Centre for Efficiency and Productivity Analysis

Working Paper Series
No. WP09/2022

The University of Queensland International Comparison Database, UQICD V3.0

D.S. Prasada Rao, Alicia N. Rambaldi,
Gholamreza Hajargasht and William E. Griffiths

Revision Version: February 2023

School of Economics
University of Queensland St. Lucia,
Qld. 4072 Australia

ISSN No. 1932 - 4398

THE UNIVERSITY OF QUEENSLAND INTERNATIONAL COMPARISONS DATABASE, UQICD V3.0

User Guide

D.S. Prasada Rao, Alicia N. Rambaldi
School of Economics, The University of Queensland

Gholamreza Hajargasht
Research School of Economics, Australian National University and School of Economics, The University of Queensland

William E. Griffiths
Department of Economics, The University of Melbourne

February, 2023.
Brisbane, QLD Australia

Link to the Website and Database: UQICD V3.0

Acknowledgments:

The principal researchers acknowledge research funding support received from the Australian Research Council through the projects DP0557606, DP0985813 and DP170103559.

The inequality module draws from the work undertaken by Prof. Duangkamon Chotikapanich, Monash University, Prof. William E. Griffiths, The University of Melbourne, Dr Gholamreza Hajargasht, and Prof. D. S. Prasada Rao on fitting income distributions with limited data. Their work received funding from ARC DP1094632 and DP140100673.

Dr Leighton Brough designed and created the UQICD website from the inception of the project and continues to provide IT support.

Dr Donjie Wu and Mr Ramazan Bora expert research assistance with UQICDV3.0 is gratefully acknowledged. The contributions of Dr K. Renuka Ganegodage, Dr L. T. Huynh and Dr Howard E. Doran to the UQICD project are gratefully acknowledged.

Cite this guide as:
Rao, D.S.P, A.N. Rambaldi, G. Hajargasht and W. E. Griffiths (2022), **UQICD V3.0 User Guide**. School of Economics, The University of Queensland

Cite UQICDV3.0 data:
Rao, D.S. Prasada, A. N. Rambaldi, G. Hajargasht, D. Chotikapanich and W.E. Griffiths, UQ International Comparisons Database: UQICD V3.0, School of Economics, The University of Queensland, St Lucia, QLD 4072, Australia. 2022.

Contents

Chapter 1

Background and the New Structure of UQICD

The release of Version 3.0 of the University of Queensland International Comparison Database marks a significant milestone in the development of UQICD. This version is a culmination of several years of intensive research dedicated to the development and compilation of panels of statistics on internationally comparable measures of purchasing power parities of currencies; real gross domestic product and its components; income distributions; and measures of inequality for 185 countries and for selected regional groupings covering the period 1970 to 2019. The UQICD V3.0 is the only database of its kind which combines and provides data on real incomes and income inequality – two components that determine economic welfare of a society. At the core of the development of V3.0 is the conceptual framework for measuring social welfare developed by Amartya Sen (1976). An additional feature that distinguishes this database is the regional modules that provide data for: (i) geographical regions of the world, (ii) for groups based on levels of per capita income as classified by the World Bank, and (iii) the administrative regions, OECD and EU.

UQICD V3.0 is designed to complement internationally comparable price and income data available from various versions of the Penn World Table (PWT), the latest version being PWT 10.0. This database also complements a number of income distribution databases currently available – World Inequality Database (WID); Income Distribution Database, OECD; Standardized World Income Inequality Database (SWIID); and the World Income Inequality Database (WIID).

UQICD V3.0 has four major modules in contrast to a single module on purchasing power parities and real incomes available in the earlier versions of UQICD. The modules of UQICD V3.0 are:

- *Module 1: Real Incomes by Country;*
- *Module 2: Inequality by Country;*
- *Module 3: Real Incomes by Region;*
- *Module 4: Inequality by Region;*

Details of the data series available in each of the modules are described in dedicated chapters of this Guide. Links to additional information on each of the countries/regions included in each module can be accessed from the corresponding website page of UQICD.

1.1 Organization of the Guide

The remaining sections of this Chapter describe the number of countries and years, as well as the geographic, administrative and income-based groups of countries covered by UQICD V3.0. Basic national accounts aggregates and relevant concepts such as: purchasing power parties of currencies (PPPs); nominal and real expenditures; and current and constant price series are explained in Section 1.4. Chapter 2 is devoted to a description of Module 1: Real Incomes by Country. The section briefly describes the new econometric methodology used in the construction of panels of PPPs and real expenditures, an approach that is different from the one used in databases like the PWT and the World Development Indicators (WDI) of the World Bank. Various data series available in the module, along with short descriptions, are listed. Chapter 3 explains the UQICD approach to the construction of a panel of income distributions available for users. Section 3.2 presents features of the four income distributions available in UQICD: log-normal; generalised beta distribution of the second kind; Pareto-log-normal; and mixture of log-normal distributions. Methodology for fitting distributions to aggregate data, and extrapolation of distributions to years with unobserved income distribution data is discussed in Sections 3.3 and 3.4. Income distribution related series available in UQICD are listed in Section 3.5. Real income and inequality measures at the regional level are a novel feature of UQICD V3.0; these are presented Chapters 4 and 5. The technical appendices of this Guide provide additional insights into the methodologies used in generating different series available in UQICD V3.0.

1.2 Coverage

UQICD V3.0 covers 185 countries from all geographical regions of the world. It provides a range of series for the years 1970 to 2019 and therefore covers the pre-Covid-19 pandemic years.

Users can download desired series for countries/regions and years of their choice using the friendly interface developed for UQICD. In order to facilitate analysis based on country groupings, UQICD makes it possible to download data for selected country groupings.

1.2.1 Dealing with newly formed countries

As UQICD covers the period 1971 to 2019, the data series provided recognize that a number of countries currently in existence had not existed through this period. There are 24 countries in our data set that were not designated as separate countries over the complete period. For these countries data are available only from 1990 (or later) when they came into existence as separate entities after the dissolution of USSR and former Yugoslavia.

To be able to maintain time-space consistency and reference-country invariance properties for the majority of countries, estimation is conducted in two steps, both use the Rao-Rambaldi-Doran (RRD) method described in Rao, Rambaldi, and Doran (2010b), Rao, Rambaldi, and Doran (2010a), and Rao and Rambaldi (2015). In step 1, the RRD estimation method is implemented for 159 countries that were in existence throughout the full period covered. In this step all the parameters of the model are estimated. Given the parameters estimated in Step 1, the RRD method is implemented for all the 183 countries for the period when they were in existence. See Appendix A for further details of the RRD method and the method implemented for newly formed countries.

1.3 Regional classifications in UQICD

The UQICD regional modules provide series for three sets of regional groupings of countries, geographical, income and administrative.

Regional groupings available on UQICD are based on the World Bank World Development Indicators classification and the membership of these regions can be found in the latest WDI World Bank (2022). The available groupings are:

Geographical

- South Asia
- East Asia & Pacific
- Sub-Saharan Africa
- North America
- Europe & Central Asia
- Latin America and Caribbean
- Middle East & North Africa

Income Groups

- Low Income
- Lower Middle Income
- Upper Middle Income
- Upper Income

Administrative Groupings

- OECD
- European Union

1.4 Basic Concepts

The users of this Guide need to be familiar with the concepts that underpin the series available as a part of UQICD. A good understanding of what each series means would help the user in selecting the most suitable series for the type of analysis envisaged.

Following the general framework of the International Comparison Program (ICP) (see World Bank (2013)), the main series available from UQICD relate to the expenditure side of gross domestic product (GDP) and its components – Consumption (C); Investment (I); Government Expenditure (G); and Net Exports.

1.4.1 Gross domestic product (GDP) in local currency units

Data on GDP and its components are regularly published by respective national statistical organizations as a part of their national accounts publications. These aggregates are expressed in local currency units (LCUs). At the country level, let $GDP_{s,i}$ and $GDP_{t,i}$ represent GDP of country j in current prices in periods$s$ and t respectively. These GDP series are not comparable across countries as the currency units are different, and not comparable over time as they are expressed in prices prevailing in the current period.

GDP at constant prices For comparting GDP over time, the concept of GDP at constant prices relative to a base year, b, is used. Let $\{Def_{b,t,i}\}$ represent GDP deflator for period t with period b as the base. Then constant price GDP, denoted by $CGDP_{b,t,j}$ is obtained by deflating GDP in period t with the GDP deflator. Thus,

$$CGDP_{b,t,j} = \frac{GDP_{t,j}}{Def_{b,t,j}} \text{ for all } t \text{ and } j \tag{1.1}$$

In any given country, constant price GDP is used to measure growth in GDP over time.

Country-specific growth rate: The growth rate of GDP in a given country from year s to t, denoted by $GR_{s,t,j}$, is given by:

$$GR_{s,t,j} = \frac{CGDP_{b,t,j}}{CGDP_{b,s,j}} \text{ for all } s, t \text{ and } j \tag{1.2}$$

The same concepts apply to the components of GDP – C, G and I.

1.4.2 GDP in common currency units

The GDP figures published by national statistical offices, expressed in LCUs, need to be converted into a common currency unit before comparing them across countries. Two types of exchange rates are used for converting GDP data – market exchange rates (XR) and purchasing power parities – leading respectively to *nominal* and *real* GDP for each country.

Exchange Rates and Nominal GDP Let XR_j represent the market exchange rate, the number of units of LCUs per one unit of a reference currency. In UQICD, like in the International Comparisons Program (ICP) at the World Bank, the US dollar is used as the reference currency or the numeraire. Then, the nominal GDP, $NGDP_{t,j}$, is given by:

$$NGDP_{t,j} = \frac{GDP_{t,j}}{XR_{t,j}} \text{ for all } t \text{ and } j \tag{1.3}$$

Nominal GDP can be added across countries as they are all in US dollars. NGDP provides a basis for constructing regional and global aggregates.

As exchange rates are determined by the demand and supply for a given currency, they may not adequately reflect price level differences across countries. That is the reason for using the label *nominal GDP* indicating that no adjustment for price level differences has been made in converting LCU GDP into a common currency.

Purchasing Power Parities of currencies and Real GDP The notion of purchasing power parities (PPPs) of currencies is a key concept that underpins international comparisons compiled as a part of the International Comparison Program (ICP).

The PPP of currency of a country in the ICP, in concept, represents the number of units of the currency of the country needed to purchase a basket of goods and services that can be purchased using one unit of the reference currency. The framework used in the ICP can be found in Rao (2013).

In UQICD, all the PPPs are expressed relative to the US dollar. For example, if the PPP of Indian rupee is 20 rupees per US dollar, it means that what can be purchased with one dollar in the United States can be purchased in India with 20 Indian rupees. PPPs vary over time and across countries. Let $PPP_{t,i}$ represent the PPP exchange rate for the currency of country j in year t.

Real GDP or GDP converted using PPP exchange rates Real GDP for a country at a given point of time, denoted by $RGDP_{t,j}$, is given by converting GDP in LCUs using PPP exchange rates. Thus

$$RGDP_{t,j} = \frac{GDP_{t,j}}{PPP_{t,j}} \text{ for } t \text{ and } j \tag{1.4}$$

This measure is termed *real* since GDP converted using PPP exchange rates not only expresses GDP in a common currency unit but also adjusts for price level differences across countries. Obviously, for the United States the PPP exchange rate is 1 as it is the reference country and, therefore, real GDP is the same as the GDP in LCUs.

Real GDPs of all the countries in the world or the region, at a given point of time, t, can be summed to provide real GDP for the world or the region at that point of time.

As real GDP is measured at a given point of time t, it is also referred to as *real GDP of country j in period t prices or in current prices* as it is the case with national accounts. Therefore, real GDP, RGDP, of a country in one period, say 2011, cannot be compared to its RGDP in another year, say 2017.

Real GDP at current prices versus at constant prices UQICD provides a series of real GDP at constant prices. Thi series is $CRGDP_{b,t,j}$, denoting real GDP of country j in period t expressed in constant year b *prices*. In the case of UQICD, these series are all expressed in constant year 2011 prices. The constant price real GDP series are obtained by updating the base period real GDP using country specific growth rates. Thus, for example, $CRGDP$ of Japan of 2017 in 2011 constant prices is given by:

$$CRGDP_{2011,2017,Japan} = RGDP_{2011,Japan} \times GR_{2011,2017,Japan} \tag{1.5}$$

where the growth rate, $GR_{2011,2017,Japan}$ is computed using equation (1.2).

1.4.3 Nominal and Real measures for components of GDP

Users may be interested in measures of household consumption (C), investment (I) or gross capital formation (GCF) and government expenditure (G). Nominal and real aggregates for each of the components of GDP are defined similar to those defined for GDP equations (1.1) to (1.5) applied to specific components. It is important that PPPs used in measuring real consumption, investment and government expenditure must make use of PPPs specific to those components and not those for GDP. UQICD provides separate PPP series for each of these components.

