

1/4/2017

A Memoir of My Work with the International Comparison Program (ICP) and the Penn World Table (PWT)

Alan Heston*

Preface

Chapter 1: Phase 1: Antecedents and Early Meetings

Background of the ICP

A. ICP Predecessors

- (1) Before Colin Clark's 1940 *The Conditions of Economic Progress*
- (2) The period 1940-65
- (3) Early Methods of Comparing PPP

B. Phase I of the ICP: 1968-70

- (1) Some Early Decisions on the Framework of the ICP Comparisons
- (2) New Bolton Meeting, June 1969
- (3) Bellagio I, October 1969

Chapter 2: Organization and Implementation of ICP Phase I

A. Phase I Activity 1970-71

- (1) Division of Labor with the UNSO
- (2) Robert Summers on Board in 1970
- (3) Gathering and Processing Prices and Expenditures: 1970-71
- (4) Extended Visit to India and Short Visit to Japan
- (5) Obtaining US Expenditures and Prices
- (6) Equipment Goods
- (7) Construction

B. Experimenting with Aggregation Methods on Preliminary Data: 1972-4

- (1) Below the Basic Heading Level
- (2) Aggregating the Basic Heading PPPs: GEKS and Geary-Khamis
- (3) To Obtain More Support You Need to Show Results
- (4) A digression on *Men of Affairs*
- (5) Organization of Phase II and Completion of Phase: SOEC
- (6) The Rest of the ICP World
- (7) A Digression on the 1967 Comparison

Chapter 3: Responses to the Phase I Report

A. Response of Reviewers. Scholars and the Media

- (1) Book Reviews
- (2) The Media

B. Response of the International Organizations

- (1) The World Bank
- (2) The European Union
- (3) The International Monetary Fund
- (4) The Regional Banks and Economic Commissions of UN System

Chapter 4: Completion of the Phase II Report

- A. Measures of Imprecision**
- B. Putting Together the 1970 and 1973 ICP Estimates**
- C. Consistency Over Time of ICP Estimates**
- D. Geary International Prices**

Chapter 5: Signing Off- Phase III Report

- A. Services in the ICP**
- B. The Contributions of in-kind Aid in the form of Advisors**
- C. Augmented Binary Comparisons**
- D. Analysis of the 1975 Results**
 - (1) Bridge Country Binaries
 - (2) Demand Analysis
- E. Estimates for non-benchmark years**
- F. A Digression on 1980**
 - (1) China's Purchasing power
 - (2) The Hill Report
 - (3) Other Meetings
 - (4) The Latin American Region
 - (5) The 1975 EU Report

Chapter 6: Estimation of GDP of Non-benchmark Countries and Sectoral Output

- A. Origins of the Penn World Table (PWT)**
- B. The Beginnings of the ICOP Studies at Groningen**

Chapter 7: ICP after Penn - Regionalization and 1980 and 1985 Comparisons

- A. The Integration of the OECD into the 1980 ICP**
 - (1) The Problem Posed by the US CPI Revision in 1978
 - (2) Putting Together the OECD and the 1980 ICP World
- B. The 1985 ICP Round; Plans and Implementation**
 - (1) The Best Laid Plans
 - (2) The Latin America Problem
 - (3) Leadership
 - (4) Picking up the pieces of the 1985 ICP

Chapter 8: Exit and Voice

- A. The World Bank and the ICP in the 1980s**

- B. The Philadelphia Meeting, August 1990**
- C. A Digression on Kravis, Summers and Penn**
- D. The Unhappy Life of the Phase VI Global Comparison**
 - (1) The Regional Comparisons
 - (2) Official Global Report on 1996 ICP: a no-Show
- E. The Castles and Ryten Reports of 1997**
 - (1) The Castles Report
 - (2) The Ryten Report

Chapter 9: Finding a Home for PWT, Part I

- A. Initial Efforts to Pass the Torch**
- B. China and the International Monetary Fund: Any Port in the Storm**
- C. Rise of Empirically Tested Growth Theories**
- D. Aside on the Big Mac Index**

Chapter 10: Getting the 2005 ICP Aloft and Landed

- A. The First Two Years: Reality Sets In**
 - (1) The Initial TAG Members
 - (2) The Coding Issue
- B. Price Validation in the Regions**
 - (1) The World Bank Price Validation Software
 - (2) The *Tool-Pack*
- C. Linking the Regions at the Basic Heading Level in ICP 2005**
 - (1) The CPD and GEKS Methods at the Basic Heading Level
 - (2) Using the Ring Country CPDs to Link the Regions
 - (3) Aggregating the Regions in the 2005 ICP
 - (4) China's 2005 Participation
- D. The Final Results of the 2005 ICP**
 - (1) The Form of Presentation
 - (2) The Ikle Method of Aggregation used in Africa
 - (3) Did the 2005 ICP Results Produce Surprises?

Chapter 11: Reflections on the ICP and the PWT

- A. Transferring the PWT to Groningen**
- B. The 2011 ICP**
 - (1) Digression on Regions
 - (2) The 2011 Results and the Poverty Fracas
 - (3) China Redux
- C. Moving the ICP Towards Annual Updates**
 - (1) You've Come a Long Way Baby
 - (2) Rolling Benchmarks
 - (3) Governance
- D. In Good Hands**

****Emeritus Professor, University of Pennsylvania***

ABBREVIATIONS

ADB	Asian Development Bank
AfDB	African Development Bank
BEA	(US) Bureau of Economic Analysis
BLS	(US) Bureau of Labor Statistics
CAR	Country Aggregation and Redistribution (method)
CPD	Country Product Dummy (method)
CPI	Consumer Price Index
CSO	(Indian) Central Statistical Office
DDG	(World Bank) Development Data Group
DGINS	Directors General of Institutes of National Statistics
DGTT	(Indian) Director General of Technology and Trade
ECA	(UN) Economic Commission for Africa
ECE	(UN) Economic Commission for Europe
ECIEL	Program of Joint Studies on Latin American Economic Integration

ECLAC	(UN) Economic Commission for Latin America and the Caribbean
ECSC	European Coal and Steel Community
EKS	Éltető Köves Szulc (method)
ELIs	Entry Level Items
ESCAP	(UN) Economic and Social Commission for Asia and the Pacific
ESCWA	(UN) Economic and Social Commission for Western Asia
EU	European Union
Eurostat	Statistical Office of the European Communities (SOEC)
FAO	Food and Agricultural Organisation
GDP	Gros Domestic Product
GEKS	Gini Éltető Köves Szulc (method)
GGDC	Groningen Growth and Development Centre
GK	Geary Khamis (method)
IADB	Inter-American Development Bank
IARIW	International Association for Research on Income and Wealth
ICOP	International Comparisons of Output and Productivity (program)
ICP	International Comparison Program
ICSC	International Civil Servants Commission
IMF	International Monetary Fund
INSEE	Institut national de la statistique et des études économiques

ILO	International Labor Organisation
MPS	Material Product System
NSF	National Science Foundation
NBER	National Bureau of Economic Research
NSS	(Indian) National Sample Survey
OEEC	Organisation for European Economic Cooperation
OECD	Organisation for Economic Cooperation and Development
OU	Oriental (currency) Unit
PISA	Programme for International Student Assessment
PPP	Purchasing Power Parity
PPS	Purchasing Power Standard
PWT	Penn World Table
SNA	(UN) System of National Accounts
SOEC	Statistical Office of the European Communities
TAG	Technical Advisory Group
UN	United Nations
UNESCO	United Nations Educational Scientific and Cultural Organisation
UNSC	United Nations Statistical Commission
US	United States
USAID	US Agency for International Development
UNSD	United Nations Statistical Division
WWI	World War I
WWII	World War II

WDI	World Development Indicators
WHO	World Health Organisation

Preface

My children, Alex and Laura, called the International Comparison Project (ICP)¹ *The Project* because it was my principal research activity during their school years and since. Their mother, Wilma Heston, (Pic 1)² contributed to the first report on the ICP before she pursued a series of translations from Persian and Pushtu literatures, both popular and literary. The first ICP report was authored by Irving B. Kravis, Zoltan Kenessey, Alan Heston and Robert Summers (1975) with the assistance of many others named in that volume and discussed below. There are two related aspects of the project that had caught my interest prior to 1968 when I joined the ICP.

I had taken a course on the Russian Economy at the University of Washington taught by Frank Holtzman in 1956. Frank did his undergraduate work at the University of North Carolina on a tennis scholarship graduating in 1940 and stationed in World War^oII in a US air base in Poltava, Ukraine, where his interest in Russian studies began. He then did his graduate work at Harvard where he was a student of Alexander Gershenkron and Wassily Leontief, both experts on the Soviet Economy. Leontief received the Nobel prize in 1973 for his work on input-output tables and Holtzman was associated with that project while finishing his Ph.D.

I mention Holzman and tennis because he was a frequent partner on the University courts while I was in Seattle. He was a better player but the points were always competitive and where he was nip and I was tuck. An East Coast product after taking a position at Tufts, he spent part of each summer on Cape Cod where my colleague Summers and son Larry came to know him on the tennis court. What I remember most from Holzman's course was the Gershenkron effect that illustrated the inherent index number problem that weighting matters if you are computing measures of change over time. Gershenkron had compared the differences in the growth in the Production Index of Russia from 1919 to 1939 using the official version employing 1919 sectoral weights with the his version using 1939 weights. Neither is a true measure but the differences were large.

