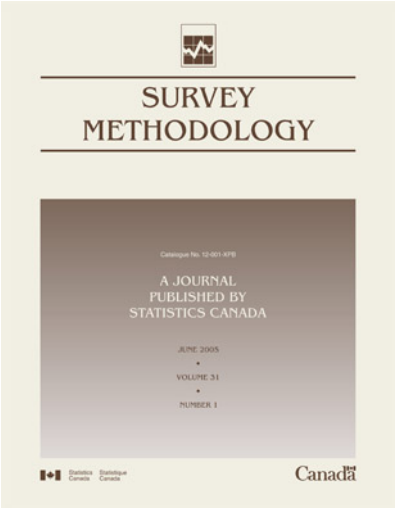
Catalogue no. 12-001-XIE

Survey Methodology

2005



Statistics Canada Statistique Canada

Canada

How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to: Business Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone: 1 800 263-1136).

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our website.

National inquiries line	1 800 263-1136
National telecommunications device for the hearing impaired	1 800 363-7629
Depository Services Program inquiries	1 800 700-1033
Fax line for Depository Services Program	1 800 889-9734
E-mail inquiries	infostats@statcan.ca
Website	www.statcan.ca

Information to access the product

This product, catalogue no. 12-001-XIE, is available for free. To obtain a single issue, visit our website at www.statcan.ca and select Our Products and Services.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service that its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136. The service standards are also published on www.statcan.ca under About Statistics Canada > Providing services to Canadians.

Statistics Canada
Business Survey Methods Division

Survey Methodology

2005

Published by authority of the Minister responsible for Statistics Canada

November 2005

Catalogue no. 12-001-XIE
ISSN 1492-0921

Frequency: semi-annual

Ottawa

Cette publication est disponible en français sur demande (nº 12-001-XIF au catalogue).

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Vol. 29, No. 1, pp. 5-17
Statistics Canada, Catalogue No. 12-001

Methodological Issues in the Development and Use of Statistical Indicators for International Comparisons

David Holt [1]

Abstract

International comparability of Official Statistics is important for domestic uses within any country. But international comparability matters also for the international uses of statistics; in particular the development and monitoring of global policies and assessing economic and social development throughout the world. Additionally statistics are used by international agencies and bilateral technical assistance programmes to monitor the impact of technical assistance.

The first part of this paper describes how statistical indicators are used by the United Nations and other agencies. The framework of statistical indicators for these purposes is described and some issues concerning the choice and quality of these indicators are identified.

In the past there has been considerable methodological research in support of Official Statistics particularly by the strongest National Statistical Offices and some academics. This has established the basic methodologies for Official Statistics and has led to considerable developments and quality improvements over time. Much has been achieved. However the focus has, to an extent, been on national uses of Official Statistics. These developments have, of course, benefited the international uses, and some specific developments have also occurred. There is however a need to foster more methodological development on the international requirements. In the second part of this paper a number of examples illustrate this need.

Key Words: Official Statistics; Statistical Indicators; International Comparisons.

1. Introduction

Official Statistics matter in national life. They are used to develop and monitor public policies, allocate resources, support public administration and decisions made by businesses. They are used too by citizens as a window on the work of government and to monitor its performance.

As important are the international uses of Official Statistics. They are used by national governments to monitor the country's performance against comparators; to ensure that economic competitiveness is maintained or enhanced; to monitor economic and social developments in other countries and the outcome of alternative economic or social policies that other states may adopt. Increasingly in some regions they are used for national participation in international decision-making and resource allocation. For these purposes internationally comparable statistics are needed. They are required too by international agencies to monitor national performance and to make comparisons. The World Bank, IMF and bilateral funding agencies depend heavily on Official Statistics to monitor the impact of policies and technical assistance programmes.

Increasingly statistics and statistical indicators are being used to set and monitor global policies. For example a review of UN Summits and major conferences during the 1990's identified over 280 statistical indicators needed to monitor UN policies made through conference decisions.

Hence the need for internationally comparable statistics has never been greater. This paper has two purposes:

- To describe the current need for internationally comparable statistical indicators for UN and related agency purposes, and
- To suggest that despite the huge investment in methodological research and development to support national statistical needs, there has not been as much emphasis on methodological research supporting the international uses. Some examples will illustrate this.

2. UN Statistical Indicators

What is an indicator?

The term "statistical indicator" has come into use particularly in relation to monitoring global policies. One might try to establish what characterizes a "statistical indicator" and what distinguishes it from the range of statistics published daily by National Statistical Offices. There are indicators, such as the Human Development Index, that are artificial constructs that combine disparate measures (GDP per capita, life expectancy at birth, literacy and educational attainment) into a single composite index number. Such indicators are not a statistical estimate of any single population characteristic and are intended only as a very broad and general measure. But most statistical indicators used by the UN, for example, are not of this kind. Rather they are simply statistical estimates of population characteristics (*e.g.*, fertility rate, life expectation at birth,

1. David Holt, Department of Social Statistics, University of Southampton United Kingdom. Former Director of the Office for National Statistics and Head of the UK Government Statistical Service.

GDP per capita). Each of these characteristics can be precisely defined even though the concept may be complex and the measurement difficult. Such statistics are important for both national and international purposes.

Since the statistical indicators are everyday statistics one may question the need for a different terminology. The reason is not based on the statistical properties but may reflect the way that the indicators are used. Indicators are meant to be high level (usually outcome) measures that are perceived to be related to some aspect of economic or social well-being. For example a low life expectancy at birth in a country is an indication of unsatisfactory living prospects and of health problems in particular. But two countries with similar life expectancies may have very different health situations and the policies needed to address these may be quite different. The statistic used as an indicator points at a problem but one would require much greater understanding of age-specific mortality rates, causes of death, the quality and range of health services and possible differences between sections of the population to formulate a policy response. That policy may be based on improved medical provision, preventative public health or social policies, greater education for those at risk or a combination of all of these. The statistical indicator is a high level monitoring instrument but policy development and monitoring require a much wider and richer statistical picture.

The fact that the indicator is used as a general measure of economic or social well-being does not imply that the methodology and sources used to measure it need not be tightly specified. The requirement is to get comparability both between countries and within a country at different points in time. Loosely specified sources and methods can give rise to inconsistencies that would invalidate the monitoring required. Indeed one of the problems of indicator use is that small changes that have no statistical or substantive significance but cause the ranks of countries to change are given far too much prominence particularly by national policy makers and commentators.

