INTERNATIONAL COMPARISON PROGRAM

RESEARCH AGENDA FOR THE 2017 CYCLE AND BEYOND

MAY 2017

Table of Contents

1. Introduction

The International Comparison Program (ICP) celebrates its Golden Jubilee next year marking fifty-years since the Program was launched in 1968 as a joint research program between the United Nations (UN) Statistical Office and a research team at the University of Pennsylvania led by Irving Kravis, Alan Heston and Robert Summers[1]. The last half-century has seen the ICP grow into a global statistical program conducted under the auspices of the UN Statistical Commission (UNSC) with the World Bank acting as the Global Implementing Agency. The last round of ICP, with 2011 as the benchmark year, had a global coverage with 177 countries participating in full-scale gross domestic product (GDP) comparisons and an additional 22 countries involved in comparisons of private household consumption.

Estimates of purchasing power parities (PPPs) of currencies and estimates of real GDP and its components, published through various rounds since the first report in 1975 to the latest report of the 2011 ICP, have been a major data source for analysis of economic performance of the world economy and its regions; and the measurement of regional and global inequality and poverty. Results from the ICP helped gain a better understanding of the economic geography of the world (Deaton, 2013). ICP 2011 results confirmed the status of China as the world's second largest economy behind USA and that the Indian economy in real terms was larger than that of Japan. Indonesia was ranked in the top 10 world economies.

Over the last fifty years, the ICP has recorded significant achievements including: establishing an effective partnership and an efficient governance structure for the ICP; designing survey instruments for price collection; developing methods dealing with comparison resistant services like dwelling, health and education; and identifying multilateral index number methods suitable for international comparisons of prices and real expenditures. Notwithstanding the impressive progress made, the ICP continues to strive for improvement and refinements in its approach with the goal of providing reliable and timely estimates of PPPs and real expenditures. The 2005 and 2011 rounds of ICP are particularly notable for a number of innovations that have placed ICP on a sound methodological foundation (World Bank, 2008 and 2015).

The 2005 ICP has seen the introduction of: the structured product descriptions (SPDs); formalization of the regionalization of ICP allowing for region-specific baskets of goods and services; a range of procedures for data validation and editing at the regional and global level; new methodology to measure PPPs for construction; productivity adjustment for government compensation in few regions; recommendation of the country-product-dummy method (CPD) for aggregation below basic heading level and the use of Gini-Èltetö-Köves-Szulc (GEKS) method for aggregation above basic heading level; and the use of ring countries and ring product lists to link regional comparisons leading to global comparisons that satisfy the principle of fixity[2]. A

---

[1] For a more detailed history of ICP, see Appendix A in World Bank (2015), the final report on the 2011 ICP.

[2] The principle of *fixity* ensures that relative sizes of the countries within a particular region within the global comparison are identical to the relative sizes of the countries obtained in comparisons within the region. For example, the relative size of Malaysia and the Philippines within the Asia-Pacific region comparisons would be maintained when global comparisons are compiled.

detailed account of these methodological innovations are detailed in what is referred to as the *ICP Book,* World Bank (2013), *Measuring the Real Size of the World Economy: The Framework, Methodology, and Results of the International Comparison Program (ICP).*

Innovations and changes introduced in the 2011 round of ICP are equally impressive and significant. Introduction of a new procedure to link the regional comparisons using prices collected by all participating countries, instead of prices from a small group of ring countries used in the 2005 ICP, marked a significant improvement over the 2005 approach. A global core list of products was established for this purpose. Importance indicators reflecting the representativeness of products priced in different countries were introduced and this information was utilized through the weighted country-product-dummy method (CPD-W) with suitably nominated weights. Fixity of comparisons from the regions was achieved through the use of *country aggregation with redistribution (CAR) – volume* method instead of the method used in 2005, which was not invariant to the choice of the reference country. The 2011 ICP introduced productivity adjustments for countries from most regions. A new and simplified method to deal with construction was introduced along with refinements to comparisons of dwelling services, and machinery and equipment. The 2011 ICP is also recognized for achieving a truly global coverage with the participation of 177 countries for comparisons at the GDP level and an additional 22 countries for comparisons of private household consumption.

There were also significant improvements in governance that have contributed to the success of the 2011 ICP. These include the establishment of a PPP Computation Task Force to ensure transparency and replicability of ICP results, and the introduction of the Results Review Group to assess the results prior to their release. Both groups worked closely with the Technical Advisory Group, which had overseen the methodological aspects of the program and their implementation.

Like all major statistical programs in the world, the ICP undergoes regular evaluations and close scrutiny by all the stakeholders with the objective of identifying avenues for improvements in the program that would ultimately result in a cost-effective and timely program delivering reliable internationally comparable macroeconomic data, for purposes of economic analysis by researchers, analysts, and policy makers at the national and international levels.

At the conclusion of the 2011 ICP, the UNSC established the Friends of the Chair (FOC) group at its 45th Session held in 2014, to carry out an evaluation of the 2011 ICP. The FOC group presented its preliminary report to the UNSC at the 46th Session held in 2015, and its final report at the 47th Session held in 2016. The final FOC report provides a review of the technical aspects of the ICP and describes, in detail, a concept for gradually transforming the ICP into a more frequent process, in order to tackle the issues stemming from infrequent standalone benchmark rounds.

2. Evaluation of the FOC Group and its Recommendations

After conducting an evaluation of the 2011 ICP, the FOC strongly endorsed the ICP through the statement:

"ICP 2011, with its considerably expanded coverage (from 150 to 199 countries), brought a much higher acceptance compared to earlier exercises. Following their improved availability, the use of PPPs around the world has increased. In addition, the applied methodology has significantly improved over that of the 2005 round. ICP 2011 has put ICP on a firm methodological basis by introducing approaches such as the global core list and applying major technical innovations such as the ICP toolkit. Specifically, the broad documentation of metadata and the further development of ICP operational guides and handbooks contributed much to the knowledge of staff conducting the work. A certain challenge arose from the cumulative effect of the two ICP rounds (2005 and 2011) which took the ICP from a one-time "snapshot" created by each solitary benchmark into a kind of time series-like environment with the requirement of time consistency." (Paragraph 12, E/CN.3/2016/9, UN, 2015).

