Statistical methods

This section describes some of the statistical procedures used in preparing the *World Development Indicators*. It covers the methods employed for calculating regional and income group aggregates and for calculating growth rates, and it describes the World Bank's Atlas method for deriving the conversion factor used to estimate GNP and GNP per capita in U.S. dollars. Other statistical procedures and calculations are described in the *About the data* sections that follow each table.

Aggregation rules

Aggregates based on the World Bank's regional and income classification of economies appear at the end of most tables. The World Bank's regional and income classifications are shown on the front- and back cover end flaps of the book. Specialized classifications, such as high-income OECD countries, are documented in *About the data* of the tables in which they appear.

Because of missing data, aggregations for groups of economies should be treated as approximations for unknown totals or average values. Regional and income group aggregates are based on the largest available set of data, including values for the 148 economies shown in the main tables, other economies shown in table 1.6, and Taiwan, China. The aggregation rules are intended to yield estimates for a consistent set of economies from one period to the next and for all indicators. Small differences between the values of subgroup aggregates and overall totals and averages may occur because of the approximations used. In addition, compilation errors and data reporting practices may cause discrepancies in theoretically identical aggregates such as world exports and world imports.

There are four principal methods of aggregation. For group and world totals noted in the tables by a *t*, missing data are imputed using a suitable proxy variable in a benchmark year, usually 1987. The imputed value is calculated so that it (or its proxy) bears the same relationship to the total of available data as it did in the benchmark year. Proxy variables are selected from a set of variables for which complete data are available for 1987. Imputed values are not calculated if missing data account for more than one-third of the total in the benchmark year. The variables used as proxies are GNP in U.S. dollars, GNP per capita in U.S. dollars, total population, exports and imports of goods and services in U.S. dollars, and value added in agriculture, industry, manufacturing, and services in local currency.

Aggregates marked by an *s* are sums of available data. Missing values are not imputed. Sums are not computed if more than one-third of the observations in the series or a proxy for the series are missing in a given year.

Aggregates of ratios are generally calculated as weighted averages of the ratios (indicated by *w*) using the value of the denominator or, in some cases, another indicator as a weight. The aggregate ratios are based on available data, including data for economies not shown in the main tables. Missing values are assumed to have the same average value as the available data. No aggregate is calculated if missing data account for more than one-third of the value of weights in the benchmark year. In a few cases the aggregate ratio may be computed as the ratio of group totals after imputing values for missing data according to the above rules for computing totals.

Aggregate growth rates are also generally calculated as a weighted average (and indicated by a *w*) of growth rates. In a few cases growth rates may be computed from time series of group totals. Growth rates are not calculated if more than one-third of the observations in a period are missing. For further discussion on methods of computing growth rates, see below.

Aggregates noted with an *m* are medians of the values shown in the table. No value is shown if more than one-third of the observations are missing.

Exceptions to the rules occur throughout the book. Depending on the judgment of World Bank analysts, the aggregates may be based on as little as 50 percent of the available data. In other cases, where missing or excluded values are judged to be small, irrelevant, or randomly distributed aggregates are based only on the data shown in the tables.

Growth rates

Growth rates are calculated as annual averages and represented as percentages. Except where noted, growth rates of values are computed from constant-price or real-value series. Three principal methods are used to calculate growth rates: the least-squares method, the exponential endpoint, and the geometric endpoint. Rates of change from one period to the next are calculated as proportional changes from the earlier period

Least-squares growth rate. Least-squares growth rates are used wherever there is a sufficiently long time series to permit a reliable calculation. No growth rate is calculated if more than half of the observations in a period are missing.

The least-squares growth rate, *r*, is estimated by fitting a linear regression trendline to the logarithmic annual values of the variable in the relevant period. The regression equation takes the form

$$\ln X_t = a + bt,$$

which is equivalent to the logarithmic transformation of the compound growth equation,

$$X_t = X_o (1 + r)^t.$$

In this equation, X is the variable, t is time, and $a = \log X_o$ and $b = \ln (1 + r)$ are parameters to be estimated. If b^* is the least-squares estimate of b, then the average annual growth rate, r, is obtained as $[\exp(b^*)-1]$ and is multiplied by 100 to express it as a percentage.