UN Statistical Indicators

In the last decade or so United Nations summits and major conferences (averaging almost two per year) have covered a wide range of economic and social issues. These meetings have resulted in declarations related to future goals and targets that have been endorsed by member states and are intended to improve the well-being of the world's population. Goals and targets call for a commitment to monitor progress towards them and, consequently indicators have been identified in relation to each goal. The intention is to monitor and report on these so that progress towards the declared goals and targets can be measured. The Millennium Development Goals, for example, subscribed to by 164 Heads of State or their representatives have resulted in 8 goals, 18 targets and 48 statistical indicators that will be monitored over the coming decades. For example there are two indicators for Goal 1, Target 2:

GOAL 1: ERADICATE EXTREME POVERTY AND HUNGER	
Target 2: Halve, between 1990 and 2015, the proportion of people who suffer from hunger	1. Prevalence of underweight children under-five years of age 2. Proportion of population below minimum level of dietary energy consumption

In total over 280 indicators had been identified from UN Summits and major conferences in the last 10 years.

This process has gone on with too little co-ordination between officials concerned with the separate UN summits and major conferences in terms of the number and choice of indicators to be monitored. The result is a plethora of indicators of different levels of importance in policy terms. The meetings have varied considerably in terms of the number of resulting indicators (ranging from a handful or less to as many as 70 being identified from a single UN conference). Also there is potential for confusion among users because of an apparent inconsistency and lack of coherence among the indicators.

The UN conferences have adopted markedly different approaches to identifying the need for indicators. In most areas the number of indicators is relatively small and these focus on outcomes. In other areas the indicators are detailed and seek to measure many different facets of policy and service delivery. For Health for example the death rate for a specific disease may be required. Additionally the required indicators may include the disease prevalence rate, the inoculation rate, the proportion of cases treated under a specified treatment regime, public health preventative measures and public understanding of the causes of the disease.

The cumulative effect of indicators added at each conference has resulted in a large demand for statistical information from each member state: a demand that has to be set alongside the demands for statistical information for national policy purposes. For countries with less well-developed statistical infrastructure this total demand can be disproportionate to the resources available to meet it. Indeed some have a concern that the whole global indicator movement has gained too much momentum and the pressure from the UN and international agencies is distorting national priorities and reducing the provision of statistics to support public policy and sound public administration in some developing countries.

Attempts have been made to distil core sets of indicators that might be afforded higher priority. The United Nations Statistical Commission (UNSC) identified the Minimum National Data Set (MNDS: 15 indicators). The OECD Development Assistance Committee – in co-operation with the UN, World Bank and IMF – identified the International Development Goals (IDG: 21 indicators). This set drew heavily on international summits up to 1995. The United Nations Development Group identified indicators to support

Common Country Assessment again based on an analysis of the requirements of UN summits (UNDAF–CCA: 57 indicators). Similarly the need to promote and assess sustainable development has resulted in an additional set (CSD: 57 indicators). There is also Basic Social Services for All (BSSA: 12 indicators). Most recently the UN has espoused the Millennium Declaration Goals and associated indicators (MDG). These sets have some common components and some differences as one might expect. Even these attempts illustrate the vagaries of the political process. For example the fact that the IDG indicators were repackaged and replaced within 5 years by the MDG indicators suggests a lack of constancy and political purpose.

In 2002 the UN Statistical Commission (UNSC 2002) adopted proposals to create a framework containing three levels of priority. The 123 most important indicators are allocated to 7 Domains:

- Demography,
- Health and Nutrition,
- Environment and Energy,
- Economics and Poverty,
- Employment and Labour,
- Education, and
- Other Social Indicators.

The Domains represent major divisions of policy responsibility that are commonly reflected by separate Ministries in many countries. Additionally important cross-cutting policy areas such as Poverty, Child Welfare or Gender that are distributed across these Domains are taken into account. Sub-Domains are identified within each Domain as being relatively self-contained policy areas. Indicators are allocated to the three priority tiers:

- First tier priority indicators reflected the need to monitor policies of the highest global and national importance. They represent the indicators that, no matter how limited the statistical capacity available, countries and international agencies would find essential for top-level monitoring of policy effectiveness. There are 2–6 tier 1 indicators per Domain.
- Tier 2 priority indicators mainly covered different policy objectives (different subdomains) from those covered by the highest priority indicators. These policy objectives should be of sufficient importance to merit a tier 2 priority indicator. Not all subdomains would necessarily do so. There are 0–13 tier 2 indicators per Domain with most Domains having much less than 13.
- Tier 3 priority indicators supported policy needs that are, albeit important, either subsidiary or judged to be less important than others. There are 2–8 tier 3 indicators per Domain.

The Criteria for Allocating Priorities to Indicators

Allocating priority must be grounded in the policy need but involves balancing a number of criteria surrounding the relevance to policy, the technical properties and current availability (or the feasibility, resource and statistical capacity implications of achieving an acceptable measure in a high proportion of countries). While one may aspire to the situation in which an indicator fully satisfies all of the criteria, in practice this will not be the case. The extent to which the indicator meets the criteria needs to be considered and a judgement made about whether any shortcomings are of such overriding concern as to disqualify a particular indicator from use.

A large number of criteria may be identified but the most important are:

Policy Relevance

- Indicators must be relevant to the policy requirement.
- Indicators should measure the real policy objective (or provide a proxy measure that is adequate for policy monitoring).
- Indicators should normally have global policy relevance.
- Indicators should be straightforward to interpret: changes over time in any direction should not be ambiguous in relation to the policy interpretation and significant differences between countries should be meaningful in terms of the policy goal.

Technical Properties

- Technical properties of the indicator should be adequate for the purpose, recognising that change over time is often more important than the level of the indicator.
- Indicators that fail to cover the target population fully should have sufficient coverage to ensure that the indicator values are unlikely to mislead policy users (*i.e.*, the potential bias as a measure of the true policy objective should be small).
- If possible, where indicators are difficult to measure for countries with less well-developed statistical capacity, simplified alternatives should be provided for use until the statistical capacity can support the more demanding measure.
- Indicators should be robust to institutional and cultural differences between countries and over time.
- Indicators should exhibit change over time at a rate that would support policy monitoring.
- Indicators should be produced with sufficient frequency and timeliness to support policy monitoring.
- Indicators should conform to international standards if these exist.

In a number of cases the application of these criteria to create the proposed framework revealed examples where the policy objective suggests allocation to a particular tier, but the inherent conceptual or statistical weaknesses of the proposed indicator and/or measurement problems cause the indicator to be allocated to a lower tier.

The numbers for each tier reflect the fact that the indicators are not intended to substitute for the mass of detailed statistical outputs from national statistical systems that support users' needs. They are intended as high level indicators for monitoring purposes.

3. General Issues

The process described in the previous section identified a number of general issues most of which have a technical dimension.

Choice of Indicators and Targets

There are two facets: first the precise form and definition of the indicator needs to be decided together with a methodology for measuring it. In practice, both national and international policy makers are inclined to express their goals directly in terms of a statistical indicator without particular concern for the definitional and measurement issues. Too often an indicator is identified with too little thought. The reality is that identifying statistical indicators for monitoring purposes should be neither a pure policy nor a pure statistical issue. The basic expression of the policy goal must drive the monitoring requirement but turning that expression into a statistical indicator that will be relevant, reliable and acceptable to the various stakeholders is a statistical function. The tension between the policy view of what is needed and the statistical view of what is feasible and technically sound needs resolution.