The single most concern raised by the FOC is the six-year gap between the 2005 and 2011 ICP benchmarks and they emphasized the need for more frequent comparisons. Moreover, the FOC recommended that the previous approach to extrapolate PPPs from the benchmark using single price indices be replaced. Instead, it proposed the use of a mix of survey and extrapolated data to provide more frequent results, starting from the next comparison cycle for reference year 2017. In order to implement this plan, the FOC recommended that a rolling survey concept be adopted, which is based on spreading price surveys over a three-year comparison cycle centered on the reference year. The rolling survey approach requires a set of reliable extrapolation indices for detailed GDP categories in order to extrapolate survey data forward and backward to a respective reference year.

The FOC acknowledged in its report the ongoing interim activities in most of the ICP regions designed to produce regional PPP updates for years within the 2012-2016 time-frame. The FOC noted that the feasibility to combine these regional interim PPP updates into a global set of PPPs be investigated.

The FOC pointed out that the rolling survey approach requires collection of the necessary information on a more regular basis than once in five or six years, and called for the integration of ICP activities into the annual work programs of the national implementing agencies. In the same vein, it also pointed out that that the integration of the ICP and CPI activities can lower the overall costs of the ICP, as well as enhance the quality, coverage and reliability of the generated price indices.

The FOC recommended the establishment of a Technical Advisory Group to develop, conduct and monitor a research agenda. In line with the main recommendation that no major methodological innovations are introduced for the 2017 ICP cycle to ensure comparability of

results over time, the FOC identified a list of research topics that are essentially limited to fine-tuning methods and procedures:

- Implementing a rolling survey approach and building PPP time series;
- Integrating ICP and CPI activities;
- Streamlining the process of establishing product lists and the use of importance indicators;
- Improving the availability and quality of input data for rents, government services, and construction;
- Streamline the use of productivity adjustments for government services;
- Fine tuning global linking procedures; and
- Quality assurance of resulting PPPs and measures of reliability.

The 47th session of the UNSC noted the findings of the FOC group and strongly endorsed its recommendations.

3. ICP 2017 and Beyond: Challenges and Opportunities

The past two rounds in 2005 and 2011 have firmly established the ICP as a global statistical program by extending its coverage to 199 countries. The enthusiastic reception of the ICP results worldwide by analysts, researchers and international organizations, and the strong endorsement from the UNSC, have no doubt provided a major boost to the program and ensured a strong future for ICP.

Having established itself as a major statistical program, the ICP has reached a stage where it needs to identify its future direction and sustainability. The evaluation and recommendations of the FOC and recommendations from UNSC clearly pose major challenges and also create opportunities for the ICP.

The opportunities for the ICP are exciting in that the program now has a permanent home through the establishment of a Global ICP Unit at the World Bank, which offers continuity and certainty for the program. This assured status of the program creates an environment for consolidation of the ICP and gives it an opportunity to modernize, by exploiting the rapid advances in information and communication technologies and by exploring how these advances can help improve the collection and validation of price data collected from the participating countries.

With the opportunities arise challenges for the program. These are: (i) adopting a new governance structure and establishing ICP within the work program at the World Bank; (ii) securing continued support and active participation of National Statistical Offices and the Regional Implementing Agencies including the African Development Bank (AfDB), Asian Development Bank (ADB), Interstate Statistical Committee of the Commonwealth of Independent States (CIS-STAT), Eurostat-Organisation for Economic Co-operation and

Development (OECD), United Nations Economic Commission for Latin America and the Caribbean (UN-ECLAC), United Nations Economic and Social Commission for Western Asia (UN-ESCWA) and others; (iii) conducting frequent and regular cycles of ICP through the introduction of a rolling survey approach to international comparisons, eventually leading to annual time series of PPPs; (iv) ensuring and improving the quality and reliability of PPPs and international comparisons of GDP and its components; and (v) securing adequate and continued funding for the program.

In addition to these operational challenges for the program, there are two other challenges of practical nature. The first challenge is to adequately demonstrate the relevance and importance of the ICP data and results to the participating countries, international organizations and researchers around the world. There is a degree of urgency to prepare an *advocacy* document which clearly articulates how ICP generated data and results can be used for policy making at the national and international levels.

The second challenge for the ICP is to provide the users with increased access to the data collected and detailed results generated by the ICP. The view often expressed is that the ICP is a very resource intensive exercise, with budgets in tens of millions of dollars, but at the end it produces a single PPP number for each country. While this is a grossly simplified statement – the ICP produces a vector of PPPs for GDP and various aggregates for each country, and allows researcher access to all underlying PPPs and average price data–, there is a degree of truth in that the results released by the ICP are quite limited as far as users are concerned. Discussions with the stakeholders reveal *data access* as the single most important concern for users.

The next section presents the research agenda proposed by the World Bank's Global ICP Unit, based on the FOC recommendations endorsed by the UNSC, for consideration by the Technical Advisory Group. The proposed agenda is set within the parameters of the UNSC that no major methodological changes are introduced for the 2017 cycle. The items in this *research agenda* can be broadly classified as topics that are relevant for the 2017 ICP cycle in the immediate term and those topics that are designed to improve the program in the medium and longer term. The Annex to this document provides the research agenda items, priorities and implementation timeframe.

4. Research Agenda for the Immediate Term

The 2017 ICP cycle is currently underway and there are a few measurement issues that need resolution. Recommendations from the following research agenda items would feed directly into the process. In addition, the research agenda for the immediate term includes: compilation of annual series of PPPs; reliability measures for PPPs; and research into the uses of ICP results for national and international purposes leading to an advocacy document. The Global ICP Unit plans to synchronize the release of annual PPP series and the advocacy document with the release of 2017 ICP results.

