

IZA DP No. 1263

**Data Issues and Databases Used in Analysis
of Growth, Poverty and Economic Inequality**

Almas Heshmati

August 2004

Data Issues and Databases Used in Analysis of Growth, Poverty and Economic Inequality

Almas Heshmati

*MTT Economic Research
and IZA Bonn*

Discussion Paper No. 1263
August 2004

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
Email: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit company supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Data Issues and Databases Used in Analysis of Growth, Poverty and Economic Inequality*

This paper focuses on the importance data issues to the analysis of growth, poverty and economic inequality. We introduce a number of major databases frequently used in applied research on growth, poverty and global and international inequality. A discussion of data quality, data consistency, variable definitions and measurement, changing population and household size leading to various necessary scale and price adjustment procedures will follow. The sampling design and various dimensions of sample dependency are also discussed. Based on the existing experience we identify a number of factors that are important in applied research. The focus is on these factors which impacts on reliability, precision, sensitivity and consistency of the results and conclusions drawn.

JEL Classification: C10, D31, D63, I32, N30

Keywords: income inequality, poverty, growth, decomposition, databases, adjustment,
Gini coefficient

Almas Heshmati
MTT Economic Research
Luutnantintie 13
00410 Helsinki
Finland
Email: Almas.Heshmati@mtt.fi

* An earlier version of this paper was completed while I was working at the World Institute for Development Economic Research, UNU/WIDER. I am grateful to Amit Kumar Bhandari for valuable comments and suggestions on an earlier draft of this paper.

1. INTRODUCTION

Availability of more and better quality data on income distribution play a major role to the ongoing and increasing interest in measuring and understanding the level, causes and development of income inequality and poverty. The 1990s was signified by a shift in research previously focused on economic growth, identification of the determinants of economic growth and convergence in per capita incomes across countries to analysis of distribution of income, its development over time and identification of factors determining the distribution of income and poverty reductions. This shift is among others a reflection of increased availability of data and awareness of the growing disparity and the importance of redistribution and poverty reductions. The growing disparity calls for analysis of various aspects of income inequality and poverty including their measurement, decomposition, causal factors and issues related to data underlying such analysis.

Literature on income inequality refers to the inequality of the distribution of individuals, household or some per capita measure of income. The most widely used index of inequality is the Gini coefficient (for reviews of economic inequality see Subramanian 1997, Cowell 2000 and Heshmati 2004a). The other notable measures of inequality are: the range, the variance, the squared coefficient of variation, the variance of log incomes, the absolute and relative mean deviations, and Theil's two inequality indices. Inequality can have many dimensions. Economists are concerned specifically with the monetarily measurable dimension related to individual or household incomes. However, this is just one perspective and inequality can be linked to inequality in skills, education, opportunity, happiness, health, life expectancy, welfare, assets and social mobility. There are several studies focusing on analysis of the interrelationship between inequality in earnings, poverty, education, growth and health. The measurement of inequalities is beyond the scope of this paper. Heshmati (2004a) reviews the recent advances in the measurement of inequality and gives attention to the interrelationship between income and non-income dimensions of inequality.

The extensive literature emerging in recent years are focused on the study of how the distribution of incomes across countries and globally has developed over time. The increased interest in income inequality may be both cause and effect of the availability of data. Availability of household surveys has been improved and several standardized databases have been created. These allow analysis of income distribution at the most disaggregate individual or per capita household levels. Income distribution is otherwise often analyzed at four levels of aggregation, namely global, continental or regional, international and intra-national levels. Income inequality can also be measured at a within-country regional level. Here the focus is on inequality in income distribution in large countries like China, India, Russia and US where regions include states, provinces, federations or distinct geographic locations within a country. Empirical studies at each level are extremely data intensive and which are based on the second half of the twentieth century.

A number of databases are frequently used in applied global, international, regional, intra-national growth, poverty and income inequality research. The descriptions of such databases are found in Summers and Heston (1991), Barro and Lee (1996), Deininger and Squire (1996) and Atkinson and Brandolini (2001). Despite the difference in objectives, country coverage, sources, units of measurement and income definitions,

each plays an important and complementary role in growth and inequality studies. Each set of data has of course benefits and limitations. The choice of data seems to affect the conclusions drawn about trends in inequality. Quality, consistency, income definition and measurement, changing population, household size and various variable adjustment procedures are among important issues that are discussed. Consistency across countries and over time is a key issue to the reliability of results. The measurement of income, its decomposition by source and sub-groups with satisfactory coverage, sampling and representation of population are important in the generation of consistent information about the state of income inequality. An absence of such properties might result in misleading or spurious results.

This paper is on the importance data to the analysis of growth, poverty and economic inequality. We introduce a number of major databases frequently used in applied research. A systematic discussion of various aspects of each of these data follows. Based on existing experience we identify a number of factors that impacts on reliability, precision, sensitivity and consistency of the results and conclusions drawn. This review of its comprehensive analyses of various issues of data makes an important contribution to the literature of economic inequality.

Rest of the paper is organised as follows. First we introduce a number of major databases frequently used in applied growth, poverty and global and international inequality research. In Section 3 a discussion of data quality and data consistency is given. Section 4 is on variable definitions and measurement. The issues of changing population and household size leading to various necessary scale and price deflator adjustment procedures, adjustment for unit characteristics and price regimes are discussed in Section 5 to 7. Sampling design and various dimensions of sample dependency are explored in Section 8. The last Section summarises the findings.

2. MAJOR DATABASES

Secondary cross-national data sets play an important role in empirical economic research offering cross-country panel data analysis. The Summers and Heston (1991), Baro and Lee (1996), Deininger and Squire (1996), World Income Inequality Database, and other by World Bank, United Nations and Luxembourg Income Study (LIS) are among the widely used secondary data sets containing information on education, population, trade, finance, health, politics and income distribution. The premise and pitfalls in the use of such data-sets and how they can be best assembled and used in income inequality (in OECD countries as a case study) is discussed by Atkinson and Brandolini (2001). The primary concern of Atkinson and Brandolini is with data quality and data consistency. These are two important but in their view neglected issues in most applied research. In the following we provide a brief description of each of the important data sets listed above.

The Barro and Lee Data Set

Many researchers have attempts to construct measure of educational attainment for international comparison. Education, as measure of human capital, has been a major variable used intensively in the specification of growth models and in explaining the variations in countries and regions the growth rates. Most widely used such data is given by Barro and Lee (1996). This data has been updated to cover 109 countries with

complete information observed at five years intervals for the years 1960-1995 and projections for 2000. The data set provides the estimates of age groups over age 15 and 25 and also breakdown by sex. It also includes the number of years of schooling achieved by the average person at the various levels and at all levels of schooling combined. A full description of the updates and implications together with discussion about the estimation methods for the measure of educational attainment is found in Barro and Lee (2000). These estimated values are also related to alternative international measures of human capital stocks (international test scores by students, international adult literacy test, estimates of market value of human capital, and OECD estimates of educational attainment). The complete data is available and freely downloadable from the web-site of the Center for International Development at Harvard University (<http://www.ksg.harvard.edu/CID>).

