initiative for Global Health, Working Paper No. 1. http://www.hsph.harvard.edu/pgda/working/working_paper1.pdf

Clark, R., York, A. and Anker, R. 1999. "Economic development and labour force participation of older persons". Population Research and Policy Review, 18: 411-432.

Crespi, G. 2004. "Imputation, estimation and prediction using the Key Indicators of the Labour Market (KILM) data set". Employment Strategy Paper 2004/16, (Geneva, ILO).http://www.ilo.org/public/english/employment/strat/download/esp16.pdf

Falcão, B. and Soares, R. 2005. "The demographic division and the sexual division of labor". CDDRL, Stanford Institute on International Studies, No. 50, September, 2005. http://iis-db.stanford.edu/pubs/20967/sexualdivision-08-03-05.pdf

Fortin, M. and Fortin, P. 1998. "The changing labour force participation of Canadians, 1969-1996: Evidence from a panel of six demographic groups". Applied Research Branch, Strategic Policy, Human Resources Development Canada, October 1999. http://www11.hrsdc.gc.ca/en/cs/sp/hrsdc/arb/publications/research/1999-000074/iw-99-07e(new).pdf

Fullerton, H. 1999. "Labor force participation rates: 75 years of change, 1950-98 and 1998-2025". Monthly Labor Review, Dec. 1999. US Bureau of Labor Statistics, Washington, DC. http://www.bls.gov/opub/mlr/1999/12/art1full.pdf

Heckman J. 1979, "Sample selection as a specification error", Econometrica, vol. 47, 1, 153-161.

Horowitz J.L., and Manski C.F. 1998. "Censoring of outcomes and regressors due to survey nonresponse: Identification and estimation using weights and imputation", Journal of Econometrics, vol. 84, 37-58.

ILO, Key Indicators of the Labour Market (KILM) 4th Edition. CD-ROM. Geneva, ILO. http://www.ilo.org/KILM

ILO, Laborsta Yearly Statistics 2007. Bureau of Labor Statistics, Geneva, ILO. http://laborsta.ilo.org

International Monetary Fund. 2005. World Economic Outlook Database. September 2005. http://www.imf.org/external/pubs/ft/weo/2005/02/data/index.htm

Kapsos, S. 2006, Forthcoming. "Harmonized labour force participation rates in the KILM dataset: Methodologies and key trends". Employment Strategy Paper, 2006, Geneva, ILO.

Little, R. and Hyonggin, A. 2003. "Robust likelihood-based analysis of multivariate data with missing values". University of Michigan Department of Biostatistics Working Paper Series 2003/5. http://www.bepress.com/umichbiostat/paper5/

Mammen, K. and Paxon, C. 2000. "Women's work and economic development" Journal of Economic Perspectives, Volume 14: Number 4, 141-164.

McMahon, P. 1986. "An international comparison of labour force participation, 1977-84" Monthly Labor Review, Vol 109, No. 5. http://www.bls.gov/opub/mlr/1986/05/art1exc.htm

Ngai, L. and Pissarides, C. 2005. "Trends in labour supply and economic growth" Center for Economic Performance, London School of Economics, 2005. http://www.iza.org/conference_files/prizeconf2005/pissarides_c316.pdf

Nicoletti, C. 2002. Correcting for sample selection bias: Alternative estimators Compared. ISER, University of Essex http://iser.essex.ac.uk/seminars/mondays/2003/spring/papers/poverty3.pdf

O'Higgins, N. 2003. "Trends in the youth labour market in developing and transition countries". Social Protection Discussion Paper Series No. 321. World Bank, 2003. http://wbln0018.worldbank.org/HDNet/HDDocs.nsf/0/42971c17d2adaaea85256df1005c5767/$FILE/0321.pdf

Scheible, W. 2000. "Methods for producing world and regional estimates for selected key indicators of the labour market" http://www.ilo.org/public/english/employment/strat/download/ep6.pdf

Scheible, W. and Mahadevan-Vijaya, R. 2002. "World and regional estimates for selected key indicators of the labour market", ILO Employment Paper 2002/36. Geneva, ILO http://www.ilo.org/public/english/employment/strat/download/ep36.pdf

Standing, G. 1978. Labour Force Participation and Development. Geneva, ILO.