4.1 Compilation of PPP Time Series and the Rolling Survey Approach

Extrapolations of PPPs at the aggregate level over long periods intervening the benchmark comparisons, coupled with major methodological innovations introduced in the benchmark comparisons, severely limit the ability to reconcile extrapolations from different benchmarks. This was certainly true in the case of the 2005 ICP results published after a gap of 12 years from the 1993 comparisons; and a relatively smaller gap of six years in the case of the 2011 and 2005 ICP comparisons. The significant differences between the extrapolations and the benchmarks led to significant revisions to estimates of the size of the world economy in real terms and its distribution across regions. The World Bank had to make major revisions to its estimates of global poor and global inequality in 2005 and 2011.

Inconsistencies between benchmark and extrapolated results arise due to a number of factors including changes in expenditure weights, commodity baskets, aggregation methods, linking regional comparisons and other considerations (Deaton and Heston, 2010; McCarthy, 2013; Annex D by Inklaar and Timmer to McCarthy, 2013; Deaton and Aten, 2017; and Inklaar and Rao, 2017 for details). A way to resolve this issue is to conduct more frequent comparisons as is the case with the Eurostat-OECD PPP program.

This issue of inconsistences between benchmarks was given prominence in the recommendations of the UN Statistical Commission:

"One of the key conclusions of the initial report on the 2011 round was that a six-year interval between ICP rounds was too long. This view was expressed by almost all stakeholders during the evaluation process and was also reflected in the reactions at the session of the Commission. It was expressed that the acceptance and relevance of ICP results depend on their frequent and timely availability (at least every two to three years with extrapolations to annual results). The Commission at its forty-sixth session expressed its preference to shorten considerably the interval of ICP worldwide comparisons with the intention of having the next benchmark occur as soon as possible, preferably in 2017". (Paragraph 45, E/CN.3/2016/9, UN, 2016)

Establishing a framework for the production of PPPs and international comparisons of GDP and its components on an annual basis is a complex and challenging task. The Global ICP Unit has taken this task seriously and had undertaken some preliminary work in this direction. In the process, the Global ICP Unit has identified four major components to its work program as it moves towards the implementation of rolling price surveys and the compilation of annual time series of PPPs.

Component 1. Moving Towards Rolling Price Surveys

The notion of rolling price surveys is not entirely new to ICP. During the implementation of the 2005 and 2011 ICP benchmarks a pseudo-rolling approach, whereby prices collected in different years were moved backwards or forwards using appropriate deflators, was used. While this was

commonly used in different regions, the approach was somewhat ad hoc and procedures were crafted on a case by case basis. The current thrust towards a more formal approach to rolling price surveys will ensure a consistent approach across all regions and thereby enhances quality of the comparisons.

The rolling survey approach, to be implemented starting in 2017, involves spreading the price data collection over a three-year time period around the reference year. PPPs for the basic headings not based on the reference year are to be extrapolated to the reference year. The Global ICP Unit identified the following data related issues:
- Determine the frequency of data collection (annual, biannual, quarterly) for the rolling survey for each aggregate of the GDP;
- Assess availability and quality of national account deflators, CPIs and other price indices by GDP component to be used in the extrapolation to the reference year;
- Determine how to link PPPs from regions with different timetables for data collection; and
- If a three-year rolling survey approach is implemented then it is necessary to assess methods of extrapolating PPPs for the two years between individual benchmarks, before the full rolling survey method is implemented (i.e., when PPPs are computed every year). This is also related to Component 3 below.

Component 2. Linking Interim Regional Updates into a Global Comparison

Since the completion of the 2011 ICP, various regions conducted interim updates. However, these regional efforts were not mandatory and largely depended on the initiatives of the Regional Implementing Agencies. The Global ICP Unit faces the challenge of linking these regional updates and make them useable in compiling meaningful global comparisons. The following issues need to be investigated and addressed:
- Effect of different timings of regional surveys ranging from 2012 to 2016;
- Impact of different degrees of coverage of the GDP;
- Effect of changing regional composition and of countries not participating in the regional interim updates at all; e.g., several Arab countries from Africa are also participating in the Western Asia comparisons, whereas China is not part of the Asia update;
- Effect of discrepancies in global core list coverage by region on inter-regional linking; and
- Effect of methodological differences between regions on the linking process.

Component 3. Building PPP Time Series for the Interim Period

As work on 2017 ICP is underway, the Global ICP Unit recognizes the need to fill the gap for the intervening period. The current timetable is to release comparisons for the years 2012 to 2016 around the same time as the release of results from 2017 ICP. Simultaneous release of the 2017 benchmark and comparisons for the intervening years will avoid scrambling by users to fill this

gap by themselves. Thus, the PPP time series will be built using the 2011 benchmark results along with data generated by regional interim updates (see Component 2 above). The main tasks involved are:

- Assess availability and quality of national account deflators, CPIs and other price indices by GDP component (as in Component 1 above);
- Examine consistency of CPI weights and national account weights, and assess the weights' effects on PPPs;
- Suggest steps for statistical capacity-building to ensure that CPIs and national account deflators are consistent over time and comparable across countries; and
- Ensure consistency between the 2011 and 2017 results and the comparisons for the intervening years.

Given the tight timeframe for achieving these goals, the Global ICP Unit has initiated work in this direction. Progress has been made in setting up an extrapolation database and establishing procedures to standardize the data provided by the countries through their respective regional implementing agencies. Detailed time-series data for expenditures, deflators, CPIs, population and exchange rate data are being collected and consistency checks are being conducted. A bi-annual submission schedule was agreed at the 2nd Inter-Agency Coordination Group (IACG) meeting. All the six regions (Africa, Asia and Pacific, CIS, Eurostat-OECD, Latin America and Caribbean and Western Asia) have reported data, but with differing degrees of coverage and quality. The unbalanced reporting from countries and sparse coverage of detailed time-series data for many countries call for careful integration of data in the compilation of time series of PPPs and in ensuring the quality of linked comparisons.

