EMPLOYMENT PAPER
2002/36

World and regional estimates for selected key indicators of the labour market

Wesley Schaible
Ramya Mahadevan-Vijaya

Employment Sector
INTERNATIONAL LABOUR OFFICE GENEVA

Foreword

The International Labour Office is often called upon to supply global and regional estimates of labour force indicators. Global and regional estimation would be a straightforward exercise except that at present some countries are able to provide the required data whereas others are not. The problem becomes, therefore, a question of what methodology to use to estimate missing values. The issue was taken up by the ILO Key Indicators of the Labour Market (KILM) programme in 1999 with Wes Schaible, author of the current document, commissioned to study various methods of producing world and regional estimates with a mind toward recommending the best methodology to be used to produce global and regional estimates for selected key indicators of the labour market. The resulting paper was published as Employment Paper 2000/6.

The current paper is the second phase of the ILO KILM programme's efforts. After touching briefly on the totality of methods considered, the paper presents global and regional estimates for four key indicators – labour force participation rate, employment-to-population ratio, unemployment rate and youth unemployment rate – produced based on the initial study's recommendation for use of a post stratified approach, using sub-regions as the poststrata within regions.

The development of a methodology for global and regional estimation is an important matter on the current agenda of the ILO, and we hope that this paper will contribute to reaching consensus for the production of basic global estimates that meet the information demands of our constituents. The KILM programme will continue to pursue innovative means of strengthening its existing database, expanding country coverage to the greatest degree possible, so as to minimize the difficulties associated with missing value estimation.

Rashid Amjad
Director a.i.
Employment Strategy Department

Table of contents

Foreword

1. Introduction and definition of the problem 1

2. Methods considered 3

2.1 Mean of the responding units 3

2.2 Post-stratification from the responding population to the total population 4

2.3 Imputation with a regression model 4

2.4 Application to key labour market indicators 5

3. Evaluation methods, results and recommendations 7

3.1 KILM 1: Labour force participation rate 8

3.2 KILM 8: Unemployment rate 10

3.3 KILM 9: Youth unemployment rate 12

3.4 KILM 2: Employment-to-population ratio 13

3.5 Consistency of total employment, unemployment and labour force estimates 15

4. The ILO Projections of the economically active population 16

5. World and regional estimates 18

Appendix 22

1. Introduction and definition of the problem

Estimates of world and regional labour force, employment and unemployment can serve a number of useful purposes. First, they can provide a comprehensive view of the labour force situation in the world and its major geographical regions. Second, the evolution of employment and unemployment over time may be analysed in conjunction with other economic and social variables to draw conclusions on future trends in various aspects of the global economy, and of regional and national labour markets. Third, they can provide global benchmarks against which the economic and labour market performance of individual economies may be compared. This aspect should not be underestimated, as economies are increasingly in competition for foreign investment and local production of global goods, and information on an economy's relative performance plays a significant role in many trade and financial decisions. Finally, world and regional estimates often receive high visibility in the media and may, therefore, affect public policy. The single statement: "160 million persons in the world are unemployed," is sufficient to draw significant global attention to the issue of unemployment and the world of work.

The International Labour Organisation's Key Indicators of the Labour Market (KILM) Team has been involved in developing procedure for producing world and regional estimates for selected indicators for more than three years. The indicators of most interest are the labour force participation rate, employment-to-population ratio, the unemployment rate and the youth unemployment rate. The KILM project also plans to produce these estimates within age groups and for males and females (see Appendix). The fact that a substantial number of countries do not provide the required information, however, presents a dilemma. Previous research within the ILO has treated the production of world and regional estimates as a small area or indirect estimation problem.[1] The research described in this paper addresses the problem as a standard missing data problem.

For these indicators, this report identifies and evaluates several methods for making estimates in the presence of missing data. For each indicator, the evaluation leads to a recommended method. Computer programmes that produce regional and world estimates along with related statistics to help evaluate the estimates are provided for each recommended method.

The following tables indicate the extent of the missing data problem for the indicators under consideration. The un-weighted response rate, which represents the percentage of countries that report the data, indicates that all the regions have some degree of non-response. Therefore, a simple sum of the responding country values would provide an undervalued regional total, and methods to estimate the missing data have to be adopted.

For the indicators of interest here, the impact of missing data depends to a large extent on the size or population of the non-responding countries. Non-response by countries with large populations leads to a greater proportion of a total being missing than non-response by countries with smaller populations. The tables, therefore, also provide a population-weighted

[1] F. Mehran: "Estimation of world and regional unemployment", *ILO Bulletin of Labour Statistics*, 1999-4 (Geneva, 1999).

response rate. The weighted response rates are calculated by dividing the population of the responding countries by the total population of all the countries.

The weighted response rates for all the indicators are greater than the un-weighted response rates. This indicates that the available data represents a larger proportion of the real total than might be indicated by the un-weighted rates.

KILM 1. Labour force participation rate

	Number of countries	Un-weighted response rate	Weighted response rate
Developed (industrialized) economies	34	82.35	99.97
Transition economies	30	90.00	99.96
Asia and the Pacific	39	71.79	99.96
Latin America and the Caribbean	46	67.39	99.85
Sub-Saharan Africa	49	89.80	94.63
North Africa and the Middle East	21	95.24	99.81

Note: The research carried out on labour force included estimates from the LABPROJ file which contributed to the higher response rates.

KILM 2. Employment-to-population ratio

	Number of countries	Un-weighted response rate	Weighted response rate
Developed (industrialized) economies	34	79.41	99.93
Transition economies	30	50.00	83.36
Asia and the Pacific	39	25.64	57.24
Latin America and the Caribbean	46	32.61	40.70
Sub-Saharan Africa	49	0	0
North Africa and the Middle East	21	9.52	28.92

KILM 8. Unemployment rate

	Number of countries	Un-weighted response rate	Weighted response rate
Developed (industrialized) economies	34	94.12	99.90
Transition economies	30	70.00	92.62
Asia and the Pacific	39	33.33	82.67
Latin America and the Caribbean	46	52.17	54.15
Sub-Saharan Africa	49	14.29	10.25
North Africa and the Middle East	21	23.81	40.28

KILM 9. Youth unemployment rate

	Number of countries	Un-weighted response rate	Weighted response rate
Developed (industrialized) economies	34	82.35	99.97
Transition economies	30	50.00	79.41
Asia and the Pacific	39	23.08	11.07
Latin America and the Caribbean	46	39.13	73.89
Sub-Saharan Africa	49	2.04	0.20
North Africa and the Middle East	21	9.52	20.12

2. Methods considered

2.1 Mean of the responding units

Given a population census with missing data, use of the mean of the responding units as an estimate of the population mean would seem to be a risky, but possible, way to proceed. With complete response, we would calculate the population total, T, as $T = \sum^{N} Y_i$.