The second facet is the choice of a target. These are chosen in relation to the indicator (for example to halve the death rate due to a particular disease by a stated year). There are two views about such targets. One is that they should be based on policy analysis and set to reflect what effective policies might be expected to achieve. In this view it is unlikely that the same target is achievable or demanding enough in every country. The second view is that the targets are simply something to aspire to and not based on any reasoned analysis. In this view target setting is entirely a political process for binding countries into a political commitment.

From a statistical perspective the danger of aspirational targets is that they will not be met (or sometimes even approached) and the process of statistical monitoring itself may fall into disrepute as a result. There is also a threat to statistical integrity if the political pressure to show progress against an unrealistically set target is too strong.

Whichever view prevails targets that are framed in terms of improvements from a given base year do require that indicator values are available at that point. Given the lack of statistical capacity in many developing countries this is problematic and for a number of the Millennium Declaration Goals for example the global statistical picture for the baseline year from which progress will be measured is seriously inadequate.

Statistical Capacity

The ability to produce consistent, reliable statistical information requires a sustained statistical capacity. This requirement is not a one-off capability but implies the ability to produce statistics on a regular basis and with the timeliness needed.

In particular a sound statistical infrastructure is essential. By this is meant:

- Underpinning systems to create and maintain sampling frames for business and household surveys.
- A critical mass of ongoing statistical activities: survey design, data collection and analysis in order to nurture the basic professional skills.
- The technical and methodological capacity to maintain and develop systems in accordance with international standards as these are developed over time.
- A developed analytic capacity.
- Adequate statistical frameworks and IT infrastructure.
- Good management to make the most use of the resources that are available.
- All of the above embedded within a wider legal and administrative structure that recognises the importance of good statistical information and the need to sustain the conditions in which it can be produced with high professionalism and integrity, consistent with the *UN Fundamental Principles of Official Statistics*.

Without this core capacity and the ongoing resources to support it, neither the statistical needs of the country nor those of the international community will be reliably served. In many countries adequate ongoing financial support is a key issue. Where this core capacity is fragile the sporadic provision of additional funds from international or bilateral funding agencies to satisfy a particular statistical need will be much less effective and is no substitute for developing what one might term "statistical sustainability".

In this regard, statistical indicators need to be viewed as the end product of often complex statistical infrastructures that are essential if the indicators are to be produced with adequate quality. Too much emphasis has been placed on the indicators and too little on the statistical sources and infrastructure that underpin these.

Indicators as Rates and Ratios

International comparisons require that statistics be put on a basis that is immediately comparable and for this reason almost all of the indicators are presented as rates,

proportions or in per capita terms. This places population estimates as a cornerstone of most of the statistical indicators. These depend on periodic Censuses to provide benchmarks and on systems of vital registration or other sources to permit inter-censal population estimates. Different statistical indicators call for population estimates for various age-sex groups as the appropriate denominator.

A particular difficulty is that the numerator of such indicators and the population denominator are often provided from different sources within a country and may be inconsistent. Hence the rates, when calculated, may not be recognised within the lead policy Ministries and can be challenged by them leading to a loss of confidence in the statistics. Population estimates from the National Statistical Institute, a policy Ministry and the UN Population Division may all differ. In extreme cases different population denominators may be used for different policy areas. This is clearly unsatisfactory and when it occurs implies a systemic problem of consistency and quality assurance and a lack of statistical co-ordination within a country.

For economic measures indicators are often expressed as per capita measures (in which case the comments above apply) or as ratios of expenditure (*e.g.*, for health or education) in relation to GDP. Complex measures such as GDP require an extensive framework of business surveys, administrative sources and underpinning infrastructure if the statistics are to be of adequate quality.

The pervasive use of GDP and of population estimates in this way underlines the importance of the quality of these estimates if other indicators are to be sufficiently reliable. Both require a strong statistical capacity and infrastructure if they are to be regularly produced.

Inadequate Administrative Sources

There are a large number of indicators that are derived from administrative systems in countries where these are well established (*e.g.*, mortality rates by cause, fertility rates, gross and net enrolment rates in education and many health indicators concerned with health services and provision). For some kinds of information often relating to public services (*e.g.*, numbers of teachers, doctors or nurses and qualifications) the only realistic sources of information are administrative and where these are inadequate they need to be strengthened. For other measures a household survey may be an alternative although there can be conceptual and measurement differences between information obtained from administrative and survey sources.

Nonetheless, in countries where the administrative systems are inadequate survey based measures are widely used in which both the numerator and denominator of the indicator may be derived consistently from survey estimates. In this case a special survey devoted to one particular area of interest (*e.g.*, health and fertility history) can provide a wide range of statistics. This is a viable possibility (at a cost) particularly when countries want a more comprehensive picture of a situation.

However, *ad hoc* surveys cannot provide the ongoing information needed to track important indicators. To ensure that critical information will be available on an ongoing basis it is necessary to invest resources into the statistical infrastructure so that surveys can be repeated regularly.

In general, even when they purport to measure the same thing, both administrative sources and surveys have strengths and weaknesses. The administrative source is often large and provides the opportunity to provide regional or local figures. However the concept contained is often not ideal for the statistical purpose. Also the source may not cover the whole population or may suffer from various inadequacies. Surveys can often measure the concept required but sample sizes are often small and there may be differences between the surveyed population and that intended because of inadequate sampling frames, response problems and measurement error.

The real methodological challenge is not to decide that one source is preferred to the other but to use all of the information available to produce the highest quality estimates possible. This will often require strong methodological effort if the statistics are to command confidence. However these data reconciliation problems often occur in countries where the methodological expertise is not strong.

Measuring Levels

There are some topics particularly concerning environmental indicators where the very idea of a measure of level may be very difficult to frame. It is often not the absolute level of the indicator so much as the trend over time within each country that is the key focus of policy.

For example there is no real meaning in measuring the average toxicity in Canada's coastal waters. One would need to define coastal water precisely and the sampling methods to achieve a representative sample of coastal water together with appropriate methods of statistical inference. In particular there would be a methodological question as to whether the sample should be weighted to represent the distribution of coastal water or that of the adjacent coastal population. In practice samples taken on a consistent basis from the same locations on repeated occasions will not provide a measure of toxicity level but will, under some strong assumptions, allow trends to be monitored. However development (such as new towns and industrial sites) will lead to new sources of toxicity over time and the location of sample sites may need to be reviewed to reflect this. At the same time data analysis will be needed to avoid the measured trends exhibiting discontinuities. The development of sample designs and methods of inference for populations of people and businesses has been one of the great achievements of Official Statistics. But there are some substantial unresolved methodological issues in designing and analysing samples of physical populations to an equivalent rigorous standard. The methods applied generally in Official Statistics may offer some contribution.