The Penn World Tables

The Penn World Tables (PWT) is probably the data set that most frequently used in Growth and income studies. It is also known as the Summers and Heston (1991) data. The PWT (<http://www.nber.org/pwt56.html>) provides statistical information in form of a set of national accounts economic time series covering 152 countries and 29 subjects including macro, industry, international trade, patent and others. The National Bureau for Economic Research (NBER) provides interface for retrieving subsets of data from the tables. A menu guides researchers through the process of selecting countries, data types, and date ranges.

The PWT database is particularly valuable for its historical depth. Another value is the relatively high quality and comparability of the data across countries. It follows the International Comparison Programme (ICP) effort to make cross-country comparisons possible on a consistent basis. Available data includes national income account statistics, balance of payments, populations, exchange rates, transportation, and standard of living indices. One important aspect of PWT is that it is based on ICP benchmark studies which provide expenditure data in a common set of parity prices including an overall purchasing power parity (PPP). An ICP benchmark study starts with a price survey of an identical set of goods and services in many countries. Drawing on the survey results, price parities for about 150 categories covering all of GDP are computed. These are used as weights in a complicated aggregation procedure to get country-specific price parities for various components and the overall PPP. The price parities and PPPs are used to convert individual countries' national currency expenditures to a common currency unit (most commonly US dollar) to make real quantity comparisons both between countries and over time. The estimates of capital stock like the human capital, factor inputs prices and price parities for each country's currency are seen as an invaluable contribution to researchers round the world.

Potentials and pitfalls in international price and quality comparisons with reference to PWT is discussed in Heston and Summers (1996). The ICP benchmark pitfalls are summarized as a number of difficulties. First, the quality of the benchmark comparisons is determined by the quality of surveys and the national accounts of the country. Second, heterogeneity in price comparisons of identical or equivalent goods across countries creates substantial matching problems. Third, the choice of an aggregation method for putting together the benchmark prices is based on consumption and national accounts data. Finally, the benchmark estimates are limited to the expenditure side of

national accounts which does not allow for sectoral (value added based) productivity comparisons. Despite the pitfalls, PWT provides a unique platform for working with both economic and non-economic variables to study development. Among important potentials of the PWT is that it allows real inter-country comparisons at different time periods. For a few applications based on PWT (Duro and Esteban, 1998) in the context of decomposition of cross-country income inequality, Bhargava, Jamison, Lau and Murray (2001) who modelled the effects of health on economic growth, and Park (2001) analyzes the trends in global distribution of income.

The Luxembourg Income Study

The Luxembourg Income Study (LIS) was created in 1983 under the joint sponsorship of the government of Luxembourg and the Center for Population and Policy Studies. The primary goal of this project is to promote comparative research on the economic status of populations in different countries. LIS is a compilation of large, national microdata on a broad array of employment income, taxes, transfers, demographic, expenditure, and occupation data. There are now more than 20 OECD countries that are supplying data to the LIS. Work on adding data from more developed countries is in progress. However, the member countries do not supply data on an annual basis, as survey data are not collected consecutively every year. These data are therefore arranged in a number of five years waves, 1979-1983, 1984-1988, 1989-1994, and 1995-1999. It permits both cross sectional and cross temporal analysis. The LIS data are available (<http://www.lisproject.org/>). User support files are available via an anonymous FTP site to acquire data information. For the use of LIS data and description of various dimensions of the data see several recent studies such as: Jäntti and Danziger (2000) who studied income poverty in advanced countries, Atkinson and Brandolini (2001) evaluating the use of secondary data in the study of income inequality, Iacoviello (1998) on income dynamics, Mahler (2001) on income inequality and Ritakallio (2001) on trends of poverty and inequality among OECD countries.

The World Income Inequality Database

The WIDER-UNDP World Income Inequality Database (WIID) is an expanded version of the Klaus Deininger and Lyn Squire 1997 World Bank database.¹ The full description of the data is found in Deininger and Squire (1996). The expanded version contains information on income inequality, income shares, and a number of variables indicating the source of data, and quality classification for 151 developed, developing and transition countries observed on an irregular basis mainly covering the period post 1950 until 1998. The current version of the database, Version 1.0 released on September 2000, is in an easily retrievable, exportable, and analyzable format. It is publicly available in the WIDER web-site (<http://www.wider.unu.edu/wiid/wiid.htm>) for browsing and free downloads to facilitate further analysis and debate on global income inequality. The document fully describes the data, background information, quality guidance and instructions is available for download on the above web-site.

For a limited number of applications for using this data set and its earlier versions are: Deininger and Squire (1996 and 1998) for description of the data, Cornia and Kiiski

¹ The data sources distributed by their share of total number of observations are as follows: WIDER (42.7%), Deininger and Squire (44.8%), and remaining sources (UNICEF/ICDC, LIS and WB HEIDE) 12.5%.

(2001) on inequality, growth and poverty in the era of liberalization and globalization, Cornia and Court (2001) who studied trends in income distribution in the post World War II period, Atkinson and Brandolini (2001) on the use of secondary data sets, Wan (2002) on income inequality and growth in transition economies, Biancotti (2003) on polarisation of distribution of income, and Li, Squire and Zou (1998) on variations in income inequality. The later data has drawn on Deininger and Squire (1996) was also used by Ravallion (2003) to test inequality convergence. The WIID data has been utilised by Heshmati (2003) to study the impact of globalizations on income inequality. Other cases of application based on WIID are analysis of the world distribution of income and income inequality (Heshmati 2004c) and analysis of continental and sub-continental income inequality (Heshmati 2004d).

Galbraith and Kum (2003) suggest the data published by the University of Texas on Inequality Project (UTIP) which is based on the Industrial Statistics database published annually by the United Nations Industrial Development Organization (UNIDO) as an alternative global inequality dataset. Deininger and Squire (2002) present an update of their 1996 dataset. The new data cover more observations with increase coverage for developing and transition countries and is based on micro level surveys.

The World Development Indicators

World Development Indicators (WDI) is the premier data source of World Bank available on global economy from 1960 onwards. However, several series of large number of developing countries are missing prior to 1970. WDI is annual compilation of data about global development. WDI 2002 includes approximately 800 indicators in 87 tables, organized in six sections: World View, People, Environment, Economy, States and Markets, and Global Links. The tables cover 152 economies and 14 country groups with basic indicators for another 55 economies. The Economy data part contains statistical data for over 550 development indicators and time series data from 1960-2001 for 207 countries and 18 country groups. The data includes social, economic, financial, natural resources, and environmental indicators. More details on the data and annual subscription are found in the World Bank web site (<http://www.worldbank.org/data/>). For a few applications using WDI see: Anand and Kanbur (1997) in the context of inequality and development, Bhargava (2001) who examine stochastic specification of the international GDP series, and Addison and Heshmati (2003) who studied the global determinants of FDI flows to the developing countries.