A second study that caught my attention while doing my graduate course work at Yale was that of William Hollister (1958) on the size of the Chinese Economy circa 1952 when the United States did not recognize the new Communist government. Mainland China was supporting the North in the Korean War and the US Defense agencies wanted to know the size of the Mainland economy. Hollister carried out his study with intelligence financial support and gave an answer that they liked but Hollister, being well trained, also gave an alternative answer they would prefer not to have seen. Essentially Hollister estimated sectoral PPPs weighting them first by US weights, which made the Mainland economy to be significant in size relative to the United States, the result wanted by the

¹ It was renamed *International Comparison Program* in 1989.

² In a file labeled *Images Memoir* there are images of individuals with an annotation file describing the people and the context.

US defense establishment to justify larger military budgets.³ But Hollister also provided estimates using Chinese weights placing China's total GDP as relatively small compared to the United States, much less attractive to his sponsors. None of this changed the focus of my dissertation which was in Monetary Economics.⁴ But the index number problem remained in the back of my mind as something to which I would like to return one day.

My qualification for writing this history is that I was there at the beginning of both the ICP and the Penn World Table and I am still on the Advisory Group of the current 2017 round of the ICP. During this almost 50 years the number of countries participating in the ICP has risen from ten in 1970 to almost 200 in 2011. Two data sets evolved as offshoots of the ICP, namely the Penn World Table that was the inspiration of my late colleague, Robert Summers and with which I have been directly associated; and the International Comparisons of Output and Productivity program of the University of Groningen, a vision of Angus Maddison (Pic 2) that is celebrating its 25th year in 2017.

Fortunately I do not need to rely solely on my memory in undertaking this memoir. Our ICP mentor, Irving or Irv Kravis, had us all write file memos about meetings, price collection, national accounts problems, issues in construction, rents and public consumption, visits to participating countries, methods of estimation and other subjects. We divided up the work between a unit at Penn and a unit at the United Nations Statistical Division (UNSD)⁵ so these memos were crucial to keeping everyone informed. These memos were filed by subject matter and chronologically and along with the published reports the eight binders written between 1970 and the early 1980s have been a principal source. Unfortunately there are some time gaps along the way, notably the years 1970 and 1974 where only part of the record made it from Philadelphia to Albuquerque where I now live. It is a personal history because I have given my impressions as well as provided images of some of the participants. It is not exhaustive because there have been decisions and discussions in Europe, New York, Washington and elsewhere not known to me. The memoir offers my take on some of the issues faced and decisions made along the way, and gives my perspective on some of our questionable decisions as well as high-fives to a few of our small triumphs.

Chapter I begins with a discussion of some of the applied studies prior to 1968 and the intellectual background to the ICP. Succeeding chapters cover the methods chosen and implementation of the 1970 Phase I of the ICP and reactions to the 1974 draft report on Phase I and to its publication in 1975. Phases II (Kravis, Heston and Summers 1978) and III (Kravis, Heston and Summers 1982) are the last rounds of the ICP in which the

³ The CIA during the Cold War carried out parallel studies for the USSR GDP but only reported US weighted estimates, substantially overstating the real size of the Soviet economy. Coming full circle Holzman continually pointed out the errors of the CIA methodology designed to mislead Congress and the Administration in op-ed pieces and other publications particularly in the 1980s.

⁴ However, during my period at Yale our family spent a year in India while I was teaching at the University of Bombay based upon my interests in monetary economics.

⁵ At that time it was *United Nations Statistical Office* but its present title, *United Nations Statistics Division*, is used throughout this memoir.

group at Penn were directly involved. Around the time Phase III was published in 1982 we had moved on to extending the country coverage at an aggregate level in a research effort termed the *Penn World Table* or *PWT*. Throughout this document Phase I will be referred to as the 1970 benchmark or ICP 1970 interchangeably, in part to distinguish it from extrapolations, like the PWT. The remaining chapters sketch the story of the ICP and the PWT up to 2016 from a personal perspective. An Appendix provides images of early ICP meetings in various locales and many of the players in *The Project*.

I would like to thank Sultan Ahmed for providing me a number of pictures and for filling in some dates and events when he was at the World Bank. Wilfred Beckerman kindly supplemented my memory with respect to early events at the OECD, when Gilbert and Kravis undertook their 1950 comparisons. And my daughter Laura and my partner, Bettina Aten for their gentle, supportive and frequent encouragement. A very special thanks goes to David Roberts (Pic 9) who has firmly edited my early drafts and as staff of the OECD a witness to much of the action from 1980 until his retirement in 2009. His continued encouragement and good humor have been invaluable. Fred Vogel and Angus Deaton have also kindly commented on an earlier draft leaving full responsibility for remaining errors with me.

Chapter 1: Phase I Antecedents and Early Meetings

Background of the ICP

The United Nations Statistical Commission (UNSC) meets every year (prior to 2000 it was every two years) bringing together staff from national and international statistical offices to give priority to various activities of the UNSD. The UNSD submits papers to the UNSC reporting progress on different projects, like manuals on international trade statistics, on the system of national accounts, on trade, expenditure and production classifications and the like. Discussions ensue, a summary of the discussions is provided to the delegates, and a final draft of the decisions is made, often after a good deal of wordsmithing. In 1965 the UNSC approved some research that would report on the suitability of the often over-valued exchange rates of that era for converting economic aggregates of member countries. Potentially this report could raise contentious issues because country contributions to the UN budget were based on exchange rate converted per capita Gross Domestic Product (GDP).

The resulting report, *International Comparisons of Production, Income and Expenditure Aggregates*, was submitted to the 15th Session of the UNSC in early 1968. It recommended that a project be initiated by the UNSD beginning with a small but economically diverse set of countries. The UNSD was able to fund some staff with the help from the World Bank, USAID, and country contributions; the Ford Foundation made a significant grant to fund a unit at Penn.⁶ Irving B. Kravis was director of the project and

⁶ The support of these economic statisticians from the United Kingdom, Canada and the United States respectively was crucial for getting the ICP off the ground. They were visionary in their views of international economic statistics at the time. Eurostat, then known as the

the unit at Penn and Zoltan Kenessey was associate director overseeing the operations at the UNSD. In terms of overall management of the ICP the Director of the UNSD, beginning with Patrick Loftus (1962-72) and Simon Goldberg (1972-79), was responsible for organizing the first three phases of the work. Abraham Aidenoff was Assistant Director of the UNSD and took part in a number of meetings during the completion of the 1970 ICP. The UNSD was within the Department of Social and Economic Affairs and under the direction of Jacob L. Mosak during Phase I of the ICP.

A. ICP Predecessors

(1) Before Colin Clark's 1940 *The Conditions of Economic Progress*

If the Nobel Prize in Economics had been awarded in the 1920s, Paul Samuelson reflected that one of the winners would certainly have been Gustav Cassel (Pic 3), the Swedish economist who wrote,

"I propose to call this parity '*the purchasing power parity*'. As long as anything like free movement of merchandise and a somewhat comprehensive trade between two countries take place, the actual rate of exchange cannot deviate very much from this purchasing power parity." Gustav Cassel (1918, p. 413).

This became known as the *Purchasing Parity Doctrine* and was a major contribution to the understanding of the international economy at the time. It strongly supported the notion of a world where prices would tend to converge to one common level, as transport costs and other obstacles to trade declined both within and between countries. The law of one price, an implication of Cassel's doctrine, remains a powerful basis for modeling a variety of international and domestic explanations of price differences across space.

It is important to keep in mind the context in which Cassel was writing. Major trading countries went off the gold standard in the World War (WW) I period during which they also experienced significantly different rates of inflation. At the end of WWI countries had to decide if and when they would return to a precious metal standard, and at what price they would set their currency in terms of gold or silver. In his article Cassel was addressing a very important policy question and his answer was that Sweden (and other

Statistical Office of the Economic Communities or *SOEC*, was also an important source of support because its predecessor, the European Coal and Steel community had carried out price and expenditure comparisons and were familiar with the issues. (De Micheis and Chantraine, 2003). Much of the activity of the UNSD through the 1980s was to prepare manuals of different types of statistics, to request data from country statistical offices in a standard form, and to bring together country tables for dissemination, a useful enough program. The ICP was different in that it required countries to submit price data and national accounts in a specified form that were in turn processed centrally into statistical outputs that were outside the control of the countries except for the two-yearly sessions of the UNSC. This potential conflict between ICP and the countries was, and remains, a continuing tension from the beginning of the ICP until the present.

countries too) should set the price of their currency in terms of gold and silver based on the purchasing power of their currency relative to its trading partners. The important point here is that Cassel stressed the purchasing power parity (PPP) was a guide to setting the exchange rate; not that the exchange rate should determine the PPP by dragging countries through unsettling periods of deflation/inflation if the exchange rate was overvalued/undervalued.⁷

International trade text books describe the Purchasing Parity Doctrine in its *absolute* and its *relative* version. The *absolute* version suggests that the price ratio of all goods in two countries would equal the exchange rate, truly the law of one price. While this is a useful starting point, in the real world it can be misleading. With the introduction of the Euro some argued that there would be no reason to carry out PPP research like the ICP.⁸ But a single currency does not assure that prices will be similar. Witness the large regional price differences within China and the United States, or countries like Costa Rico that use the US dollar but have prices well below those in the United States.