Meta-Data

This is essential if users are to understand any particular issues affecting the statistical indicator values for any country. Good meta-data (such as is required by the IM's SDDS and GDDS) is a general requirement but there are specific situations when countries should ensure that specific meta-data is provided.

- When national priorities result in an indicator which is not fully comparable with those produced by other countries. Failure to provide informative meta-data will fail those users who seek to use the indicator for comparative purposes.
- Where national standards or targets are adopted (for example in setting a national poverty standard) the basis of this measure needs to be available to users.
- Population forecasts (and inter-censal estimates in countries where vital registration systems are unreliable or unusable) will depend crucially on the data sources and assumptions made about age-specific fertility rates for example. A clear specification of the underpinning assumptions is essential to users.

Distributional Measures

A number of indicators call for separate analyses by sex and as a general rule if the data source can support it then this should be routinely provided. The same applies to analysis by subgroups (*e.g.*, region, age-group, ethnic or social classifications). There is a broader issue about providing indicators that measure inequality and distribution within each country. There are a rather small number of indicators that focus on distributional issues (*e.g.*, share of consumption by lowest quintile of population) but the large majority of indicators are based on national averages. This is a significant deficiency in the existing indicator list. Much deprivation and inequality in the world will be masked by indicators based on national averages. Analyses by subgroups (*e.g.*, by gender, region, age group, income groups, ethnic or social classifications) where feasible would illuminate this issue much more. Similarly, additional measures of inequality, such as the ratio of consumption by the highest 20% of households to the lowest 20% have much to commend them.

4. Some Selected Methodological Issues

In the second part of this paper a small number of methodological issues are discussed specifically in the context of international comparisons.

4.1 The Methodological Paradigm

In general the paradigm adopted by Official Statisticians to ensure comparability is based on several components:

- Conceptual clarity of the item to be measured.
- Precise definitions of relevant terms that can be applied in practice.
- And precisely defined classification systems.
- A clear specification of the target population to which the estimates apply.
- Development of appropriate sources and methods, even questionnaires, to obtain the data and compile it into the estimates required.
- Often, international standards, manuals and descriptions of best practice to cover all or most of the above.

The basic assumption is that if the measuring instrument and related methodology can be defined precisely enough then it can be applied independently in different countries and the resulting statistics will be internationally comparable. Hence: control the measurement process and the outputs will be comparable.

This approach generally yields relatively comparable statistics but not absolutely so and not all of the time.

4.2 Literacy

It is, of course, well known that the translation of some measures from one language and culture to another is fraught with difficulty and measuring functional literacy is an example. In any one country one can test comprehension of a text that is grounded in everyday experiences and the requirements of daily life. But the task of transferring this into another language and culture and getting a precisely comparable measure of functional illiteracy is very difficult. Even when great effort has been made to achieve this (*e.g.*, the International Adult Literacy Study 1999) it may be that only approximate comparability can be achieved especially if the same measures are used over time so that within country changes may be monitored. In practice literacy measures for almost all countries are much cruder; for example a self-assessed response to a Census question such as "Can the person read a letter? This approach may provide a broad estimate of the number of people who can read to a certain level in some circumstances but is unlikely to provide comparability either between countries or within a country over time. Large changes in the level of literacy within a country may provide evidence of real change but small changes may simply reflect the unreliability of the measure. In order to monitor literacy levels for global policy emphasis is placed on 15–24 year olds since these reflect the flow of newcomers to the adult pool and improvements in educational access and attainment will show larger changes to literacy levels for this group than for the adult population as a whole. Hence the inherent weaknesses in the measure may, to an extent, be mitigated by focussing on a group for which large change may be expected. Such an approach will, however, miss the effect of adult literacy programmes.

4.3 Interactions Between the State and the Citizen

International comparability is made more difficult whenever we seek to measure something that is affected by the interaction between the state and the citizen because the way in which the state provides for particular services may differ from country to country. In these cases precisely the same measuring instrument applied in different countries may give different results. Consider for example the case of housing provision for low income families. In some countries this is provided free or for very low rent. In others the rental cost is at market levels but families get state benefits to allow the payments to be covered. Hence the mechanisms by which the state interacts with the individual will affect important economic measures. As a consequence the international comparability of statistics collected and compiled under precisely the same conceptual framework can be impaired. In some circumstances money flows are imputed to reduce the discrepancies but this is impractical if the provision of cheap housing is very widespread.

Similar issues can arise for medical provision. In some countries medical services are provided by the state, free at the point of access and directly funded from taxation. In others the system is funded through a system of medical insurance which may have elements of both state and personal contributions. When medical services are provided to an individual the real flow of payments may vary. The medical practitioner may make a direct claim (to the state or to a medical insurance fund) for the services provided. In other systems the individual may be the formal claimant but with the payment made directly to the medical practitioner. Or the individual may be required to pay the costs and to claim these back from the state or insurance fund as a payment back to the individual. To some extent these arrangements may be regarded as alternative ways of achieving the same end: a state-facilitated system to ensure that individuals have good access to medical services. In practice money flows are imputed to eliminate most of the institutional differences.

A third example is the estimation of tax revenue which has a direct impact on the estimation of public expenditure and government deficit. Under SNA93 this assessment is made on an accruals basis and in the year in question will be based on the tax assessments made to individuals and businesses. In countries that use well-established self assessment methods and a high level of tax collection through employers the difference between the estimate of tax to be collected and that which is subsequently achieved in the following years may be very small (there will be companies that cease to function and default on the tax liability and people who may die without leaving an estate sufficient to cover the tax due). In other countries with different forms of tax assessment and recovery practice there may be much larger differences between the tax assessed and that which is eventually recovered. Where a shortfall occurs this will in due course be written off against the financial account. But this write-off will have no impact on the estimates of public expenditure and government deficit. Hence a system that "optimistically" estimates the level of taxation that will finally be recovered will result in a lower estimate of government deficit that will never be corrected when the tax shortfall is written off against the financial account. Given the importance attached by international bodies to levels of public expenditure and government deficit a lack of comparability in these key measures matters. In this third example there is no universally agreed method to eliminate the differences although in the European Union specific (non-SNA) rules have been introduced in the debt and deficit manual to eliminate the discrepancies.