4.2 PPPs and Real Expenditures for Dwelling Services

Estimation of PPPs for dwelling services posed formidable problems in the 2005 and 2011 rounds of the ICP. Cross-country comparisons in this area are difficult because of the varying mix between rented and owner-occupied dwellings, absence of reliable and regular rental surveys, and unreliability of housing expenditures and the absence of adequate imputation of expenditures for owner occupied housing in many countries. Consequently, methods based on rental survey data, as well as that based on the quantity or direct volume methods, were canvassed and used in the last two rounds of the ICP (Heston, 2013). However, the poor quality of rental data, housing volume indicators and housing expenditures was a major factor why the Asia-Pacific region had abandoned using data from these sources. Instead the real expenditures for that region were imputed using the reference quantity approach (ADB, 2013).

The problem with dwelling services was recognized in the FOC group report and its evaluation states:

"… Several aspects make the comparison of housing expenditures challenging, for instance, the recording of housing expenditures and the significantly varying market situations from one country to another. There were substantial efforts from the regional coordinators and the

Global Office to improve the methodology as well as input data (prices as well as national accounts figures) in this complicated area; however, the actual progress was rather moderate. Most likely the dual-approach for collecting both rental and quantity data needs to be maintained. Improvements in this area should primarily come from improvements in data quality and availability (price and dwelling stock data as well as national accounts data)." (Paragraph 36, E/CN.3/2016/9, UN, 2015).

The poverty economists at the World Bank have expressed their concern with lack of reliable data on dwelling services as it is an important item in household budgets. Thus, the poverty economists are keen to see reliable PPPs and real expenditures for dwelling services in the 2017 ICP.

The Global ICP Unit is making special effort to improve quality of data on dwelling services and identified the following items for immediate attention:
- Assess the quality of housing input data;
- Recommend additional housing quality indicators for data collection, where possible;
- Propose imputation techniques in cases of missing housing data;
- Suggest refinements and improvements to the existing methods for housing estimates to better account for quality differences; and
- Investigate the use of new methods for estimating housing expenditures, such as the user cost approach.

The Global ICP Unit hopes to significantly improve dwellings estimates for inclusion in the 2017 comparison cycle.

4.3 Productivity Adjustment for Government and Construction Labor

Adjustment for productivity differentials across countries has been a major ICP initiative in the 2005 and 2011 rounds, designed to improve PPPs for government compensation. There are a number of outstanding issues for consideration in the 2017 ICP cycle.

The following excerpt from FOC report succinctly summarized the state of play with productivity adjustments:

"The use of a productivity adjustment for global linking of salaries in government services in ICP 2011 was an obvious improvement. There were, however, several weaknesses: the productivity adjustment calculated using capital-labor estimates for the whole economy for such specific areas as health, education and collective services is a very rough approximation (also, the accuracy of productivity adjustment factors for different countries seem to be different), and not all regions used productivity adjustment in the regional comparisons. In addition, the regions did not use the same methodology (e.g., OECD/Eurostat used the "output" approach for health/hospital services and education but other regions used the "input cost" approach). In effect, the results of the countries depend, in some cases heavily, on the regional methodology, and interregional comparability was decreased." (Paragraph 37, E/CN.3/2016/9, UN, 2016)

A review of the compensation data by the Global ICP Unit revealed uneven quality and data gaps. There is scope to improve the productivity adjustment factors and the following topics are identified for research in this area in order to improve estimates of PPPs for government services for the 2017 comparison:

- Review employee compensation data, examine data gaps, provide advice on filling them, and review gap-filling procedures;
- Assess the quality of data for labor inputs and capital stocks portion of the productivity adjustment and evaluate them for consistency across countries;
- Update productivity adjustment factors for use in the 2017 ICP cycle;
- Examine effects of the adjustment factors not being used in some regions on the global results; and
- Examine the scope for productivity adjustment for labor inputs in construction and make a recommendation as to whether productivity adjustments should be made in the 2017 ICP.

4.4 Fine Tuning Global Linking Procedures

The global linking in the 2011 ICP based on the global core approach and CAR procedure is rather straightforward. However, the actual linking had many exceptions: dual regional participation of some countries (Russia, Egypt and Sudan); singleton countries, Cuba, the Caribbean and Pacific islands were linked indirectly (not as a part of global linking) to the global comparison via other regions and individual countries; and the CIS countries were linked to the Eurostat-OECD comparison via Russia. Some GDP components also posed challenges for linking, as the Eurostat-OECD approach for those components differed from that followed by ICP regions. These components included health, education and construction. There is an urgent need to revisit global linking procedure and fine-tune it to address these issues:

- Assess impact of individual countries on global and regional results, and investigate if the reliability of the global comparison can be improved with a two-stage linking where the core countries are linked first, then the rest is linked to them without disturbing the core;
- Assess the possibility of linking the CIS to other regions directly through the global core approach and CAR procedure; and
- Develop a strategy to accommodate special participation cases and specific GDP components such as health, education and construction.

4.5 Quality and Reliability of PPPs

The ICP publications provide estimates of PPPs and real expenditures at the GDP level and for major components of GDP. However, there is seldom any indication on the quality of the published figures and no measures of reliability or measures of uncertainty associated with PPPs are provided. Currently researchers have no option but to consider all the published results to be of similar quality though intuition suggests that comparisons between some

countries (UK and Germany for example) would be intrinsically more reliable than some others (e.g. USA and Mozambique).

Uncertainty associated with published PPPs could be due to product selection, coverage and quality of price surveys, sampling errors, quality of national accounts expenditure data, and also due to the index number methods used in the aggregation process. There have been some attempts to identify a framework to construct measures of reliability. For example, Deaton (2012) and Rao and Hajargasht (2015) consider deviations of observed prices from the predictions generated using the *law of one prices* as a basis for computing standard errors for PPPs. Needless to say, further research is necessary to establish an analytical framework to construct reliability measures associated with PPPs from ICP.

The FOC report makes a passing reference to reliability measures as it deals with quality assurance.