The calculated growth rate is an average rate that is representative of the available observations over the entire period. It does not necessarily match the actual growth rate between any two periods.

Exponential growth rate. The growth rate between two points in time for certain demographic data, notably labor force and population, is calculated from the equation:

$$r = \ln(p_n/p_1)/n,$$

where p_n and p_1 are the last and first observations in the period, n is the number of years in the period, and ln is the natural logarithm operator.

This growth rate is based on a model of continuous, exponential growth between two points in time. It does not take into account the intermediate values of the series. Note also that the exponential growth rate does not correspond to the annual rate of change measured at a one-year interval, which is given by $(p_n - p_{n-1})/p_{n-1}$.

Geometric growth rate. The geometric growth rate is applicable to compound growth over discrete periods, such as the payment and reinvestment of interest or dividends. Although continuous growth, as modeled by the exponential growth rate, may be more realistic, most economic phenomena are measured only at intervals for which the compound growth model is appropriate. The average growth rate over n periods is calculated as

$$r = \exp[\ln(p_n/p_1)/n] - 1.$$

Like the exponential growth rate, it does not take into account intermediate values of the series.

World Bank Atlas method

In calculating GNP in U.S. dollars and GNP per capita for certain operational purposes, the World Bank uses a synthetic exchange rate commonly called the Atlas conversion factor. The purpose of the Atlas conversion factor is to reduce the impact of exchange rate fluctuations in the cross-country comparison of national incomes.

The Atlas conversion factor for any year is the average of a country's exchange rate (or alternative conversion factor) for that year and its exchange rates for the two preceding years, after adjusting for differences between the rate of inflation in the country and the G-5 countries (France, Germany, Japan, the United Kingdom, and the United States). A country's inflation rate is measured by its GNP deflator. The inflation rate for G-5 countries is measured by changes in the SDR deflator. (Special drawing rights, or SDRs, are the International Monetary Fund's unit of account.) The SDR deflator is calculated as a weighted average of the G-5 countries' GDP deflators in SDR terms. The weights are determined by the amount of each currency included in one SDR unit. Weights vary over time because the IMF changes the composition of both the SDR and the SDR exchange rate for each currency changes. The SDR deflator is first calculated in SDR terms and then converted to U.S. dollars using the SDR to dollar Atlas conversion factor.

This three-year averaging smooths annual fluctuations in prices and exchange rates for each country. The Atlas conversion factor is then applied to a country's GNP. The resulting GNP in U.S. dollars is divided by the midyear population for the latest of the three years to derive GNP per capita. When official exchange rates are deemed to be unreliable or unrepresentative of the effective exchange rate during a period, an alternative estimate of the exchange rate is used in the Atlas formula (see below).

The following formulas describes the computation of the Atlas conversion factor for year t:

$$e_t^* = \frac{1}{3}\left[e_{t-2}\left(\frac{p_t}{p_{t-2}} \Big/ \frac{p_t^{SS}}{p_{t-2}^{SS}}\right) + e_{t-1}\left(\frac{p_t}{p_{t-1}} \Big/ \frac{p_t^{SS}}{p_{t-1}^{SS}}\right) + e_t\right]$$

and for calculating GNP per capita in U.S. dollars for year t:

$$Y_t^{\$} = (Y_t/N_t)/e_t^*,$$

where e_t^* is the *Atlas* conversion factor (national currency to the U.S. dollar) for year t, e_t is the average annual exchange rate (national currency to the U.S. dollar) for year t, p_t is the GNP deflator for year t, p_t^{SS} is the SDR deflator in U.S. dollar terms for year t, $Y_t^{\$}$ is the Atlas GNP in U.S. dollars in year t, Y_t is current GNP (local currency) for year t, and N_t is midyear population for year t.

Alternative conversion factors

The World Bank systematically assesses the appropriateness of official exchange rates as conversion factors. An alternative conversion factor is used when the official exchange rate is judged to diverge by an exceptionally large margin from the rate effectively applied to domestic transactions of foreign currencies and traded products. This applies to only a small number of countries as shown in the *Primary data documentation*. Alternative conversion factors are used in the Atlas methodology and elsewhere in the *World Development Indicators* as single-year conversion factors.