The *absolute* version found little support after WWII for many reasons. First, there were barriers to trade across countries that impose wedges between national and world prices of tradables. Second, in the 1950s, most countries maintained capital controls under Bretton Woods fixed exchange rate system. Third, as the regime of fixed exchange rates and restrictions on capital flows gradually dwindled in the early 1970s, the world became much more financially integrated than it was when Cassel wrote. Exchange rates in the short-run began to be more influenced by relative interest rates than by trade in goods and services. The volume of capital movements across countries today is over fifty times larger than the international trade in goods and services.

Today the *relative* version of the Purchasing Power Parity Doctrine prevails in most trade models and in extrapolations of PPPs. If prices in a reference currency, A, increase for some aggregate like GDP by 10 percent and only 5 percent in country B then the

⁷ Keynes was the editor of the *Economic Journal* where Cassel's article, which fit his views, was accepted for publication. Keynes was heavily involved in the debate over whether the United Kingdom should return to its pre-war exchange rate of \$4.80 per pound or a lower rate. Proponents of the pre-war rate primarily made the case that it was important to maintain the United Kingdom's reputation in international finance. Keynes argued that the \$4.80 rate should not be maintained and if it was maintained it would force a downward adjustment of prices and production. The chair of the Committee said if you ask five economists a question, you get six answers, two from Mr. Keynes. The United Kingdom did maintain the pre-war rate in 1924 and the UK economy did not enjoy the prosperity that other countries did in the remainder of the 1920s.

⁸ When the Euro was introduced Denmark recommended abandonment of the PPP program in Eurostat precisely on the grounds the prices would be the same in the Euro zone. It was pointed out that within a single currency zone in the United States consumption prices varied by 30 to 40 percent between cities of the South and Midwest compared to the Northeast or West Coast. And within the Euro zone price levels today have diverged, not converged after the Euro was introduced.

relative version says the PPP of B will fall by about 5 percent. In addition much modelling assumed that in the medium run there would be tendency for the PPPs and exchange rates to move together. There are many advantages to analysts working with these assumptions. And they are the most common assumptions made when organizations like the World Bank make extrapolations between benchmark studies.

Long before the development of Cassel's contribution to the literature on international trade and regional science there were anecdotal and more systematic references to price differences across space. A watershed event was the influx of precious metals into Spain, and subsequently Northern Europe and even more gradually Asia over the period 1500-1700. Silver was the main import/plunder sent back to Spain, the European price of gold in terms of silver rose from 9:1 to 15:1 as New and Old World monetary stocks merged. Spain's leading export to its European trading partners became silver devastating its traditional export industries like wool, at the time a puzzling phenomenon analogous to present day *Dutch Disease*.⁹ The Dutch and British East India companies in turn found that their most profitable export to Asian markets was silver where the price of silver to the price of gold was initially 9:1. The law of one price is a convenient model for thinking about the arbitrage incentives that gradually drove the price of gold in terms of silver to the same level around the world. In the centuries from 1580 to 1800 the gold-silver arbitrage worked slowly through the world system¹⁰; today, it would take seconds.

The trading companies used *Bazaar Walkers* to obtain prices in Middle East markets to send back overland to the European ports to inform the home offices what suitable items to ship out and home on new voyages. The systematic collection of prices became common but records are often episodic. Cities and business journals began to collect prices over time in the 19th centuries that permitted some spatial comparisons of price differences. The Aldrich (Nelson Aldrich, Statistician) Report to the US Senate Finance Committee in 1893 examined wholesale prices, wages and transport rates within the United States over the period 1840 to 1890. In addition price statistics from similar reports for the United Kingdom and continental cities, including for example Hamburg and Vienna, were appendices to the Report. The Board of Trade surveys in the United

⁹ Explanations for the decline in former export industries in Spain ranged from the loss of moral fiber of the population as illustrated by the outrageous dress of the women and men to the general laziness of the working class.

¹⁰ *The Flaw of One Price* is an article in the *Economist* (October 18, 2003, p. 73) reporting on the Euro-price of a number of items in Euro-area countries. Clearly the expectation was that one price should prevail, but in fact very significant discrepancies remained for years after the Euro was introduced. The percentage difference between high and low prices across the countries was 160 percent for a cinema ticket and 100 percent for a cup of coffee (non-tradables), and 75 percent for milk and jeans and over 50 percent for Pampers and Nurofen, some typical tradables. The consistent message of empirical studies is that the law of one price does not even prevail for tradable goods.

Kingdom similarly covered different cities within the United Kingdom and other countries including the United States.¹¹

Time series on consumer prices, wages and wholesale prices became common before WWI. By the late 1930s economic statistics collected by governments were more systematically reported within a national accounting framework with respect to output and expenditures. Price and wage statistics became formalized, frequently on a national basis including both larger and smaller urban areas. Catalogue sales became an increasing share of the market for durable items complementing regular price collection surveys by government statistical agencies.

¹¹ A very good survey of source materials is provided by R. C. Allen (1994).

(2) The period 1940 to 1965

One of the major contributors to the national accounts literature in the 1930s was Colin Clark (Pic 4), then serving as a statistician for the Economic Advisory Council from 1929-31 and as a Lecturer in Statistics at Cambridge from 1931-37. Keynes was on the Economic Advisory Council when Clark was there and was instrumental in obtaining Clark his position at Cambridge. Keynes used the national income framework of Clark and Kuznets and some of their estimates in *The General Theory of Employment, Interest and Money*.¹² Clark's 1937 book *National Income and Outlay* covered the year 1924 and quarterly estimates for 1929-36 using the now common identity that national income can be calculated from the income, product and expenditure sides. It became the model for the national accounts of the United Kingdom and also traced them back to the 1688 estimates of Gregory King. Keynes was influential in the founding of the Department of Applied Economics at Cambridge and Richard Stone became its first Director where Clark was a researcher. Clark left Cambridge in 1937 eventually taking an advisory position in the Department of Industry of the Labor Government of Queensland in Australia where he remained until 1952.

In the midst of those responsibilities, Clark managed to complete *The Conditions of Economic Progress* in 1940 without any research assistance and only a slide rule and an adding machine as computational aids. This was a very influential study that made a binary comparison for consumption between each or a number of countries in Europe, Asia and elsewhere with the United States. The sources of data ranged from "reasonable" for 16 countries to "quite scratchy" for the other 34 countries. He converted national currency totals to a common *international unit* that was conceptually similar to the *international dollar* in Phases I-III of the ICP.¹³ The PPP for consumption was used for the conversion of totals that included investment and government. With the benefit of hindsight this approximation of the GDP PPP is not far off for many countries in various ICP rounds. Clark went on to create a world total of income for his 50 countries. And, in terms of welfare measures, he chose as his denominator hours of work when available and otherwise per capita income.

WW II was an economic shock for most economies, so it was not propitious timing for Clark's book to receive the attention it deserved. He would subsequently bring out a second edition in 1947 and an updated edition in 1957 that used a reference year of 1950. The 1957 edition introduced the notion of an *Oriental Currency Unit (OU)*. Clark came to view the *international unit (IU)* as only appropriate for countries relatively close economically: a total of 26 economies with a reference year of 1929, which Clark felt was an improvement over an average of years that was used in his 1940 edition. Countries from Asia, Africa or Latin America were not included among the twenty-six.

¹² For a very rich discussion of Clark's career and contributions, see Angus Maddison (2004).

¹³ The ICP *international dollar* was defined as having the same purchasing power as the US\$ over GDP in a benchmark and other years. Clark's *international unit* had a similar definition but over a series of years, 1925-34, was the reference period for his analysis of economic growth of the 16 countries with the more reliable data.

In Table XV (Clark, 1957, pp. 58-59) he provides estimates for 19 countries for 1950 and an additional four countries for 1948 using the OU.¹⁴ A main feature of the OU was that it excluded investment and all government services except education and health, and allowed a flat 6 percent for house rent expenditures in all countries. He extended these estimates in OU to 102 countries in Table XIX in the following way. M.K. Bennett (1951) had used a set of physical indicators for a set of 102 countries to derive an index of development. Clark estimated a regression for the 23 countries that had a value of the Bennett index and OU value and used it to estimate OU values for the remaining countries for which Bennett had provided an index. This procedure is a good example of the ingenuity Clark displayed in deriving the most from limited available data. His efforts to translate his OU results to international unit measures are reported but apparently were not viewed with enough confidence to total up the results to a total for the 102 countries.