4.4 Comparing Economic Measures – Purchasing Power Parities (PPP)

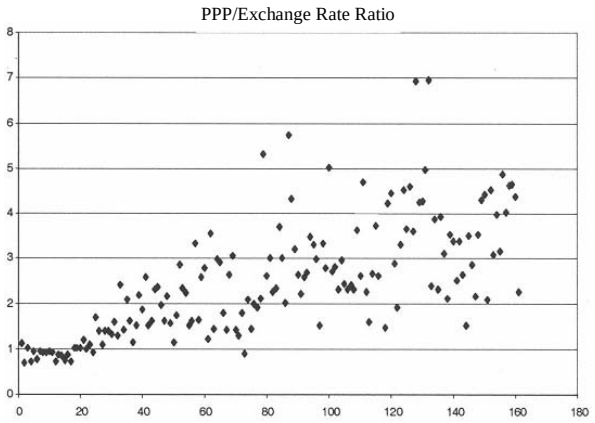
For comparative purposes economic measures (*e.g.*, GDP, per capita income or expenditure on Health or Education, living standards) that are measured in national currencies must be converted to a common unit of measurement.

The point at issue is whether conversion from national currencies to a common unit (say US$) should be made using the comparative exchange rate values of different currencies, or should be made on the basis of equalizing the purchasing power of the currency. This is an important issue that can have a profound effect on international comparisons. For example in 1999 the Human Development Report (HDR) claimed that "the gap in per capita income (GNP) between the countries with the richest fifth of the world's population and those with the poorest fifth widened from 30:1 in 1960, to 60: 1 in 1970, to 74:1 in 1995." These statistics are based on exchange rate conversion and yet the corresponding PPP ratios are about 12:1 in 1960, 18:1 in 1990 and 16:1 in 1997. Not only are the ratios much smaller but the clear upward trend presented in the HDR figures is not apparent in the measure expressed in PPP.

The exchange rate conversion values of any currency are determined by the international financial markets and reflect the market forces in those institutions. Indeed, in the modern world, exchange rates are little affected by international trade and the exchange of goods and services in world markets. The second approach uses Purchasing Power Parities (PPP) to reflect domestic prices on an internationally comparable basis. The value of national income or economic output in any country is equated to others on this basis. In this approach, the PPPs provide an international valuation of what the local currency will buy within the country (United Nations 1992).

Figure 1 shows a plot of the ratio of exchange rate conversion to PPP conversion for most of the countries of the world. The x–axis is the 1997 Human Development Index (HDI) rank of each country. The most industrialized countries occupy the lowest 20 places at the left of the graph

and the further right one goes the lower the level of development of the country as measured by the HDI. For the industrialised countries the ratio of PPP to US$ exchange rate conversion factors is fairly close to 1. However, for less developed countries the ratio is greater – in many cases much greater – than 1. The upwards slope of the plot shows that the ratio of PPP to US$ exchange rate conversion is generally larger, the lower the HDI rank. For the least developed countries the ratio can be as much as 4 or more. Hence, because the ratio is close to 1, a comparison of economic measures between the United States and a major European country, for example, would be fairly similar using either exchange rates or PPP conversions. However, a similar comparison between the United States, or any of the most industrialized countries, and a least developed country would be very different. In such a case, the conversion of per capita income, for example, using PPP conversion could be as much as 4–6 times larger than the conversion using exchange rates (an exchange rate measure of GDP per capita of $1,000 would be $4,000–6,000 in PPP terms). Hence, the choice of conversion factor has a significant effect across the developed/developing spectrum.



Source: UNSC (2001)

Figure 1.

Powerful reasons exist for using PPP conversion rather than US$ exchange rate conversions for real economic (rather than purely financial) phenomena such as standard of living comparisons (as reflected by per capita GDP) and, by extension, for comparisons of economic output (GDP) and national income (GDP or per capita GDP). (UNSC 1998).

Table 2
International Comparisons: Ratios of Per Capita Measures of Output or Use of Goods and Services

Comparison	Daily per capita suply of calories, 1996	Daily per capita supply of fat, Total (grams), 1996 (a)	Daily per capita supply of protein, Total (grams), 1996 (a)	GDP Index	TVs, per 1,000 people, 1996	Carbon dioxide emissions per capita (metric tons), 1996	Com'l energy use (oil equiv'nt) per capita (kgm), 1996	Per capita electricity consum'n, 1996	Main telephone lines, per 1,000 people, 1996	Intern'l tourism departs per 1,000 people, 1996	Personal computers per capita	Real GDP per capita (PPP$), 1997	Per capita GDP (US$) 1997
Japan/China	1.0	1.3	1.3	1.6	2.8	3.3	4.5	9.1	10.9	32.6	42.7	7.7	45.9
Sing/Indonesia				1.6	1.6	16.3	11.7	18.8	24.4	111.0	45.2	8.2	26.8
Korea/Vietnan	1.3	2.3	1.5	1.7	1.8	0.8	8.0	23.1	26.9		39.9	8.3	29.8
Mexico/ Nicaragua	1.3	1.8	1.6	1.5	1.1	5.3	2.9	3.9	3.7	1.6		4.2	10.0
SA/Mozbique	1.6	2.4	2.1	2.2	41.0	69.0	5.2	58.9	33.3		47.1	10.0	21.9
SA/C African Rep	1.5	1.2	1.6	1.7	24.6	69.0		125.4	33.3	0.8		5.5	11.3
Brazil/Ecuador	1.1	0.8	1.3	1.1	2.0	0.8	1.4	2.6	1.3		4.7	1.3	3.0
T&T/Haiti	1.5	2.4	1.5	1.7	63.6	86.0	22.9	40.2	21.0	11.5		5.4	12.6
Sey'lls/Sri Lanka	1.1	1.5	1.5	1.4	2.3	5.8		7.2	14.0	98.0		3.3	6.1
Sey'lls/India	1.0	1.6	1.3	1.6	3.0	2.1		3.0	13.1		10.3	4.9	12.7
Kuwait/Jordan	1.1	1.2	1.4	1.6		10.1	7.9	14.2	3.9		3.4	7.3	15.6
Lebanon/Jordan	1.2	1.4	1.2	1.2		1.8	1.1	1.7	2.5	19.0		1.7	4.2
Egypt/Ethiopia	1.8	2.6	1.5	2.1	31.5		2.2	36.4	16.7	41.9		6.0	10.6
Maur's/Madag'r	1.5	2.6	1.7	2.1		15.0		25.3	54.0			10.0	16.7

The approach due to Castles (2000) is illustrated using a range of bilateral comparisons of countries from the same region in Table 2. The ratio of per capita consumption of various items in each pair of countries is presented, together with ratios of the per capita GDP for the two countries based on PPP conversion and exchange rate conversion. A general pattern may be discerned. For items such as food consumption, which are price inelastic, the bilateral comparisons are relatively close to 1, with countries with higher per capita GDP having somewhat higher consumption. The ratios are much larger for items (*e.g.*, televisions or personal computers) that depend on disposable income and are much more price elastic. In general, the PPP comparison for any pair of countries falls within this pattern, having a larger value than the ratios for food consumption but smaller than those for the technological items. This is what one would expect. The exchange rate comparisons, however, are generally much larger and often lie outside the range of consumption even for items such as PCs and televisions.