But in the presence of non-response, we must, either implicitly or explicitly, estimate the total of the non-responding units if we are to arrive at a value for the total. That is, if R is the number of units that respond and, $\tilde{R}$, the number of units that do not respond (R + $\tilde{R}$ = N) then we may write the estimated population total as $\hat{T} = \sum^{R} Y_i + \sum^{\tilde{R}} \hat{Y}_i$. When the sample mean of the observed units is used to impute for each of the unobserved units, we have

$$\hat{T} = \sum^{R} Y_i + \sum^{\tilde{R}} \left(\sum^{R} Y_i \Big/ R \right),$$

When written in this form it is clear that the variable of interest associated with each unobserved unit is estimated by the sample mean of the observed units. It should be noted that this estimator can also be written as $\hat{T} = N\left(\sum^{R} Y_i \Big/ R \right)$ and the estimator of the population mean as $\hat{T}/N = \sum^{R} Y_i \Big/ R$. When written in this form the fact that the mean of the responding units is being used to impute for the non-responding units is disguised.

For the responding mean not to be misleading as an estimator of the population mean, the average value of the variable of interest for the non-responding units must be approximately the same as that for the responding units. More precisely, under a model-based theory, the incomplete sample mean is unbiased as an estimator of the population mean under the model which specifies that, for each unit in the population, the expected value of the random variable representing the unobserved variable of interest is a constant. That is, $EY_i = m$, for i = 1,2…N.

The assumption that the expected value of the variable of interest for the non-responding units is equal to that of the responding units is specific to each variable of interest and is rarely verifiable. This is not an assumption to be made lightly and is one that a survey practitioner would rarely make unless there is simply no other option.

2.2 Post-stratification from the responding population to the total population

In post-stratification an auxiliary variable correlated to the variable of interest is used to create post-strata. Within each post-strata, the mean of the responding units is used to impute for each of the non-responding units. A post-stratified estimator of the population total that uses post-strata denoted by h = 1, 2, . . ., H, may be written as

$$\hat{T}_p = \sum_{h=1}^{H}\left(\sum^{R_h} Y_{hi} + \sum^{\tilde{R}_h}\left(\sum^{R_h} Y_{hi} \Big/ R_h\right)\right).$$

This estimator can also be written in another form that leads to the post-strata being more commonly called weighting or adjustment cells.

$$\hat{T}_p = \sum_{h=1}^{H}\left(N_h\left(\sum^{R_h} Y_{hi} \Big/ R_h\right)\right).$$

Even though post-stratification is a technique designed to reduce the variance of an estimator, there is no guarantee that it will do so. Similarly, even though the non-response adjustment using weighting cells is designed to reduce non-response bias, there is absolutely no guarantee that it will achieve this purpose. As above, a reduction, if any, depends on the relationship between the variable of interest and the variable(s) used to create post-strata or weighting cells.

Under model-based theory, $\hat{T}_p \big/ N$ is an unbiased estimator of the population mean under a model which specifies that, for each population unit within each post-stratum, the expected value of the random variable representing the unobserved variable of interest is a constant, i.e., for post-stratum j, $EY_{ji} = m_j$, for i = 1,2… N_j. Of course, if the variable used to create the post-strata is related to Y, this assumption is somewhat more palatable than the one needed for the mean of the responding units estimator described in 2.1 to be unbiased.

Even though the validity of the model needed for this method to be unbiased cannot always be tested in practice, it should be noted that this is a common method of non-response adjustment used in population based surveys.

2.3 Imputation with a regression model

Regression imputation is possible when, like the weighting cell adjustment, data correlated with the variables of interest are available on the complete sample. However, instead of using this information to create cells and impute the cell mean for missing values, the data are used in a regression model. The incomplete sample data (dependent and independent variables) are used to estimate the parameters of the model and then the missing value (dependent variable) for each unit in the incomplete sample is predicted using the independent variables known for the unit. Although it is not necessary in most applications, for convenience in the application considered in this paper assume that the auxiliary variables are

available for all non-observed units in the population, so the estimator of a population total with regression imputation may be written as follows:

$$\hat{T}_r = \sum^{R} Y_i + \sum^{\tilde{R}} x_i \hat{b} ,$$

where x_i is the row vector of known auxiliary variables and $\hat{b}_s$ is the column vector of the usual best linear unbiased estimators of regression model parameters. As in the estimators above, the sum, $\sum^{R} Y_i$, is known so that the problem of estimating $\hat{T}_r$ is simply one of estimating the sum of the Y values for the non-responding units. In the regression estimator this is accomplished by using the sum (over the non-responding units) of the predicted values from the regression equation. For the estimates produced in this paper, the model parameter was estimated under the assumption of constant variance of the error term.

This estimator that uses regression imputation to represent the unobserved values is model unbiased under the regression model used for imputation. In addition, the variance of this estimator is readily available under the model assumptions. However, there is often little or no evidence outside the sample data that the model is valid so that caution must be exercised when interpreting estimates subject to non-response and also if error measures are to be produced.

2.4 Application to key labour market indicators

Each of the three methods discussed above were used to generate world and regional estimates for indicators (KILM Nos. 1, 2, 8 and 9), using the data for 1995 as reported on the KILM (1999) CD-ROM.[2] For the post-stratification method, countries within each region were grouped into levels or strata-based on auxiliary variables. Four different auxiliary variables were identified and used to create different sets of post-strata. Based on the criteria of data availability for a majority of the countries and correlation with the KILM variables, population, Gross Domestic Product (GDP per capita at purchasing power parity) and the Human Development Index (HDI) were the three independent auxiliary variables identified. The fourth set of post-strata was created on the basis of the sub-region classification of countries within each region that the ILO currently uses.

The auxiliary variables GDP, HDI and population generated two sets of post strata each – one where countries were grouped into three levels and the other where the countries were grouped into four levels. Along with the sub-region stratification, there were seven sets of post-strata. They were used to generate seven post-stratified estimates for each indicator.