The Human Development Indicators

Recently World Bank also produces tables on Poverty and Hunger for countries which included in the HDI database (<http://www.developmentgoals.org/Data.htm>). The tables provides information about people living on below \$1 a day, poverty gap at \$1 a day, percentage share of income or consumption held by 20% of the poorest and prevalence of child malnutrition. Description of sample survey data from 91 countries has used by Milanovic (2002a) is also found on the World Bank web-site on available Research Data sets (<http://econ.worldbank.org/>). For studies of income mobility based on the World Bank World Tables has produced by Parker and Gardner (2002), Ravallion (1997) on economic growth and human development, and Noorbakhsh (1998) who computed human development index.

The Millennium Indicators Database

On an annual basis the United Nations Development Program (UNDP) in its Human Development Report (HDR) also publishes Tables on various issues like poverty, income inequality, education, health, etc. The 2003 UNDP HDR (2003) published Tables on Millennium Development Goals (MDG) indicators.² The United Nations Statistics Division maintains the recently created global Millennium Indicators Database (<http://millenniumindicators.un.org>). The data is compiled from international data series provided by international data agencies. The database forms the statistical basis for annual report to the UN General Assembly on global and regional progress towards the Millennium Development Goals and their targets. The database provides updated data to the UN system organizations, various research institutes and individual researchers to be used in research related to the MDGs. For a recent study of the progress towards the MDG in Africa see Sahn and Stifel (2003).

3. DATA QUALITY AND DATA CONSISTENCY

The original sources of data described above are often found in the national household surveys, administrative income tax record data, social security and labour market agency records. Each source has its benefits and limitations. For instance, income tax record is often limited to incomes exceeding a certain level, while household surveys may not be representative of the population studied.

Some data sources like the LIS data is related to standardised and internationally comparable micro-data-sets. Despite an increased data availability and improved data quality, there is a gap between theoretical advancement on determinants of income and wealth and their distribution and the empirical investigations due to lack of sufficiently detailed high quality data on distribution. To mention an example, Elbers, Lanjouw and Lanjouw (2003) provides micro-level estimation of poverty and inequality based on expenditure per capita using Ecuadorian data that have levels of precision comparable to those of commonly used survey based welfare estimates. This is an improvement over survey based estimates which are only consistent for very large number of households.

In the following section we focus on secondary data sources. A comparison of different data sources show that the choice of data affects the conclusions drawn about trends in income inequality for a single country over time. For instance, differences in income concept, income source, measurement unit and methodology may affect not just the level but also the trend in income inequality. In an attempt to link secondary data sets Atkinson and Brandolini (2001) compare four secondary data sets including the studies of United Nations in the 1950s and 1960s, the World Bank compilation in the 1970s, the Deininger and Squire data-set, and the WIID data-set for the period 1945 to 1970 but limiting their attention to the (European) OECD countries.

² In the 2000 UN Millennium Declaration, countries committed to work together to meet concrete 8 goals consisting of 18 targets for advancing development and reduction of poverty by 2015 or earlier. The MDGs are to: (i) eradicate extreme poverty and hunger, (ii) achieve universal primary education, (iii) promote gender equality and empower woman, (iv) reduce child mortality, (v) improve maternal health, (vi) combat HIV/AIDS, malaria and other disease, (vii) ensure environmental sustainability, and (viii) develop a global partnership for development.

Atkinson and Brandolini (2001) analyzed the secondary data sets on income distribution which make a number of suggestions regarding the construction, development and the use of secondary data sets. They conclude among others that such data should be a fully documented cumulation of earlier work and provide a complete picture of available information to their users. Guidelines are given regarding the choices to be made in defining the distribution under consideration, sources of the data and ways to deal with differences in definition across countries or over time. The authors find the differences in definitions quantitatively important, but doubt whether simple adjustments are satisfactory solution to the heterogeneity of the available statistics.

Country heterogeneity is an important issues that analyst should account for. Estimation of unobservable country heterogeneity may not be a satisfactory measure in pooling data. Lloyd, Morrissey and Osei (2001) in testing validity of pooling countries in large panel in empirical analysis of the aid and trade relationship find significant differences in results for distinct panels. If the nature of the time series relationship underlying specific observation differs, these should not be pooled into one panel. Results from such data might be misleading if not spurious. Going back to micro-data authors highly recommended the use of data sets where the observations are as fully consistent as possible. Several arguments for the use of consistent micro-data can be found in for instance: Heady, Mitrakos and Tsakloglou (2001) who analyse the comparative effects of cash transfers on inequality and poverty by using consistent household data, Foster and Shorrocks (1991) on subgroup consistent measures of poverty, and Salas (1998) on welfare-consistent inequality indices in changing populations.

Many secondary data sets contain multiple observations. The presence of multiple observations for the same country and at the same period might be due to differences in definition, levels, coverage, weights or methods of calculation. The definition of income may differ by source across country and time. The problem of consistency across country and over time is a key issue to the reliability of results in country and comparison studies based on panel data. The issue of consistency is in particular emphasised in the WIID database. However, accounting for consistency in a very strict way leave researchers with a very small sample of countries and observations inadequate for analysis of income inequality at the global level.

4. VARIABLE DEFINITIONS AND MEASUREMENT

In this section the focus is mainly on key variable containing the WIID database. The choice is motivated by the fact that WIID embodies several other data sources. The WIID database contains information on Gini coefficient (measured in percentage points), income shares (accruing to the specified part of the population), and the number of variables indicating the source of data including area coverage, population coverage, income definition, source of the data³, and quality classification. Some of the Gini

³ A bibliography on the original sources of each data point is provided on the WIDER web-site. For time-series or cross-sectional analysis based on WIID, one should verify the consistency of the data sources for chosen countries across time and make necessary corrections to make the national data comparable across countries in terms of income concepts and reference units. However, it should be noted that Atkinson and Brandolini (2001) in their review of the secondary data sets doubt whether simple adjustments are satisfactory solution to the heterogeneity of the available statistics.

estimates originated from the source, while others have been computed from grouped data at UNU/WIDER using mainly POVCAL software.⁴

Different concepts of income are used in the WIID database. For instance, income is measured as: (i) gross/factor/market income, (ii) net income including/excluding non-taxable transfers, (iii) gross/net monetary income, (iv) gross/net taxable income, (iv) gross/net earnings, (v) gross/net expenditure, (vi) disposable income, and (vi) pre/post tax and transfer income.⁵ The concepts of multidimensionality and complexity of the income is a good indicator of the many consistency and quality problems associated with the use of such databases. Lack of consistent and high quality micro data across countries does not leave a better choice to the researchers.