"The ICP quality assurance framework was derived from the Data Quality Assessment Framework developed by IMF, which focuses on the quality-related features of governance of statistical systems, core statistical processes and statistical products. The ICP quality assurance framework covered six topics: prerequisites of quality; assurance of integrity; methodological soundness; accuracy and reliability; serviceability; and accessibility." (Paragraph 33, E/CN.3/2016/9, UN, 2016)

Furthermore, one of the FOC recommendations is:

"Quality assurance of resulting PPPs and measures of reliability". ((g), Paragraph 83, E/CN.3/2016/9, UN, 2016)

The Global ICP Unit identified two strands of research in regards to ICP quality assurance framework.

Component 1. Quality of Data
- Investigate sources of data variability at each level of estimation to assess the overall quality of the results and where data quality can be improved;
- Examine variability in basic heading PPPs before and after linking and identify countries and basic headings appearing as outliers before and after linking;
- Examine variability in basic heading expenditure shares and differences in consumption estimates stemming from household surveys and national accounts; and
- Analyze effect of the countries with high variability in expenditure shares and prices on regional and global results.

Component 2. Reliability Measures for PPPs
- Establish a conceptual framework that underpins reliability measurement;

- Develop a statistical framework for the computation of standard errors or confidence intervals;
- Implement the framework and compute standard errors or confidence intervals for PPPs from the 2005, 2011 and 2017 comparisons as well as for PPP changes over time; and
- Provide guidance to users as to how these measures of reliability can be used in practice.

The Global ICP Unit and regional implementing agencies seek to resolve data quality issues and have processes in place to ensure improved data quality for the 2017 ICP cycle.

4.6 Uses of PPPs for National and International Policy Making

Though the ICP has been in existence for nearly 50 years, there has never been any concerted effort to prepare a document outlining the uses of ICP results and demonstrate their relevance and usefulness for policy making at the national level and show how policy makers can gain better understanding of their respective economies from a global perspective. The paper by Ward (2009) is one of very few papers on this subject. The paper articulates uses of PPPs for policy making at the national level but it does not provide illustrations of how published ICP results can be linked to the uses described in the paper. The *ICP Book* (World Bank, 2013) includes papers by Chen and Ravallion; Deaton and Dupriez; Thomas et al., Silver; and Inklaar and Timmer that demonstrate the use of ICP results at the international level.

The need for a strong advocacy document was recognized in the FOC report.

 "It is recommended that the Statistical Commission authorize the Governing Board to reach out and demonstrate the value of ICP data to policymakers and other important users and donors, in particular showcasing to donors that ICP generally responds to user needs and strives for further openness with regard to access to data and metadata. ICP stakeholders should be invited to accelerate, combine and coordinate outreach efforts at the global, regional and national levels." Paragraph 70, E/CN.3/2016/9, UN, 2016)

The Global ICP Unit recognizes this immediate need and hopes to have such an advocacy document ready for release coinciding with the release of the 2017 ICP results. The document may cover the following aspects but not necessarily be limited to:
- A primer on the basic concepts of PPPs, price levels and real expenditures;
- Major uses at the international level: real size of the world economy and its distribution across regions, inequality, poverty, growth, competitiveness, spatial price level analysis, and other uses;
- Policy-making uses at the national level: real size and structure of GDP, the size of government, relative national competitiveness, price and consumption relationships, and other uses; and
- Other uses: price structures and globalization, price distortions and market structures, the real value of aid, energy efficiency and energy intensity.

5. Research Agenda for the Medium Term

This section lists a number of important topics that are not of immediate relevance to the 2017 ICP round. These agenda items are critical for strengthening ICP and ensuring its long term viability. It is expected that work on these topics will begin as soon as it is practicable and recommendations emanating from research on these topics will be implemented in collaboration with the national statistical offices in the participating countries and the regional coordinating agencies.

5.1 CPI-ICP Synergies and Sub-National PPPs

The need for a closer integration of CPI and ICP activities has long been recognized and discussed at various ICP meetings and conferences on international comparisons. However, little progress has been achieved in this direction. A major factor limiting progress in this direction is the lack of a clear notion of what CPI-ICP integration of activities actually means. Initially such integration was taken to imply a closer alignment of product lists used in the ICP and the products priced for CPI price surveys. Needless to say, such a suggestion was met with stiff resistance from participating countries as the national CPIs have a specific purpose and are central to policy making within the countries.

Over the last decade, a better understanding of benefits from CPI-ICP synergies has emerged. The following statement from the FOC report explains this point clearly.

 "A rolling benchmark concept for ICP implies that national statistical institutions collect the necessary information on a more regular basis than every five years. Such frequency calls for the integration of ICP into the work programs of national statistical institutions. While the production of price indices (especially CPI) and the compilation of expenditure-based GDPs is a permanent process and normally part of each national statistical institution's work program, the production of, for example, nationwide average prices very often falls outside of the official price statistics. In the case of price surveys, use of ICP-relevant surveys at regular intervals can also coincide with the need for price-level comparisons across regions (mainly for large countries) and can overlap with CPI surveys. Synergies with related statistics are certainly higher in the case of more frequent surveys, which positively affect cost-effectiveness. Similarly, to make ICP more regular and sustainable, it is essential that it become institutionalized at the global, regional and national levels. It is important that global and regional coordinating agencies incorporate ICP work into their work programs as an established business line." Paragraph 60, E/CN.3/2016/9, UN, 2016)

As ICP moves towards annual compilation of PPPs, it is important to limit the burden on the participating countries. Achieving potential synergies by integrating ICP and CPI survey activities is likely to reduce the burden on the national statistical agencies involved in the ICP. Recognizing these synergies, the FOC urged that participating countries to integrate and

harmonize ICP activities with their regular price statistics and national accounts work to ensure that the ICP becomes a sustainable and truly permanent program.

The CPI-ICP integration process received impetus from the recent surge in interest in the compilation of PPPs at the sub-national level. A number of countries, including China, India, Indonesia, Italy, the Philippines, South Africa and Vietnam have expressed interest in compiling PPPs for comparisons of prices and real incomes across regions within these countries. Through this process, the countries gain an increasing appreciation of the ICP methods and survey instruments for making spatial price comparisons.