[2] ILO: *Key Indicators of the Labour Market 1999*, CD-ROM (Geneva, 1999). The 2001-2002 edition is currently available.

Two regression estimates were generated using population and GDP as the explanatory variables respectively. In sum, the following 10 specific estimates were generated for the indicators (Nos.1, 2, 8 and 9):

1. Imputing the mean of the responding countries within each region

Post-stratified estimators:

2. Using sub-region to create post-strata
3. Using per-capita GDP level to create three post-strata
   Stratum 1: countries with GDP less than 3 000
   Stratum 2: countries with GDP greater than or equal to 3 000 and less than 10 000
   Stratum 3: countries with GDP greater than or equal to 10 000
4. Using per-capita GDP level to create four post-strata
   Stratum 1: countries with GDP less than 3 000
   Stratum 2: countries with GDP greater than or equal to 3 000 but less than 6 000
   Stratum 3: countries with GDP greater than or equal to 6 000 but less than 100 000
   Stratum 4: countries with GDP greater than or equal to 100 000
5. Using HDI to create three post-strata
   Stratum 1: countries with HDI values less than .5
   Stratum 2: countries with HDI values greater than or equal to .5 but less than .9
   Stratum 3: countries with HDI values greater than or equal to .9
6. Using HDI to create four post-strata
   Stratum 1: countries with HDI values less than .5
   Stratum 2: countries with HDI values greater than or equal to .5 but less than .7
   Stratum 3: countries with HDI values greater than .7 but less than .9
   Stratum 4: countries with HDI values greater than or equal to .9
7. Using population levels to create three post-strata
   Stratum 1: countries with population values less than 7 000
   Stratum 2: countries with population greater than or equal to 7 000 but less than 100 000
   Stratum 3: countries with population greater than or equal to 100 000
8. Using population levels to create four post-strata
   Stratum 1: countries with population values less than 7 000
   Stratum 2: countries with population greater than or equal to 7 000 but less than 40 000
   Stratum 3: countries with population greater than or equal to 40 000 but less than 100 000
   Stratum 4: countries with population greater than or equal to 100 000

Regression estimators:

9. Using GDP as the independent variable
10. Using population as the independent variable

The estimation procedure used for each method is briefly described below:

Post-stratified estimators: For the post-stratified estimators, the countries within each region are grouped into strata based on the auxiliary variables. The mean value of the KILM variable for each stratum within a region is calculated by averaging the values for the responding countries in that stratum. Similarly, the means for the *region* is also calculated by aggregating the responding countries in all the strata within the region.

For each stratum a population weighted response rate is also calculated. This weighted response rate is the ratio of the population of the responding countries to the population of all

the countries in the strata. The weighted response rate is used as a quality indicator in order to avoid imputation based on extremely low post-strata response rates.

If the weighted response rate for the stratum is greater than 10 per cent, the average value for the strata is substituted for missing country data. If however, the weighted response rate for the stratum is less than 10 per cent, the *regional* mean is substituted for the missing country value. Similarly if the stratum has zero response, the *regional* mean is used to impute the missing values. For countries that do not have data for the auxiliary variables and therefore cannot be grouped into strata, the regional means are used for imputing missing values. Once the missing data have been imputed the country values are aggregated to arrive at the regional totals. The regional totals are aggregated to calculate the world estimate.

Regression estimators: Data for the non-responding countries are imputed using the predicted values from regressions using the two auxiliary variables as independent variables. Linear regressions through the origin are estimated for each region using data only for the responding countries. However in order to avoid unreliable parameter estimates, we have arbitrarily decided not to estimate a regional parameter if the number of responding countries in the region is less than five. Instead, the estimated parameter of a world regression model is used to calculate the predicted values and impute missing values for that region.

After imputing the missing KILM values for countries within each region, the country values are then aggregated to arrive at regional totals. The regional totals are added up to generate the world estimate.

Some of the non-responding countries also do not have data for the auxiliary variables, GDP and population. It is therefore not possible to generate a predicted value based on the regression coefficients. In the GDP-regression model, the regional mean value of the indicator is used to impute data for the non-responding countries with missing GDP data. That is, the mean imputation method is used in combination with the regression method. Resorting to the mean imputation method can be avoided for the population-regression model since it is possible to get an estimate for the missing population values using United Nations data.

The United Nations provides population estimates for all countries with a population greater than 200 000 and also the regional total (UN region) for all the countries in the region. By subtracting the sum of the population values of countries with populations over 200 000 from the regional population total, the sum of the population for the countries with populations less than 200 000 can be obtained for each United Nations region. The United Nations regional figures are then added to obtain a world figure and the average population size across the countries with populations less than 200 000 is calculated. This average population size is used to impute for the missing population figure in countries with population less than 200 000.

3. Evaluation methods, results and recommendations

For each of the indicators under consideration 10 different estimates were generated using the 10 different estimators. A comparison of the actual estimates is generally not useful since there is no true population value for comparison. However in this application it is possible to establish an upper and a lower bound for the estimates. The sum of the indicator

values of the responding countries forms a lower bound. Since this total excludes all the missing countries, any imputation method should produce a larger estimate.

In a given country, the total employment, unemployment and labour force are all, by definition, less that the population of the country. So an upper bound can be formed by using the population value instead of the missing employment, unemployment or labour force value for non-responding countries. However, for the purposes of this research, a different, approximate upper bound was used instead. The upper bound is created by the fact that the response rate for all the indicators is higher among countries with larger populations. Most of the non-response occurs in countries with small populations. The mean value of all the responding units therefore has an upward bias. Imputing the missing values for the smaller non-responding countries with this inflated mean would produce an inflated estimate. The mean imputation estimate therefore can be used as an approximate upper bound for comparison with estimates from other estimators. That is, an estimator with reasonably small non-response bias should produce estimates that are smaller than the mean imputation estimates.

Comparing the other estimates against this upper and lower bound is one method of evaluating the estimators. A second method of evaluation is based on the r-square statistic. The r-square is a measure of how well a regression line fits the data in a plot of the variable of interest and an independent variable. If the regression using an independent variable fits the data well, the post-strata created with that independent variable can be expected to produce relatively homogeneous groups. The estimates for each of the indicators are evaluated separately:

3.1 KILM 1: Labour force participation rate

The labour force participation rate is defined as the total labour force divided by the working age population. For each responding country the total labour force and the labour force participation rate is provided. For the non-responding countries the total labour force is imputed using the various estimators. The labour force is then divided by working age population to generate the participation rate.