Income can be decomposed into different income sources used in income inequality measurement by income sources (for examples, such decomposition and comparison see Shorrocks (1982), Lerman and Yitzhaki (1985), Cheng (1996), Liebbrandt, Woolard and Woolard (2000), Milanovic (2000), Wan (2001), Morduch and Sicular (2002), and Podder and Chatterjee (2002)). Understanding the key sources of inequality is important. The inequality can be linked to the cause of income sources. Blacklow and Ray (2000) in their comparison of income and expenditure inequality estimate used Australian unit record data found inconsistencies between the two inequality movements. The within household groups inequality dominated the between groups component.

A decomposition analysis by income sources throws new lights on the identification and the process of generating income inequality. Lerman and Yitzhaki decomposed income into: head wages and salaries, head self-employment earnings, spouse earnings, other family earnings, transfers and property income sources. Cheng used household survey data from 1994 collected from 5 grain producing provinces in China to study the impacts of agricultural activities on inequality in non-agricultural cases by province and various sources of household income. Wan studied changes in regional inequality in rural china by the measures of decomposing inequality change by the Gini index. He divided the Gini index into: structural effects, real inequality effects, interactive effects and aggregate full contribution from wages, farming and residual income components. Milanovic in testing the median-voter hypothesis, income inequality and income redistribution used household budget survey data on factor, gross and disposable income definitions. Results showed that different concepts highlight different aspects of distribution.

Liebbrandt, Woolard and Woolard (2000), in their analysis of the contribution of various income (remittances, wage, capital, state transfers, agriculture and self-employment) components to income inequality in the rural former homelands of South Africa found that wage income is both the main income component and also the most important source of inequality for the African population. To improve wage earnings

⁴ POVCAL is a statistical program designed by Shaohua Chen, Gaurav Datt, and Martin Ravallion at DEC-RG, World Bank. It is an easy to use and reliable tool for routine poverty assessment work. It uses sound and accurate methods for calculating poverty and inequality measures with only a basic PC and grouped distributional data. For an overview of the program, data requirement, and free downloads see the web-site: <http://www.worldbank.org/html/prdph/lmsms/tools/povcal/>

⁵ Market income includes the earnings of all household members and all incomes from interest, dividends, rents and other market sources. Disposable income is equal to market income plus transfers minus taxes.

the policy induced change is that they should target the rural labour market and wages. The linkage between the labour market and inequality are analysed by using a panel of African households in Kwazulu-Natal studied by Liebbrandt and Woolard (2001) showed that the contribution of wage income is uneven across different levels of aggregation and across time suggesting complex patterns of inequality generation.

The reference unit when referring to an income distribution, often labelled as unit of measurement or income recipients can be different from one data set to another. It is often defined as: (i) household or family, (ii) household or family equivalent (scale squared, social assistance, OECD, Oxford and HBAI), (iii) person, household per capita, head of household, or simply the tax unit. Here family is defined as family plus unrelated individual family excluding non-dependent.

The sampling unit can be a person, household or family. The enumerated unit determining the inequality ranking by person, household, family, or head of household. In some cases the person is both sampling and enumeration units, while in other the household is the sampling unit, but following per capita equalisation procedure the person is enumeration unit. The distribution in the former is the true person distribution, while in the later inconsistency may appear as a result of the fact that the distribution consists of multiple observations within a household.

The area coverage is another key factor affecting the sample representativeness of the population. It refers to the land area included in the original sample surveys: (i) total land, (ii) urban, capital, cities, metro, (iii) rural, rural interior or regions.

The population coverage refers to the population covered in the sample surveys in the chosen land area. It include: (i) all, (ii) different shares of all, all excluding different sub-groups, (iii) different sub-sectors, (iv) income recipients, (v), taxpayers, or other sub-groups of the population. Inequality can be decomposed by different subgroups (Shorrocks 1984) where the overall inequality can always be calculated from the size, mean and inequality value of each population subgroup. For instance Li (2001) examined changes in poverty profile in China by localities and population sub-groups. Chakravarty (2001) studied the variance of a subgroup decomposable measure of inequality. At more aggregate level inequality decomposition can be made by rural-urban or agriculture-non-agricultural industrial and service sectors.

In order to improve the reliability of results from secondary data sets, the WIDER data points have been assigned to two broad categories 'Reliable' and 'Less Reliable' data observations. The assignments were based on various sources of information obtained from the primary information. The primary information concerns: survey composition, sampling methods, time period, non-response, weighting methods, income definitions, coding and various statistical adjustments. The Reliable earnings, income or expenditure data refer to entire national, rural or urban populations of a country. It further refers to the employed, economically active, or other such subsets of the entire, rural or urban populations of a country. Labelling an observation as Less reliable data reflects presence of missing information, inconsistencies, errors in grouping or estimation methods, small population coverage, or other factors limiting the data quality.⁶

⁶ Detailed information on micro data collection, sampling, data quality, documentation and various aspects of data sources can be found for instance at: <http://psidonline.isr.umich.edu/> on Panel Study of

The variable describing the data source is a variable indicating the source from which the observation value has been obtained. It is divided into primary and secondary sources. Most databases used in cross-country comparisons belong to the second class of data sources.

5. EQUIVALENCE SCALES

Households differ in many respects other than income like the size and the composition which affect the well being of the well being of the household members. Heterogeneity in actual and estimated consumption needs by age of member of household and the presence of economies of scale due to the household size, makes it necessary that some equivalence methods of scales be used. The method allows comparison of households of different sizes and age compositions and calculation of effective household size by assigning the household members a weight. The assigned weight is based on representative consumption amounts. The Gini coefficient will differ by the applied equivalized households or scale methods of individual members. In addition to data comparability, the equivalence scale reflects the cost of additional members for a household by comparing the before and after household size increases.

There is a number ways to construct an equivalence scale. The first measure is an overestimated extreme equivalence scale, where income equal to the total amount of household income assigned to each household member, i.e. a weight of 1 is given to all members. A second extreme measures is an underestimated scale obtained by dividing the aggregate income by the number of households members (n), i.e. giving a weight of $1/n$ to each member. In the later case in the estimation of the income distribution adjustment is made for the household size. Several other but intermediate scale measures are possible to use. One smooth measure is the square-root-scale where the weight assigned is $1/\sqrt{n}$. Another intermediate possibility is to assign heterogeneous weights, $\sum_{i=1}^n w_i$, to different household members (i) by age and gender. For instance, to assign the first adult in a household a weight of 1, the second (or other) adults 0.7 and each child a weight of 0.5, or alternatively a range of weights according to the ages and needs of the children. Blacklow and Ray (2000) using Australian data finds income and expenditure inequality estimates sensitive to the equivalence scale used as the household size deflator.