Chapter 2

Module 1: Real Incomes by Country

This module, Real Incomes by Country is the most important module of UQICD as the series available in this module are subsequently used in compiling series available in Modules 2, 3 and 4. The Real Incomes module has been available in all the earlier versions of UQICD starting from Version 1.0 released in 2012.

The main series available in this Module are panels of purchasing power parities (PPP), real GDP, and its three main components: Consumption; Investment; and Government Expenditure. A special feature of UQICD is the availability of charts showing the evolution of price levels (PPP/XR) over time for all the countries in the database. In addition to the PPP series available, a number of standard series of interest from secondary sources are made available to users to facilitate basic analyses as well as to derive other measures from the series that can be downloaded from UQICD.

2.1 Methodology used in the construction of panel of PPPs

All the existing sources of panels of PPPs such as the Penn World Table (PWT) and the World Development Indicators (WDI) are anchored on extrapolations of PPPs available for selected benchmark years from the International Comparison Program. The UQICD constructs panels of PPPs using the methodology described in Appendix A. The main task underlying panels of PPPs is to extrapolate and interpolate PPPs for all the countries covered using the data on PPPs available from the primary source which is the International Comparison Program (ICP). For a history of ICP and how the Program is conducted and to find the methodology used in compiling PPPs for different benchmarks and for participating countries, the reader can refer to World Bank (2013) and Rao and Balk (2013).

The ICP data represents an incomplete panel of PPPs with coverage of: 10, 16, 34, 60, 64, 117, 146, 177 and 176 participating countries respectively in benchmark years 1970, 1973, 1975, 1980, 1985, 1993, 2005, 2011 and 2017. In addition to PPPs from ICP, PPPs for OECD and EU countries are available for additional years. The main task undertaken for UQICD is to derive a complete panel of PPPs from 1970 to 2019 using the all the data available from ICP.

The World Bank World Development Indicators (WDI) simply take the latest benchmark year PPPs and extrapolates them backwards and forwards using the price deflator series available from the countries. For example, World Bank (2014)'s WDI PPP series are generated using only the 2011 ICP benchmark data that were released in April, 2014. The other and the most popular source of data on PPPs for non-benchmark years is the Penn World Table (PWT).

The latest version, PWT 10.0, uses two consecutive benchmarks to interpolate PPPs for intervening years for countries that have participated in both benchmarks. For those countries that have participated in only one benchmark, PPPs from that year are extrapolated forward and backward using the price deflator series available from the countries.

A radically different methodology underpins the UQICD series. The methodology, described in detail in Appendix A, uses an econometric approach to combine PPP data available from all the participating countries in all the benchmarks along with price deflator data available from the countries. The UQICD approach also makes use of predictions generated from an econometrically estimated panel regression model that relates price levels (ratios of PPPs to exchange rates) to a number of explanatory variables. Details of the model are available in Appendix A. All these three sets of information are combined using a state-space approach to generate optimal extrapolations of PPPs that make up the panels of PPPs available.

There are several the advantages of using an econometric approach to extrapolate PPPs instead of a simple updating with only on one or two benchmarks

1. The first and foremost advantage is that all the information available from all the sources are used in producing optimal extrapolations of PPP to make up the panels from UQICD.
2. The second advantage is that it is now possible to attach measures of reliability, in the form of standard errors, with each of the PPPs available from UQICD.
3. The econometric methodology is flexible enough to produce extrapolations to suite various needs. It is possible to impose a constraint that PPPs in the panel must equal the PPPs from the ICP for those benchmark years and for the participating countries.
4. Many analysts believe that panels of PPPs must reflect accurately the movements in prices as observed in the countries and reported as implicit price deflators in national accounts. UQICD's approach makes it possible to produce series under this constraint as well.
5. Panels of PPPs from UQICD are invariant to the choice of the reference country. Proof of this important property is provided in Rao and Rambaldi (2015)

2.2 Purchasing Power Parity (PPP) Series

The PPP series from this Module are critical to the series constructed and made available in Modules 2,3 and 4. They are used to compile series on real GDP and its components as well as series at current and constant prices.

Unless stated otherwise, all PPPs are expressed relative to the US dollar. The following are the major series available.

2.2.1 PPPs at the GDP level

This series provides estimates of PPPs that can be used to convert GDP into a common currency unit.

PPP_GDP: This series is constructed without imposing any constraints to track ICP benchmarks or the domestic price deflators. These PPPs are the optimal predictors of PPPs based on information from all the benchmark years and any other auxiliary information available. This is the series recommended for computing real GDP series.

SEPPP_GDP: Standard errors associated with the PPP_GDP series.

2.2.2 PPPs for components of GDP

These are the PPPs that can be used, in any given year, for converting national accounts aggregates into a common currency unit (US dollar in this case) which can be compared across countries. As PPPs refer to the aggregate under consideration, users must select the PPP series that is most appropriate for their analysis or application.

The following PPP series are available for the components:

1. ***PPP_CONS*** – Unconstrained extrapolated series for the Consumption aggregate (and ***SEPPP_CONS*** associated standard error)

2. ***PPP_GCF*** – Unconstrained extrapolated series for the Investment aggregate (and ***SEPPP_GCF*** associated standard error)

3. ***PPP_GOV*** – Unconstrained extrapolated series for the Government expenditure aggregate (and ***SEPPP_GOV*** associated standard error)

The methodology used in the estimation of PPPs for components is described in Appendix B.

2.3 GDP and its Components in Common Currency Units

UQICD provides series on Nominal GDP (see equation (1.3)) and its components expressed in per capita terms. The Real GDP (see equation (1.4)) and component series are obtained by converting GDP in local currency units using unconstrained PPP exchange rate series as described in Section 1.4.2.

2.3.1 Current Price Series available from UQICD

1. **Real (in PPP Exchange Rate) Series**
   (a) RGDP_PC - per capita real income in USD
   (b) RCONS_PC- per capita real household consumption expenditure in USD
   (c) RGOV_PC- per capita real general government expenditures USD
   (d) RGCF_PC- per capita real gross capital formation USD

2. **Nominal (in Market Exchange Rate) Series**
   (a) NGDP_PC - per capita nominal income in USD
   (b) NCONS_PC- per capita nominal household consumption expenditure in USD
   (c) NGOV_PC- per capita nominal general government expenditures in USD
   (d) NGCF_PC- per capita nominal gross capital formation in USD

These series are useful for making comparisons across countries in any given year but cannot be used for comparisons across years. In order to make comparisons over time and space, it is necessary to have GDP, Household Consumption, Investment and Government Expenditures series which are expressed at constant prices. For example, if the objective is to compare real per capita income of India in 2012 with real per capita income of the US in 1995, this task can be accomplished only using aggregates at constant price series. In compiling series at constant prices, it is necessary to fix the reference country and also a reference year. In the UQICD series at constant prices, United States is the reference country and the year 2011 is the reference year.

2.3.2 Constant Prices Series available from UQICD

UQICD provides Constant Prices Real GDP and its Components. Details of the procedure and the derivations involved are shown in Appendix C.

1. CRGDP_PC- real incomes per capita in USD at 2011 prices
2. CRGDPC_PC- real household consumption expenditures per capita in USD at 2011 prices
3. CRGDPG_PC- real general government expenditures per capita in USD at 2011 prices
4. CRGDPI_PC- real gross capital formation per capita in USD at 2011 prices

2.3.3 Auxiliary series

A number of series that provide basic information will by default be included when data is downloaded by a user. Full definitions are provided in UQICD. These are:

1. CODE_WB - World Bank 3-letter code
2. YEAR - Year
3. COUNTRY - Country name
4. CODE_IFS - IMF International Financial Statistics (IFS) Code
5. CURRENCY - The currency unit used in 2019
6. CURRENCYCODE - The currency code (ISO 4217) by country used in computing the PPP
7. REGION - The country's geographical region

Series for a number of economic variables useful in international comparisons are also available for users. These series are drawn from standard sources such as the World Development Indicators, UN and IMF. A full description of the variables along with their sources are available on UQICD. Users need to ensure the choice of right variables for the purpose of analysis to be conducted. The supplementary series available on UQICD are listed below.

1. XR - The exchange rate
2. CURRENCYCODE_HIST - The historical currency code (ISO 4217) by country by year
3. PPP_GDP_ICP - PPP Data for GDP from ICP and OECD for benchmark years.
4. PPP_CONS_ICP - Data for consumption from ICP and OECD for benchmark years.
5. PPP_GOV_ICP - PPP Data for government from ICP and OECD for benchmark years.

6. PPP_GCF_ICP - PPP Data for investment from ICP and OECD for benchmark years.
7. POPULATION - Total population
8. XR2 - UN National Accounts Analysis of Main Aggregates (AMA) exchange rate
9. XR1 - IMF exchange rate
10. CONS_SHARE - share of Household final consumption expenditure, etc. (% of GDP).
11. GCF_SHARE - share of Gross capital formation (%GDP)
12. GOV_SHARE - share of General government final consumption expenditure (% of GDP)
13. GDP_PC_LCU_CURRENT - GDP per capita in local currency units at current prices
14. GOV_PC_LCU_CURRENT - Government expenditure per capita at current prices - LCU
15. CONS_PC_LCU_CURRENT - Consumption per capita at current prices - LCU
16. GCF_PC_LCU_CURRENT - Investment per capita at current prices - LCU
17. GDP_2010DEF - The GDP implicit deflator - 2010 base year.
18. GOV_2010DEF - GDP deflator for government expenditure_2010 base year
19. CONS_2010DEF - GDP deflator for private (household) consumption_2010 base year
20. GCF_2010DEF - GDP deflator for Investment_2010 base year

Additional information on each of the 185 countries included can be accessed from UQICD by following the buttons labeled ”Notes on Countries/Regions" from the entry page of each of the four modules.

Chapter 3

Module 2: Inequality and Income Distributions by Country

Inclusion of the Inequality Module in the UQICD data series represents a major step forward in providing researchers and users with panels of real incomes as well as the distribution of incomes – two major components of economic welfare of a society. The real GDP series provides estimates of the size of the economies with real per capita GDP or income serving as in indicator of standard of living in these economies. It is well recognized (Sen (1976); and Stiglitz, Sen, Fitoussi, et al. (2009)) that inequality in the distribution of income is a critical determinant of economic welfare. A series of research advances achieved over the last fifteen years have enabled the UQICD team to compile this panel of income distributions for 180 countries over the period 1970 to 2019.

Availability of income distribution data is largely determined by the frequency with which household income and expenditure surveys are conducted. These data play a major role in the estimation of incidence and severity of poverty at the national, regional and global level. Despite the importance attached to inequality data, availability is sparse. More importantly, detailed unit record data are usually not available for the users and often researchers rely on aggregated distribution data in the form of income shares for decile or ventile groups. While aggregated data are used in approximating inequality measures like the Gini coefficient and the Theil's measure, details of the underlying income distributions are seldom available.

The UQICD V3.0 module on inequality serves to fill this gap. Using extrapolation/interpolation methods to compile a tableau of income shares of decile groups for all the countries for the period 1970 to 2019, the UQICD makes use of research undertaken in Australia (Chotikapanich, Griffiths, and Rao (2007) and Chotikapanich et al. (2012), Chotikapanich et al. (2018)) and the techniques developed by these researchers in constructing panels of fitted parametric income distributions.

The UQICD series in the inequality module complement a number of income distribution databases currently available – World Inequality Database (WID); Income Distribution Database, OECD; Standardized World Income Inequality Database (SWIID); and the World Income Inequality Database (WIID).

3.1 Inequality data series used in the module

The income distribution series available in this UQICD module are estimated using aggregated income distribution data drawn from the World Bank through its POVCAL interface. In order to ensure consistency, UQICD relies on the World Bank POVCAL as the only source for all the series made available to users. The country coverage is necessarily limited by availability of data from this source. The data are in the form income shares for decile groups of population – income share of the poorest 10 percent; share of the next 10 percent; so on until the income share of the richest 10 percent of the population. The link to the World Bank website is: `http://iresearch.worldbank.org/PovcalNet/povDuplicateWB.aspx`. POVCAL has more detailed data in the form of expenditure shares for ventiles and centiles, but decile shares are prefereed for the UQICD inequality module as the cell sizes and reliability of income shares at disaggregated level, for example share of the top 1 percent of population, are likely to be problematic. Data used in the construction of the UQICD series were downloaded on 21^{st} March, 2021 before the final compilation.

3.2 Fitted Distributions available through UQICD

The inequality literature highlights numerous statistical distributions that can be used to model distribution of income. Kleiber and Kotz (2003) provide a comprehensive description of statistical size distributions for use in economics. After a careful consideration of a range of distributions, four distributions were selected for inclusion in the series available through UQICD. These are briefly discussed below.

3.2.1 Log-normal distribution

The log-normal distribution has a long history of applicationsstudying the distribution of income. It is a distribution determined by two parameters, denoted by μ and σ^2, that relate to the location and dispersion of the log income. The Gini and Theil's measure of inequality are completely determined by the parameter σ^2. The log-normal distribution is known to fit income distribution data well for lower levels of income.