The European Coal and Steel Community

The Treaty of Paris of 1951 established the European Coal and Steel Community (ECSC) consisting of six countries: France, Italy, West Germany, Belgium, Luxembourg and the Netherlands. It introduced a common market in 1953-4 for coal and steel in part to remove from post-war Europe a catalyst for conflict and to create an institution that would be one step towards the European Economic Community created by the Treaty of Rome in 1957.¹⁵ An activity of the ECSC was to survey prices and wages in the member countries that were aggregated to overall measures of real consumption across their membership.¹⁶ Jan Van Yzeren (1956) was hired as a consultant to the ECSC to look at ways to aggregate the price and wage surveys. Van Yzeren produced several different methods, one of which, his preferred *balanced* method, was illustrated, but not ultimately used in the first ICP report. (Kravis, et. al., 1975, pp.67-8 and Table p.75).

¹⁴ The 1957 edition runs to 720 pages and is less focused than the 1940 edition with many different threads running through the different chapters including a very thoughtful treatment of the agricultural, industrial and service shares in production, income distribution, Russia and an Excursus on economies in the ancient world. In contrast to the 1940 edition the 1957 edition does number its Tables. However, the former had a summary chapter, but the latter did not. It should also be noted that the Gilbert-Kravis (1954) study published when the 1957 edition was in press, so Clark was only able to note that it had much better data coverage for its countries.

¹⁵ The European Economic Community became the *European Union* with the signing of the Maastricht Treaty in November 1993. European Union is used throughout the rest of the memoir.

¹⁶ Bert Balk kindly pointed out that the ECSC surveys were not mentioned in an earlier draft of this section. For a fuller treatment, see Balk (2008). Pp. 42-4. Balk (Pic 30) was a thesis advisor to van Yzeren, who late in his career was awarded a Ph.D. that was very well celebrated.

The Bretton Woods Regime

The Bretton Woods agreement established the International Monetary Fund (IMF) and a system of fixed exchange rates that ruled from 1947 to 1971. Many at the time realized that those exchange rates were further from PPPs than at the time Cassel first used the term purchasing power parity in 1918. This was because there were more barriers to the free flow of trade, capital and people compared to WWI. In addition, most countries instituted foreign exchange controls during WWII that remained in effect long after the IMF was established. But comparisons among countries were needed to make assessments to the various parts of the United Nations (Food and Agriculture Organization (FAO), World Health Organization (WHO) and the like), for quotas for the IMF and World Bank (then known only as the International Bank for Reconstruction and Development or IBRD). Outside of Clark's work, the exchange rate was the only wheel in town circa 1947.

The Council of Mutual Economic Assistance

The Council of Mutual Economic Assistance (CMEA) was established in 1949. From 1960 it included a Working Group to undertake binary purchasing power comparisons between the USSR and other members. The comparisons were conducted within the Material Product System (MPS) of national accounts.¹⁷ The Working Group completed comparisons for 1959 and 1966. Beginning in 1971 the Standing Commission of the CMEA on Statistics took over this work. Four comparisons were carried out by this body: for 1973, 1978, 1983 and 1988.¹⁸ Some of this work is described in Syzilagi (1966) and Ivanov (1978) (Pic.6,7). The experience and advice of statistical staff in Hungary and Poland became an important input into the initial ICP round.

The Gilbert-Kravis Study

Because exchange rates among their member countries were often artificially maintained, the Statistics Division in the Economics and Statistics Department of the Organisation of Economic Cooperation and Development (OECD)¹⁹ under the leadership of Milton Gilbert initiated PPP binary comparisons between the United States and Italy,

¹⁷ The MPS differed in two major ways from the UN System of National Accounts (SNA): it did not include services in their total as does the SNA and it did not net out purchases from other firms in totaling up gross sales of enterprises as in the SNA. Often the totals of MPS and SNA were not greatly different from each other.

¹⁸ Youri Ivanov, formerly at CISSTAT where PPP comparisons are carried out across the Commonwealth of Independent States and who currently teaches Statistics at the University of Moscow, kindly provided background on previous comparisons.

¹⁹ At the time it was the *Organisation of European Economic Cooperation (OEEC)*. It became the Organisation of Economic Cooperation and Development (OECD) in 1961. It is referred to as OECD throughout this memoir.

the United Kingdom, France, Belgium and Germany in the early 1950s.²⁰ Gilbert invited Irving B. Kravis from the University of Pennsylvania to join him in the study.²¹ The work was initiated in 1952 and published in 1954. (Gilbert and Kravis, 1954).²² This was followed up with a study of nine European countries (the six original countries plus Belgium, Denmark and the Norway) and the United States (Gilbert and Associates, 1958).

The Economic Commission for Latin America

Another set of cross country comparisons was carried out by the UN Economic Commission for Latin America and the Caribbean or ECLAC (1967, then ECLA) and S.N.Braithwaite (1968). These studies priced the same items in all cities employing the same team in the capital cities of each country. Further the prices were aggregated using average expenditure shares of all countries, the Walsh (1910) method, which is also illustrated in the first ICP report. A linking was made to the United States through a set of prices for Houston obtained from the US Bureau of Labor Statistics (BLS). These methods are discussed further in the next section.

²⁰ In this period expenditures were grouped into categories for which there were expenditure weights or shares. Price comparisons were then made for specified goods and services within a category that were summed and averaged first to the category level and then to various aggregate levels like all meats, food and beverages, or household consumption. On the production side, comparisons would be for productive sectors like agriculture or construction using both direct and indirect quantity comparisons to arrive at aggregates like manufacturing or all of GDP.

²¹ Two aspects of this initial study stand out from my conversations with Kravis. First, in their discussions Gilbert always focused on the quantity comparisons and Kravis the price side of purchasing power comparisons. Necessarily they would end up with the same conclusion, but they always argued from a different starting point. A second recollection provides a sense in which Kravis always gave great attention to detail and to checking results, much due to Simon Kuznets, a mentor to both Gilbert and Kravis at Penn. The initial results became available few days before Kravis was to return to Philadelphia to begin teaching. He was so surprised by the Paasche-Laspeyres spreads for Italy and the United States of 75 percent that he postponed his return from Paris several days to redo the calculations. Happily there were no mistakes, and the Italy-United States spread turned out subsequently to be unremarkable for many pairs of countries. However at that time Kravis was familiar with Paasche-Laspeyres spreads from time to time price indexes where differences over several years were rarely over 5 percent, hence his initial surprise at the Italy-US result.

²² Wilfred Beckerman, who worked on (and in fact, authored) the update of Gilbert-Kravis (Gilbert and associates, 1958) kindly provided some background on the roles that each played in the work. Gilbert had more of a conceptual role while Kravis saw through implementation and both shared in the writing. Angus Maddison was also at the OECD at the time and was later to pursue PPP studies from the production side at the University of Groningen developing international comparisons of productivity. He established the Groningen Growth and Development Center (GGDC), where he mentored a number of graduate students. The GGDC is currently involved in many projects such as KLEMS, supply chains, a center of an international consortium, the Maddison Project that continues his tradition of historical output and productivity studies.

(3) Early Methods of Comparing PPPs

Country comparisons are of interest to assess the total and per capita volume of goods and services that each economy is producing and consuming. Direct quantity comparisons can often be made, like hospital bed days or kilograms of meat consumed. In both these examples, there can be differences in quality that do not make direct quantity ratios very comparable as measures of volume. Moreover for much of GDP we have only expenditures for a diverse set of items within a basic heading²³. Examples are household textiles, condiments and spices, and restaurant meals. In these basic headings quantity comparisons are indirect as in (1) below, where Q is quantity, E refers to expenditures in a basic heading and PPP is the purchasing power parity for a basic heading. The subscripts A and B refer to countries.

$$(1) \quad Q_B / Q_A = (E_B / E_A) / (PPP_B / PPP_A)$$

In a two country comparison the basic heading PPP is the geometric mean of the price ratios of each matching item within a basic heading.²⁴ Above the basic heading level countries typically have different expenditure shares so it does matter whether the basic heading quantities of country A or B or some average of the two are used. When the quantity weights of A, the reference country, are used the resulting PPP is called a Laspereyes price index and if the quantities of B are used, a Paasche price index. Usually the Laspereyes price index is larger compared to the Paasche price index.²⁵ In time to time indexes it has been common to publish a Laspereyes price index or a chain index where weights change each year. Across space it is common to use a Fisher index, the square root of the product of the Paasche and Laspereyes indexes.

²³ A basic heading is the lowest level of aggregation in the breakdown of expenditure on GDP for which PPPs are calculated. This level of aggregation is generally determined by the lowest level of final expenditure for which explicit expenditure weights can be estimated. Hence, while in principle a basic heading would consist of a group of similar well-defined goods or services, in practice it can cover a broader range of products than is theoretically desirable. Basic headings are the building blocks of a comparison. It is at the level of the basic heading that expenditures are defined, products selected, prices collected, prices edited and PPPs first calculated and averaged.

²⁴ The geometric mean is preferred because the result is independent of which country is taken as the reference. Where the average is unweighted these basic heading PPPs are referred to as *elementary indices* in the price index literature. In general there are no item weights below the basic heading. In the 1980s the EU countries began to distinguish between goods and services that were *more* or *less* representative. For a more detailed discussion see the ICP Handbook (2008, Ch. 11) and the EU Methodological Manual (2012).

²⁵ In time to time indexes A would be the early year and B a more recent year. Using early year weights for headings where prices have risen will push up the Laspereyes price index compared to the Paasche index using later year weights to the extent relative quantities go down when prices go up. This *law of demand* is rarely violated in empirical studies. The same principle applies when A and B refer to countries.