The PPP measure seems more consistent with the other measures and more relevant for the purposes intended.

There are, of course applications for which exchange rates are appropriate, such as the expression of a country's international debt relative to its GDP.

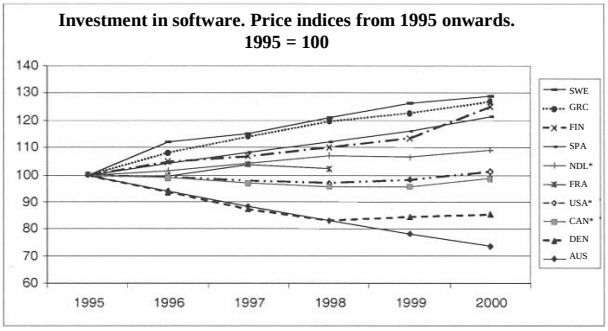
4.5 Price Indexes for International Market Prices

For some goods and services (particularly in Information and Communication Technologies – ICT) the rapid rate of technological change has made it much harder to estimate price changes by normal methods. The element of quality change in simultaneous price and product changes is significant and National Statistical Offices have responded to this by greater use of hedonic regression methods to adjust for quality changes. Even when these methods are applied independently by different countries there can still be large differences in the price deflators that are arrived at and yet, to a large extent the goods and services are traded in an active international market. Similarly it is possible for producers within the same NSO who compile national import and export price indexes to use different price deflators for the same type of goods and services.

These differences matter: within a country they can lead to significant impacts on key statistics such as the balance of trade and fixed capital formation. Between countries they distort the levels of ICT investment being made and the productivity analyses aimed at measuring the impact of ICT investment on growth and economic performance.

Wyckoff (1995) observed that, in the case of computer price indices in OECD countries, large differences in the prices were more likely to reflect methodological differences than real price differences between countries. Lequiller (2001) found significant country differences in the attribution of software expenditure between fixed capital formation and intermediate consumption. The question is whether these differences are due to methodological differences and hence distort international comparisons.

If we consider the case of computer software, for example, Figure 2 illustrates the range of price indexes applied to software by a range of countries. The differences in national estimates of the price indices are dramatic and will have a significant effect on the international comparability of statistics that depend on the price indices.



Source Edwards, Comisari and Johnson (2002)
citing Ahmad (2002)

Figure 2.

In a separate analysis Colecchia and Schreyer (2001) collate estimates of average annual percentage growth in software investment (1990–95) for a range of OECD countries. These estimates depend on nationally estimated price indices. They also recalculate the values using an internationally harmonised price index. The results are given in Figure 3. The latter raises the mean growth from 6.3 to 8.2. More significantly in terms of international comparisons it lowers the standard deviation from 4.8 to 2.9 making the national estimates more similar.

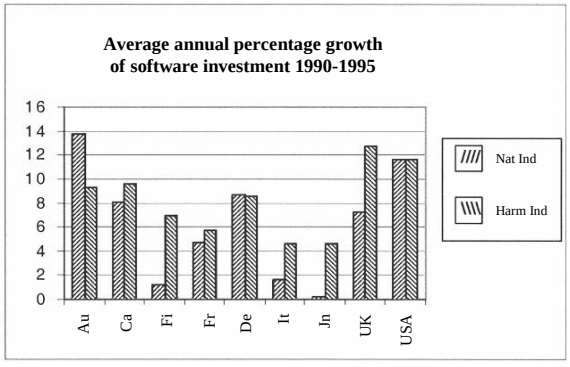


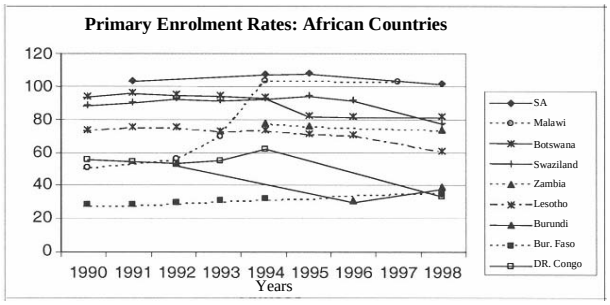
Figure 3.

This is an example where broadly comparable procedures applied independently in different countries give such different measures of something that ought to be much the same in all countries that one must question the international comparability of the economic statistics that depend on the measure. In this case the methodological paradigm breaks down.

The basic question is whether it is appropriate for each country to independently apply somewhat similar methods to matters such as price indices for goods and services that have a strong international market. Alternatively it could be argued that to improve international comparability countries should cede an element of national statistical sovereignty by using internationally estimated indices. The issues are what methodology should be applied; to what data (presumably collected on a collaborative basis from a range of countries) and what are the consequential issues for economic analyses of national data. Using coherent estimates of price indices for import and export prices would also need to be considered.

4.6 Imputation and Aggregation

For the purposes of monitoring international policies it is not enough to measure statistical indicators at the national level. Most of the statistical series comprise rates, ratios or proportions. The country level measures need to be aggregated to provide measures at the regional and global level. This requirement generates a number of methodological problems that need further investigation and development.



Source: UN Millennium Indicator Database

Figure 4.

Figure 4 contains the primary level Net enrolment rates for some African Countries. Although Education statistics have been chosen here these illustrate a number of features that are common to a wide range of statistical indicators and countries:

- The series are incomplete with missing values in some years for all countries and the level of completeness varies from one country to another. Indeed there can be countries with only one figure in the recent past or, for some series, with none at all.
- The last figures available are for 1998.
- The data show different trends with participation rates increasing in some countries and decreasing in others.
- Some countries exhibit sudden changes in the participation rate from one year to the next (*e.g.*, Botswana Malawi). Countries may exhibit erratic series (*e.g.*, Rwanda).

The objective for inference

The objective for an aggregate statistic at the Regional or Global level needs to be clear. For a regional rate for Africa for example one might naturally assume that the objective is to estimate $Y_{R,T}$ the rate for the region *R* at time *T*:

$$Y_{R,T} = \sum_{j \in R} Y_{j,T}\, w_{j,T} \Big/ \sum_{j \in R} w_{j,T,} = \sum_{j \in R} Y_{j,T}\, \mu_{j,T}. \quad (1)$$

$Y_{j,T}$ is the corresponding rate for country *j* and year *T* and $\mu_{j,T} = w_{j,T} / \sum_{j \in R} w_{j,T}$.