If income, Y, is lognormally distributed with parameters μ and σ^2, then the density function is:

$$f_Y(y) \;=\; \frac{1}{y\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln y \;-\; \mu)^2}{2\sigma^2}\right) \quad ; \; y > 0$$

and the distribution function is given by:

$$F_Y(y) \;=\; \Phi\left(\frac{\ln y \;-\; \mu}{\sigma}\right)$$

where $\Phi(.)$ is the distribution function of the standard normal distribution.

3.2.2 Pareto-lognormal distribution

The Pareto distribution is known to fit higher incomes well and in contrast the lognormal distribution fits lower income levels better. The Pareto-lognormal distribution combines virtues of both of these distributions, it is a modification of the lognormal distribution which exhibits Pareto behaviour in the right tail This distribution was suggested by Colombi (1990) for modelling income distributions. Hajargash and Griffiths (2013) and Griffiths, Chotikapanich, and Hajargasht (2022) provide details of properties of the Pareto-lognormal distributions and provide formulae to compute various inequality and poverty measures associated with this distribution.

Following Griffiths and Hajargasht (2013), the density functions is given by:

$$f_Y(y, \mu, \sigma, \alpha) = \frac{\alpha}{y}\phi\left(\frac{\ln y - \mu}{\sigma}\right) R(y_1) \qquad y > 0$$

where $\phi(.)$ is the density function of a standard normal random variable; $R(t) = [1 - \Phi(t)]/\phi(t)$ is a Mill's ratio and

$$y_1 = \alpha\sigma - \frac{\ln y - \mu}{\sigma}$$

The distribution function is given by:

$$F_Y(y; \mu, \sigma, \alpha) = \Phi\left(\frac{\ln y - \mu}{\sigma}\right) - \phi\left(\frac{\ln y - \mu}{\sigma}\right) R(y_1)$$

3.2.3 Generalized Beta distribution of the second kind – GB2 distribution

The third distribution included in the UQICD series is the generalized beta distribution of the second kind (GB2). It is a four-parameter distribution that nests many popular three-parameter specifications of income distributions including the generalized gamma, beta2, Singh-Maddala and Dagum distributions and also many two-parameter distributions (see McDonald and Xu (1995)). Chotikapanich et al. (2018) provide full technical details and demonstrated the versatility of the GB2 distribution. Chotikapanich et al. (2012) estimate beta2 distributions for 91 countries for the years 1993 and 2000. In an application involving 10 countries, Hajargash and Griffiths (2013) find that the GB2 distribution compares favorably with the four-parameter double Pareto-lognormal distribution.

The probability density function (pdf) of income Y isdefined as

$$f_Y(y|a, b, p, q) = \frac{ay^{ap-1}}{b^{ap}B(p, q)\left(1 + \left(\frac{y}{b}\right)^a\right)^{p+q}} \quad y > 0 \tag{3.1}$$

The cumulative distribution function (cdf) corresponding to (3.1) is given by

$$F_Y(y|a, b, p, q) = \frac{1}{B(p, q)}\int_0^w t^{p-1}(1-t)^{q-1}dt = B(w|p,\, q) \tag{3.2}$$

where $w = (y/b)^a/[1 + (y/b)^a]$, where $a > 0,\ b > 0,\ p > 0$ and $q > 0$ are its parameters and $\mathrm{B}(p, q) = \int_0^1 t^{p-1}(1-t)^{q-1}dt$ is the beta function. We have found that this parameterization does not converge well during the estimation. We replace b with $\frac{mB(p, q)}{B(p + 1/a + q - 1/a)}$ which reparametrizes the distribution in terms of m where m is the mean of the distribution, and $B(w|p,\, q)$ is the cdf of the standardized beta distribution with parameters $(p,\, q)$.

3.2.4 Mixture of lognormal distributions

Mixtures of lognormal distributions provide a flexible function form to model income distributions. The number of distributions in the mixture and the weights accorded are determined by the income distribution data. Using mixtures of lognormal distribution has been a common practice which has recently been employed in the study of inequality in Africa by Chotikapanich et al. (2014).For estimation of this distribution using grouped data see Griffiths and Hajargasht (2012).

For a mixture of lognormal distributions, UQICD specifies one with three components labelled as MLN32. Its density and distribution functions are written as:

$$f_Y(y) = (1/y)\left(w_1\phi(ln\,y;\mu_1,\sigma_1) + w_2\phi(ln\,y;\mu_2,\sigma_2) + w_3\phi(ln\,y;\mu_3,\sigma_2)\right) \qquad y > 0$$

$$F_Y(y) = w_1\Phi(lny;\mu_1,\sigma_1) + w_2\Phi(lny;\mu_2,\sigma_2) + w_3\Phi(lny;\mu_3,\sigma_2)$$

where w_1,w_2, and w_3 are the weights of each component adding to 1. $\phi(\ .\ ;\mu,\sigma)$ and $\Phi(\ .\ ;\mu,\sigma)$ are the pdf and cdf of a normal distribution with mean μ and variance σ^2. Note that the variance of the third component is assumed to be equal to that of the second component. This was done since estimation with three distinct variances led to very large standard errors in many cases. In some cases all the variances are assumed to be equal.

3.3 Fitting income distributions to limited data

Real per capita income in current US dollars In order to fit income distributions using the income share data described in Section 3.2 above, it is necessary to supplement the data with observed per capita income. To make the fitted distributions comparable across countries, UQICD makes use of real per capita income in US dollars derived by converting per capita incomes in local currency units using PPP exchange rates described in Module 1.

The panel of income distributions for the countries and periods covered by UQICD are fitted using econometric methodology developed by Chotikapanich, Griffiths, Hajargasht and Rao over the last fifteen years (Chotikapanich, Griffiths, and Rao, 2007; Chotikapanich et al., 2012; Chotikapanich et al., 2014; Chotikapanich et al., 2018; Hajargasht et al., 2012; Hajargash and Griffiths, 2013; Griffiths and Hajargasht, 2012; Griffiths and Hajargasht, 2015; and Hajargasht and Griffiths, 2020). A brief outline of the method moments technique used in fitting these distributions is described in Appendix D. A comprehensive description of the methodology used can be found in Chotikapanich, Griffiths, and Hajargasht (2022).

3.4 Construction of Panel of Income Distributions

The decile group income share data from the World Bank represents an unbalanced panel with data available for different countries for different sets of years. The following steps have been used to fill gaps in income distribution data:

3.4.1 Interpolation between years

For any country, if income share data are available in periods t and $t{+}k$ with gaps in between, the income shares for decile groups for periods in between are interpolated using a weighted average of shares at the two end points, with weights reflecting the distance from the end points. Let the income shares for *i-th decile group* in periods t and $t{+}k$ be denoted by $\{s_{i,t}; s_{i,t+k} : i = 1, 2, ..., 10\}$. Then the income shares for years in between, denoted by $\{s_{i,r} : r = 1, 2, ..k - 1\}$ are obtained by interpolating income shares as follows:

$$s_{i,t+r} = s_{i,t} + \left(\frac{r}{k}\right)(s_{i,t+k} - s_{i,t}) \quad = \quad \left(1 - \frac{r}{k}\right) s_{i,t} + \frac{r}{k} s_{i,t+k}$$

3.4.2 Extrapolation at the end points

After experimenting with several alternatives, UQICD implemented an extrapolation strategy which simply maintains the observed shares at the end points – but with extrapolation limited to five years. For example, if the earliest and the latest years for which income share data are available are 1980 and 2013, then the observed income shares in the year 1980 are maintained for the years 1975 to 1979. Similarly, income shares observed in 2013 are maintained for the years 2014 to 2018. Assumption of constancy of income shares beyond five years was deemed untenable.

3.5 Series available in the Inequality Module

In this module the users make selections from the four distributions listed below. All the associated parameters and standard errors are automatically provided.

3.5.1 Parameters of Distributions

Parameters of Lognormal (LN) distribution:

1. LN_MU: Parameter μ of lognormal distribution
2. LN_MU_SE: Standard error of estimated μ
3. LN_SIG: Parameter σ of lognormal distribution
4. LN_SIG_SE: Standard error of Estimated σ
5. LN_MEAN: Estimate of mean income
6. LN_MEDIAN: Estimate of median income

Parameters of Pareto Lognormal (PLN) distribution

1. PLN_MU: Parameter μ of Pareto-lognormal distribution
2. PLN_MU_SE: Standard error of estimated μ
3. PLN_SIG: Parameter σ of Pareto-lognormal distribution
4. PLN_SIG_SE: Standard error of Estimated σ
5. PLN_ALPA: Parameter α of Pareto-lognormal distribution

6. PLN_ALPHA_SE: Standard error of Estimated α
7. PLN_MEAN: Estimate of mean income
8. PLN_MEDIAN: Estimate of median income

Parameters of GB2 Distribution

1. GB2_M: Parameter m of GB2 distribution
2. GB2_M_SE: Standard error of estimated m
3. GB2_P: Parameter P of GB2 distribution
4. GB2_P_SE: Standard error of Estimated P
5. GB2_Q: Parameter q of GB2 distribution
6. GB2_Q_SE: Standard error of Estimated q
7. GB2_A: Parameter a of GB2 distribution
8. GB2_A_SE: Standard error of Estimated a
9. GB2_MEAN: Estimate of mean income
10. GB2_MEDIAN: Estimate of median income

Parameters of Mixture of Lognormal (MLN32) distribution. Note that estimates of w_3 are given as $1-(w_1+w_2)$

1. MLN32_MU1: Parameter $\mu 1$ of the first component of mixture of lognormal distributions
2. MLN32_MU1_SE: Standard error of estimated $\mu 1$
3. MLN32_MU2: Parameter $\mu 2$ of the second component of mixture of lognormal distributions
4. MLN32_MU2_SE: Standard error of estimated $\mu 2$
5. MLN32_MU3: Parameter $\mu 3$ of the third component ofmixture of lognormal distributions
6. MLN32_MU3_SE: Standard error of estimated $\mu 3$
7. MLN32_SIG1: Parameter $\sigma 1$ of the first component of mixture of lognormal distributions
8. MLN32_SIG1_SE: Standard error of Estimated $\sigma 1$
9. MLN32_SIG2: Parameter $\sigma 2$ of the second component of mixture of lognormal distributions
10. MLN32_SIG2_SE: Standard error of Estimated $\sigma 2$
11. MLN32_W1: Parameter $w1$ of the first component of mixture of lognormal distributions
12. MLN32_W1_SE: Standard error of Estimated $w1$
13. MLN32_W2: Parameter $w2$ of the second component of mixture of lognormal distribu-

tions

14. MLN32_W2_SE: Standard error of Estimated $w2$
15. MLN32_MEAN: Estimate of mean income
16. MLN32_MEDIAN: Estimate of median income

3.5.2 Inequality Measures

Inequality measures computed from specific functional forms of the income distribution (DIST - GB2; LN; MLN32; and PLN)

1. DIST_GINI: Gini measure of inequality from specific functional form of the income distribution (DIST - GB2; LN; MLN32; and PLN)
2. DIST_THEIL: Theil's L-Measure of Inequality for the selected distribution
3. DIST_SH10: Share of the poorest 10 percent of the population of the country based on the selected distribution
4. DIST_SH30: Share of the poorest 30 percent of the population of the country based on the selected distribution
5. DIST_SHTOP10: Share of the richest 10 percent of the population of the country based on the selected distribution
6. DIST_SHTOP01: Share of the richest 1 percent of the population of the country based on the selected distribution

3.5.3 Inequality Measures for different distributions

The following formulas were used to calculate inequality measures for the various distributions

$$\text{LN\_GINI} = 2\Phi\left(\frac{\sigma}{\sqrt{2}}\right) - 1$$

$$\text{LN\_THEIL} = \frac{\sigma^2}{2}$$

$$\text{PLN\_GINI} = \frac{2\exp\{\alpha(\alpha-1)\sigma^2\}}{(2\alpha-1)}\Phi\left(\frac{(1-2\alpha)\sigma}{\sqrt{2}}\right) + 2\Phi\left(\frac{\sigma}{\sqrt{2}}\right) - 1$$

$$\text{PLN\_THEIL} = \ln\left(\frac{\alpha-1}{\alpha}\right) + \frac{1}{\alpha-1} + \frac{\sigma^2}{2}$$

The Gini for GB2 distribution doesn't have a closed form solution; therefore, it is computed using numerical integration

$$\text{GB2\_GINI} = \frac{2\int_0^\infty yf(y)F(y)dy}{\mu} - 1$$

$$\text{GB2\_THEIL} = \ln(\mu/b) - \psi(p)/a + \psi(q)/a \ \text{ with } b = \frac{mB(p,q)}{B(p+1/a+q-1/a)}$$

where $\psi(x) = d \log \Gamma(x)/dx$ and $\Gamma(.)$ is a gamma function.

$$\text{MLN32\_GINI} = -1 + \frac{2}{M}\sum_{i=1}^{3}\sum_{j=1}^{3} w_i w_j M_i \Phi\left(\frac{\sigma_i^2 + \mu_i - \mu_j}{\sqrt{\sigma_i^2 + \sigma_j^2}}\right)$$

$$\text{MLN32\_THEIL} = \ln M - \sum_{j=1}^{3} w_j \mu_j$$

with $M_i = \exp\{\mu_i + 0.5\sigma_i^2\}$ and $M = \sum_{j=1}^{3} w_j \exp\{\mu_j + 0.5\sigma_j^2\}$

3.5.4 Plots

In addition to the numerical measures listed, UQICD also provides the following plots:

`Country_LN` provides plot of density functions for every country for years 1970,1980,1990, 2000, 2010, 2019 based on the estimated lognormal distributions.