The approach of Gilbert and Kravis was to make binary comparisons of Fisher indexes for each European country with the United States in what is called a *star system*.²⁶ The reason for putting the United States at the center was that the dollar was not subject to exchange controls as opposed to the European countries. The indirect relation between say the United Kingdom and France could be easily derived from $\text{France/United Kingdom} = (\text{France/United States}) / (\text{United Kingdom/United States})$. The CMEA countries also used a star system with the USSR as the country conducting binary comparisons with members like Poland or Bulgaria. It is important to keep in mind that the indirect comparisons will only equal a direct comparison under special conditions²⁷ so there was always an interest in producing multilateral results that were independent of the reference country.

An important set of Latin American comparisons were carried out under the auspices of ECLAC from 1955-62 with a reference year of 1950 (Braithwaite, 1968). In the late 1950s many countries in the region had multiple exchange rates, most frequently for different export and import groupings. This made choosing an exchange rate for comparing consumption or other aggregate between Latin American countries fairly arbitrary. ECLAC made a number of simplifications to balance data availabilities, computing capacity, and resources to produce their estimates.

One simplification was to have each country provide capital city prices for each item so that basic heading PPPs were simply the geometric means with respect to any reference country. To simplify the aggregation an average of the expenditure shares were taken for the countries weighted by their per capita GDP converted at exchange rates for 1950, an assumed normal year.²⁸ The aggregation of the basic heading parities became quite straightforward even with just electric calculators. Another important aspect of the study was to link the results for 1960 and 1962 with the United States using special price surveys conducted in Houston and Los Angeles to obtain a better item overlap with Latin American countries. This allowed the results of the studies to be expressed in US dollars,

²⁶ I did not know Milton Gilbert personally so I asked another survivor, Wilfred Beckerman, his recollections of working at OEEC and elsewhere with both Gilbert and Kravis. He likened Gilbert to the composer who has the conceptual framework, Kravis as the conductor seeing the performance through to the end, and Beckerman who was in the violin section of the orchestra, first violinist, of course.

²⁷ A sufficient but not necessary condition is that all countries price the same items and the expenditure shares are the same across countries.

²⁸ The average expenditure shares used by ECLAC is similar to that of Walsh (1910). The Walsh method was a simple average of the shares of all countries participating in a comparison for each basic heading. A follow up comparison (Salazar-Corrilla, 1978) was carried out under the auspices of the Program of Joint Studies of Latin American Economic Integration (ECIEL). ECLAC carried out its studies normally with the statistical offices of the member countries whereas ECIEL worked with research organizations in each country, a later source of irritation to the countries. Members of ECLAC included the United States that at that time regarded ECLAC as too far to the left, which led the United States to support ECIEL economic studies in preference to ECLAC's.

which were well understood at the time, as opposed to the currency of any particular Latin American country.

A few other studies attempting to measure volumes across countries prior to the ICP should also be mentioned. Cost of living and labor productivity studies for autos, and textiles were carried out between WWI and WWII by the International Labour Organization (ILO), country statistical offices and industry associations. Paige and Bombach (1959) undertook a detailed comparison of productivity from the output side between the United Kingdom and the United States under the auspices of the OECD. Beckerman (1966) and Beckerman and Bacon (1966) attempted to make volume comparisons across a wide variety of countries based on physical indicators in a similar way to Bennett (1951). This approach was especially designed to include developing countries where conventional national accounts were not well developed but where many physical indicators were available. The Beckerman-Bacon method was one of a number of short-cut type approaches that have evolved to estimate real product across countries.

B. Phase I of the ICP: 1968-1970

The year 1968 is cited as the beginning of the ICP because the UNSC approved this activity in their February session. A lot of groundwork went before in which I did not participate. The groundwork included a paper to the UNSC in 1965 that recommended that the UNSD consider how the problem of developing comparable measures of income and production across member countries could be addressed and to draw up a project proposal for presentation at the 1968 UNSC. The interim period was devoted to developing a plan and to obtaining resources for such a project. Some countries were willing to contribute resources in kind, usually seconded statisticians, who it turned out, proved quite effective. Other countries contributed money most of which went into a trust fund administered by the World Bank.²⁹

Leadership of the ICP was the next issue. While Patrick Loftus, the UNSD Director at the time, would be overall supervisor, what was needed was intellectual leadership for the project. In the end Irving Kravis agreed to direct the project with a counterpart at the UNSD, Zoltan Kenessey of the Hungarian Statistical Office. Hungary and Poland had been involved with purchasing power comparisons with the USSR and their staffs were up to date on methodology, an important input for the ICP. The unit at Penn was headed by Kravis and financed by a grant from the Ford Foundation to cover the first years of the ICP. This gave the project an academic input that was important in formulating a basic framework for the initial comparison but also subsequent phases of the work.³⁰

²⁹ There was some reluctance of donors to give grants directly to the United Nations. This applied to the Ford Foundation, some country donors and the World Bank. In retrospect this provided some beneficial checks on all groups involved.

³⁰ I once asked Irv why he took on the project, given how highly regarded the Gilbert-Kravis report had become. If not his exact words, in effect he answered he thought he could do a better job than Colin Clark. However, Kravis envisaged the ICP taking less of his time than it

The UNSC approved the plan for comparisons on the expenditure side of GDP, the framework for Phase I of the ICP. But it also endorsed comparisons from the production side as had been carried out by Paige and Bombach (1959). The advantage of production side comparisons is that they permit productivity comparisons for sectors of the economy as well as for all of GDP. Kravis agreed with this point but argued that the production side, *if done thoroughly*, required collecting not only all the prices needed on the expenditure side, but also additional collection of input prices into production. *Done thoroughly* meant double deflation of final output values, which did require collection of much more data. In the end, the ICP adopted the expenditure side approach and did not attempt to collect data on the production side. However, others did assign selected ICP expenditure heading PPPs to output sectors to make productivity comparisons, an ad hoc and rough method. Happily Angus Maddison and his group at Groningen observed the benign neglect of the output side by the ICP. And they did not make the “best the enemy of the good” but rather began to implement more pragmatic output side comparisons that have become much firmer over time.

Initially a select group of countries were invited to participate, Colombia, Kenya and India to represent developing countries, Hungary and Poland to represent planned economies, France, Germany and Italy from the European Union, Japan, the United Kingdom and the United States to represent the developed countries. Because the Netherlands, Belgium and Luxembourg were also in the Economic Union, there was always the possibility that they would participate. At the time the project began, it was recognized that some potential countries had major statistical issues, for example expenditures in the rural sector of Colombia. In addition it was not clear that Poland would participate. It was originally planned that there would be comparisons for at a minimum six countries for 1967 as well as for the 1969 benchmark price comparison. In retrospect a benchmark for 1969 was a quite unrealistic expectation. Further, six additional countries were contacted during the planning period about participation if not immediately, relatively soon thereafter. They were Belgium and the Netherlands from the European Union, and Iran, Korea, Malaysia, and the Philippines from West and East Asia.

I joined the project in spring 1968 and took part in an early meeting with a Japanese delegation of Sadanori Nagayama, Tsu-tomu Noda and Mitsuru Ide who visited Penn. Ide was an academic and both Nagayama and Noda would teach when they retired from government. All three had been involved with binary comparisons of Japanese cities with those in China and India. The Japanese were enthusiastic about comparisons for household consumption, but they were reluctant to do all of GDP. Kravis strongly pushed for all of GDP and the Japanese delegation agreed to think further about government and capital formation, and all agreed to move forward on consumption first. The capital formation discussion with Japan was in effect postponed to a later time.

did. He had hoped that his long-time collaborator, Robert Lipsey, who taught International Economics at Queens College and was to head the National Bureau of Economic Research (NBER) office in New York, would do the day to day running of the ICP at Penn. In the end Lipsey chose not to give up his Greenwich Village way of life.

(1) Some Early Decisions on the Framework of the ICP Comparisons

One major decision was discussed above, namely to undertake the Phase I comparison from the expenditure side. In striking contrast to the ICP rounds in 2005 and 2011, Phase I was a very top down affair with representatives of the UNSD or Penn dealing directly with countries, the exception being members of the European Union. The Statistical Office of the European Communities (SOEC) in Luxembourg was responsible for determining the cost of living relative to Brussels of EU employees in various locales in Europe, such as the European Centre for Nuclear Research (CERN). So a price collection mechanism already existed for member countries, although their specifications were not necessarily the same as would be used in the ICP. Kenessey and Kravis visited many of the countries during 1968 and 1970 establishing a time line for country submission of consumption prices, and discussing the worksheets being developed for capital formation and government.

Another decision that impacted the resources available to the UNSD related to post-adjustment allowances. Prior to 1974 the UNSD provided cost of living adjustments relative to New York for the various locales of employees of FAO, UNESCO, WHO, the Regional Commissions and related organizations. The resources available in this unit were complementary to the ICP and might have been better coordinated with other price work of UNSD, like the ICP. When this activity was transferred to the International Civil Service Commission (ICSC), not only did the UNSD lose some staff familiar with pricing in different countries, they lost the increased budget that ICSC received accompanying the transfer. The fact that the European Union did carry out these surveys was a major reason that SOEC had budgets for price work that in turn was important for early ICP work to gain momentum.