Harmonizing CPI and ICP activities will reduce the data collection burden resulting from the implementation of the rolling survey method, while making the CPI and ICP prices more consistent and comparable, which would help improve extrapolations. The integration of the CPI and ICP activities will enhance their application for additional uses at the national and sub-national levels. The following topics are identified for research under this agenda item:
- Assess areas where harmonizing CPI and ICP processes will create synergies;
- Prepare a set of guidelines for countries to facilitate the process of harmonizing CPI and ICP activities;
- Conduct a case study of selected countries to evaluate the use of SPDs to better define their national CPI product specifications;
- Examine the Asian and African experience with sub-national PPPs and formulate guidelines on the use of the CPI product lists and prices to compute sub-national PPPs;
- Analyze temporal consistency of CPI and sub-national PPPs; and
- Expand work on sub-national PPPs to more countries.

5.2 PPPs for Exports and Imports

In the ICP, foreign trade sector is traditionally incorporated using simple exchange rates. However, there has been a rich literature on incorporating terms-of-trade gains and losses due to the export-import price differentials into international comparisons. It is obvious that the terms-of-trade effect can be very significant in measuring GDI (gross domestic income) growth, given wide swings in oil and other raw material prices. Put simply, an improvement in the terms of trade means a country gets more for less. In the SNA, a change in the terms-of-trade is treated as a price effect, rather than a real effect, hence the terms of trade are not captured by the real GDP measure. Thus, using the terms-of-trade adjustment should improve quality of extrapolation, and help bridge the gap between ICP benchmarks.

Moreover, the terms of trade considerations can be important in comparing GDP across space as well. Even when approximating PPPs for exports and imports by the exchange rate, implicitly assuming that the law of one price holds for tradables, the computed effect can reach 10% of GDP. Some researchers compute much larger effects when using unit value ratios for exports and imports. However, the computation results could become susceptible to the quality of

information on foreign trade transactions, and data gaps could become wide outside the OECD area.

Given the importance of this topic for the extrapolation, as well as productivity comparisons the following is identified for further research:
- Investigate data availability and reliability for the terms-of-trade computation, both across space and across time;
- Evaluate various methodologies to incorporate the term-of trade effect into the real GDP and PPP computation, on an experimental basis; and
- Incorporate the terms-of-trade effects into the 2005, 2011 and 2017 ICP results, on an experimental basis.

5.3 PPPs for Construction

In the 2011 round, regions, expect Eurostat-OECD and CIS, estimated PPPs for Construction based on input prices for a set of materials, labor, and equipment hire, which were subsequently weighted based on resource mixes data. Eurostat-OECD compiled construction PPPs using a model-based Bills of Quantities approach, while the CIS region used a modified version of this method. The ICP 2011 input-based approach was a compromise that made the best of less than uniform measurement standards for this tough-to-measure activity. A model-based Bills of Quantities approach clearly would be better, but in general it can be argued that the 2011 input-based approach was and still is the best possible and achievable option, given the limited availability of data, as it is much less costly.

The 2011 ICP results for some countries were not fully plausible. Input data quality remains an issue, and the methodology needs to be investigated as well. Input costs (materials, labor and equipment) do not take into account the outstanding gap among economies for profits, taxation and contractor mark-ups, and therefore do not fully reflect market prices. Thus, further research regarding these problems is needed in future cycles of ICP.

Given the importance of construction in many countries (29% of GDP in China, for example), the following topics are identified for further research:
- Investigate input data for construction, including variability;
- Look into methodological improvements, in particular evaluate the use of productivity adjustments for labor (as in 4.3 above);
- Improve estimates of the resource mixes weights; and
- With the help of construction experts, review the availability and quality of other construction-related information, such as mark-ups and professional fees, and devise methods to incorporate this information in the compilation of construction PPPs.

5.4 PPPs for Health and Education

Compilation of PPPs for health and education is a challenging task given the complexity of the provision and consumption of health and education goods and services. In a vast majority of countries health and education services are provided through a mixture of government-run and privately run schools, universities, hospitals and other institutions. As the services provided by governments are not sold at market prices, an input-based approach is used. This has been the approach used in the ICP, a description of the ICP approach to the compilation of PPPs for health and education is provided in Blades (2013). The Eurostat and OECD have started using output-based measures of prices and volumes for health and education. Given these new developments, it is appropriate to explore the feasibility and effect of introducing this approach in other regions. The challenges associated with implementing the output-based approach in other regions were also discussed in the FOC report.

"No doubt the OECD/Eurostat experience with the output approaches in these areas should be carefully analyzed by all regions. The challenge with the output approach, however, is that it requires data that are not always available or are of a questionable nature. Outputs (and their quality) are more difficult to define and could be less comparable. For the time being, the input-based approach with productivity adjustment is most likely the best possible option for the worldwide ICP. However, the process for calculating productivity adjustment factors, as well as the application of productivity adjustment factors, should be streamlined." (Paragraph 38, E/CN.3/2016/9, UN, 2016)

Accordingly, the following aspects need to be researched:
- Articulate an output-based approach for health including classification of services;
- Design an output based approach to education by defining outputs and quality measures;
- Assess the availability of data in participating countries, including sources from the World Health Organization (WHO), the United Nations Educational, Scientific and Cultural Organization (UNESCO) and other organizations;
- Evaluate effects of implementing the output-based approach compared to the regular input-based method.

6. Ongoing and Longer-term Research Agenda

6.1 ICP PPPs and Global Poverty Measurement – Post Atkinson Commission

For more than three decades, the World Bank has relied on PPP measures from ICP in its calibration of the international poverty line (IPL) and in producing estimates of incidence of global poverty. The initial IPL of $1/day was based on results from the 1985 ICP and the Bank has subsequently revised the lines to $1.08, $1.25 and $1.90 based, respectively, on PPP data from the 1993, 2005 and 2011 rounds of the ICP. The influence of successive rounds of ICP on the IPL and the resulting changes in global poverty estimates have been the subject of intense debates. These debates have led the World Bank to seek advice from a group of distinguished researchers in this area led by late Professor Atkinson.