`Country_PLN` provides plot of density functions for every country for years 1970,1980,1990, 2000, 2010, 2019 based on the estimated Pareto-lognormal distributions.

`Country_GB2` provides plot of density functions for every country for years 1970,1980,1990, 2000, 2010, 2019 based on the estimated GB2 distributions.

`Country_MLN32` provides plot of density functions for every country for years 1970,1980,1990, 2000, 2010, 2019 based on the estimated mixture of lognormal distributions.

Chapter 4

Module 3: Real Incomes by Region

This module is unique in that there are no other sources or databases providing the type of information and series that are made available in Module 3 of UQICD. The published reports of the International Comparison Program (ICP) provide estimates of GDP for administrative groupings of countries participating in the program. The 2011 and 2017 phases of the ICP had a global coverage with 176 participating economies. The 2017 ICP report (World Bank, 2020) provides estimates of the size of the global economy or world GDP in PPP terms. In contrast, the series available in this module of UQICD not only provide such estimates for all the years from 1975 but also provide estimates of global growth and inflation for five year intervals beginning from 1970. The only other source with estimates of year-on-year global growth and inflation is the *World Economic Outlook* (WEO) regularly published by the International Monetary Fund (International Monetary Fund (2022)). However, the WEO estimates of global growth and inflation are not related to any identifiable measures of the size of the global economy for different years. This Module of UQICD starts with a clearly established series of global GDP in real terms over time and offers estimates of global growth and inflation consistent with such series. These UQICD series are based on the recently established research on this topic by Balk, Rambaldi, and Rao (2022).

There are two distinct sub-classes of series provided in this module of UQICD. These are:

1. Estimates of the size of the economies of different regions and the world; and

2. Decomposition of changes over time into growth, inflation and PPP exchange rate change effects;

UQICD V3.0 provides regional aggregates only at the gross domestic product level. Series for different components of GDP (Household Consumption, Government Consumption and Gross Fixed Capital Formation) will be included in future versions of UQICD.

4.1 Regions: Composition

Users can download aggregates for the World and for any of the geographical or administrative regions in ICP for any year in the range 1970 to 2019. Aggregates for country groups based on the World Bank income groupings are also available. The available groupings are:

- South Asia; East Asia & Pacific; Sub-Saharan Africa; North America; Europe & Central Asia; Latin America and Caribbean; Middle East & North Africa; and the World.

- Low Income; Lower Middle Income; Upper Middle Income; and Upper Income.

In addition, users can select and download country groupings for two administrative regions

- OECD; and the European Union.

4.2 Regional Aggregates

The regional aggregates are simply the sum total of the respective measures over all the countries belonging to the region. For example, real GDP of a region R is the sum total of real GDP of all the countries belonging to region R. Thus, using the notation established in Section 2, the real GDP of region R is given by:

$$RGDP_R = \sum_{j \in R} RGDP_j = \sum_{j \in R} \frac{GDP_j}{PPP_j}$$

Similarly, nominal GDP is the sum of nominal GDP of all the countries belonging to region R is given by:

$$NGDP_R = \sum_{j \in R} NGDP_j = \sum_{j \in R} \frac{GDP_j}{XR_j}$$

The following aggregates are available in Module 3 of UQICD:

1. POPULATION – Total population of the region
2. RGDP – Real GDP of the region - sum of real GDP (US dollars, PPP terms) of countries belonging to the region.
3. NGDP – Nominal GDP of the region - sum of nominal GDP (US dollars, exchange rate terms) of countries belonging to the region.
4. CRGDP – Constant real GDP of the region in Year 2011 prices - sum of constant real GDP in 2011 prices (US dollars, PPP terms) of countries belonging to the region.

4.3 Decomposition of change in Real GDP over time

The decomposition of changes in real GDP at the regional level provided by UQICD V3.0 is based on the framework developed by Balk, Rambaldi, and Rao (2022). It clearly identifies the aggregate and its change that needs to be decomposed. In a global context UQICD decomposes the change in real GDP for the world (the 185 countries in UQICD).

$$
\begin{aligned}
\frac{RGDP_{2017,W}}{RGDP_{2011,W}} &= \frac{\sum_{j=1}^{185} RGDP_{2017,j}}{\sum_{j=1}^{185} RGDP_{2011,j}} \\
&= \frac{\sum_{j=1}^{185} GDP_{2017,j}/PPP_{2017,j}}{\sum_{j=1}^{185} GDP_{2011,j}/PPP_{2011,j}} \\
&= \prod_{j=1}^{185} [GR_{2011,2017,j}]^{w_j} \times \prod_{j=1}^{185} [Def_{2011,2017,j}]^{w_j} \times \prod_{j=1}^{185} \left[\frac{PPP_{2011,j}^{USA}}{PPP_{2017,j}^{USA}}\right]^{w_j} \\
&= \text{Average of domestic growth } \times [\text{Average of domestic inflation rates } \times \text{PPP change effect} \\
&= \text{Global growth } \times \text{Global inflation}
\end{aligned}
$$

This equation provides a formal framework for computing global growth and inflation as weighted averages with weights w_j (the weights used in the decomposition are discussed in Appendix E) of growth rates in different countries. It establishes a clear connection between measures of the size of the world economy in 2011 and 2017 and global growth – this connection is somewhat unclear and possibly missing in the compilation of global growth and inflation measures by the IMF.

Appendix E provides a detailed exposition which shows how changes in RGDP of the world, or any region, can be decomposed into: *a growth component*; a *price change or domestic inflation effect*; a *PPP exchange rate change effect*; and, finally, a *regional inflation* measure which is the composite effect of the domestic inflation and the PPP exchange rate effects.

For any given period *t to t+5,* the followings series for growth and inflation components are available from UQICD Module 3:

1. RGDP_GR: The change in Real GDP for the countries in the region
2. DOM_GR: Weighted average of domestic growth rates of countries in the region
3. DOM_INF: Weighted average of domestic inflation rates of countries in the region
4. PPP_EFFECT: Effect of changes in PPPs over time on changes in real GDP of the region
5. PRICE_EFFECT: Total contribution of price related effects - product of DOM_INF and PPP_EFFECT.

The series RGDP_GR, and its components, are available for each year starting from 1975, and are computed for five-year periods. For example, **the growth in real GDP (RGDP_GR) for 1975 measures the growth in RGDP over the period 1970 to 1975**. This is the reason why regional series are available only from 1975.

Chapter 5

Module 4: Inequality by Region

This module focuses on inequality statistics for different regions of the world. It complements Module 3 on economic aggregates such as real, nominal and constant 2011 price real GDP for different regions and regional groupings. The levels of inequality along with data on real per capita incomes can provide an indication of the level of welfare at the regional level. Inequality measures available here are linked to income distributions and inequality measures in Module 2. The standard Gini and Theil's L measures of inequality along with estimated income shares of bottom and top income groups of population in these regions are also available.

5.1 Regions: Composition

Users can download aggregates for the World and for any of the geographical or administrative regions in ICP for any year in the range 1970 to 2019. Aggregates for country groups based on the World Bank income groupings are also available. Membership of these regions can be found in the latest WDI World Bank (2022).

Users may note that the countries belonging to different regions included in the compilation of inequality measures may differ from the membership of regions in Module 3. These differences reflect the fact that inequality data are not available for all the countries in all the regions. Consequently, users of Module 4 are advised to download real income series available as a part of this module (see below).

The available groupings are:

- South Asia; East Asia & Pacific; Sub-Saharan Africa; North America; Europe & Central Asia; Latin America and Caribbean; Middle East & North Africa; and the World.
- Low Income; Lower Middle Income; Upper Middle Income; and Upper Income.

In addition, users can select and download country groupings for two administrative regions

- OECD; and item European Union.

There are two sub-classes of the series provided in this module of UQICD. These are:

1. Estimates of the size of the economies of different regions and the world;
2. Inequality measures at the regional and global level along with decompositions measuring within and between regional inequalities.

5.2 Population and Real Income Measures at the Regional Level

In this Version of UQICD, all regional aggregates refer only to gross domestic product. Series focusing on the components of GDP will be included in future versions of UQICD. Series available here are:

1. POPULATION: Total population of the region

2. RGDP: Real GDP in PPP terms and current prices

5.3 Inequality Measures at the Regional Level

The following inequality related series are available from UQICD for all the years starting from 1970 until 2019. Consistent with the Inequality Module (Module 2), inequality measures are provided separately for all the four income distributions: Generalized-beta 2 (GB2); lognormal (LN); Mixture of log-normal distributions (MLN32 - mixture of three lognormal distributions with two distinct variances); and the Pareto-lognormal distribution (PLN). Each of these measures are associated with a distribution shown as a pre-fix. For example, GB2_GINI refers to the Gini measure of inequality computed using the GB2 (generalized beta 2) distribution fitted to income distribution data of each country. The user can select any of the four distribution options available.

The inequality measures for the regions are computed using the regional income distribution which is the population share weighted average of the distributions of the countries belonging to the region.

Let $f_j(y)$ and $F_j(y)$ represent, respectively, the density and distribution functions of income distribution of country j for $j \in R$. Then these functions for a given region are given by:

$$f_R(y) = \sum_{j \in R} \lambda_j f_j(y)$$

$$F_R(y) = \sum_{j \in R} \lambda_j F_j(y)$$

where λ_j for $j \in R$ is the population share of country j.

The inequality measures available from UQICD, with separate distributional prefixes, are:

1. DIST_GINI: Gini measure of inequality with specific functional form of the income distribution (DIST - GB2; LN; MLN32; and PLN)

2. DIST_THEIL: Theil's L-Measure of Inequality for the selected distribution

3. DIST_THEIL_W: Theil's within-country inequality for the selected distribution

4. DIST_THEIL_B: Theil's between-country inequality for the selected distribution

5. DIST_SH10: Share of the poorest 10 percent of the population of the region based on the selected distribution

6. DIST_SH30: Share of the poorest 30 percent of the population of the region based on the selected distribution
7. DIST_SHTOP10: Share of the richest 10 percent of the population of the region based on the selected distribution
8. DIST_SHTOP01: Share of the richest 1 percent of the population of the region based on the selected distribution

Chapter 6

Concluding Remarks

The main purpose of this Guide is to provide the users with background information regarding the series available in UQICD V3.0. The series provided here are generated using methodology developed by the research team and, therefore, are distinct from any comparable series on PPPs, real incomes and inequality available elsewhere. The appendices provide the readers with details of the methodology that underpins the extrapolation of PPPs and construction of panels of PPPs and real incomes; fitting income distributions with limited aggregate data; and computation of regional growth, inflation and inequality series.

The UQICD by its very nature will undergo constant updating as new data become available. In addition, the research team is working on further improvements to inequality series both in terms of coverage and the methodology to incorporate additional data that may be available from other sources. The team is also working on adding functionality to UQICD that facilitates users to undertake compilation of auxiliary series, such as series for country groupings other than those available in UQICD, designed for specific research projects.

Appendix A

Construction of panels of PPPs at current prices

The material in this section is a revised version of that available from UQICD User Guide Version 2.1.2.

A.1 The RRD Method

The material in this Appendix is based on Rao, Rambaldi, and Doran (2010b), Rao, Rambaldi, and Doran (2010a), Rambaldi, Rao, and Ganegodage (2010), and Rao and Rambaldi (2015). The econometric problem is one of signal extraction. That is, there are a number of sources of "noisy" information that can be combined to extract the signal. At any time period t the N countries can be placed in one of three groups when t is an ICP benchmark year or in one of two groups otherwise. In an ICP year, the groups are: the reference country (the US is used in UQICD), the non-participating countries and the participating countries. In a non-ICP benchmark year there are only two groups: the reference country and all others.