(2) New Bolton Meeting, June 1969

During the 1968-70 period there were two important meetings devoted to discussing implementation issues and overall methodology of the project: the first June 1-4, 1969 at the New Bolton Center of Penn and the second at the Rockefeller Center at Bellagio in Italy in October 1969. The New Bolton Center is Penn's Veterinary facility in the Pennsylvania countryside that caters small workshops, and is conducive to working, walking and talking. The purpose of the meeting was to obtain reactions from the various countries or regional representatives to initial proposals on the expenditure classification and treatment of issues like own production. A quick review of the Minutes of the meeting will give a flavor of these early discussions. Loftus, Kenessey, Kravis and I were there with 13 others of whom Mrs. Mod, director of the Hungarian Statistical Office should be especially mentioned because of her background in international comparisons. Andrew Flatt represented the UN Economic Commission for Africa (ECA), and would later be at the UN Economic and Social Commission for Asia and the Pacific (ESCAP), and Philip Goybet the European Union. Ms. Maurice represented the United Kingdom, Janet Norwood, Milton Moss and J. Rottenberg, the US government and J. Salazar, the Brookings Institute.

One perennial issue, including planning for the 2017 ICP, is whether one size fits all in terms of the number of basic headings for which countries provide prices and expenditures. The conclusion at New Bolton was that countries or regions may reduce or increase the number of basic headings so long as the overall framework of basic headings was preserved. The point was raised as to whether triangles of countries, like India, Japan and the United States might pursue a common list. The United Kingdom indicated that it would like to pursue a common list with the United States. In terms of timing it appeared that Hungary, the European Union, Colombia and the United States might be able to simultaneously prepare a list and all agreed that would be useful. In the end India, Japan and Kenya were to develop their pricing lists somewhat later than the other countries.

Some typical classification and pricing questions.

- Should food consumed in hospitals and restaurants be removed from their expenditure group and transferred to food? Consensus was no, but they could be separated within hospitals, say, and anyone could add together all food.
- Should the service and food part of restaurant meals be separated? Consensus was to try pricing comparable meals in national chain restaurants. As an aside, the BLS at the time collected restaurant menus on a quarterly basis.
- Should expense account meals, which loom large in Japan, be added to GDP? Strong objections to changing GDP were expressed by the United Kingdom and the United States but no consensus was reached.
- How should subsidies be handled? Mrs. Mod carried the day arguing strongly that the price used for comparison should include both consumer expenditures and the housing subsidy, but she was more skeptical about this adjustment for other expenditure headings.
- Should an imputation be made to government for the implicit return on government buildings? Thought feasible except for the military.

Other decisions at New Bolton included asking countries for annual average prices including seasonal items. Substitution of comparable items where there are known national differences, for example, sporting events might be baseball and basketball in the United States and cricket and soccer in the United Kingdom. In construction, substitution of materials may be acceptable and price per square meter for different building types - commercial, institutional and residential - should be the unit. For roads, different types by grade and terrain were recommended, again different materials may be substituted. To conclude on New Bolton, it was very much focused on issues of implementation.

(3) Bellagio I, October 1969

The first meeting of the ICP Advisory Board was at Bellagio in the Lake country of Italy where a document, *Plans for International Product and Purchasing Power Parity Comparisons*, was discussed. The Rockefeller Center at Villa Serbelloni on a hill above the village of Bellagio overlooking Lake Como was a major contrast to New Bolton. At the time Villa Serbelloni provided rooms and meals for small workshops and conferences as

well as longer stays for writers and scholars working on particular projects. The rooms themselves were large and had ceilings I would guess that were fifteen feet high including the bathrooms that had floor to ceiling windows that had no curtains as I remember because nothing could be seen from the ground. Those attending meetings had lunch together and in the evening all residents shared a cocktail hour and dinner and wider ranging conversations. The Villa grew some of its produce and had its own vineyard and winery.

Among the Advisory Board members, Milton Gilbert, Simon Kuznets and Richard Stone could not attend. Nor could S.Tsuru from Hitotsubashi University, nor Loftus and Mosak from the United Nations because the opening of the General Assembly corresponded with the only dates that Villa Serbelloni had available. Mrs. Mod from Hungary, Madam Kreczkowska from Poland, Uttam Chand from India, Wilfred Beckerman from the United Kingdom, Guy Bertaud from the European Union and Kenessey, Kravis and Heston were in attendance. Between the New Bolton meeting and Bellagio, the reference year for Phase I of the ICP had moved from 1969 to 1970.

A great deal of the time was spent on wording, which was appropriate because the document would receive wide distribution and would be a background paper at the 1970 UNSC. There was an extended discussion on what was meant by representative item as opposed to an item that was available in the market but with a price that might be high because of low volume. This difference has subsequently become part of the EU-OECD methodology and was introduced in a more limited form in the 2011 ICP. Beckerman raised the question of whether countries should be encouraged to choose items that had less price variability within and across countries, even if the item was not a volume seller. This consideration arose in the context of footwear in India where sandals (chappals) were most commonly worn but much less so in Europe and the United States. In practice Bata shoes were chosen for comparison in 1970 because India exported Bata shoes to Canada and other countries so even if much less common in India, the prices were representative for men's shoes.

Several exchanges occurred dealing with the own production for own consumption as well as own production that the household sold like home spun cotton thread. The consensus was to treat own production as recommended in the 1968 UN System of National Accounts (SNA), namely to value such production at the prices in the first stage of marketing. A major departure of the ICP from the SNA was in the treatment of health and education services paid for by governments or non-profit institutions. In the SNA only education and health expenditures paid for by households was counted in household consumption. Mod, Kenessey and Kreczkowska argued that this would lead to major non-comparability of the quantities of these services consumed between the Socialist countries and the United States. This discussion led to the distinction between private and public provision of education and health services labeled ICP and SNA concepts of consumption in Phase I of the ICP. The ICP distinction was subsequently adopted in the 1993 SNA where the two concepts are termed *Actual Consumption of Households* and *Household Final Consumption Expenditure*.

Additional discussion of health services centered on whether quantity measures like hospital bed days (Beckerman) should be used or (as Kravis argued) attempts should be made to get closer to consumer valuation by making some price comparisons. Forty years later in 2011 the European Union and the OECD moved towards the Kravis position but the data demands remain beyond most of the other ICP countries.

In discussing equivalent products Uttam Chand pointed out that it was common in India to have a tailor make clothes for a family because it was cheaper than buying ready-made clothing at a store. In this case it was agreed to use a price for a given specification that was most common in a country, in the Indian case, cost of materials plus tailoring charges. Put another way, in 1970 it would make no sense for India to price a ready-made shirt specification or for the United States to price a tailored shirt. This type of substitution was common in construction specifications involving type of material, and in a number of consumption items.

Short-cut and simplified methods of comparison were discussed at some length.³¹ Previous work concluded there was considerable error in using these methods for low income countries, so that at least for getting the levels of income right, it will be necessary to complete purchasing power studies that provide a benchmark. Kreczkowska pointed out that while indicators may not get levels right, they may still be useful for moving benchmark PPP estimates over time between benchmarks. Simplified methods of obtaining PPPs would aim to work within the SNA framework to make estimates for a relatively small number of aggregates. Mod suggested trying a larger number of countries collecting as few as 30 prices covering main aggregates of expenditure. This is a type of reduced information approach that is gaining support for the 2017 comparisons. The conclusion of this discussion was that the physical indicator approach had promise, but until benchmark PPP estimates are available it will not be possible to judge the merits of one short-cut or simplified method of estimation versus another.

³¹ Beckerman (1966) and Beckerman and Bacon (1966) made estimates of real product using physical indicators on the right hand side and per capita consumption at exchange rates on the left for 22 industrial countries. Estimating equations were generated using from 1 to 5 indicators, and depending on indicators available in less developed countries estimates of real product were made for another 60 countries. Researchers in Eastern Europe also used indicators but the regressions were typically time series of national output on a single indicator, generating an estimating equation for each indicator. This procedure generated for each country real product estimates equal to the number of countries and indicators of which some average was taken. See for example, Janossy (1963) and Szilagyi (1964).

Chapter 2 Organization and Implementation of ICP Phase I

This chapter covers a period when the broad framework of the ICP was established in terms of the expenditure classification, the detailed item price list and the special worksheets for education and government but many details had to be worked out. A multilateral comparison was not new, *per se*, but because there were alternative methods that could be used, each with its merits and limitations, there were important decisions to be made. The organization of the work between New York, Philadelphia, Luxembourg (the European Union) and the countries was a continuing work in progress.

A. Phase I Activity 1970-1

(1) Division of Labor with the UNSD

There was an agreed upon division of responsibility between the group that Kenessey supervised in New York and the group that Kravis supervised at Penn. In terms of the collection of prices New York worked on public expenditures and equipment goods. The Penn group worked on consumption and construction. Contacts with countries were not rigid but New York concentrated on Hungary and Poland, when its participation was still not decided, as well as Kenya and Colombia while Philadelphia did Japan, India, and the United States. With respect to the EU countries both groups were involved which was also the case for the United Kingdom. Because of Kravis' previous work Penn took the lead in matters of methodology and computation. Visitors from participating countries typically visited both New York and Philadelphia. On country visits all aspects of the data needs would be discussed with those responsible, sometimes at several offices including central banks, planning offices, and ministries of construction and the like.