There is a desire that the differences among individuals ability in and needs be reflected in the designing of tax-and-transfer systems. As a measure Atkinson and Bourguignon (1987) and Jenkins and Lambert (1993) developed sequential Lorenz dominance in multi-dimensional framework to rank income distributions when needs differ. Mayshar and Yitzhaki (1995) suggested a method for identifying indirect tax reforms that adhere to uni-dimensional criterion for a Dalton's principle of transfers, approving a small enough transfer from rich to the poor. Mayshar and Yitzhaki (1996) extended the approach to a two-dimensional criterion for a Dalton-improving tax reform, when there is social approval for transfers from the more able to the less able and from the less needy to the more needy households. Households' size, age composition, or health

Income Dynamics, at <http://www.diw.de/english/sop/index.html> on The German Socio-Economic Panel Study, and at the World Bank data-site <http://www.worldbank.org/poverty/data/index.htm> for various data sources and from different countries.

status are the indications of heterogeneous needs and are distinct from heterogeneous abilities. The later approach and its application and comparison with other criteria is carried out by using the UK household expenditure and tax data. Cowell (2000) in his analysis of income distribution as a multivariate problem, proposes that the households or families should be distinguished by characteristics other than income referring to income sources and needs. Recently, Ebert and Moyes (2003) proposes a new and normative approach for adjusting household income by introducing an equivalent income and size adjusting functions in order to compare the living standards of households with different needs.

Despite various adjustments and modification of household data like those discussed above, few researchers explore motivations for measurement of inequality. Kaplow (2002) criticises the large body of literature devoted to various measures or indices of income inequality, in which little attention is given to the question to measure inequality. The reasons for measurement will determine whether and how measurement should be done. Descriptive measures (used for instance in the regression analysis of relationship between inequality and growth) and Normative measures (designed to have direct policy relevance) are examined by Kaplow. The latter depending on the economic theory and field of empirical application may be useful and appropriate measure of inequality. Kaplow emphasis that more attention to be devoted to the purposes of measurement, which affects how practice of measurement should be undertaken.

6. ADJUSTMENTS FOR UNIT CHARACTERISTICS

At the micro level changes in the composition of families is an important factor that must be accounted for in the analysis of distribution of income of households. Recently, US income inequality by sub-groups like gender, marital status, full/part-time employment, over time, and contribution of growing wage disparities and changing family composition on the overall income inequality between 1979 and 1996 estimated by Burtless (1999). Burtless examines the trend in overall inequality using the concept of adjusted equivalent personal income. The definition of equivalent income used is written as: $Y_A = Y_U / (F / H) \times H^\theta$ where Y_A is the adjusted money income per person, Y_U is the unadjusted total family income, F is the number of persons in nuclear family unit, H is the number of persons in household, and θ is the assumed adjustment parameter for family size. It can also, as previously mentioned, to reflect the needs of individual family members. If family and household sizes are equal, $F=H$, then $Y_A = Y_U / H^\theta$ and in a single-person household, $F=1$, $Y_A = Y_U$. The adjustment resulted in a rise in the Gini coefficient of family income inequality from 0.365 to 0.425 or about 16.4% increases. While growing pay disparities especially among the men is the direct contributor to the trend in overall inequality, much of the rise is found to be due to the shifts in family composition and other causes. The impact of growing correlation of husband and wife earned income and the increasing percentage of persons who live in single-adult families and with more unequal incomes on overall inequality are found to be significant. Burtless used US Current Population Survey (CPS) files show that the higher gender earnings disparity, growing positive correlation of income within families, and growing proportion of families with single adults explains 33-44%, 13% and 21-25% of the increase in overall inequality, respectively.

There are numerous empirical cases of adjustment procedures to incomes. One such case is Biewen (2002) on dynamics of equivalent incomes in Germany. Another is Jäntti and Danziger (2000) on equivalent scale for aggregate and different population subgroups in industrialized countries. Paraje (2002) adjusted income for adult equivalent and economies of scale to compute polarisation measures in the Greater Buenos Aires. Finally, Schmidt (2002) analyzed income polarisation using cross-national equivalent files.

In earlier discussion we showed that the age composition in a household resulted in heterogeneity in the needs by households requiring adjustments and scales. In addition to the heterogeneity impacts of the within household age composition, the distribution of age at the aggregate level has implications for cross-country studies. Based on micro data on self-assessment of health status Gerdtham and Johannesson (2000) estimated income-related inequality in Sweden with respect to life-years, health status and quality-adjusted life-years (QALYs). Survival and quality-adjusted survival years in different income groups is estimated. Results indicate that inequalities in health favouring the higher income groups. Health inequality increases with age. Income affects health and the other way around. Income is correlated with other factors that affect health such as education and unemployment. Gerdtham and Johannesson's results show that the inequality comparisons between countries may be affected by differences in the countries age distribution.

At the macro level changing population and its impacts on the performance of inequality indices in a single dynamic analysis and in cross-country static welfare analysis explored by Salas (1998). The standard welfare-consistent inequality indices satisfy the marginal population replication axiom or marginal entrants. These indices are found to perform well when comparing inequality over different population distributions. As a further line of research Salas recommends analyses of the performance of population subgroup decomposable inequality indices under changing subgroup populations.

Various studies have found large differences and conflicting results in terms of how income distribution in the world has changed. Svedborg (2002) attributes the differences to the fact that those studies employ: (i) different measurements of income, (ii) different measurement of income distribution, and that (iii) they weight/select countries and populations differently. In comparison of countries many studies as measurement of income have used: national per capita GNP converted to US dollars according to current official exchange rates or national GNP per capita adjusted for differences in purchasing power across countries. The later is more reliable by taking into consideration for the differences in the relative prices of goods and services sold locally or internationally by the poor or rich countries.

The most frequently used alternative measurement of income distribution are: the ratios of the per capita incomes in the richest and poorest countries at certain times called Kuznets ratios, and the income distribution over the entire range of countries, such as Gini coefficient. None of the two measures in Svedborg's view however, capture transitions or changes in position of countries in per capita income ranking over time. Estimation of changes in world income distribution must be taken into consideration for countries population size and its variations over time. Then the Gini coefficient will reflect changes in per capital incomes across countries, as well as changes in respective

countries' share of the world's population. Most studies have used different measurement of the relative distribution neglecting how the absolute differences (gap) in income between countries have developed over time.