Since most non-responding countries do not have working age population figures, these also have to be imputed. However, since working age population by itself is not an indicator variable and since it is very highly correlated with the total population, it is only imputed once using a regression model with population as the dependent variable. The working age populations are then used with all the different estimates of total labour force to calculate the participation rate.

Response rates by population group

Population (millions)	Response	Non-response	Total	Response rate
Less than 7	88	40	128	69
Greater than or equal to 7, less than 20	45	0	45	100
Greater than or equal to 20, less than 100	36	1	37	97
Greater than 100	9	1	10	90

The response table indicates that the response rates for KILM 1 are higher for countries with larger populations. The mean imputation estimates can therefore be used as an upper bound for the total labour force. The following table provides the upper and lower bounds for regional total labour force.

	Lower bound - total of responders	Upper bound - mean imputation
Developed (industrialized) economies	454,433	551,812
Transition economies	200,191	222,435
Asia and the Pacific	1,484,896	2,068,248
Latin America and the Caribbean	185,960	275,940
Sub-Saharan Africa	223,397	248,783
North Africa and the Middle East	99,165	104,124

The results from the other estimators can be judged on the basis of their distance from the two boundaries. More specifically, a better estimator can be expected to lie closer to the lower bound than the upper bound since most of the missing values belong to relatively small countries. The results from the other estimators are listed below.

Post-stratification estimates

	Sub-region	GDP		HDI		Population	
		3 levels	4 levels	3 levels	4 levels	3 levels	4 levels
Developed (industrialized) economies	460,013	551,812	551,812	551,812	551,812	479,160	479,160
Transition economies	214,534	222,435	222,435	222,435	222,435	205,008	205,008
Asia and the Pacific	2,108,585	2,068,248	2,068,248	2,082,851	2,140,207	1,497,947	1,497,947
Latin America and the Caribbean	211,581	297,040	297,767	284,176	294,583	198,333	198,333
Sub-Saharan Africa	102,875	248,783	248,783	246,434	244,829	237,065	234,205
North Africa and the Middle East	102,662	104,124	104,124	106,005	106,005	99,953	99,953

Regression estimates

	GDP	Population
Developed (industrialized) economies	551,812	454,576
Transition economies	222,435	200,266
Asia and the Pacific	2,068,248	1,485,452
Latin America and the Caribbean	221,474	186,266
Sub-Saharan Africa	248,783	235,853
North Africa and the Middle East	104,124	99,351

The estimates for most of the regions from the GDP post-stratification estimator are the same as the mean imputation estimates. This is due to the fact that all the non-responding countries in these regions are also missing GDP data. Therefore as explained in the procedure section, the regional mean is used for imputation. This might suggest that more complete GDP data might improve the GDP post-stratification estimates considerably.

However, given the population bias in the response rate and the lack of correlation between GDP and population, the GDP estimates can be expected to be inflated even with complete GDP data. This does prove to be the case with the HDI estimates, which also suffer from the similar lack of correlation between HDI and population. Moreover the low r-square values for GDP also indicate that it may not be the best auxiliary variable.

R-squares

	GDP	Population
Developed (industrialized) economies	0.30	0.95
Transition economies	0.16	0.99
Asia and the Pacific	0.01	0.97
Latin America and the Caribbean	0.20	0.98
Sub-Saharan Africa	0.09	0.99
North Africa and the Middle East	0.12	0.98

The higher r-squares indicate that population is a better choice than GDP for the independent variable. The estimates from the population regression are closer to the lower bound than the estimates from the population post-stratification. The population post-stratification estimator groups together all the countries with small populations. Given the high incidence of non-response amongst the small population group, the imputations are based on limited data and may therefore be unreliable. In fact some strata may have no response and in that case, the regional mean will be used, once again leading to inflated estimates.

The high r-square and the population bias in the response rate point to the superiority of the population regression estimator over the post-stratification population estimator. The population regression estimator is recommended for KILM 1.

3.2 KILM 8: Unemployment rate

The unemployment rate is defined as unemployed divided by the labour force. For each responding country the total unemployment and the unemployment rate is provided. For the non-responding countries the total unemployment is imputed using the various estimators, and is then divided by the labour force to calculate the unemployment rate.

Since most non-responding countries also do not have labour force figures, these also have to be imputed. However since total labour force by itself is not a KILM variable and since it is very highly correlated with the total population, it is only imputed once using a regression model with population as the independent variable. The labour force values are then used with all the different estimates of total unemployment to calculate the unemployment rate.

The following tables of results for KILM 8 estimators indicate the same patterns as the other indicators. The response rate is higher for larger population groups, leading to larger post-stratification estimates. The population r-square is also high for most regions. The population regression estimator is therefore recommended for KILM 8 as well.

Response rates by population group

Population (millions)	Response	Non-response	Total	Response rate
Less than 7	52	76	128	41
Greater than or equal to 7, less than 20	21	24	45	47
Greater than or equal to, 20 less than 100	23	14	37	62
Greater than 100	6	4	10	60

	Lower bound - total of responders	Upper bound - mean imputation
Developed (industrialized) economies	33,648	35,751
Transition economies	14,049	20,070
Asia and the Pacific	48,280	144,842
Latin America and the Caribbean	7,384	14,153
Sub-Saharan Africa	427	2,989
North Africa and the Middle East	5,328	22,379

Post-stratification estimators

	Sub-region	GDP		HDI		Population	
		3 levels	4 levels	3 levels	4 levels	3 levels	4 levels
Developed (industrialized) economies	33,725	35,751	35,751	34,478	34,478	34,083	34,083
Transition economies	20,361	18,867	18,867	16,725	15,036	15,783	15,693
Asia and the Pacific	155,724	196,151	197,758	173,507	188,222	83,632	82,746
Latin America and the Caribbean	11,360	15,266	14,612	11,345	11,854	12,135	11,459
Sub-Saharan Africa	3,358	3,221	3,067	1,211	873	3,853	3,622
North Africa and the Middle East	22,910	21,528	21,388	14,919	12,228	24,235	23,649

Regression estimators

	GDP	Population
Developed (industrialized) economies	3,364	33,651
Transition economies	14,727	15,300
Asia and the Pacific	49,448	57,332
Latin America and the Caribbean	10,189	12,800
Sub-Saharan Africa	1,316	3,714
North Africa and the Middle East	29,976	12,521

R-squares

	GDP	Population
Developed (industrialized) economies	0.36	0.84
Transition economies	0.15	0.95
Asia and the Pacific	0.00	0.51
Latin America and the Caribbean	0.35	0.80
Sub-Saharan Africa	0.60	0.79
North Africa and the Middle East	0.13	0.86

3.3 KILM 9: Youth unemployment rate

The youth unemployment rate is defined as youth unemployed divided by the total youth labour force. For each responding country the total youth unemployment and the youth unemployment rate is provided. For the non-responding countries the total youth unemployment is imputed using the various estimators. This is then divided by the youth labour force to calculate the unemployment rate.