Recommendation 10 of the Atkinson Commission Report (World Bank, 2016) deals directly with the use of PPPs from ICP. The recommendation is that cross-country price comparisons (or exchange rates) be fixed at the 2011 PPP values, and that the IPL should be adjusted over time for each country, in local currency, using national CPI, or suitable alternatives, until 2030.

This is a far reaching recommendation as far as ICP is concerned in that the recommendation implies no role for PPPs from ICP rounds between 2011 and 2030 on IPL determination and on estimates of global poverty. However, in its response to Atkinson Commission Report, the World Bank left the scope for use of PPPs from ICP wide open. The following excerpt from the World Bank response adequately summarizes the state of play.

"While we welcome this recommendation, we recognize the trade-off that underlies it: incorporating new PPP values from future ICP rounds is desirable to the extent that real changes in purchasing power across countries are captured; and undesirable to the extent that changes arising only from methodological changes between rounds are captured instead. In the past, vigorous debates have ensued as to whether changes in global poverty when new PPP data became available were real, or induced by methodological instability. A concern that such perceptions undermine the credibility of the entire exercise were paramount in shaping this recommendation. Recognizing this trade-off, we plan to follow this recommendation, but leave open the possibility that future PPP rounds might be used again to inform the construction of the IPL, even before 2030, if and only if we are satisfied that the ICP methods have substantially stabilized over at least two ICP rounds. This would require that our own researchers, and the broader scholarly community, are largely of the view that changes in PPPs are then driven by real changes in cost-of-living parities, rather than in data collection or index-number methodology." (page 3, A Cover Note to the Report of the Commission on Global Poverty, chaired by Prof. Sir Anthony B. Atkinson October 18, 2016)

In summary, Recommendation 10 of the Report clearly fixes the IPL at $1.90 based on PPPs from the 2011 ICP and the national currency equivalents in 2011 are projected forward based on the CPIs in respective countries. However, the World Bank response leaves the door open to the use of PPPs from ICP exercises between now and 2030. Any future decision to use PPPs from any of the subsequent rounds of the ICP has to rely on sound arguments demonstrating the superiority of the new PPPs to the use of 2011 PPPs as recommended by the Commission. Monitoring and close scrutiny of PPPs from the 2017 ICP and beyond and their impact on poverty measures is an essential task for the ICP and the World Bank. The following can be undertaken:
- Assess the impact of 2017 ICP PPPs on global poverty measures;
- Compare poverty estimates from the 2017 ICP PPPs with those resulting from the implementation of Atkinson Report based on the 2011 ICP PPPs;
- Examine the implications of the shift towards compilation of annual series of PPPs which are planned for production from 2019 onwards;
- Assess implications of availability of sub-national PPPs for large countries like China, India, and Indonesia for poverty measurement at the World Bank;

- Identify strategies to make use of data underlying the headline PPP estimates from ICP; and
- Investigate whether poverty-specific PPPs are more pertinent for poverty estimates. The Global ICP Unit has initiated work on this topic by examining whether the use of household survey expenditures and prices of products more relevant to the poor would yield significantly different PPPs than those produced by the ICP.

6.2 Exploring Innovations in Technology and Data Sources for PPP Measurement

This is the last of the topics that deserves serious consideration. The current approach to price collection in the participating countries is based on standard survey approach used in the CPI compilation. Prices are collected from various outlets in different locations and then averaged and used in the computation of basic heading PPPs.

The world is witnessing a major technological revolution which is influencing both by the way consumers purchase goods and services, and also the availability of large scale databases with information on prices and quantities associated with each transaction recorded at the point of sale. Scanner data are increasingly available and are now being considered by national statistical offices in their CPI compilation methodologies. Given that internet shopping and the use of services like Uber are becoming popular, prices from supermarkets and standard taxi fares may not provide reliable estimates of price level differences.

There is scope and need to modernize ICP and it is important to develop a vision as to how ICP will be implemented in five or ten years. Availability of scanner data, possibility of collecting prices through web-scraping, use of portable devices that can record the location and also make it possible to record product characteristics at the time of price collection, crowdsourcing of price data collection, etc. offer a range of alternative sources that need to be explored in future ICP cycles.

The World Bank undertook pilot projects to capture price data using crowdsourcing and smart phones. The basic concept is to have a network of price collectors across each country collecting prices, and recording observation place and date, along with some additional information. The following need to be investigated:
- Assess the quality of the prices from alternative sources using standard ICP data validation methods; and
- Compare price levels and dynamics between CPI and ICP price collections with the alternative sources.

6.3 Accounting for Product Quality Differences in PPP Measurement

The ICP approach to international comparisons is founded on the basic principle of comparing prices of like with like across countries. Accordingly, considerable effort goes into the determination of the basket goods and services for price comparisons at the regional level; and

in the calibration of global core product list for global linking. Achieving a balance between *comparability* of the goods priced and at the same time ensure that the products are *representative* or *important* in the participating countries is well recognized, and ultimately a judicious compromise between these two competing considerations is reached. In order to ensure pricing of identical/similar products across countries, the ICP in its 2005 benchmark comparisons introduced SPDs, which attach a list of price determining characteristics associated with each product or service priced. SPDs for products are more narrowly defined in the case of Eurostat and OECD and the SPDs are more flexible in other regions like Africa.

There is a growing realization that despite the care taken in product specification and at the time of price collection, it is still difficult to ensure that there are no differences in qualities of products actually priced in different countries. Often this aspect is noticed at the price validation workshops where some countries may report lower than *expected* prices and ensuing discussions raise the issue as to whether the observed differences are a reflection of a lower quality.

A related issue is that even when a product is priced according to exact product specifications, there may be certain service elements that may have influence on price recorded. When it comes to services like medical and educations services, differences in quality are clearly apparent. Similarly, when it comes to transport sector prices, prices for rail and bus transport, the ICP specifications do not account for quality differences in terminal services, frequency and reliability of services. Similar differences in quality differences are generally acknowledged but are placed in the *too hard basket* at this stage.