As mentioned previously, data sources and income definitions may result into different and conflicting inferences. Capeau and Decoster (2003) explain the driving forces behind the differences in the two extreme positions on rise (Sala-i-Martin 2002a and 2002b) and fall (Milanovic 2002a and 2002b) of inequality by income measures, population weighted inequality measures and inequality among citizens irrespective of location. See also Atkinson and Brandolini (2001) for discussion of the use of several different databases and their implications for inferences about the level and temporal patterns of income inequality.

In regression analysis a typical way to deal with differences in definition across countries, or across time is by introducing dummy variables to adjust (additively or multiplicatively) for differences in say income definitions. There is a risk that selection of countries and years may bias the result. The Deininger and Squire (1996) propose comparison of paired estimates, i.e. comparison of gross and net incomes for the same country at the same period as preferred method. Furthermore, the dummy adjustment may also be unsatisfactory for demographic factors. Atkinson and Brandolini (2001) conclude that simple dummy variable adjustments mentioned above are unsatisfactory solutions to country, income definition and time period heterogeneity in the existing data. They prefer the use of data sets where observations are as fully consistent as possible.

7. ADJUSTMENTS FOR PRICES REGIMES

The cost of living varies across regions which has implications for comparisons of welfare levels of households situated in different regions within a country. Such cost heterogeneity has been neglected in welfare programs. Construction of bilateral and multilateral spatial cost of living and welfare indices developed by Kakwani and Hill (2002) is applied to construct urban and rural cost of living and welfare indices for five regions of Thailand. This price adjustment has implications for poverty and inequality measurement. Maasoumi, Diamond, Nieswiadomy and Slottje (1994) also studied the effects of relative price changes and cost of living adjustment on welfare indices.

Among other examples of empirical cases of price adjustment to mention are Ravallion and Chen (1999) who used Chinese provinces micro data find that two-thirds of the conventionally measured increases in the post reform inequality vanishes when market-based valuation methods are used and allowances are made for regional cost-of-living differences and structural changes in the rural economy. Non-farm income, grain production and higher returns to farmland increased inequality, while lower returns to physical capital and private transfers reduced inequality. Declich and Polin (2002), in analysing absolute poverty in Italy have taken into account for possible effects of different costs of living in the northern and southern areas. The regional variation in the cost of living is accounted for the cost the regional inequality.

Pendakur (2002) finds the standard adjustment for price regimes by using expenditure-independent price deflators and price-independent equivalence scale not fully adequate adjustment measures. Pendakur shows that more flexible expenditure-dependent price

deflators and price-dependent equivalence scale affects both the level and trend in measuring family expenditure inequality in Canada over 1969-1997. In the later rich and poor and large and small families are allowed to respond to price changes differently. The price deflators vary across regions and overtime. Unlike the cases described above Blacklow and Ray (2000) using Australian data did find income and expenditure inequality estimates sensitive to the equivalence scale used as the household size deflator, but not to the cost of living index which used as the price deflator.

8. SAMPLING DESIGN AND SAMPLE DEPENDENCY

The techniques of simple random sampling are seldom appropriate in the empirical analysis of income distribution. Cowell and Jenkins (2000) in this relation has draw attention to potential problems and proposes various types of weighting schemes required from the point view of welfare-economics, sampling design or estimation problem considerations. The different types of weights can have different implications for the sampling distribution of estimated welfare indices. The weights reflect endogenous and exogenous attributes that may affect the way in which we choose to aggregate the information about the income distribution. The attributes here refer to household size and partitioning households into internally homogenous subgroups. The extent of bias on the estimated welfare index will thus depend on the degree of heterogeneity of population by size, asymmetry in the composition of sample and the sample variance.

Conventional inference procedures assume independent samples when testing for inequality changes over time. The issue of inference for testing the interpolated Gini coefficient and the generalised entropy class indices with dependent samples is examined by Zheng and Cushing (2001). They establish inference tests for changes in inequality indices with completely independent samples (PSID), and then generalise it to cases with partially dependent samples (CPS). The effects of sample dependency on standard errors of inequality changes are examined through simulation studies and applications to the CPS and PSID data.⁷ An inspection of the results suggests that sample dependency has substantial effects, in the range of 3.3 to 17.1 per cent, on the standard errors if the samples are strongly correlated/or have significant overlap requiring a method to correct for sample dependency. Yitzhaki (1991), Karagiannis and Kovacevic (2000), Ogwang (2000) and Giles (2002) are a few recent examples of proposing methods to compute the Gini index and to calculate the (Jackknife) variance estimator for the Gini coefficient. In the latter case the Gini measure is obtained from an artificial OLS or seemingly unrelated regressions. For the empirical part the Penn World Tables five-year interval data on 133 countries in the years 1970-1985 is used.

The issue of statistical inference for inequality and poverty measurement with dependent data is also discussed by Schluter and Trede (2002). This concerns another dimension of dependency, namely data contemporaneously dependent across members of the same household. Much empirical research is based on methods assuming that income is an independent and identically distributed random variable. Application of

⁷ For recent examples of sample dependency and block bootstrapping in testing for first and second order stochastic dominance based on balanced panel of PSID and Swedish household data-sets see Maasoumi and Heshmati (2003).

such methods to contemporaneously dependent data produces biased results. Monte Carlo results show that standard errors in the presence of dependence are larger regardless of the nature of the intra-household correlation. The authors develop distribution-free non-parametric methods of statistical inference when there are contemporaneous dependencies and consider Generalized Entropy index of inequality, Gini coefficient, Lorenz curves and common poverty indices, FGT (Foster, Geer and Thorbecke (1984)). In comparison of the performance of the proposed method to the standard ones assuming various degree of dependency, the conclusion is that there is a large gain in precision if one takes dependencies into account without any loss at the absence of dependencies. The sample size also matter for the biasedness of the Gini coefficient. Deltas (2003) finds a negative relationship between Gini coefficient and populations size. The Gini coefficient is downward biased in small populations when income is generated by common distributions like log normal and exponential. This is relevant especially when comparing Gini coefficient across sub-populations of small sizes.

9. SUMMARY

There exists number of databases that frequently are being used in applied global, international and regional growth and income inequality and poverty research. Although, these databases differ by objectives, country coverage, sources, unit of measurement and income definitions, yet each one play an important role in inequality studies and by serving as complement to each other. Each data have of course their benefits and limitations and the choice of data affects the conclusions drawn about trends in inequality for a single or groups of countries and over time.

Quality, consistency, income definition and measurement, changing population and household size and various variable adjustment procedures are among important data issues discussed here. Consistency across countries and over time is a key issue to the reliability of results in panel studies. However, an invalid pooling due to differences in the nature of the time series relationship underlying the data might result in misleading or spurious results.

Measurement of income, its decomposition by source and sub-group with satisfactory coverage, sampling and enumeration units and representative of population is important in generation of consistent information about the state of income inequality. Heterogeneity in ability, in actual consumption and consumption needs by age, gender and size of households and presence of economies of scale requires calculation of representative consumption amounts by using some measures of equivalence methods of scales. The cost of living varies by location primarily divided into urban, rural or regions. Empirical results show that estimates of income and consumption inequalities are sensitive to the cost of living index as the price deflator.