In equation (1) the natural value for $w_{j,T}$ is the population size for the relevant age group in country *j* at year *T*. Thus for the enrolment rate data presented above the national enrolment rates would be aggregated to produce the regional (or global) rate. Corresponding estimates of change Δ_{T_1,T_2} between years T_1 and T_2 may be similarly defined at the national, regional or global level. For example:

$$\Delta_{R,T_1,T_2} = Y_{R,T_2} - Y_{R,T_1}. \quad (2)$$

$$= \sum_{j \in R} Y_{j,T_2}\, w_{j,T_2} \Big/ \sum_{j \in R} w_{j,T_2} - \sum_{j \in R} Y_{j,T_1}\, w_{j,T_1} \Big/ \sum_{j \in R} w_{j,T_1}. \quad (3)$$

Similarly annualized change $\Delta_{T_1,T_2}/(T_2 - T_1)$. may be defined.

The regional statistics defined by equations (1) and (2) will be dominated by the national rates (and changes) for the larger countries. In a region that contains China or India for example smaller countries may have national rates that are quite different and these will have little impact on the regional rate. The same is true for estimates of change. Similarly the variance of the regional statistics will tend to be dominated by the largest countries because of the impact of the weights squared $\mu^2_{j,T}$. For the regional estimate of the level for example:

$$V(Y_{R,T}) = \sum_{j \in R} V(Y_{j,T})\, \mu^2_{j,T}. \quad (4)$$

An alternative emphasis may be required if the global target is meant to impose a commitment on each country regardless of size. Here the country might be regarded as the unit of analysis (rather than the person as is implicit in the aggregate statistics described above). One possibility could be to define $Y_{R,T}$ and Δ_{T_1,T_2} by setting $w_{j,T}$ equal for all countries so that all countries contributed equally to the summary statistic. Clearly there are other alternatives such as summarizing the countries performance at the regional level by reporting

on the number of countries that exceed/fall short of the population weighted regional statistic by a threshold amount together with the range of country levels (or changes) observed.

Constant or time-dependant weights for estimates of change

Many of the statistical series call for national statistics that incorporate a changing population structure over time. For example the proportion of people below a poverty threshold will be changing because of changes to household income (expenditure, or consumption) but also because the population size itself is changing. Indeed over a period of 15–20 years fertility rates in many developing countries imply very significant population growth. Hence the denominator implicit in $Y_{j,T}$ at different years T will properly reflect this change. When producing the aggregate measure $Y_{R,T}$, it is natural to use the population weights $w_{j,T}$ and hence the relative proportions $\mu_{j,T}$ relating to year T. It is less obvious whether the weights $w_{j,T}$ (and hence $\mu_{j,T}$) used to produce the aggregate measure of change Δ_{R,T_1,T_2} for a region or the whole world should change with time. The measure of change Δ_{R,T_1,T_2} may be decomposed as follows:

$$\Delta_{R,T_1,T_2} = \sum_{j\in R}(Y_{j,T_2} - Y_{j,T_1})\,(\lambda\,\mu_{j,T_2} + (1-\lambda)\,\mu_{j,T_1}) + \sum_{j\in R}(\mu_{j,T_2} - \mu_{j,T_1})\,((\lambda)\,Y_{j,T_1} + (1-\lambda)\,Y_{j,T_2}). \quad (5)$$

The measure of change is thus a composite measure involving both the change in Y over time and the change in weights. Since the weights $\mu_{j,T}$ simply provide the linear combination of the country measures it is arguable that these should be held constant between T_1 and T_2 so that the second term in equation (5) is made zero. The first term in equation (5) is arguably a better measure as an index of change since it represents a linear combination of the country changes.

The same rationale may be applied when the national measures are economic and measured in the local currency and these have to be converted to a common basis using PPP conversion for example. It may be argued that a constant value of PPP conversion should be applied to all values in local currency whatever the time period to which they apply.

There still remains the choice of λ and values of 0 or 1 would use the weights for one of the reference years only.

Of course a measure of change based on the first term in equation (5) implies that this is not arithmetically equal to the difference between the measures of level in the two years.

Missing Values

Most of the statistical series used for monitoring global policy have gaps of various kinds. For some series most countries are represented with data for most years. For other series data may be available for a smaller subset of UN member states, but with perhaps only one or two data points for some of the countries represented and these related to different years for different countries. If inference is required for year T and this is missing then the question of imputation method arises. Figure 4 illustrates the situation for the primary education enrolment rates.

Much research and development on missing data in Official Statistics has focussed on the raw micro-data and causes such as non-response. In calculating aggregate measures for statistical indicators it is the national statistic for an entire country in a particular year that is unavailable. Common assumptions such as that data are missing at random are inapplicable in this case. In general the lack of completeness of statistical series for each country may often be related to the statistical capacity of the country to produce the range of statistics required. This in turn is often related to the level of development generally and to some extent the size of the country since the per capita statistical effort required is generally greater for small countries. This has two general consequences:

– If we consider the regional estimate $Y_{R,T} = \sum_{j\in R} Y_{j,T}\mu_{j,T}$ and only a small proportion of country values are missing (and if these relate to countries with small weights $\mu_{j,T}$), then the regional estimate will be relatively robust to any reasonable imputed value for the missing values. Moreover the weights associated with the imputed values and measures such as $\mu^2_{j,T} / \sum_{j\in R}\mu^2_{j,T}$ will provide diagnostic information about the extent to which the regional (or global) estimate may be dependent on imputed values.
– It must be recognised however that if many of the statistical series are related to economic and social development and if countries with missing data are generally low in statistical capacity (and by extension development generally) then this is a case of informative non-response. Hence the term "reasonable imputed value" in the previous bullet point needs to take account of this.

In general there are three levels of information that might potentially assist with imputation for the missing values in a time series. These are (a) values for other years in the same series for the same country, (b) associated series from the same country that may convey information about the series in question and (c) statistical series from other countries that might be considered "similar" in some sense to the country for which the value needs to be imputed.

In addition the range and sophistication of available methods may vary greatly (see for example Chatfield 1996). The objective is to predict the value $Y_{j,T}$ of the trend at time T for the series in question. The length of the time series available are generally short and the series are non-stationary. Since the series are annual, for many of them

seasonal effects may be less important. A simple method may involve naïve trend fitting (and then prediction for the missing values) using least squares fitting on the values available in the series. More sophisticated methods will remove the trend to arrive at a stationary series and then apply various approaches such as weighted averages and modelling the correlation structure of the series to arrive at predicted values for the missing values of the stationary time series. These are then combined with the initially removed trend estimates to yield predicted values for the missing values in the original series.

This general problem could benefit from some substantial methodological investigation perhaps taking account of some of the following:

- The ultimate objective is not to model the series, nor even to predict the missing value for use as an inference at the country level. The imputed value will be combined with observed values from other countries to produce the aggregate measure which is the ultimate objective.
- The time series available are often short.
- So long as the statistical series is not too noisy the highest quality predictive information will likely come from the values for other years in the same series and the same country. For many situations, since the objective is to predict the trend level at $t = T$, this may imply that simple trend estimation methods such as regression using year as an explanatory variable for the series in question may be adequate. For example:

$$Y_{j,\,t} = \alpha_j + \beta_j\ \mathrm{t} + \varepsilon_{j,\,t}, \qquad (6)$$

where $\mathrm{V}(\varepsilon_{j,\,T}) = \sigma^2$.