Once it is recognized that quality differences may be driving some of the price differences then the PPPs observed are biased as they fail to account for price differences. It is indeed quite possible that the price levels in low income countries may need to be adjusted upwards in order to account for the likely lower quality associated with goods and services. This in turn means real per capita expenditures in these countries would be adjusted downwards leading to increased magnitudes of measures of inequality and poverty. The relationship between quality in manufacturing and services and the Balassa–Samuelson relationship is explored in a recent paper by Zhang (2016).

At this stage, the effect of quality differences on PPPs and the flow-on effects is a matter of conjecture. When researchers use unit values from household expenditure surveys, invariably attempts are made to adjust for possible quality differences in goods and services consumed by households in different income groups and regions. Similar attempts need to be explored and it is important for the ICP to devote resources to devise methods to account for quality differences across countries. The following aspects could be researched:
- Use of extended SPD approach to capture quality differences;
- Explore the possibility of incorporating quality indicators in the CPD regressions – similar to hedonic regressions but using broadly based quality indicators appropriate for each basic heading; and

- Conduct a pilot study in one or two regions for selected basic headings to measure the extent of bias induced by quality differences.

7. Conclusions

The research agenda presented in this note focuses on improving the methodology and approaches underlying the ICP. An attempt has been made in this document to identify areas of research that will provide valuable input into the 2017 ICP cycle currently underway and beyond. However, given that the ICP is becoming a permanent statistical exercise, it is essential to separate the "statistical production" function from the "methodological development" function, as stressed at the ICP Next Steps meeting held after the 47th Session of the UNSC. Moreover, there is a need to ensure that with each methodological development adopted, the impact of this change on the results is measured and documented, to facilitate the use of the results.

References

ADB (2013), *Purchasing Power Parities and Real Expenditures*, Asian Development Bank, Manila.

Blades, D. (2013), "Health and Education", Chapter 11, *Measuring the Real Size of the World Economy*, World Bank, Washington DC.301-318.

Deaton, (2010), "Price indexes, inequality, and the measurement of world poverty," American Economic Review, 100(1): 5–34.

Deaton, (2012), "Calibrating measurement uncertainty in purchasing power parity exchange rates", Research Program in Development Studies, mimeographed, Princeton.

Deaton, A. (2013), "Reshaping the World", Introduction, World Bank (ed.), *Measuring the Real Size of the World Economy*, World Bank, Washington DC. Washington DC. pp. 1-12.

Deaton, Angus, and Alan Heston. 2010. "Understanding PPPs and PPP-Based National Accounts." *American Economic Journal: Macroeconomics* 2 (4): 1–35.

McCarthy, Paul (2013), "Extrapolating PPPs and Comparing ICP Benchmark Results" in World Bank (ed.), *Measuring the Real Size of the World Economy*, Chapter 18, World Bank, Washington DC. 474-505

Heston, Alan (2013), "Dwelling Services" in World Bank (ed.), *Measuring the Real Size of the World Econom*y, Chapter 16, World Bank, Washington DC. pp. 319-342.

Inklaar, R and M.P. Timmer (2013, "A Note on Extrapolating PPPs" Annex D to Chapter 18 by McCarthy, P, World Bank (ed.), *Measuring the Real Size of the World Economy*, World Bank, Washington DC. pp. 502-505.

Rao, D.S. Prasada and G. Hajargasht (2015), "Stochastic Approach to Computation of Purchasing Power Parities in the International Comparison Program (ICP)", *Journal of Econometrics.* Bol. 191(2), pages 414-425. Elsevier

UN (2016), *Final report of the Friends of the Chair group on the evaluation of the 2011 round of the International Comparison Programme*, UN Economic and Social Council, 47th Statistical Commission Report, United Nations, New York. E/CN.3/2016.9

Ward (2009), "Purchasing Power Parities and their Policy Relevance" in Prasada Rao (ed.) *Purchasing Power Parities: Recent Advances In Methods And Applications*, Edward Elgar Publishing Company, 86-120.

World Bank (2008), *Global Purchasing Power Parities and Real Expenditures. 2005 International Comparison Program*, World Bank, Washington DC.

World Bank (2013), *Measuring the Real Size of the World Economy* (Edited book), World Bank, Washington, DC.

World Bank (2015), *Purchasing Power Parities and the Real Size of World Economies – A Comprehensive Report of the 2011 International Comparison Program*, World Bank, Washington DC.

World Bank (2016), *Monitoring Global Poverty: Report of the Commission on Global Poverty,* World Bank, Washington DC

Zhang, Q (2017), "The Balassa–Samuelson relationship: Services, manufacturing and product quality", Journal of International Economics, Volume 106, May 2017, Pages 55–82

Annex: Research Agenda Priorities and Timeframe

					2020 CYCLE		
					Data Collection		Results
			2017 CYCLE				
			Data Collection		Results		
IMMEDIATE TERM	2016	2017	2018	2019	2020	2021	2022
Compilation of PPP Time Series and the Rolling Survey Approach							
Component 1. Moving Towards Rolling Price Surveys	■	■	■				
Component 2. Linking Interim Regional Updates into a Global Comparison		■	■	■			
Component 3. Building PPP Time Series for the Interim Period	■	■	■				
PPPs and Real Expenditures for Dwelling Services	■	■	■	■			
Productivity Adjustment for Government and Construction Labor			■				
Fine Tuning Global Linking Procedures			■				
Quality and Reliability of PPPs							
Component 1. Quality of Data			■				
Component 2. Reliability Measures for PPPs			■	■			
Uses of PPPs for National and International Policy Making	■						
MEDIUM TERM	2016	2017	2018	2019	2020	2021	2022
CPI-ICP Synergies and Sub-National PPPs	■						
PPPs for Exports and Imports		■	■				
PPPs for Construction					■	■	
PPPs for Health and Education					■		
ONGOING AND LONG-TERM	2016	2017	2018	2019	2020	2021	2022
ICP PPPs and Global Poverty Measurement – Post Atkinson Commission		■	■	■			
Exploring Innovations in Technology and Data Sources for PPP Measurement				■			
Accounting for Product Quality Differences in PPP Measurement					■	■	■

Left axis: PRIORITY — Higher (top) to Lower (bottom)