Most non-responding countries do not have youth labour force figures and these have to be imputed. Similar to the total labour force in KILM 8, the youth labour force is only imputed once using a regression model with population as the independent variable. The youth labour force values are then used with all the different estimates of total youth unemployment to calculate the youth unemployment rate.

Response rates by population group

Population (millions)	Response	Non-response	Total	Response rate
Less than 7	36	92	128	28
Greater than or equal to 7, less than 20	15	30	45	33
Greater than or equal to 20, less than 100	17	20	37	46
Greater than 100	5	5	10	50

	Lower bound – total of responders	Upper bound – mean imputation
Developed (industrialized) economies	9,489	11,523
Transition economies	4,101	8,201
Asia and the Pacific	2,725	11,809
Latin America and the Caribbean	5,014	12,812
Sub-Saharan Africa	25	1,210
North Africa and the Middle East	1,315	13,810

Post-stratification estimators

	GDP		HDI		Population	
	3 levels	4 levels	3 levels	4 levels	3 levels	4 levels
Developed (industrialized) economies	11,523	11,523	11,523	11,523	9,663	9,663
Transition economies	8,137	8,255	8,201	7,387	5,763	5,464
Asia and the Pacific	13,620	13,620	15,786	14,779	7,694	7,679
Latin America and the Caribbean	12,867	12,893	13,479	12,360	9,181	8,035
Sub-Saharan Africa	1,210	1,210	1,210	1,210	1,210	1,210
North Africa and the Middle East	13,156	13,156	13,810	14,464	20,348	17079

Regression estimators

	GDP	Population
Developed (industrialized) economies	11,523	9,492
Transition economies	6,479	5,132
Asia and the Pacific	8,137	21,765
Latin America and the Caribbean	11,494	6,649
Sub-Saharan Africa	1,665	5,798
North Africa and the Middle East	7,522	3,888

R-squares

	GDP	Population
Developed (industrialized) economies	0.30	0.86
Transition economies	0.17	0.97
Asia and the Pacific	0.03	0.71
Latin America and the Caribbean	0.22	0.96
Sub-Saharan Africa	…	…
North Africa and the Middle East	0.03	1.00

… = Not available

The overall response rate for KILM 9 is lower than the response rates for the KILM Nos. 1 and 8. The number of responding countries for sub-Saharan Africa is less than three and therefore a regional regression is not generated and the r-square is not reported. However the population-response table indicates that the pattern of higher response rates from countries with larger populations continues. This leads to a similar interpretation of the results as in the case of KILM Nos. 1 and 8. The GDP and HDI post-stratification estimators generate inflated estimates. Population once again is the better auxiliary variable.

Given the relatively high population r-squares, the population regression method is preferable to the post-stratification method. The population regression estimator is therefore also recommended for KILM 9.

3.4 KILM 2: Employment-to-population ratio

The employment-population rate is defined as employed divided by the working age population. For the non-responding countries the total employment is imputed using the various estimators. This is then divided by the working age population to calculate the employment-population rate.

Since most non-responding countries do not have working age population figures, these also have to be imputed. The working age population is only imputed once using a regression model with population as the independent variable. The working age population values are then used with all the different estimates of total employment to calculate the employment-population rate.

The overall response rate for KILM 2 is the lowest of all the set of indicators considered in this paper. There is, in fact, no response for sub-Saharan Africa. Post-stratification estimates are not generated for this region, since neither a stratum nor a regional mean can be calculated. The regression estimators do produce an estimate for sub-Saharan Africa by using the coefficient from a world regression equation.

Response rates by population group

Population (millions)	Response	Non-response	Total	Response rate
Less than 7	30	98	128	23
Greater than or equal to 7, less than 20	14	31	45	31
Greater than or equal to 20, less than 100	19	18	37	51
Greater than 100	6	4	10	60

The population response table indicates that besides having the lowest overall response rate, KILM 2 also has the lowest response rate for the small population group. This magnifies the tendency of the regional and stratum means to be inflated. The post-stratification estimates for GDP and HDI reported below do reflect this inflation. The population r-square once again confirms that the population regression model is the most suitable estimator.

	Lower bound - total of responders	Upper bound - mean imputation
Developed (industrialized) economies	420,410	529,402
Transition economies	146,209	292,416
Asia and the Pacific	88,727	3,460,334
Latin America and the Caribbean	66,438	203,744
Sub-Saharan Africa	…	…
North Africa and the Middle East	20,272	212,858

…= Not available

Post-stratification estimators

	Sub-region	GDP		HDI		Population	
		3 levels	4 levels	3 levels	4 levels	3 levels	4 levels
Developed (industrialized) economies	425,763	529,259	529,259	522,167	522,167	449,592	449,592
Transition economies	269,386	306,529	306,529	292,416	289,822	191,140	175,540
Asia and the Pacific	2,754,033	3,285,428	3,285,428	3,536,765	3,658,868	1,692,433	1,880,025
Latin America and the Caribbean	158,567	227,395	218,751	219,772	214,106	171,629	142,836
Sub-Saharan Africa	…	…	…	…	…	…	….
North Africa and the Middle East	212,858	176,401	186,818	212,858	160,777	212,858	212,858

…= Not available

Regression estimators

	GDP	Population
Developed (industrialized) economies	524,920	420,714
Transition economies	250,390	176,560
Asia and the Pacific	2,467,813	1,601,339
Latin America and the Caribbean	182,297	165,236
Sub-Saharan Africa	81,741	302,451
North Africa and the Middle East	184,680	139,958

R-squares

	GDP	Population
Developed (industrialized) economies	0.20	0.95
Transition economies	0.10	0.99
Asia and the Pacific	0.01	0.99
Latin America and the Caribbean	0.27	0.99
Sub-Saharan Africa	…	…
North Africa and the Middle East	0.55	0.98

…= Not available

3.5 Consistency of total employment, unemployment and labour force estimates

The total labour force in an economy is defined as the sum of total employment and total unemployment. Of course, it is desirable for regional and world estimates of these three variables to have this same relationship. However, several issues needed to be addressed in the production of the world and regional estimates so that the three estimates would be consistent in this sense.