Higher gender earnings disparity, growing positive correlation of income within families, and growing proportion of families with single adults explains large parts of the increase in overall inequality. An incorrect assumption about independent and identically distributed income over time and contemporaneous dependent across members of household produces biased results. Few existing studies show that there is a large gain in precision if one takes dependencies into account.

REFERENCES

- Addison T. and A. Heshmati (2003), The new global determinants of FDI to developing countries, *Research in Banking and Finance*, forthcoming.
- Anand S. and S.M.R. Kanbur (1997), Inequality and development: a critique, in S. Subramanian (ed) *Measurement of inequality and poverty*, Oxford University Press, pp. 128-158.
- Atkinson A.B. and F. Bourguignon (1987), Income distribution and differences in needs, In: Feiwel G.R. (ed.), *Arrow and Foundations of the Theory of Economic Policy*, Macmillan, London.
- Atkinson A.B. and A. Brandolini (2001), Promise and pitfalls in the use of “secondary” data-sets: income inequality in OECD countries as a case study, *Journal Economic Literature* 39, 771-799.
- Barro R.J. and J-W Lee (1996), International measures of schooling years and schooling quality, *American Economic Review* 86(2), 218-223.
- Barro R.J. and J-W. Lee (2000), International data on educational attainment: updates and implications, Centre for International Development at Harvard University, CID Working Paper 2000:42.
- Bhargava A. (2001), Stochastic specification and the international GDP series, *Econometric Journal* 4, 274-287.
- Bhargava A., Jamison D.T., Lau L.J. and C.J.L. Murray (2001), Modeling the effects of health on economic growth, *Journal of Health Economics* 20, 423-440.
- Biancotti C. (2003), A polarization of polarization? The distribution of inequality 1970-1996, Bank of Italy.
- Biewen M. (2002), The covariance structure of the East and West German incomes and its implications for the persistence of poverty and inequality, *IZA Discussion Papers* 2002:459.
- Blacklow P. and R. Ray (2000), A comparison of income and expenditure inequality estimates: the Australian evidence, 1975-76 to 1993-94, *Australian Economic Review* 33(4), 317-329.
- Burtless G. (1999), Effects of growing wage disparities and changing family composition on the U.S. income distribution, *European Economic Review* 43, 853-865.
- Capeau B. and A. Decoster (2003), The rise or fall of world inequality: big issue or apparent controversy?, Unpublished manuscript.
- Chakravarty S.R. (2001), The variance of a subgroup decomposable measure of inequality, *Social Indicators Research* 53(1), 79-95.
- Cheng Y-S. (1996), A decomposition analysis of income inequality of Chinese rural households, *China Economic Review* 7(2), 155-167.
- Cornia G.A. and J. Court (2001), Inequality, growth and poverty in the era of liberalization and globalization, Policy Brief No. 4, Helsinki: UNU/WIDER.
- Cornia G.A. and S. Kiiski (2001), Trends in income distribution in the post WWII period: evidence and interpretation, WIDER Discussion Paper 2001/89, Helsinki: UNU/WIDER.
- Cowell F.A. (2000), Measurement of inequality, in Atkinson A.B. and Bourguignon F. (Eds), *Handbook of Income Distribution*, Volume 1, North Holland, chapter 2, 87-166.

- Cowell F.A. and S.P. Jenkins (2000), Estimating welfare indices: household weights and sample design, London School of Economics
- Declich C. and V. Polin (2002), Absolute poverty and the cost of living: an experimental analysis for Italian households, ISAE DP.
- Deininger K. and L. Squire (1996), A new data set measuring income inequality, *World Bank Economic Review* 10(3), 565-591.
- Deininger K. and L. Squire (1998), New ways of looking at old issues: inequality and growth, *Journal of Development Economics* 57, 259-287.
- Deininger K. and L. Squire (2002), Revisiting inequality: new data; new results, World Bank.
- Deltas G. (2003), The small-sample bias of the Gini coefficient: results and implications for empirical research, *The Review of Economics and Statistics* 85(1), 226-234.
- Duro J.A. and J. Esteban (1998), Factor decomposition of cross-country income inequality, 1960-1990, *Economics Letters* 60, 269-275.
- Ebert U. and P. Moyes (2003), Equivalence scales reconsidered, *Econometrics* 71(1), 319-343.
- Elbers C., J.O. Lanjouw and P. Lanjouw (2003), Micro-level estimation of poverty and inequality, *Econometrics* 71(1), 355-364.
- Foster J.E., Greer J. and E. Thorbecke (1984), Notes and comments: a class of decomposable poverty measures, *Econometrica* 52(3), 761-766.
- Foster J.E. and A.F. Shorrocks (1991), Subgroup consistent poverty measures, *Econometrica* 59(3), 687-709.
- Galbraith J.K. and H. Kum (2003), Estimating the inequality of household incomes: filling gaps and fixing problems in Deininger & Squire, UTIP Working Paper 2003:22.
- Gerdtham U-G. and M. Johannesson (2000), Income-related inequality in life-years and quality-adjusted life-years, *Journal of Health Economics* 19, 1007-1026.
- Geweke J. and M. Keane (2000), An empirical analysis of earnings dynamics among men in the PSID: 1968-1989, *Journal of Econometrics* 96, 293-356.
- Giles D.E.A. (2002), Calculating a standard error for the Gini coefficient: some further results, University of Victoria, Department of Economics, Working Paper EWP0202.
- Heady C., T. Mitrakos and P. Tsakloglou (2001), The distributional impact of social transfers in the European Union: evidence from the ECHP, *Fiscal Studies* 22(4), 547-565.
- Heshmati A. (2003), Measurement of a multidimensional index of globalization and its impact on income inequality, WIDER Discussion Paper 2003:69.
- Heshmati A. (2004a), Inequalities and their measurement, IZA Discussion Paper 2004:1219.
- Heshmati A. (2004b), A review of decomposition of income inequality, IZA Discussion Paper 2004:1221.
- Heshmati A. (2004c), The World distribution of income and income inequality, Unpublished manuscript.
- Heshmati A. (2004d), Continental and sub-continental income inequality, Unpublished manuscript.
- Iacoviello M. (1998), Inequality Dynamics: evidence from some European countries, Working Paper No. 191, Maxwell School of Citizenship and Public Affairs, Syracuse University.