The following theoretical considerations underline the econometric model,

1. The log of the observed PPPs from the ICP in the benchmark years, $\tilde{p}_{it}$, are related to the log of the true and unobserved PPPs, p_{it}, through the following equation:

$$\tilde{p}_{it} = p_{it} + \xi_{it} \tag{A.1}$$

where ξ_{it} is a random error accounting for measurement error with the properties:

$$E(\xi_{it}) = 0;\ E(\xi_{it}^2) = \sigma_\xi^2 V_{it} \tag{A.2}$$

The measurement error variance-covariance is of the form

$$V_t = \begin{bmatrix} 0 & 0 \\ 0 & \sigma_{1t}^2 jj' + diag(\sigma_{2t}^2, \ldots, \sigma_{Nt}^2) \end{bmatrix}$$

where,

$V_{it} = \sigma_{1t}^2 + \sigma_{it}^2$, for $i \neq 1$ and $V_{1t} = 0$,

j is a vector of 1's,

σ^2_{it}, is the variance of the PPP from the ICP benchmark for country i in period t[1].

σ^2_{it} captures the accuracy of the national statistics compilation in country i. More developed countries that provide more resources to their national statistical agencies are expected to have a lower variance. In the empirical implementation of the method it is assumed to be inversely related to the GDP of country i in period t[2].

2. The numerical value of the PPP for the reference/numeraire country, PPP_{1t} is set to 1, and thus, its log form is given by,

$$p_{1,t} = 0;\ t = 1, 2, \ldots, T \tag{A.3}$$

3. The key element of the approach is the regression model used in extrapolating PPPs to non-participating countries using PPP data from the ICP benchmarks. The regression model draws on the literature on the explanation of national price levels (Kravis and Lipsey (1983); Clague (1988) and Bergstrand (1991, 1996)). A linear model in logarithms of price levels is postulated as below:

$$r_{it} = ln(PPP_{it}/XR_{it}) = \beta_{0t} + \mathbf{x'}_{it}\beta_s + e_{it} \tag{A.4}$$

for all $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$

Deviating from the usual assumptions on the disturbance term, we assume that errors in (A.4) are spatially autocorrelated . The following specification is used

$$e_t = \phi W_t e_t + u_t \tag{A.5}$$

where $|\phi| < 1$ and $W_t(N \times N)$ is a spatial weights matrix and $u_t \sim N(0, \sigma^2_u I_N)$. The term spatial in the present contexts refers to socio-economic distance rather than the traditional geographical distance. It follows that $E(u_t u_t')$ is proportional to $\Omega_t = (I - \phi W_t)^{-1}(I - \phi W_t)^{-1'}$. Inspection of equation (A.4) shows it is possible to obtain estimates of the parameters by using the unbalanced panel available from ICP, with the outcome (dependent) variable $\hat{r}_t = \tilde{p}_t - ln(XR_t)$.

With estimates of the parameters in (A.4), predictions of PPPs consistent with price level theory can be generated for any country in any period. These are given by:

$$\hat{p}_t = \hat{\beta}_{0t} + \mathbf{x'}_t\hat{\beta}_s + ln(XR_t) + \hat{\phi} W_t \hat{e}_t \tag{A.6}$$

These predictions serve as starting values for RRD estimation algorithm.

A.1.1 The RRD Estimation Model and Algorithm

Using (A.6) as a set of starting predictions, RRD embeds a re-written version as follows,

$$y_t{=}Z_t p_t{+}B_t X_t \theta{+}\zeta_t \tag{A.7}$$

where,

[1]crucial to the invariance of the method is that the variance of the numeraire country appears in the expression.

[2]In order to avoid circularity, GDP in $US adjusted by market exchange rates is used in the estimation process.

y_t is a vector of the following *observed* information (see equation (A.8)),

$\tilde{p}_t$ is a vector of log transformations of the ICP PPP benchmarks for participating countries, $P\tilde{P}P_t$.

$\hat{p}_t$ is a vector of log PPP regression predictions for non-participating countries. The predictions are based on (A.6).

$p_{1t} = ln(PPP_{1t}) = 0$, the first element of y_t is zero as this is the observation for the reference country which is a constraint in the system.

θ is a function of β_{0t} and β_s

B_t a mapping matrix to non-participating countries

Z_t is a partitioned selection matrix with components which select the reference country (country 1), S_1, the non-participating countries, S_{np}, and the participating ICP countries, S_p,

ζ_t is a random vector capturing the uncertainty arising from each set of sources of observed values of PPP_{it}.

In an ICP benchmark year, the vectors in (A.7) are defined as follows,

$$y_t = \begin{bmatrix} 0 \\ \hat{p}_t \\ \tilde{p}_t \end{bmatrix};\ Z_t = \begin{bmatrix} S_1 \\ S_{np} \\ S_p \end{bmatrix};\ \zeta_t = \begin{bmatrix} 0 \\ S_{np}\upsilon_t \\ S_p\xi_t \end{bmatrix} \tag{A.8}$$

In a non-ICP year there are no observations from the ICP, thus the only observations are those produced by the predictions from the price level model and the constraint, the system reduces to

$$y_t = \begin{bmatrix} 0 \\ \hat{p}_t \end{bmatrix};\ Z_t = \begin{bmatrix} S_1 \\ S_{np} \end{bmatrix};\ \zeta_t = \begin{bmatrix} 0 \\ S_{np}\upsilon_t \end{bmatrix} \tag{A.9}$$

The variance-covariance matrix of ζ_t is given by,

$$E\left(\zeta_t\zeta_t'\right) \equiv H_t = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \sigma_u^2 S_{np}\Omega_t S_{np}' & 0 \\ 0 & 0 & \sigma_\xi^2 S_p V_t S_p' \end{bmatrix} \tag{A.10}$$

the first row is zero as it represents the reference country constraint. The non-participating countries have error υ_t, and the ICP measures have error ξ_t (see equations (A.8) and (A.9)).

The RRD is set as a state-space system. The time evolution of PPPs, that is, the updating of PPPs from period $t-1$ to t, is using the standard definition, through the GDP deflators in the country concerned and in the reference country. Thus,

$$PPP_{i,t} = PPP_{i,t-1} \times \frac{GDPDef_{i,[t-1,t]}}{GDPDef_{1,[t-1,t]}} \tag{A.11}$$

Taking logarithms on both sides of (A.11), and assuming the updating equation (A.11) holds on average due to measurement error, we have

$$p_{it} = p_{i,t-1} + c_{it} + \eta_{it} \tag{A.12}$$

where $c_{it} = ln(\frac{GDPDef_{i,[t-1,t]}}{GDPDef_{US,[t-1,t]}})$; and η_{it} is random error accounting for measurement error in the growth rates. Equation (A.12) is commonly used in constructing panels of PPPs including the PWT and in the construction of the Maddison series[3]. The variance covariance matrix of η_{it} is assumed to be similar to the matrix in equation (A.2).

The RRD estimation algorithm involves the Kalman filter algorithm. Upon convergence, $\theta \rightarrow 0$ and (A.7) reduces to y_t=$Z_t p_t$+ζ_t, which is then used by the Kalman filter and Smoothing algorithm to produce estimates of the latent vector $p_t = \ln PPP_t$ and an associated mean squared prediction error matrix. The point to note here is that unlike the PWT and other extrapolation methods, this approach generates predictions for all the cells (time periods and countries). However, it is trivial to limit the regression based PPPs, $\hat{p}_t$, (through Z_t and B_t) to be used by the model's predictor to only those countries and years when no ICP benchmark observations, $\tilde{p}_t$, are available.

RRD approaches the problem of finding predictions for the vectors of PPPs by combining from a variety of sources of information through the ICP benchmarks, regression predictions, the nummeraire constraint, and, the updating equation in (A.12). A state-space representation is suitable for these kinds of problems. The approach proposed combines all the information in equations (A.7) to (A.12) into a state-space representation of the problem. They form of a set of observation and transition equations on the state vector p_t which is the vector of unobserved log PPPs, $ln(PPP_t)$. Under Gaussian assumptions, the Kalman filter and Smoother predictor of the conditional mean, $\widetilde{p_{it}}$, conditional on information available at time t, is a minimum squared error predictor of the state vector, p_t[4]. The panel of PPPs is then obtained by,

$$\widetilde{PPP}_{it} = \exp(\widetilde{p_{it}}); \;\; i = 1, ..., N \text{ and } t = 1, ..., T \tag{A.13}$$

where the wide "$\widetilde{\ }$" is used to denote the RRD estimates of the log of PPP, $\widetilde{p_{it}}$, and corresponding smoothed estimated PPPs, $\widetilde{PPP}_{it}$

A.2 Analytical properties of panels of PPPs from RRD

In order to provide a better appreciation of the features of RRD, a number of analytical results are presented here. In particular, these properties demonstrate the flexibility of the method and show how it provides intuitively meaningful predictions under specific scenarios. The following properties are stated without proofs but complete proofs are provided in the referenced materials.

1. *The predicted PPPs are weighted sums of the information in* $\{y_t\}$

   Using the results from Koopman and Harvey (2003) it is possible to express the RRD estimate of the PPPs at period t, $\widetilde{p_t} = \ln(\widetilde{PPP_t})$, as a weighted sum of information immediately closest to the time period t, with the highest weight at t and decreasing

[3] Maddison (2007) presents series that are extrapolated from the 1990 benchmark year.

[4] Technical details and equations for the Kalman Filter and Smoother are provided in Appendix A.6 and Appendix B of Rao, Rambaldi, and Doran (2010a).

weights for period j further away in each direction from t. The weights w_{jT} depend on the benchmark information, regression information and measurement error structures attached to that information.

$$\widetilde{p_j} = \sum_{j=1}^{T} \omega_{jT} y_j \tag{A.14}$$

The size and shape of the weights depend on the time period. The forms of the weights are shown in the Appendix of Huynh, Rambaldi, and Rao (2014). For example, if the sample goes up to 2012, for t =2011, the highest weight will be from the 2011 information with a discounted memory back to the start of the sample; although in practice most of the non-zero weights might be from the immediate past years. Some weight, although likely to be small, will be from the 2012 information. The weights sum to one.

The adjustment provided by the weights is from information about the movement of PPPs between benchmarks *after the deflator movement has been incorporated.* This information include national accounts data, ICP benchmarks and the influence of movements in other trading partners which is brought into the weights through the cross-sectional correlation information gathered by the price level regression used by the UQICD.

2. *The predicted PPPs are "weighted averages" of benchmark-year only extrapolations.* This property provides an intuitive explanation of how the method works.

   Suppose there are $M + 1$ benchmark years. If regression based predictions are used to extrapolate PPPs to non-participating countries *only in benchmark years* and then use the implicit price deflators to extrapolate from one year to the next, then it is possible to construct a panel of extrapolated PPPs for each of the benchmark years. In this case, an obviously intuitive approach is to make use of an average of these $M+1$ panels of PPPs. An important property of the RRD approach is that, in this case, the predictions $\widetilde{p_t}$ can be shown to be a weighted average of the $M + 1$ panels of PPPs, where the weights are determined by the diagonal elements of the 'Kalman Gain' matrices, which represent the gain in information provided by an additional benchmark. The weights can be interpreted as reflecting the reliability of the $j-th$ benchmark. The proof of this important property is presented in Rao, Rambaldi, and Doran (2010b).

3. *Invariance of the Predicted PPPs to the Choice of the Reference Country*

   The relative purchasing powers of currencies of countries should, in principle, be invariant to the choice of the reference country. It can be shown that RRD extrapolated PPPs satisfy this important invariance property. The proof of this property is presented in the Appendix of Rao and Rambaldi (2015).

4. *Constraining the model to track PPPs for countries participating in the benchmarks*

   As the ICP is the main source of PPPs for countries participating in different benchmarks, and given that respective PPPs are determined using price data collected from extensive price surveys, one may consider it necessary that the econometric method proposed should generate predicted PPPs that are identical to PPPs for the countries participating in different ICP benchmarks. In RRD this can be achieved by simply setting the variance of the disturbance term in equation (A.1) (see equation (A.2)) to be equal to zero. In this case a particular property of Kalman filter predictions is that the predicted PPPs

($\widehat{PPP}_{it}$) will be identical to the ICP benchmark, $P\tilde{P}P_{it}$[5], when t is a benchmark year. By imposing this constraint we produce the series *PPP_ICP_CON*.

5. *Constraining the model to preserve movements in the Implicit GDP Deflator*

   In the currently available PWT and the Maddison's series, growth rates in real GDP and movements in the implicit price deflators are preserved. As the GDP deflator data are provided by the countries and given that such deflators are compiled using extensive country-specific data, it is often considered more important that the predicted PPPs preserve the observed growth rates implicit in the GDP deflator. This essential feature can be guaranteed in RRD by simply stipulating the variance of the error in the updating equation (A.12) be zero. It is trivial to show that the national level movements in prices are preserved using the formulae for the fixed interval Kalman Smoother[6].

We note here that it is not possible to simultaneously constrain the predictors to track the benchmark PPPs as well as the national movements in GDP deflators. One has to choose either one or none of these restrictions when generating panels of extrapolated PPPs. The recommended approach is to simply use unconstrained equations and thereby not impose either of the restrictions described above.