Economies reporting these three estimates to the International Labour Organization do not always obtain the estimates from the same source and, therefore, the labour force figure does not always equal the sum of the other two. The ILO Bureau of Statistics and information repositories used by KILM obtain information from national sources such as household surveys, establishment surveys, administrative records, etc. Priorities were set based on information on the type of source and selections were made so that in the estimation process an economy's response was used only when information on all three variables was available from the same type of source.

On some occasions economies provided only two of the three variables. In these cases, the missing variable was calculated as the appropriate sum or difference of the two responses provided. Some economies provided only one of the three variables. It is generally desirable to use all the data in the estimation process; this is especially true when unit response rates are low, as is the case in several regions. However, the use of these data in the regression imputation procedure produces imputed values that are not necessarily consistent with the reported values and also values that are not necessarily consistent with one another in those economies that did not respond to any of the three variables.

These discrepancies were addressed by adjustments motivated by post-regression imputation constraints. In the case of one actual and two imputed values, the inconsistency between the actual and imputed values was proportionally allocated to the imputed values so that the imputed values were adjusted to be consistent with the actual value. In the case of three imputed values, the difference between labour force and the sum of employment and unemployment was proportionally allocated to all three imputed values so that they were each

adjusted to be consistent with one another. The methods used are detailed in an internal report, which is available upon request.[3]

4. The ILO Projections of the economically active population

The ILO Bureau of Statistics has a programme of estimates and projections of the economically active population. This programme produces projections of the economically active population for 178 countries by sex and age groups for selected years during the period 1950–2010.[4] In addition, world and regional projections are also provided.[5] This programme has now reached its fourth edition, the first edition being published in 1971. The following is an excerpt from Volume 10 (fourth edition) of this series:

> "The purpose of this programme is to provide ILO member States, international agencies and private users with the most comprehensive, detailed and comparable working tool possible at the national, regional and international levels. It also aims to furnish coordinated and compatible demographic data on population, economically active population, agricultural population, school population, households, birth, fertility, mortality rates, etc."

One significant feature of international estimates and projections of the economically active population is that they enable comparison of levels and trends in the active population of various regions and countries over past and future years.

As for the preceding editions, the estimates and projections of active population for the fourth edition were obtained by applying to the population estimates and projections (medium variant) produced by the United Nations.[6] The ILO estimates and projections of activity rates by sex and age groups as of the middle of each reference year, that is, the years 1950, 1960, 1970, 1980, 1990, 2000 and 2010. By comparison with the three preceding editions, the fourth edition introduces innovations and changes in concepts, definitions and methodology.

More detailed descriptions of these methodologies can be found in the above referenced publications. However, the most recently implemented approach for producing country estimates can generally summarized as followed: Country estimates and projections of

[3] To see ILO: *Adjustment to produce consistency between employment, unemployment and labour force totals, ILO world and regional KILM estimates*, (internal paper, Geneva, 2001), send an email request to kilm@ilo.org.

[4] See ILO: *Sources and Methods: Labour statistics, Vol. 10, Estimates and projections of the economically active population, 1950-2010* (Geneva, 2000) and http://www.ilo.org/public/english/bureau/stat/child/actrep/ecacpop.htm.

[5] See ILO: *Economically active population, 1950-2010, Vol. 5: World* (fourth edition, STAT Working Paper No. 1996-5).

[6] UN: *World Population Prospects 1950-2050: the 1996 Revision*, St/ESA/Ser.A/145 (New York).

populations by age and sex are available from the United Nations estimates and projections program. Also, country specific labour force activity rates by age and sex are obtained by application of a detailed set of interpolation and extrapolation models chosen across time for each age and sex group within each country. Missing data issues are generally addressed in one of three ways: use of country experts to provide figures, use of data from neighbouring countries, and use of data from the same country but from an earlier time period. Country estimates (projections) of labour force are then obtained by applying the interpolated or extrapolated activity rate to the population within each age and sex group. Total country estimates are obtained by adding across age and sex groups.

Both the ILO Projections of the economically active population and the KILM team estimations deal with the missing data problem in their respective approaches. The focus of the ILO projections programme is to make country estimates. To deal with the missing data problem, this programme uses estimates from country experts, models that have parameter values obtained from neighbouring countries, and models that have parameter values obtained from different time periods within the same country. The focus of the KILM team effort has been to make regional and world estimates, not country estimates. For it's effort, the KILM team assembled additional estimates from non-responding countries so that the missing data problem within regions was reduced. Furthermore, when dealing with the residual missing data, a model that allows the regional parameter to be specific to the region and the time period of interest was used.

The main differences in these two approaches seem to be in the programme objectives and resulting model parameters. The objective of the ILO Projections of the economically active population - as the name implies - seems to be to produce country estimates and projections for the future. Interpolation and extrapolation models generally have parameters that are not specific to both the domain and time period of interest. However, if the objective is to estimate year-to-year change in the actual country population, the use of such models will generally underestimate the true variability in actual changes. Models of this type are generally appropriate when the objective is to estimate an underlying trend primarily explained by the independent variables in the models.