- Jenkins S.P. and P.J. Lambert (1993), Ranking income distributions when needs differ, *Review of Income and Wealth* 39, 337-356.
- Jäntti M. and S. Danziger (2000), Income poverty in advanced countries, in Atkinson A.B. and Bourguignon F. (Eds), *Handbook of Income Distribution*, Volume 1, North Holland, chapter 6, pp.309-378.
- Kakwani N. and R.J. Hill (2002), Economic theory of spatial cost of living indices with application to Thailand, *Journal of Public Economics* 86, 71-97.
- Kaplow L. (2002), Why measure inequality?, NBER Working Paper Series 2002:9342.
- Karagiannis E. and M. Kovacevic (2000), A method to calculate the Jackknife variance estimator for the Gini coefficient, *Oxford Bulletin of Economics and Statistics* 62, 119-122.
- Lerman R.I. and S. Yitzhaki (1985), Income inequality effects by income source: a new approach and applications to the United States, *Review of Economics and Statistics* 67(1), 151-156.
- Li H., Squire L. and H. Zou (1998), Explaining international and intertemporal variations in income inequality, *Economic Journal* 108(466), 26-43.
- Li S. (2001), Changes in poverty profile in China, WIDER Discussion Paper 2001/21, Helsinki: UNU/WIDER.
- Liebbrandt M., C. Woolard and I. Woolard (2000), The contribution of income components to income inequality in the rural former homelands of South Africa: a decomposable Gini analysis, *Journal of African Economies* 9(1), 79-99.
- Liebbrandt M. and I. Woolard (2001), The labour market and household income inequality in South Africa: existing evidence and new panel data, *Journal of International Development* 13, 671-689.
- Lloyd T., O. Morrissey and R. Osei (2001), Problems with pooling in panel data analysis for developing countries: the case of aid and trade relationships, CREDIT Research Paper 2001:14.
- Maasoumi E., Diamond C., Nieswiadomy M., and D.J. Slottje (1994), The effects of relative price changes and cost of living adjustments on some welfare indices, in *Measurement of welfare and inequality*, in W. Eichhorn (ed), Springer Verlag.
- Maasoumi E. and A. Heshmati (2003), Evaluating dominance ranking of PSID incomes by various household characteristics, Unpublished manuscript.
- Mahler V.A. (2001), Economic globalization, domestic politics and income inequality in the developed countries: a cross-national analysis, Luxembourg Income Study Working Paper 273, Luxembourg.
- Mayshar J. and S. Yitzhaki (1995), Dalton-improving indirect tax-reform, *American Economic Review* 85, 793-808.
- Mayshar J. and S. Yitzhaki (1996), Dalton-improving tax reform: When households differ in ability and needs, *Journal of Public Economics* 62, 399-412.
- Milanovic B. (2000), The median-voter hypothesis, income inequality, and income redistribution: an empirical test with the required data, *European Journal of Political Economy* 16, 367-410.
- Milanovic B. (2002a), True world income distribution, 1988 and 1993: First calculation based on household surveys alone, *Economic Journal* 112(476), 51-92.
- Milanovic B. (2002b), Can we discern the effect of globalization on income distribution? Evidence from household budget surveys, World Bank Policy Research Paper 2876, Washington DC: World Bank.

- Morduch J. and T. Sicular (2002), Rethinking inequality decomposition, with evidence from rural China, *Economic Journal* 112(476), 93-106.
- Noorbakhsh F. (1998), The human development index: some technical issues and alternative indices, *Journal of International Development* 10, 589-605.
- Ogwang T. (2000), A convenient method of computing the Gini index and its standard error, *Oxford Bulletin of Economics and Statistics* 62, 1, 123-129.
- Paraje G. (2002), Inequality, welfare and polarisation in the Greater Buenos Aires, 1986-1999. Unpublished manuscript presented at the International Workshop on Income Distribution and Welfare, Bocconi May 30 to June 2, 2002.
- Park D. (2001), Recent trends in the global distribution of income, *Journal of Policy Modeling* 23, 497-501.
- Parker S.C. and S. Gardner (2002), International income mobility, *Economics Letters* 76, 179-187.
- Pendakur K. (2002), Taking prices seriously in the measurement of inequality, *Journal of Public Economics* 86, 47-96.
- Podder N. and S. Chatterjee (2002), Sharing the national cake in post reform New Zealand: income inequality trends in terms of income sources, *Journal of Public Economics* 86, 1-27.
- Ravallion M. (1997), Good and bad growth: the human development reports, *World Development* 25(5), 631-638.
- Ravallion M. (2003), Inequality convergence, *Economics Letters* 80, 351-356.
- Ravallion M. and S. Chen (1999), When economic reform is faster than statistical reform: measuring and explaining income inequality in Rural China, *Oxford Bulletin of Economics and Statistics* 61, 33-56.
- Ritakallio V-M. (2001), Trends of poverty and income inequality in cross-national comparison, Luxembourg Income Study Working Paper No. 272.
- Sahn D.E. and D.C. Stifel (2003), Progress towards the MDG in Africa, *World Development* 31(1), 23-52.
- Salas R. (1998), Welfare-consistent inequality indices in changing populations: the marginal population replication axiom: a note, *Journal of Public Economics* 67, 145-150.
- Sala-i-Martin X. (2002a), The disturbing "Rise" of global income inequality, NBER Working Paper Series 8904.
- Sala-i-Martin X. (2002b), The world distribution of income (estimated from individual country distribution), NBER Working Paper 8933.
- Schluter C. and M. Trede (2002), Statistical inference for inequality and poverty measurement with dependent data, *International Economic Review* 43(2), 493-508.
- Schmidt A. (2002), Statistical measurement of income polarisation: a cross-national comparison, Paper presented at the Berlin 10th International conference on panel data.
- Shorrocks A.F. (1982), Inequality decomposition by factor components, *Econometrica* 50(1), 193-211.
- Shorrocks (1984), Inequality decomposition by population subgroups, *Econometrica* 52(6), 1369-1385.
- Subramanian S. (1997), (ed.), *Measurement of inequality and poverty*, Readers in Economics, Oxford University Press.

- Summers R. and A. Heston (1991), The Penn World Table (Mark 5): an expanded set of international comparisons, 1950-1988, *Quarterly Journal of Economics* 106, 327-368.
- Svedbery P. (2002), Income distribution across countries: how is it measured and what do the results show?, *IIES Seminar Papers, SWoPEc* 2002:698.
- UNDP (2003), *Human Development Report 2003*, United Nations Development Program.
- Wan G.H. (2001), Changes in regional inequality in rural china: decomposing the Gini index of income sources, *The Australian Journal of Agricultural and Resource Economics* 45(3), 361-381.
- Yitzhaki S. (1991), Calculating Jackknife variance estimator for parameters of the Gini method, *Journal of Business and Economic Statistics* 9, 235-239.
- Zheng B. and B.J. Cushing (2001), Statistical inference for testing inequality indices with dependent samples, *Journal of Econometrics* 101, 315-335.