The standard errors for the computed series in UQICD are computed as follows,

$$SE(\widetilde{PPP}_{it}) = \sqrt{exp(2 \times \widetilde{p_{it}})exp(\hat{\Psi}_{ii,t})exp(\hat{\Psi}_{ii,t} - 1)} \tag{A.15}$$

where $\hat{\Psi}_{ii,t}$ is the *ith* diagonal element of the estimated variance-covariance from RRD. The team is now working on the computation of standard errors for the UQICD constructed PPPs using bootstrapping techniques described in Huynh, Rambaldi, and Rao (2014).

[5] This result follows from the work of Doran (1992).

[6] The proof of this property is provided in Appendix B of Rao, Rambaldi, and Doran (2010a).

A.3 The Variables used in the price level regression

The variables used in the price level regression, (A.4), for GDP level are:

Variable	Description	Source
Response Variable		
PL_GDP	Price Level GDP = PPP_GDP_ICP\XR	
XR	The exchange rate	WDI and UN, although ultimate source is International Monetary Fund, International Financial Statistics.
PPP_GDP_ICP	PPP Data for GDP from ICP and OECD.	OECD; ICP for 1970, 1975, 1980, 1985 reports; World Bank downloaded from 1996 onwardsdata.
Predictor Variables:		
AGRIC	Agriculture, value added (% of GDP).	WDI and UN National Accounts database
D_APTA	Dummy for The Asia-Pacific Trade Agreement	Miscellaneous sources
D_ASEAN	dummy for ASEAN (Association of Southeast Asia) countries.	Miscellaneous sources
D_CAFTA	dummy for Central American free market.	Miscellaneous sources
D_CBERA	dummy for Caribbean Basin Economic Recovery Act	Miscellaneous sources
D_COMST2	dummy variable for countries either current or former communist rule	Miscellaneous sources
D_EU	Dummy for the European Union	Miscellaneous sources
D_EURO	Dummy for the Euro Area	Miscellaneous sources
D_EUROPEG	Dummy variable for countries with currencies (CFA_franc) pegged to European Euro	IMF
D_GAFTA	dummy for greater Arab free trade area	Miscellaneous sources
D_ISLAND	dummy variable for Islands	CIA fact book
D_LANDLOCKED	dummy variables for landlock countries	CIA fact book
D_MERCOSUR	dummy for MERCOSUR (an economic and political agreement among some south American countries) countries.	Miscellaneous sources
D_NAFTA	dummy for NAFTA (North American Free trade Agreement) countries	Miscellaneous sources
D_POUNDPEG1	Dummy for Pound peg (Shambaugh's pegtype 1 and 2)	IMF
D_USDPEG	Dummy variable for countries with currencies either pegged to the US$ for substantial amounts of time or use US$ as the legal tender - during the post-Bretton Woods era (1973 onwards).	IMF
D_WACFA	dummy variable for countries with common west African CFA franc currency	Miscellaneous sources

Variable	Description	Source
INTERNET	Internet users (per 100 people)	WDI
KA_OPEN	The Chinno-Ito index	Link–> M. Chinn and H. Ito
LABPOP1	Labour force as percentage of total population. For developing countries the labour force is simply defined as the "economically active"" population	Labour force as percentage of total population. For developing countries the labour force is simply defined as the "economically active"" population
LIFE	Life expectancy at birth, total (years): Life expectancy at birth indicates the number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life.	WDI
MOBILE	Mobile cellular subscriptions per 100 people	WDI
NONTRADBLE	Non-tradable sector value added (% of GDP): Sum of Construction, Wholesale, retail trade, restaurants and hotels, Transport, storage and communication and "Other Activities"	WDI and UN National Accounts database
PHONE	Telephone mainlines (per 1,000 people): Telephone mainlines are telephone lines connecting a customer's equipment to the public switched telephone network. Data are presented per 1,000 people for the entire country.	WDI
SECSCHOOL	School enrolment, secondary per '000 gross enrolment	CNTS/WDI
SERVICE	Service, value added (% of GDP).	Undata base
TRADE	Trade (% of GDP): Trade is the sum of exports and imports of goods and services measured as a share of gross domestic product.	WDI and UN National Accounts database

Appendix B

Construction of Panels of PPPs for the Components of GDP

The material for this section is an updated version of Huynh, Rambaldi, and Rao (2014) on the extrapolation of PPPs for components of GDP: Private Consumption (CONS), Government Expenditure (GOV), and Gross Capital Formation (GCF). In order to do the extrapolation, economic models and econometric method are needed. The econometric method chosen to estimate PPPs for GDP components is the RRD method described in Appendix A. This section is devoted to discussing the economic models for each component together with the data constraints.

There are not established models for identifying the macroeconomic considerations used in the specification of models for the price levels of the components of GDP, Private Consumption (CONS), Government Expenditures (GOV), and Gross Capital Formation (GCF). Some of the issues are considered below.

B.1 Private Consumption (CONS)

Household consumption is the largest component of GDP. On average, individual consumption constitutes 69 percent of GDP (World Bank (2008)). It constitute personal expenditures that fall under one of the following categories: durable goods, non-durable goods, and services. Examples include: food, jewelry, gasoline, medical expenses, and rent, although it does not include owner occupied housing expenses. Hence, consumption, like GDP, involves both tradables (goods) and nontradables (services). Therefore, the structural determinants of a consumption price level are similar to those of GDP. There is no ready-made model in macroeconomics to account for the price level of private consumption, but a close examination of the theoretical reasonings of the structural determinants of national price level, suggest that these are also applicable to consumption price levels. The conclusion is that in constructing an economic model to explain aggregate consumption's price level, the set of variables, X_t, included in the regression component of RRD (see equations (A.4) and (A.7)) for explaining GDP price levels (PPP/XR), can be used for consumption.

B.2 Government Expenditure (GOV)

Government expenditure contains government consumption on final goods and services and gross government investment. Examples of government consumption spending include salaries of public servants to produce and provide services to the public, such as public school education, health care, defense, justice, general administration, and the protection of the environment. Gross investment by the government consists of spending for fixed assets that directly benefit the public, such as highway construction, or that assist government agencies in their production activities, such as purchases of military hardware. It does not include any transfer payments, such as social security or unemployment benefits Burstein, Neves, and Rebelo (2004b). Therefore, government expenditure mainly consists of salary payments to government employees and purchase of tradable goods like machinery and equipment or military weapons.

From macroeconomic theory we know that salary payments or wage rates are determined by the marginal productivity of labour. As a result, labour in high-income countries with high labour productivity will earn higher wages than their counterpart in low-income countries, which postulates a positive relationship between wage rates and per capita income. The price of capital goods like equipment and military hardware are, on the other hand, likely to be negatively correlated with per capita income. Therefore, the relationship between the overall price level of government expenditure (which is the combination of wage-rates and capital-goods price levels) and per capita income depends on the proportion of service (employment) and tradable goods purchased. It is also found that the volume of military spending is positively correlated with the national price level (Bergstrand (1996)), hence positively correlated with the government expenditure price level.

In summary, an economic model explaining government expenditure price levels would ideally include variables explaining wage rates and prices of capital goods; which are variables measuring labour productivity, average income, proportion of service and goods purchased by the government, volume of military spending and investment rates of the governments.

B.3 Gross Capital Formation (GCF)

In most countries Investment or Gross Fixed Capital Formation together with government account for a third of GDP, though there are exceptions, like China. Investment mostly comprises purchases of machinery and equipment and construction services. Distribution services (wholesale, retail, and transportation) are much less important for investment than for consumption (World Bank (2008), Burstein, Neves, and Rebelo (2004a)). Examples include business investment in equipment, construction of a new mine, purchase of software, purchase of machinery and equipment for a factory or spending by households (not government) on new houses. One point to note is that investment in this context does not include exchanges of existing assets or purchases of financial products. Buying financial products is actually classified as 'saving', as opposed to investment.

From the two main categories of investment: equipment purchase and construction, it can be inferred that investment involves both tradable goods and nontradable services like GDP and consumption. However, while consumption contains in itself a higher proportion of nontradables, investment mostly involves tradable capital goods as the import content of investment is much larger than that of consumption (Burstein, Neves, and Rebelo (2004a)). It is agreed that services prices are lower in low-income countries, but it is controversial whether equipment or capital-goods prices are the same across nations.

Hsieh and Klenow (2007) claim that the absolute price of capital goods is no higher in poor countries than in rich countries. Their study, which uses data from the Penn World Tables, produces positive and mostly significant results suggesting, if anything, a higher investment price in rich countries. The authors explain that the high relative price of investment in poor countries is due to the low price of consumption goods in those countries since poor countries have low efficiency in producing investment goods and need to produce consumer goods to trade for them (Hsieh and Klenow (2007)). This result is exactly what is predicted by the Balassa-Samuelson hypothesis. On the contrary, common views and other empirical evidence seems to suggest the opposite. (Alfaro and Ahmed (2010)) use highly disaggregated data on trade in capital goods to study differences in the price of capital across countries and find that the price of imported capital goods is negatively and significantly correlated with the income of the importing country. This finding explains why in poor countries, the relative prices of capital to consumption goods are observed to be higher.

Several hypotheses have been proposed to explain why tradable capital goods are actually more expensive in poor countries. The first reason might be the measurement problems in the PWT and ICP price data set, especially in regards to developing countries. RRD also acknowledges this problem by incorporating measurement errors for the ICP into their econometric model, assuming that the variance of errors are inversely related to income per capita (see equation (A.1) and footnote 2). The second possible reason is price discrimination, which means producers set their selling prices of the same goods higher for poorer countries. Price discrimination has been discussed in the literature (Mertens and Ginsburgh (1985), Verboven (1996), and Ayres and Siegelman (1995)), speculating that it might be profitable for firms to charge higher prices to groups of consumers that have a lower average reservation price if the variance of reservation prices within the group is sufficiently large. Within the context of traded capital goods, a vendor that knows this might rationally charge higher prices to all of its customers in poor countries. The third possible reason is transaction costs. For many developing countries, high tariffs or other forms of capital controls would likely drive up the price of imported capital goods. Besides, higher costs for poor countries are associated with searching for and negotiating (directly or indirectly) foreign purchases, as well as the volume of trade. Low-income countries might also be paying more for capital goods shipped in smaller quantities (Alfaro and Ahmed (2010)).

Other factors beside income that are documented to affect capital goods prices are investment rates or growth (Alfaro and Ahmed (2010)). For example, in research using a data set for capital-goods and equipment prices covering the 1870–1950 period for 11 OECD countries, the authors argue that relative capital-goods prices are strongly negatively correlated with investment rates (Collins and Williamson (2001)).

From the discussion above, there are several groups of variables that should ideally be included in the economic model explaining investment price levels. These are variables that measure the proportion of equipment purchases (tradable capital-goods) to construction services (non-tradables), income per capita, transaction costs (e.g. capital control, volume of trade), investment rates and growth.

B.4 Data constraint and choices of variables

Data for the PPP extrapolation of CONS, GOV and GCF are the ICP benchmark PPPs for the components, the socio-economic data for each country and the bilateral trade data required to compute the spatial weights matrix (see equation (A.5)).

The benchmark PPP data for Consumption, Government expenditure and Investment were collected from two different sources for the 11 benchmarks. For 1970, 1973, 1975, 1980, 1985, 2005, 2011 and 2017, benchmark PPP data for the components were collected from ICP and the remaining years of 1990, 1993, 1996, 1999, 2002, 2008, 2014 were obtained from Eurostat-

OECD. Several features of the PPP data are noteworthy. The number of countries varies across benchmarks. The first benchmark (1970) covered only 13 countries, whereas the most recent (2017) benchmark represents truly global comparisons with 176 countries. A related point worth noting is that PPPs for all the benchmarks prior to 1990 were based on the GK method and PPPs for the more recent years are all based on the GEKS method of aggregation.

The socio-economic data and the already computed spatial weight matrix are both obtained from the UQICD database. In this database there are socio-economic variables, variables representing productivity level, the degree of openness of the economy, national resources, trade balance, currency and trade agreements.

The spatial weight matrix W_t (in equation (A.5)) used in modeling the spatial error structure is proportional to trade closeness as measured by bilateral trade flows (see Rambaldi, Rao, and Ganegodage (2010)).

The dimensions of the extrapolation were largely determined by data availability. A number of countries were excluded because of missing data and the time frame 1970-2019 was likewise chosen because of poor data availability prior to 1970.

Socio-Economic Variables (forming $\boldsymbol{x}_{it}$ in equation (A.4)) included in the regression are chosen based on the determinants of the national price level as well as the structural economic determinants of price level for each component discussed above, and, the availability of data. Details of the variables chosen for each component are discussed next.

Explanatory variables for private consumption

The economic model used to explain the consumption price level includes all the variables that explain the national price level. They are per capita income, national resources, the degrees of openness, international tourism, country size, foreign trade ratios and trade balance. Exchange-rate adjusted per capita income for each country is used to construct the matrix V_t (see equation (A.2)). Per capita income cannot be used directly as an explanatory variable since there will be an endogeneity problem. To overcome this difficulty, variables representing productivity, which are in accordance with the productivity differential model of Balassa for explaining the national price level, are used as proxies for per capita income.