In contrast, the objective of the KILM team is to produce regional estimates of year-to-year change in the regional population values. Consistent with this objective, the model behind the estimator used by the KILM team incorporates a parameter that is allowed to change and be specific to each time period and domain of interest. The assumption of a model parameter that does not change over time is, of course, more restrictive than the assumption of a model parameter that can change over time. This less restrictive model leads to implementation differences; that is, it results in the estimator using only data from the time period and domain of interest to make estimates for that time period. This approach will generally produce estimates that reflect the actual year-to-year changes in population values better than other, indirect approaches. On the other hand, such models will generally not perform very well when the objective is to produce estimates of an underlying trend. A general discussion of direct and indirect estimators and their advantages and disadvantages can be found in the ILO Employment Paper 2000/6.[7]

[7] W. Schaible: *Methods for producing world and regional estimates for selected key indicators of the labour market*, ILO Employment Paper 2000/6 (Geneva, 2000); website: http://www.ilo.org/public/english/emplyoment/strat/publ/ep00-6.htm

5. World and regional estimates

The results of an application of the procedure described in Section 3 to develop initial world and regional estimates for 1995–2000 are presented in tables 1 to 4. It is important to note that the data set used for testing methodologies was static. The data set used to calculate the world and regional estimates in this section, however, may include newly available data so as to maximize the response rate to the greatest degree possible up to the date of this publication. *Therefore, response rates, r-squares and regression results will differ from the 1995 results presented above.* The regional details are limited to those regions and years that generated acceptable rates of response for the indicators.[8]

Based on the world and regional estimates model, in 1999 the global labour force was 2.68 billion, with the number of unemployed reaching 160 million, corresponding to a global unemployment rate of 6 per cent. Between 1995 and 2000, the unemployment rate for the developed (industrialized) grouping hovered close to 7 per cent. In 1999 the highest unemployment rate observed for a region with sufficient response rate was that of the transition economies at 11.2 per cent. Continued levels of unemployment in the double digits for the transition economies probably reflect the deep transformation that these economies continue to undergo in their movement to a market-based economy.

[8] The "acceptable" unweighted response rate was 30 per cent in at least two of the three variables (labour force participation, employment and unemployment) when generating regional estimates; for generation of the world estimate, estimates had to exist for at least four regions.

Table 1. Labour force participation rate, 1995-2000

	1995	1996	1997	1998	1999	2000
World labour force (millions)	*2 505*	*2 586*	*2 604*	*2 645*	*2 680*	…
			Percentages			
World	**64.4**	**65.3**	**64.6**	**64.5**	**64.2**	…
Developed (industrialized) economies	60.7	60.7	60.7	60.8	61.1	60.3
Transition economies	59.5	59.2	57.3	57.1	56.9	58.9
Asia and the Pacific	68.8	70.6	69.4	68.9	68.8	…
Latin America and the Caribbean	59.4	59.5	61.8	62.1	62.0	…
Sub-Saharan Africa	63.3	…	…	…	…	…
Middle East and North Africa	48.1	47.0	43.8	…	…	…

… = Not available

Table 2. Employment-to-population ratio, 1995-2000

	1995	1996	1997	1998	1999	2000
			Percentages			
World	**60.8**	**61.7**	**60.9**	**56.6**	**60.4**	…
Developed (industrialized) economies	56.1	56.2	56.4	56.6	57.2	56.5
Transition economies	54.5	53.7	51.5	51.0	50.5	51.9
Asia and the Pacific	66.0	67.8	66.5	65.8	65.7	…
Latin America and the Caribbean	54.9	54.8	56.8	56.8	57.6	…
Sub-Saharan Africa	…	…	…	…	…	…
Middle East and North Africa	…	41.4	38.2	39.5	…	…

The Asia and Pacific region regularly showed the highest level of labour force participation, reaching slightly more than 70 per cent in 1996 before the onset of the Asian financial crisis and retreating to roughly 69 per cent between 1997 and 1999. For the years available, the region of the Middle East and North Africa displayed the lowest labour force participation rates, ranging from 48 per cent in 1995 to 44 per cent in 1997. Latin America and the Caribbean displayed signs of modest but consistent growth in labour force participation between 1995 and 1999 (from 59 to 62 per cent in the period). The transition economies were the only region showing a decrease in labour force participation rates (from slightly more than 59 per cent in 1995 to 57 per cent in 1999 although rising again to 59 per cent in 2000). The developed (industrialized) economies showed slight to no change between 1995 and 1999, holding steady at approximately 61 per cent throughout the period.

Nearly two-thirds of the unemployed persons in the world live in the regions of Asia and the Pacific, the Middle East and North Africa, and Latin American and the Caribbean, a logical result given that the vast majority of the world's labour force resides in these areas. When adjusted for the size of the labour force, the unemployment rates within these regions varied between 4 and 12 per cent by 1999.

Table 3. Unemployment rate, 1995-2000

	1995	1996	1997	1998	1999	2000
World unemployed (millions)	*141*	*145*	*152*	*158*	*160*	…
			Percentages			
World	**5.6**	**5.6**	**5.9**	**6.0**	**6.0**	…
Developed (industrialized) economies	7.4	7.4	7.2	6.8	6.5	6.2
Transition economies	8.5	9.2	10.1	10.7	11.2	11.9
Asia and the Pacific	4.1	3.9	4.2	4.5	4.5	…
Latin America and the Caribbean	7.6	8.0	8.0	8.6	7.1	…
Sub-Saharan Africa	…	…	…	…	…	…
Middle East and North Africa	…	11.9	12.7	…	…	…

…= Not available

In 1999 more than 40 per cent of the unemployed in the world were young women and men. The youth unemployment rate was higher than the adult rate in all regions with available information. Table 4 shows that the incidence of unemployment among young people was generally twice the rate of adult unemployment in the developed (industrialized) economies, the transition economies, Asia and the Pacific, and Latin America and the Caribbean. A higher youth unemployment rate is not surprising, however; it simply reflects the fact that the risk of unemployment is higher at labour market entry points than at any other point. But at twice the adult rate, the youth rate is revealing and points to a need to focus on youth unemployment in national labour market policies.