The use of data from the same series if it is reasonably stable will ensure that the informative non-response issue is taken into account since the parameter estimates α_j and β_j will relate to the specific country and will be estimated from data from that source.

Consider as an illustration equation (6) written in matrix form for a series of length $k\,(t = 1, \ldots, k)$ and where imputation is required for $t = k+1$. Prediction for a missing value at the end of the series is likely to be less reliable but is also likely to be realistic since it will occur in practice.

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \sum, \qquad \boldsymbol{V}(\sum) = \sigma^2 \boldsymbol{I} \qquad (7)$$

$$\begin{aligned} \boldsymbol{V}(\hat{Y}_{k+1}) &= ((\mathbf{1},\ \boldsymbol{k}+\mathbf{1})\,(\boldsymbol{X}^T\boldsymbol{X})^{-1}(\mathbf{1},\ \boldsymbol{k}+\mathbf{1})^T + \mathbf{1})\,\sigma^2 \\ &= \left[\frac{2\,(2k+1)}{k\,(k-1)} + 1\right]\sigma^2. \end{aligned} \qquad (8)$$

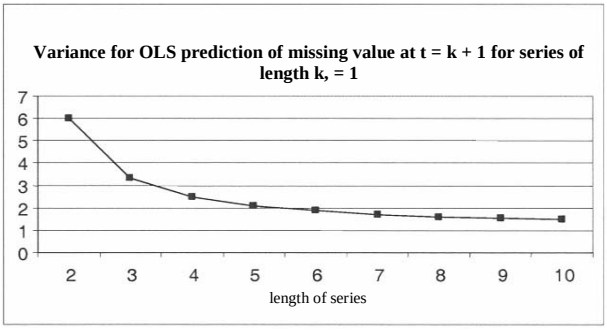


Figure 5.

Figure 5 shows the relative variance for predicting Y_{k+1} the missing value at $t = k+1$ for a series of length k under OLS assumptions. For a series of infinite length the variance will be 1. The point of interest is how quickly the variance drops for a series of 5 or 6 points and how relatively gradually further variance reduction occurs. The OLS assumptions may be replaced by some more general covariance structure such as $\mathrm{Corr}(\varepsilon_{j,t},\ \varepsilon_{j,t+r}) = \rho'$. For small and moderate values of $\rho\,(\rho \le 0.5)$, the variance based on GLS estimates is very similar to Figure 5.

Of course the assumptions above are unrealistic since most time series prediction methods would take account of the correlation structure for recent periods by using exponential weighting of the most recent observations to predict the residual associated with $t = k+1$. However depending on the extent to which recent observations are correlated with $t = k+1$, estimating the parameters α_j and β_j from a very short series will, to some extent, automatically take account of the positive correlation of the residuals at recent periods. If this is so then the decay shown in Figure 5 may be a rough approximation to the impact of the length of the series used.

Clearly a more extensive study of the impact of simple and more sophisticated methods for imputing the missing value would be of considerable benefit.

- For such methods there will be a trade off between variance and bias related to the length of the series used. A relatively short part of the series where the local linearity of the prediction model is more likely to approximate reality may yield a less biased estimate of β_j due to model misspecification but provide parameter estimates of α_j and β_j with higher variance and hence a more variable predicted value.
- Alternative methods that take account of the correlation structure of the time series can be considered and the extent to which these provide a significant improvement in the quality of the prediction would be of interest. One needs to keep in mind that the ultimate objective is to generate the regional (or global) summary measure and the imputed value may have relatively little impact on this in terms of variance.

– When there are no values at all for the statistical series in a particular country or when the series is erratic and/or has very few values for other years the situation is much more difficult. There may be greater added benefit in using the information contained in other series from the same country and from other countries. One might conjecture that a regional effect, such as a drought for example, may affect other countries in a region and that the statistical series may display similar characteristics even if the series themselves are at different levels. The potential for borrowing strength from other time series in the country of interest and others needs to be explored. Hierarchical model-based methods developed for small area estimation could be investigated in this case although the total volume of data even in a region with 30–50 countries will not be large. Also the question of establishing which countries might be suitable sources of information in any situation may require both expert judgement of the similarities and dissimilarities between countries as well as formal statistical diagnostics. If the available series are short then identifying and fitting suitable models will be a challenge.

When we consider the estimation of change between two years T_1 and T_2 the same issues surrounding missing values and imputation emerge. As for the regional estimates of level, countries with small relative weights $\mu_{j,T}$ are unlikely to have a significant impact on the regional estimate of change. However under current international practice it is quite common for the regional estimates of $Y_{R,T}$ to be based on whatever national statistics are available for year T and hence for differences between two years to be based on different sets of countries. This is clearly unsatisfactory and will cause the estimate of change to be biased. Imputation for missing values is needed and the statistical properties of the resulting estimates of change need to be explored. The question of separate or joint imputation for missing values from the same series may also be considered.

5. Conclusions

A description of the use and importance of statistical indicators and the framework in which they are produced is provided.

It is suggested that there has been less focus on methodological development for statistics used for international comparisons than there has been on statistics used for national domestic purposes. A number of examples have been provided illustrating the need for additional methodological work in this field.

References

Ahmad, N. (2002). Proposals on measuring software transactions. Paper presented to the 22-23 April meeting of the OECD Software Task Force, Paris.

Castles, I. (2000). Comments on use of PPP and exchange rate conversion in 1999 Human Development Report. Correspondence to Friends of the Chair Review Group.

Chatfield, C. (1996). *The Analysis of Time Series (5th edition)*. London: Chapman and Hall.

Colecchia, A., and Schreyer, P. (2001). ICT Investment and Economic Growth in the 1990's: Is the USA a Unique Case. OECD.

Edwards, R., Comisari, P. and Johnson, T. (2002). Beyond 1993: The system of national accounts and the new economy. *Proceedings at IAOS Conference on Official Statistics and the New Economy*, London.

United Nations (1992). *Handbook of the International comparisons programme*. United Nations, New York.

Unsc (1998). *Evaluation of the International Comparison Project*. UN Statistical Commission, New York.

Unsc (2001). An assessment of the criticisms made of the human development report, 1999. UN Statistical Commission, New York.

Unsc (2002). An assessment of the statistical indicators derived from United Nations summit meetings. UN Statistical Commission, New York.

Wyckoff, Andrew W. (1995). The impact of computer prices on international comparisons of labour productivity. *Economics of Innovation and New Technology*, 3, 2, 277-293.