The predictor variables used in the regression are the same as those used in the GDP level regression (see Table A.1). The response variable is $PL\_CONS = \frac{PPP\_CONS\_ICP}{XR}$.

Explanatory variables for government expenditure

The same procedure of variable selection for Private Consumption is used for Government Expenditure. First, the theoretical discussion by Bergstrand (1996) suggests that the government expenditure price level would ideally include variables explaining wage rates and capital-goods price those which measure labour productivity, average income, proportion of service and goods purchased by the government, volume of military spending and investment rates of the governments. However, we did not have data on the share of services to government expenditure, or the volume of military spending and investment rates. As a result, a set of variables representing productivity and average income are selected together with economic variables which include measures of trade balance and degree of openness. The predictor variables used in the regression are the same us those used for the GDP level regression (see Table A.1). The response variable is $PL\_GOV = \frac{PPP\_GOV\_ICP}{XR}$.

Explanatory variables for gross capital formation

Among the three components, Gross Capital Formation is the most difficult one to model given our current dataset. From a theoretical perspective, the group of variables that should ideally be included in the economic model explaining investment price are: variables that measure the proportion of equipment purchased (tradable capital-goods) contraction services (non-tradables), income per capita, transaction costs (e.g. capital control, volume of trade), investment rates and growth. Given the available data, a set variables which represents income per capita, transaction costs (capital control), and trade volume are used. The predictor variables used in the regression are the same us those used for the GDP level regression (see Table A.1). The response variable is $PL\_GCF = \frac{PPP\_GCF\_ICP}{XR}$. This regression has the lowest R-sq amongst those from the three components.

The benchmark PPPs of Gross Capital Formation are found to be highly correlated with the market exchange rate (with correlation coefficient of 0.95). This reflects the fact that investment goods are mostly tradables. However, we cannot use exchange rate as an explanatory variable given it is in the denominator of the dependent variable. Hopefully, the explanatory power of the regression will be improved when we can include variables that measure the tradables-to-nontradables ratio in investment and investment rates.

Appendix C

Real Incomes at Current and Constant Prices

The construction of consistent panels of comparable incomes over time and space is described below. We first provide a brief description focusing on aggregate gross domestic product (GDP).

Let GDP_{it} represent GDP in country i in period t expressed in local or national currency units. These GDP aggregate measures are not comparable across countries or over time as they are influenced by price levels in the respective countries and time periods.

Let XR_{it} and PPP_{it} respectively denote the exchange rate and the purchasing power parity of the currency of country i which is equivalent to one unit of currency of a reference or numeraire country[1]. The *nominal and real* GDP of country i in period t, respectively, denoted as NGDP and RGDP, expressed in the currency units of a reference country, are given by

$$NGDP_{it} = \frac{GDP_{it}}{XR_{it}} \tag{C.1}$$

and

$$RGDP_{it} = \frac{GDP_{it}}{PPP_{it}} \tag{C.2}$$

The *NGDP* adjusts for differences in currency units. In contrast, RGDP adjusts for differences in currency units as well as purchasing powers of currencies based on differences in price levels observed in different countries[2]. We note a few features of the real GDP series.

1. $RGDP_{it}$ is comparable and additive across countries *at a given period* t but not for countries at different points of time. It is possible to compute regional totals for the period t.

2. $RGDP_{it}$ is not comparable to $RGDP_{ks}$ for all t not equal to s. Thus $RGDP_{it}$ may be termed real GDP series at *current (period t) prices*. However, this does not necessarily mean that there is a set of prices which can be used as reference prices in deriving the real GDP series.[3]

[1]We drop the subscript for the reference country to keep the notation simple. A superscript to indicate the reference country would be useful but suppressed to facilitate notational brevity.

[2]RGDP is basically a volume measure, a concept advocated for use by the System of National Accounts (SNA) and emphasised in McCarthy (2013)

[3]See Rao and Balk (2013) for a definition of real income comparisons at a set of reference prices and for

3. $RGDP_{it}$ and PPP_{it} are typical outputs of the ICP if t is a benchmark year.

4. $RGDP_{it}$ is obtained by deflating GDP by using a suitable spatial price deflator; here it is PPP_{it}.

The PPP_{it} and $RGDP_{it}$ series for periods $t = 1, 2, 3, \ldots, T$ and countries $i = 1, \ldots, N$ are called a a panel of PPPs and real incomes at *current or period t prices* to emphasize the fact that these PPPs and real GDP aggregates are not comparable over time. The problem of construction of these series at current prices has been satisfactorily addressed by the PWT and also the RRD econometric approach discussed in Appendix A, and we denote by $\widetilde{PPP}_{it}$ the predictions of PPP_{it} constructed from either of these approaches[4].

Given these definitions and the underlying notation, the main problem is one of constructing panels of real incomes at *constant* reference country k (for example USA) period t prices (for example 2011).

C.1 Constant time-space comparisons

In the construction of constant time-space comparisons, cross-country comparisons in any given year are made using the UQICD series $RGDP_{t,k,i}$. For the rest of the presentation in this section, and in Appendix E, a re-arranged notation of subscripts is used. The current year prices will be denoted by t, the base price year will be denoted by τ, country k will be the numeraire (suppressed to simplify notation), and the country for which the aggregate is defined will be denoted by i.

1. Let $CRGDP_{\tau,t,i}$ represents real GDP of country i and period t, and in constant prices of year τ[5]

2. Let $Def_{t,i}$ represent the GDP deflator for country i at time t:

UQICD V3.0 constructs $CRGDP_{\tau,t,i}$ as follows:

$$\begin{aligned} CRGDP_{\tau,t,i} =& RGDP_{t,i} \times PPP_{t,i} \times \frac{Def_{\tau,i}}{Def_{t,i}} \times \frac{1}{PPP_{\tau,i}} \\ =& GDP_{t,i} \times \frac{Def_{\tau,i}}{Def_{t,i}} \times \frac{1}{PPP_{\tau,i}} \\ =& \frac{CGDP_{\tau,t,i}}{PPP_{\tau,i}} \end{aligned} \tag{C.3}$$

$CGDP_{\tau,t,i}$ represents the GDP of country i, in domestic currency in period t expressed in constant period τ prices. We use $\tau = 2011$ and k =USA such that $CGDP_{\tau,t,i}$ is expressed interms of 2011 US dollars.

examples where deflated series could be interpreted as real income comparisons at some reference prices. For example, the GK based real GDP figures could be considered as real income comparisons obtained at GK international prices along with a Leontief utility function. Similarly, real income comparisons obtained using the Tornqvist index as the deflator corresponding with a translog cost function.

[4]Any panel of PPPs at current prices can be used as a starting point.

[5]These series are similar to the GDP series at constant prices produced by national statistical offices except that the focus in such cases is on a single country.

Similarly, constant price real Household Consumption (CONS), Government Expenditures (GOV) and Gross capital formation (GCF) can be constructed by replacing the current prices PPPs at the GDP level by the corresponding component specific PPPs (PPP_CONS, PPP_GOV, PPP_GCF) and GDP by the corresponding expenditure aggregates (CONS_LCU_CURRENT, GOV_LCU_CURRENT, GCF_LCU_CURRENT) in the above derivations.

Appendix D

Fitting Income Distributions to Aggregate Data

The data for estimating income distribution was available in the grouped form with data on the proportion of each group and its corresponding mean. Here we describe the Generalized Method of Moments that was used to fit the income distributions. This section is based on Hajargasht et al. (2012) and Griffiths and Hajargasht (2015)

Consider a sample of T observations $(y_1, y_2, \ldots, y_T)$ which are randomly drawn from a parametric income distribution $f(y;\phi)$, and grouped into N income classes where $(z_0, z_1), (z_1, z_2), \ldots, (z_{N-1}, z_N)$ are exogenously chosen class bounds with $z_0 = 0$ and $z_N = \infty$. The grouped data are generally available in the form of population proportions $c = (c_1, c_2, \ldots, c_N)$, and class mean incomes $\bar{\mathbf{y}} = (\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_N)$. Normally, class bounds are not reported; therefore we assume that data is not available for the z_i, and our problem is to estimate ϕ, along with the unknown class limits $z_1, z_2, \ldots, z_{N-1}$. To estimate the model using GMM we set up sample moment conditions such as

$$\mathbf{H}(\theta) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{h}(y_t, \theta) \tag{D.1}$$

where $\theta = (z_1, z_2, \ldots, z_{N-1}, \phi')'$. The GMM estimator $\hat{\theta}$ is defined as

$$\hat{\theta} = \arg\min_{\theta} \mathbf{H}(\theta)' \mathbf{W}\mathbf{H}(\theta) \tag{D.2}$$

where $\mathbf{W}$ is a weight matrix. The moment condition associated with our grouped data can be written as

$$\mathbf{H} = \begin{bmatrix} c_1 - k_1(\theta) \\ \vdots \\ c_N - k_N(\theta) \\ \bar{y}_1 - \mu_1(\theta) \\ \vdots \\ \bar{y}_N - \mu_N(\theta) \end{bmatrix} = \begin{bmatrix} c - k(\theta) \\ \bar{y} - \mu(\theta) \end{bmatrix} \tag{D.3}$$

For the sample proportions c_i, we note that the corresponding population proportions, that we denote by $k_i(\theta)$, are given by

$$k_i(\theta) = \int_{z_{i-1}}^{z_i} f(y;\phi)dy \quad i = 1, 2, \ldots, N \tag{D.4}$$

For group means the we note that the corresponding population means are given by $\mu_i(\theta) = \tilde{\mu}_i(\theta)/k_i(\theta)$ where

$$\tilde{\mu}_i(\theta) = \int_{z_{i-1}}^{z_i} yf(y;\phi)dy \tag{D.5}$$

Griffiths and Hajargasht (2015) have derived the optimal weight matrix as follows

$$\mathbf{W} = \begin{pmatrix} \mathbf{D}(\mathbf{1}/\mathbf{k}) & 0 \\ 0 & \mathbf{D}(\mathbf{k}/\boldsymbol{v}) \end{pmatrix} \tag{D.6}$$

where $\mathbf{k}$ is a vector including k_is, $\mathbf{D}$ denotes a diagonal matrix, division here is element by element division, $v_i = \tilde{\mu}_i^{(2)}/k_i - \tilde{\mu}_i^2/{k_i}^2$ and $\tilde{\mu}_i^{(2)}(\theta) = \int_{z_{i-1}}^{z_i} y^2 f(y;\phi)dy$.

Estimation requires the exact forms of $k_i(\theta)$, $\tilde{\mu}_i(\theta)$ and $\tilde{\mu}_i^{(2)}(\theta)$ for each distribution. For lognormal and GB2 these can be found in Hajargasht et al. (2012), for Pareto-lognormal in Hajargash and Griffiths (2013) and for mixtures of lognormals in Griffiths, Chotikapanich, and Hajargasht (2022).

Appendix E

Decomposing GDP change

The material in this section follows Balk, Rambaldi, and Rao (2022) (BRR) and Rao, Rambaldi, and Doran (2015). The BRR decomposition of the change in Real GDP of the world (or region) into Growth, Inflation and PPP exchange rate change effects is discussed below. The same decomposition is applicable for any sub-group of countries. UQICD provides this decomposition for geographical regions, income based country groups (as per World Bank definitions in December 2019); and the administrative groups, OECD and EU.

E.1 Decomposition of change in real GDP

A useful starting point for this exposition is the decomposition of change in GDP of a country, expressed in local currency units, over a period. For purposes of illustration the years 2011 and 2017, which represent the last two ICP benchmarks, is considered.

E.1.1 Decomposition of change in GDP of a country measured in local currency units

At the country level, let $GDP_{2011,i}$ and $GDP_{2017,i}$ represent GDP of country i in current 2011 and 2017 prices respectively. Further let $\{Def_{2011,2017,i}\}$ represent the GDP deflator in country i for 2017 with 2011 as the base. Then, the change in GDP at current prices can be decomposed as follows:

$$\begin{aligned}
\frac{GDP_{2017,i}}{GDP_{2011,i}} &= \frac{GDP_{2017,i}/Def_{2011,2017,i}}{GDP_{2011,i}/Def_{2011,2011,i}} \times \frac{Def_{2011,2017,i}}{Def_{2011,2011,i}} = \frac{CGDP_{2011,2017,i}}{CGDP_{2011,2011,i}} \times \frac{Def_{2011,2017,i}}{Def_{2011,2011,i}} \\
&= \frac{CGDP_{2011,2017,i}}{CGDP_{2011,2011,i}} \times Def_{2011,2017,i} \\
&= GR_{2011,2017,i} \times \frac{Def_{b,2017,i}}{Def_{b,2011,i}} \\
&= \text{growth rate in country } i \times \text{domestic price change in country } i
\end{aligned} \tag{E.1}$$

where $CGDP_{2011,2017,i}$ represents GDP of country i in 2017 expressed in constant 2011 prices.