The staff in New York included Alphonso Pardo-Gutierrez who worked on public expenditures and national accounts and Michael McPeak who worked on equipment goods. Both took part in most discussions with visitors and shared work on Colombian data. McPeak would later join the Inter-American Development Bank (IADB). Donald and Karen Woods worked in Kenya on developing both the expenditure side of their national accounts and their consumer price list. Don would later join the BLS. Alfonso Uong and Antonio Yu took on a variety of tasks including national accounts.

Woods and Yu both began at Penn and then were added to the staff in New York. Alicia Civitella was an important member of the Phase I at Penn who became an authority on specifications of goods and services working with commodity specialists in the United States and other countries. She also was an all-rounder in terms of inputting prices and expenditures into the computing framework that had been developed at Penn. And as results began to emerge from the batch processing of that period she would share in our gloom or glee as the output warranted. She also worked closely with three others who contributed greatly to Phase I: Samvit P. Dhar, whose background was national accounts at the Central Statistical Office (CSO) in New Delhi, Sultan Ahmad (Pic 39), a graduate student from Bangladesh, who completed his dissertation on reduced information methods, and Lorenzo Perez (Pic 39), who took part in visits by visitors from Colombia and Venezuela. Alicia's skills were appreciated by the ICSC where she would eventually

end her career. Samvit worked at Penn for about 20 months in 1970-71 on leave from the CSO in India. He returned to India for a period, still involved in ICP work and would eventually work at the UNSD on Phase III of the project. Sultan Ahmed was heavily involved in the programming and computing for Phase I. Sultan went on to join the World Bank where he worked until his retirement, and did consulting in various countries during the 2005 ICP.

(2) Robert Summers on Board in 1970

Two of the Advisory Board Members who were not able to attend Bellagio met with Kravis and me earlier in 1969, namely Simon Kuznets and Richard Ruggles. Kuznets taught both Gilbert and Kravis as graduate students at Penn before he moved to Johns Hopkins³² and eventually Harvard. I knew Richard and Nancy Ruggles at Yale prior to joining Penn. They had both been involved with Latin American PPP comparisons through the Brookings Institute with Jorge Salazar and were supportive of the ICP. After these meetings and Bellagio, Irv wanted to bring in someone who could better deal with the econometric problems of moving from binary to multilateral comparisons. And that is how Robert Summers (Pic 11) came to join the Penn group in 1970, a fortuitous development.

The first problem that Bob tackled was at the basic heading level where in two country comparisons the issue was simple, price comparisons were done for items that both countries thought were representative. In Phase I the initial item list for consumption was built up from a set of BLS specifications that were modified in discussions with each country. In the end there was for each basic heading a matrix of prices and countries that had lots of blanks, but also lots of overlap.³³ Bob liked the idea of using all the price information that was available and that led him to develop the country-product-dummy (CPD) method for dealing with the issue of missing prices. He assumed the missing prices were randomly distributed as if someone had spilled coffee on a full price matrix so that some were missing, an assumption that was convenient econometrically, but often contested.³⁴ The CPD equation is:

³² Another student of Kuznets at John Hopkins, Robert Fogel (2013), has written a fine book on the Kuznets tradition. One point that Kuznets pushed in correspondence and at Penn was that there was a tradeoff between covering a lot of countries using physical indicators and doing the type of detailed price comparison envisioned for Phase I of the ICP. My memory is that Kravis paid lip service to this idea, and suggested I write a paper on indicators for the International Association of Income and Wealth meetings in Ronneby, Sweden in 1971 that was published in 1973.

³³ When all countries provide all prices then the CPD and binary geometric means are identical and the results are the same as the CPD or GEKS (discussed later) methods. As noted earlier the Latin American comparisons imposed the condition that all countries provide all prices, therefore reducing the multilateral comparison to a set of binary comparisons.

³⁴ Bob's modeling of CPD meant the error term was log-normally distributed which was convenient. However, a more plausible explanation of why there are missing prices is that the item was not representative of what was consumed in a country for that basic heading. If this is the reason, the implicit estimate that CPD provides for the missing price would be less

$$(4) \ln \text{PRICE}_{ij} = \alpha_i P + \beta_j C + \ln \varepsilon_{ij},$$

where log price is regressed on dummy or class variables P for the product and C for country and ε is the error term.

The coefficient on country, β_j , is of interest. If prices are in national currencies the exponent of β_j , is the PPP relative to the reference currency; or if all prices are divided by the exchange rate then the exponent of β_j is the price level of the country to the reference currency or currencies if a group of countries is the referent.³⁵ So long as there are at least two countries pricing each item, all prices enter into the estimation process. CPD will be discussed further below as will Bob's contribution in sorting out aggregation methods and in the development of the PWT.

Bob also was involved in the work on hedonic regressions that were used in work on automobiles, rents, and some machinery items. In his previous work with Robert Lipsey, Kravis had experimented with estimation of prices of items with a number of well-defined price determining characteristics. Transport equipment or pumps of various capacities are types of equipment goods for which a small number of characteristics, like weight and horsepower, capture most of the variation in price. Estimation of a regression equation where log of price is on the left hand side and a set of continuous or class variables (dummies taking on the value 0 or 1 as in a CPD) are on the right hand side. The coefficients of such a regression can be used to estimate the price of a standard specification for a given item, say automobiles, from the price data for a single country. The estimates of the standard specification can then be directly compared across countries.

Alternatively country prices may be pooled into a single equation, where the country coefficients provide the desired PPP or price level for the heading, like a CPD. Typically pooled equations require that all countries provide data on all characteristics, not the usual situation for the ICP. The basic headings where hedonic equations were used in Phases I to III of the ICP were house rents, automobiles, both for consumption and investment, and one or two equipment goods.

than what the actual price would be in that country. While CPD was criticized for this reason, in fact, I think the resulting country coefficients, the β_j s, are what we want.

³⁵ This second way of estimating the CPD equation is what is called price-level form. If the United States is 100 then the anti-log of the β_j s will be numbers ranging from perhaps 60 to 130 for a basic heading like rice. A value of 60 would mean the price of rice in that country was 60 percent of the US price. We found the price level form to make the CPD results easier to compare than if they were in PPPs. And circa 1970 when rounding could cause the loss of significant digits, use of the price level form reduced the variance of output compared to using PPPs that varied between less than two for the United Kingdom to over 1000 for Italy.

House rents are illustrative of some of the issues faced.³⁶ The basic framework was to estimate rental PPPs by type of structure, size, age and available amenities. In Phase I, 38 combinations of year built (from 1910 to 1960), size (from 15 to 90 square meters), and availability of amenities (including none) were used. Rarely do rental surveys have the same characteristics across a set of countries unless there has been some prior agreement to use a common survey as was the case for the European Union. Size of dwelling is most often in square meters, but in the United States and many Latin American countries it is number of rooms, which were converted roughly into square meters. Countries typically ask if there is running water, electricity and an indoor toilet, but in the United States it is number of toilets that is an important indicator of rent differences for the same size dwelling. In the case of the three EU countries in Phase I, their surveys asked rents for a given specified dwelling. There were seven EU specifications in the 38 combinations of which six were comparable with the United States.

The BLS survey of market rents had a large enough set of characteristics that it was possible to do binary comparisons of the United States with the other countries to obtain PPPs for many of the 38 combinations by size, age, and available amenities.³⁷ In addition multilateral PPPs were estimated. For each cell in this ten country by 38 matrix of rents, there were also weights that were used to obtain an aggregate PPP for rented and owner occupied dwellings using a weighted CPD.³⁸ However, even if the PPPs are well estimated, they will only produce good volume estimates if the PPPs on rents and owner occupied housing along with the appropriate quantities of housing are imbedded in the reported expenditures on housing services.

One way to check on this is to compare direct quantity estimates from housing stock surveys that are typically a part of the population census, but were not available for Colombia or Kenya. However, housing stock estimates usually capture only a small part of the quality variation in the dwellings of countries and overestimated volumes for India, Japan and the United Kingdom. The bottom line is that more effort on housing in Phases I and III of the ICP was devoted to this relatively large share of consumer expenditures than has been devoted in subsequent phases, the EU-OECD comparisons excepted. Countries are able to supply rental surveys but the quality detail available in housing censuses is weak in many countries and the national accounts expenditures are often based on these housing quantities with little allowance for quality. This remains an area

³⁶ Chapter 9 of the Phase I report is devoted to dwelling rents, which includes both market rents and implicit rent of owner occupiers. Owner-occupied dwellings are typically larger in square meters than rental units and range from 20 percent to 60 or 80 percent of housing stock. Market rents are applied to owner-occupied units allowing for differences in size and amenity differences.

³⁷ Robert Gillingham (1983) had associations with Kravis at Penn and had carried out some hedonic rent estimates at the BLS assisted us with the understanding the BLS rent surveys that were used in Phase I.