United Nations. World Population Prospects: 2006 Revision Population Database. http://esa.un.org/unpp/

World Bank, World Development Indicators 2005 Database. http://www.worldbank.org/

Annex 1: Country composition of each sub-regional grouping

Africa

Eastern Africa

- Burundi
- Comoros
- Ethiopia
- Eritrea
- Djibouti
- Kenya
- Madagascar
- Malawi
- Mauritius
- Mozambique
- Réunion
- Rwanda
- Somalia
- Zimbabwe
- Uganda
- Tanzania, United Republic of
- Zambia

Middle Africa

- Angola
- Cameroon
- Central African Republic
- Chad
- Congo
- Congo, Democratic Republic of
- Equatorial Guinea
- Gabon
- Sao Tome and Principe

Northern Africa

- Algeria
- Libyan Arab Jamahiriya
- Morocco
- Western Sahara
- Sudan
- Tunisia
- Egypt

Northern America

- Canada
- United States

Latin America and the Caribbean

Caribbean

- Bahamas
- Barbados
- Cuba
- Dominican Republic
- Grenada
- Guadeloupe
- Haiti
- Jamaica
- Martinique
- Netherlands Antilles
- Aruba
- Puerto Rico
- Saint Lucia
- Saint Vincent and the Grenadines
- Trinidad and Tobago
- Virgin Islands (US)

Central America

- Belize
- Costa Rica
- El Salvador
- Guatemala
- Honduras
- Mexico
- Nicaragua
- Panama

South America

- Argentina
- Bolivia
- Brazil
- Chile
- Colombia
- Ecuador

Asia

Eastern Asia

- China
- Hong Kong, China
- Japan
- Korea, Dem. People's Rep. of
- Korea, Republic of
- Macau, China
- Mongolia

South-Central Asia

- Afghanistan
- Bangladesh
- Bhutan
- Sri Lanka
- India
- Iran, Islamic Rep. of
- Kazakhstan
- Kyrgyzstan
- Maldives
- Nepal
- Pakistan
- Tajikistan
- Turkmenistan
- Uzbekistan

South-Eastern Asia

- Brunei Darussalam
- Myanmar
- Cambodia
- Indonesia
- Lao People's Dem. Rep.
- Malaysia
- Philippines
- Timor-Leste
- Singapore
- Viet Nam
- Thailand

Europe

Eastern Europe

- Bulgaria
- Belarus
- Czech Republic
- Hungary
- Moldova, Republic of
- Poland
- Romania
- Russian Federation
- Slovakia
- Ukraine

Northern Europe

- Denmark
- Estonia
- Finland
- Iceland
- Ireland
- Latvia
- Lithuania
- Norway
- Sweden
- United Kingdom
- Channel Islands

Southern Europe

- Albania
- Bosnia and Herzegovina
- Croatia
- Greece
- Italy
- Malta
- Montenegro
- Portugal
- Serbia and Montenegro
- Slovenia
- Spain
- Macedonia, The former Yugoslav Rep. of

Oceania

Autralia-New Zealand

- Australia
- New Zealand

Melanesia

- Solomon Islands
- Fiji
- New Caledonia
- Vanuatu
- Papua New Guinea

Micronesia

- Guam

Polynesia

- French Polynesia
- Tonga
- Samoa

Southern Africa		Western Asia	Western Europe	
Southern Africa • Botswana • Lesotho • Namibia • South Africa • Swaziland **Western Africa** • Cape Verde • Benin • Gambia • Ghana • Guinea • Côte d'Ivoire • Liberia • Mali • Mauritania • Niger • Nigeria • Guinea-Bissau • Senegal • Sierra Leone • Togo • Burkina Faso	• French Guiana • Guyana • Paraguay • Peru • Suriname • Uruguay • Venezuela, Bolivarian Rep. of	**Western Asia** • Azerbaijan • Bahrain • Armenia • Cyprus • Georgia • West Bank and Gaza Strip • Iraq • Israel • Jordan • Kuwait • Lebanon • Oman • Qatar • Saudi Arabia • Syrian Arab Republic • United Arab Emirates • Turkey • Yemen, Republic of	**Western Europe** • Austria • Belgium • France • Germany • Luxembourg • Netherlands • Switzerland	

Less developed regions

The less developed regions comprise all regions of Africa, Asia (excluding Japan), Latin America and the Caribbean plus Melanesia, Micronesia and Polynesia.

More developed regions

The more developed regions comprise all regions of Europe plus Northern America, Australia/New Zealand and Japan.

Annex 2: Country files

see http://laborsta.ilo.org