E.1.2 Decomposition of change in world GDP expressed in the currency unit of the reference country using PPPs

The decomposition shown here is equally applicable to any regional grouping of countries. The change in world real GDP, expressed in the currency units of USA using PPPs, is the ratio of sum of real GDP of all the countries in the world. As the current ICP covers 176 countries, this change is represented by the following ratio:

$$\frac{RGDP_{2017,W}}{RGDP_{2011,W}} = \frac{\sum_{i=1}^{176} GDP_{2017,i}/PPP_{2017,USA,i}}{\sum_{i=1}^{176} GDP_{2011,i}/PPP_{2011,USA,i}} \tag{E.2}$$

Balk, Rambaldi, and Rao (2022) (BRR) provide a decomposition of the ratio in (E.2) using the Sato-Vartia index number formula into global growth and global inflation components. Global inflation in turn has two components, one based on domestic inflation rates and the other on the effect of PPP change. Using the same notation as in equation (E.1), the BRR decomposition of change in world real GDP is given by:

$$\begin{aligned}\frac{RGDP_{2017,W}}{RGDP_{2011,W}} &= \prod_{i=1}^{176} \left[GR_{2011,2017,i}\right]^{w_i} \times \prod_{i=1}^{176} \left[Def_{2011,2017,i}\right]^{w_i} \times \prod_{i=1}^{176} \left[\frac{PPP_{2011,USA,i}}{PPP_{2017,USA,i}}\right]^{w_i} \\ &= \text{Global growth} \times \text{Average of domestic inflation rates} \times \text{PPP change effect} \\ &= \text{Global growth} \times \text{Global inflation}\end{aligned} \tag{E.3}$$

Each component in (E.3) is a weighted average of the corresponding country-specific changes.

Let $s_{t,i}$ represent the share of real GDP of country i in the world real GDP,

$$s_{t,i} = \frac{RGDP_{t,i}}{RGDP_{t,W}} = \frac{RGDP_{t,i}}{\sum_{i=1}^{176} RGDP_{t,i}} \qquad t = 2011, 2017 \tag{E.4}$$

then the weights, w_i in (E.3) based on the Sato-Vartia index, are given by

$$w_i = \frac{L\left(s_{2011,i}, s_{2017,i}\right)}{\sum_{j=1}^{176} L\left(s_{2011,j}, s_{2017,j}\right)}$$

and

$L(a,b)$ is the logarithmic average of numbers a and b given by:

$$L(a,b) \equiv \frac{a-b}{\ln a - \ln b} \; if \; a \neq b; \text{ and } L(a,a) \equiv a.$$

Bibliography

Alfaro, Laura and Faizal Z. Ahmed (2010). *The Price of Capital: Evidence from Trade Data*. Tech. rep. SSRN eLibrary.

Ayres, Ian and Peter Siegelman (1995). "Race and Gender Discrimination in Bargaining for a New Car". In: *The American Economic Review* 85(3), pp. 304–321.

Balk, Bert M., Alicia N. Rambaldi, and D. S. Prasada Rao (2022). "Macro-Economic Measures For A Globalized World: Global Growth And Inflation". In: *Macroeconomic Dynamics* 26.2, pp. 314–360. DOI: 10.1017/S1365100520000152.

Bergstrand, Jeffrey H. (1991). "Structural Determinants of Real Exchange Rates and National Price Levels: Some Empirical Evidence". In: *American Economic Review* 81.1, pp. 325–334.

— (1996). "in International Comparisons of Prices, Output and Productivity: Contributions to Economic Analysis Series, J. Salazar-Carrillo and D.S. Prasada Rao (eds)". In: Elsevier Science Publishers. Chap. Productivity, Factor Endowments, Military Expenditures, and National Price Levels, pp. 297–317.

Burstein, Ariel, Joao C. Neves, and Sergio Rebelo (Jan. 2004a). *Investment Prices and Exchange Rates: Some Basic Facts*. NBER Working Papers 10238. National Bureau of Economic Research, Inc. URL: http://ideas.repec.org/p/nbr/nberwo/10238.html.

— (2004b). "Investment Prices and Exchange Rates: Some Basic Facts". In: *NBER Working Papers* 10238.

Chotikapanich, Duangkamon, William E. Griffiths, and Gholamreza Hajargasht (2022). "Modelling Income Distributions with Limited Data". In: *Advances in Economic Measurement*. Ed. by Duangkamon Chotikapanich, Alicia N. Rambaldi, and Nicholas Rohde. Springer Nature Singapore Pte Ltd: Pallgrave Macmillan. Chap. 5, pp. 233–263.

Chotikapanich, Duangkamon, William E. Griffiths, and D.S. Prasada Rao (2007). "Estimating and Combining National Income Distributions using Limited Data". In: *Journal of Business and Economic Statistics* 25, pp. 97–109.

Chotikapanich, Duangkamon et al. (2012). "Global income distributions and inequality, 1993 and 2000: Incorporating country-level inequality modeled with beta distributions". In: *Review of Economics and Statistics* 94.1, pp. 52–73.

Chotikapanich, Duangkamon et al. (2014). "Inequality and poverty in Africa: regional updates and estimation of a panel of income distributions". In: *IARIW 33rd General Conference, Rotterdam*. Citeseer.

Chotikapanich, Duangkamon et al. (2018). "Using the GB2 income distribution". In: *Econometrics* 6.2, p. 21.

Clague, Christopher K. (1988). "in World Comparisons of Incomes, Prices and Product: Contributions to Economic Analysis Series, J. Salazar-Carrillo and D.S. Prasada Rao (eds)". In: Elsevier, North Holland. Chap. Explanations of National Price Levels, pp. 237–262.

Collins, William J. and Jeffrey G. Williamson (2001). "Capital-goods prices and investment, 1870-1950". In: *The Journal of Economic History* 61, pp. 59–94.

Colombi, Roberto (1990). "A new model of income distribution: The pareto-lognormal distribution". In: *Income and wealth distribution, inequality and poverty*. Springer, pp. 18–32.

Doran, Howard E. (1992). "Constraining Kalman Filter and Smoothing Estimates to Satisfy Time-Varying Restrictions". In: *Review of Economics and Statistics* 74.3, pp. 568–572.

Griffiths, William E., Duangkamon Chotikapanich, and Gholamreza Hajargasht (2022). "A note on inequality measures for mixtures of double Pareto–lognormal distributions". In: *Australian Economic Papers* 61.2, pp. 280–290.

Griffiths, William E. and Gholamreza Hajargasht (2012). *GMM estimation of mixtures from grouped data: an application to income distributions*. Tech. rep. Working Paper.

— (2015). "On GMM estimation of distributions from grouped data". In: *Economics Letters* 126, pp. 122–126.

Hajargash, Gholamreza and William E. Griffiths (2013). "Pareto–lognormal distributions: Inequality, poverty, and estimation from grouped income data". In: *Economic Modelling* 33, pp. 593–604.

Hajargasht, Gholamreza and William E. Griffiths (2020). "Minimum distance estimation of parametric Lorenz curves based on grouped data". In: *Econometric Reviews* 39.4, pp. 344–361.

Hajargasht, Gholamreza et al. (2012). "Inference for income distributions using grouped data". In: *Journal of Business & Economic Statistics* 30.4, pp. 563–575.

Hsieh, Chang-Tai and Peter J. Klenow (2007). "Relative Prices and Relative Prosperity". In: *American Economic Review* 97.3, pp. 562–585.

Huynh, Linh Tu, Alicia N. Rambaldi, and D.S. Prasada Rao (Aug. 2014). "Econometric Estimation and Aggregation of PPP Panels for Components of GDP". In: *33rd IARIW General Conference, the Netherlands (http://www.iariw.org/papers/2014/Rambaldi2Paper.pdf)*.

International Monetary Fund (2022). *World Economic Outlook*. URL: `https://www.imf.org/en/Publications/WEO`.

Kleiber, Christian and Samuel Kotz (2003). *Statistical size distributions in economics and actuarial sciences*. John Wiley & Sons.

Koopman, Siem J. and Andrew C. Harvey (2003). "Computing observation weights for signal extraction and filtering". In: *Journal of Economic Dynamics and Control* 27, pp. 1317–1333.

Kravis, Irvin B. and Robert E. Lipsey (1983). *Toward an Explanation of National Price Levels*. Princeton University Press, Princeton.

Maddison, Angus (2007). *Contours of the World Economy 1-2030 AD: Essays in Macro-Economic History*. Oxford: Oxford University Press.

McCarthy, Paul (2013). "in Measuring the Real Size of the World Economy: The Framework, Methodology, and Results of the International Comparison Program-ICP, World Bank". In: World Bank. Chap. National Accounts Framework for International Comparisons: GDP Compilation and Breakdown Process, Ch. 1, pp. 59–93.

McDonald, James B. and Yexiao J. Xu (1995). "A generalization of the beta distribution with applications". In: *Journal of Econometrics* 66.1-2, pp. 133–152.

Mertens, Yves and Victor Ginsburgh (1985). "Product Differentiation and Price Discrimination in the European Community: The Case of Automobiles". In: *Journal of Industrial Economics* 34.2, pp. 151–166.

Rambaldi, Alicia N., D. S. Prasada Rao, and Kumudini R. Ganegodage (2010). "Modelling Spatially Correlated Error Structures in the Time-Space Extrapolation of Purchasing Power Parities." In: *Price Index Numbers in Time and Space, L. Biggeri and G. Ferrari (eds)*. Contributions to Statistics. Springer-Verlag, Berlin, Germany, pp. 63–96.

Rao, D. S. Prasada (2013). "Measuring the Real Size of the World Economy: The Framework, Methodology, and Results of the International Comparison Program (ICP)". In: ed. by World Bank. World Bank. Chap. The Framework of the International Comparison Program.

Rao, D. S. Prasada and Bert M. Balk (2013). “Concepts and Terminology for International Comparisons of Prices and Real Incomes”. In: *Ottawa Group. International Group on Price Indices.* Statistics Denmark. Copenhagen. URL: http://www.ottawagroup.org/Ottawa/ottawagroup.nsf/home/Meeting+13+-+Copenhagen,+2013.

Rao, D. S. Prasada and Alicia N. Rambaldi (2015). “An Econometric Approach to the Extrapolation of PPPs to Non-Benchmark Years”. In: *3rd Post-2011 ICP RCA Meeting, International Comparison Program. World Bank.* URL: http://pubdocs.worldbank.org/en/267001488208862014/RaoRCA-2015.pdf.

Rao, D. S. Prasada, Alicia N. Rambaldi, and H.E. Doran (2010a). “An econometric approach to the construction of complete panels of purchasing power parities: analytical properties and empirical results”. In: *The Econometric Society World Congress, 17-21 August, 2010. Shanghai.* URL: http://www.webmeets.com/files/papers/ESWC/2010/2931/WorldCongress.pdf.

Rao, D. S. Prasada, Alicia N. Rambaldi, and Howard E. Doran (2010b). “Extrapolation of purchasing power parities using multiple benchmarks and auxiliary information: A new approach.” In: *Review of Income and Wealth* 56.Suppl 1, S59–S98.

— (2015). “An econometric approach to the construction of complete panels of purchasing power parities: analytical properties and empirical results”. In: *3rd Post-2011 ICP RCA Meeting, International Comparison Program. World Bank.* URL: http://pubdocs.worldbank.org/en/267001488208862014/RaoRCA-2015.pdf.

Sen, Amartya (1976). “Real National Income”. In: *The Review of Economic Studies* 43.1, pp. 19–39. ISSN: 00346527, 1467937X. DOI: 10.2307/2296597. URL: http://www.jstor.org/stable/2296597 (visited on 11/01/2022).

Stiglitz, Joseph, Amartya Sen, Jean-Paul Fitoussi, et al. (2009). “The measurement of economic performance and social progress revisited”. In: *Reflections and overview. Commission on the measurement of economic performance and social progress, Paris.*

Verboven, Frank (1996). “International Price Discrimination in the European Car Market”. In: *RAND Journal of Economics* 27.2, pp. 240–286.

World Bank (2008). *Global Purchasing Power Parities and Real Expenditures. 2005 International Comparison Program.* Washington DC: World Bank. URL: http://siteresources.worldbank.org/ICPINT/Resources/270056-1255977254560/6483625-1337016259587/2005ICPReport_FinalwithNewAppG.pdf.

— (2013). *Measuring the Real Size of the World Economy: The Framework, Methodology, and Results of the International Comparison Program (ICP.* Tech. rep. World Bank (Ed.) , Washington, D.C.

— (2014). *World Development Indicators 2014. World Development Indicators; Washington, DC. ©World Bank. License: CC BY 3.0 IGO.* URL: https://openknowledge.worldbank.org/handle/10986/18237.

— (2020). *Purchasing Power Parities and the Real Size of World Economies. Results from the 2017 International Comparison Program.* Washington DC: World Bank: The World Bank Group. DOI: 10.1596/978-1-4648-1530-0.

— (2022). *The World by Income and Region.* URL: https://datatopics.worldbank.org/world-development-indicators/the-world-by-income-and-region.html.