Table 4. Youth unemployment rate, 1995-2000

	1995	1996	1997	1998	1999	2000
World unemployed youth (millions)	*58*	*61*	*61*	*65*	*66*	…
			Percentages			
World	**10.0**	**10.4**	**9.8**	**10.2**	**10.4**	…
Developed (industrialized) economies	14.6	14.6	14.2	13.4	12.8	10.4
Transition economies	17.0	16.7	17.0	18.3	18.1	…
Asia and the Pacific	9.9	11.0	8.2	10.3	10.4	…
Latin America and the Caribbean	12.1	12.5	12.0	13.0	15.2	…
Sub-Saharan Africa	…	…	…	…	…	…
Middle East and North Africa	…	…	…	…	…	…

…= Not available

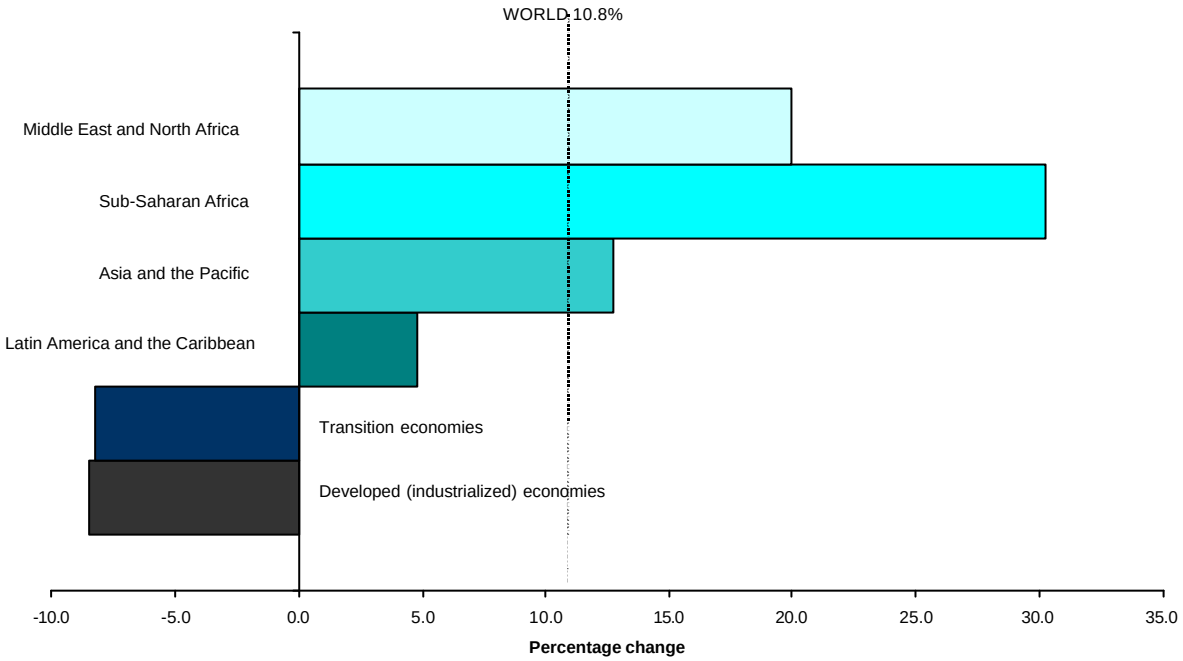
Note: The population-weighted response rate for the Asia and Pacific region was considerably greater than the unweighted response rate; it was therefore decided to publish the estimate for this region despite its unweighted response rate of slightly less than 30 per cent for selected years.

There were marked differences in the youth labour market by region. In developed (industrialized) economies and transition economies that were experiencing higher levels of economic growth, young people represented 13 and 16 per cent of the population respectively, but their share of the total population was declining in the late 1990s. Young people in these countries were not only enjoying the greater employment opportunities available in fast-growing economies, but were also benefiting to a greater degree than earlier generations because there were comparatively fewer young labour market entrants competing for jobs. In contrast, in many developing economies the share of young people in the total population was 20 to 25 per cent and growing. In these economies, young women and men faced much more limited employment opportunities; they also experienced increasing competition for jobs due to the rapid increase in the number of young labour market entrants. Since economic growth is

expected to slow down worldwide in the coming years, while the rapid expansion of the young population is expected to continue in many developing economies (especially in sub-Saharan Africa and the Middle East and North Africa), pressure on the youth labour market is likely to increase in these regions, leading to higher youth unemployment rates (see Figure A).

Figure A. Youth population growth estimates, 2000-10

Figure A. Youth population growth estimates, 2000-10



Source: United Nations: *World population prospects: The 2000 revision* (New York, 2001).

In 2000, the global labour market still appeared to be healthy in terms of employment and unemployment; however, during the course of the first few months of 2001, signs of a global economic slowdown that is likely to affect the global labour market have been witnessed. According to the World Bank, the world GDP growth rate is expected to fall from 4 per cent in 2000 to 2.2 per cent in 2001.[9] Growth in the developed economies may decline from 3.6 to 1.6 per cent while developing and transition economies may slow from 5.4 to 4.2 per cent. Slow growth in world demand and trade will hit the developing economies hardest, in particular those with strong trade ties to developed (industrialized) economies such as the United States, Canada and those in Europe. Any global economic slowdown will certainly have an effect on the level of global unemployment within the world labour market in the coming years.

[9] World Bank: *Global economic slowdown: What's the effect on developing countries*?, short note prepared by the World Bank's Global Development Finance team (Washington, DC, 2001), available at website: http://wbln0018.worldbank.org/eurvp/web.nsf/c8e353e95f273593c12569960057f517/12fa04a62354e4acc1256a380036a93b.

APPENDIX

The following tables provide the r-squares for gender and age groups with population as an auxiliary variable. The r-squares are provided separately for each region except for cases where less than two countries respond and it is not possible to calculate an r-square.

KILM 1. Population r-squares

	Ages 15-24	Ages 25-54	Male	Female
Developed (industrialized) economies	0.96	0.95	0.95	0.95
Transition economies	0.99	0.99	0.99	0.99
Asia and the Pacific	0.97	0.96	0.99	0.91
Latin America and the Caribbean	0.98	0.98	0.99	0.96
Sub-Saharan Africa	0.98	0.99	0.99	0.97
North Africa and the Middle East	0.95	0.92	0.99	0.92

KILM 2. Population r-squares

	Male	Female
Developed (industrialized) economies	0.95	0.94
Transition economies	0.99	0.99
Asia and the Pacific	0.98	0.70
Latin America and the Caribbean	0.99	1.00

KILM 8. Population r-squares

	Male	Female
Developed (industrialized) economies	0.86	0.81
Transition economies	0.95	0.94
Asia and the Pacific	0.98	0.99
Latin America and the Caribbean	0.82	0.73
Sub-Saharan Africa	0.95	0.75
North Africa and the Middle East	0.71	0.98

KILM 9. Population r-squares

	Male	Female
Developed (industrialized) economies	0.87	0.83
Transition economies	0.97	0.97
Asia and the Pacific	0.70	0.56
Latin America and the Caribbean	0.90	0.94
North Africa and the Middle East	1.00	1.00