³⁸ In the end we used the multilateral estimates because their differences from the binaries were not large and the number of bridges between countries in the 38 by 10 rent matrix used more information from the countries.

of the ICP where plausible PPPs can be estimated, but they are not necessarily consistent with the total expenditures on dwelling services so the volume estimates are not necessarily comparable. Experiments using just direct quantity estimates have not been successful either, usually because housing censuses capture very few quality differences. This remains a continuing issue for improving ICP.

(3) Gathering and Processing Prices and Expenditures: 1970-71

By 1970 worksheets had been developed for the expenditure classification, for education, public employment, population and exchange rates, and a specification book with codes and descriptions for consumption, capital formation, and public sector employee grades. These were essential to making clear the scope of the work required by participating countries. Specifications of items to be priced were continually being developed during the period 1970-71, with price collections being carried out. Countries had provided an initial set of consumption specifications based on the BLS specifications converted to the metric system and supplemented by early additions and deletions. Countries were consulted on the items they thought could be priced based upon their consumer price index (CPI) collections and visits to their markets.

It is fair to say that most of our contacts in statistical offices were not familiar in any detail with what was actually in their CPIs. Since a CPI is concerned with changes in prices over time, so long as the same item is priced in two periods, the needs of a CPI are met. Typically collectors know what they are pricing from month to month, but the actual item priced is not necessarily recorded at higher levels. When at all possible visits to markets were made with a collector often with staff of the Statistical Office for whom it was typically a learning experience.

A large part of visits to countries was taken up developing item lists for comparisons. A flavor of these discussions within the European Union is recalled by Donald Murphy, a former Director-General of the Central Statistical Office of Ireland, who recalled in 2000 a meeting on PPPs he attended in 1972.

“This was a three-day meeting of the ‘Prices’ working group in Luxembourg chaired by Silvio Ronchetti, who was a Statistical Office Director at the time and subsequently became Director-General. I remember staying in the exotic-sounding former El Dorado Hotel next door to the Luxair building near the railway station. The meeting was conveniently held in a conference room at the top of the Post Office building across the road. It was also the first meeting for colleagues from the UK Department of Employment (Finn Forsyth) and Statistics Denmark (John Jensen).

The purpose of the meeting was to scrutinize the comparability of the individual prices collected for over 700 consumer goods and services for each of the six original Member States to estimate purchasing power parities (PPPs). The painstaking approach adopted was quite a shock and raised worries about the physical endurance that would be needed in this new Community work environment. On the first day the prices for each item were scrutinised in laborious

detail. I recall 'strawberry jam' taking more than half a day — there were long discussions about the comparability of the quality of jam priced in different countries! This scrutiny process speeded up on the second day and then there was a mad sprint on the final day to finish all items.

In retrospect, this working party proved to be an enjoyable introduction to the European statistical system. Good friendships were formed as the group met frequently (as it still does) and members participated directly as observers in the national price surveys. Some of the national representatives at the time were Siegfried Guckes from Germany, Hugues Picard from France, Jan Vollebregt from the Netherlands, Luciana Tappi Giovannini from Italy. Richard Kuhner was the head of the relevant Statistical Office unit at the time. To ensure strict comparability, a complex system of overlapping multinational pricing teams operated at the time including the use of two chauffeur-driven Commission Mercedes for suburban pricing in each capital city — resources appeared to be more freely available in those days!

My main initial contribution to the PPP project was the addition of 'Guinness Stout' and 'Irish Whiskey' (insisting on the inclusion of the letter 'e') to the pricing list to ensure Irish representativeness! My family was young at the time and I also noticed that the list did not then include 'baby food', 'baby clothing' and 'nappies', which featured significantly in my household budget at the time!"

From De Michelis and Chantrain (2003, p.61)

I quote Murphy at length because he captures much of nature of such ICP meetings. In Europe they could bring the top country CPI personnel to a three day meeting for such a purpose, whereas in our country visits we were trying to accomplish the same thing in less time by being much less detailed in our specifications. Two of those named above were to work for the ICP. Hugues Picard was at the UNSD for Phase IV of the ICP a decade later and Finn Forsyth who, after retiring from the UK Department of Employment, visited a number of Phase IV countries for the UNSD as ICP foreign assistance by the UK government. Denmark, Ireland, and the United Kingdom had joined the European Union in 1967 to become effective in 1969 and to be implemented over the next few years.

Developing a classification system for basic headings and item prices was a major piece of infrastructure for the project. At Penn we developed software for the 700 IBM mainframes of the time, programming in FORTRAN. Inputs were punch cards, both for programs and data. Programs were submitted manually and output was put in bins anywhere from five to fifty minutes later depending on demand on the mainframe, a great improvement over the early 1960s, but a far cry from today. Each new price or correction of a previously entered price would require one punch card, with ample chance of human error. The programs CLEANSER and COMPARE were used to check the printouts of prices in US dollars at exchange rates and price ratios between countries. These checks turned up unit price errors and prices that were inputted incorrectly or that had to be referred back to the countries. We thought we were doing a good job, and maybe we

were for the time, but verification procedures now employed by the European Union, the OECD, the World Bank and the ICP Regional Coordinating Agencies have been much improved.

Our initial classification divided expenditures into 152 basic headings: 109 in consumption, 5 in government, 10 in construction, 26 in transport equipment and other equipment goods, plus one for change in inventories and one for the net foreign balance. Few countries had this kind of detail in their national accounts at the time, so a lot of consultation was needed, especially for countries estimating GDP only from the production side, like Colombia, Kenya and India.

(4) Extended Visit to India and Short Visit to Japan

I spent spring semester 1970 on a separate grant working at the CSO in New Delhi on the Indian price specifications and detailed expenditures and national accounts. For a time I worked closely with Samvit Dhar before he joined us at Penn for almost two years. It made sense for India to build up its GDP from the production side because estimates of agriculture, forestry and mining were relatively well developed. Censuses of manufacturing provided estimates of production for large scale firms. Construction expenditures were difficult for the expenditure or production side; government was easier because there were budgets at the state, center and local levels. The ICP asked for breakdowns of construction into type of structure and roads and other non-structures. These breakdowns were often difficult when the total of construction was estimated from inputs like concrete and structural steel and labor from employment or census surveys.

When GDP was obtained from the production side the total of consumption was often a residual after deducting investment and government. Breaking down this total consumption into basic headings is usually done on the basis of consumer expenditure surveys, often using the commodity flow method as a check on surveys or as alternative to expenditure surveys. India's National Sample Survey (NSS) circa 1970 had a very strong sampling frame, especially compared to other low and middle income countries. But even in India, the NSS expenditure surveys, while criticized for the length of their questionnaire, had less detailed expenditures than called for in the ICP classification. When this was true and commodity flow could not offer more detail, then countries supplied more aggregate groupings of expenditures to the ICP.

During my semester at the CSO I had my own office and typically dealt with economic statisticians in National Accounts either asking questions concerned with Indian expenditures or with facilitating my visits to other government offices. Early on I met periodically with a Mr. K.L.Geethakrishnan, Deputy Secretary of the Department of Statistics. He was in the Indian Administrative Service, the successor to the Indian Civil Service of the British Raj. Since he was not a statistician I did not have much occasion to seek his help, and in fact did not understand his interest in my activities. Finally, I learned Mr. Geethakrishnan had the responsibility of obtaining approval for me to be visiting CSO and to be undertaking the work that I had been doing for eight weeks already. And I learned that a Secretary or an Assistant Secretary were paid better and had more influence than the Director of CSO. Did I mention I am a slow learner in such matters?

Why would Mr. Geethakrishnan need approval for me to be doing ICP related work at the CSO? The issue appears to have been this. At the time the official exchange rate in India was 7.5 rupees per US dollar, while the secondary market was 10 to 15 rupees per dollar. Apparently the Finance Ministry or others were afraid that the result of Phase I of the ICP would be that the PPP would be larger than the exchange rate. Fortunately one of India's abler statisticians, Moni Mukerjee of the Indian Statistical Institute in Calcutta, explained to his colleagues at meetings of an ICP Advisory Board in New Delhi that the PPP will be less than the exchange rate, closer to 3.0 was his guess. In fact it turned out to be 2.2. Whether Mukerjee carried the day I do not know. What I was told is my file went up to the cabinet meeting chaired by Indira Gandhi, and my research work was approved, this in the last week of my stay.

Wilma and I took part in discussing specifications and going to shops. One strong memory is the degree to which our counterparts generalized their own regional preferences to all India for food items. In going over specifications it would have saved a great deal of time if the list began with transport equipment or personal care when minds were fresh, rather than with food and rice. Everyone in India and many Asian countries is an authority on the kinds of rice that are eaten, at least in their region. This carried over to many types of fruits and vegetables as well as clothing and footwear. To represent India it was essential to have staff from at least north and south India sitting in on the discussions.

For items like footwear we chose to go with Bata, a Canadian company that had shops in many outlets in India cities and also exported shoes from India to a number of countries. At the time Bata supplied footwear to Sears under a different name so their shoes were quite representative of footwear pricing, though not of quantities sold in India. We also collected cloth samples to compare with the type of materials produced by mills in the United States. As mentioned earlier, this was to build up the price of a shirt, blouse or trousers from the cost of materials and tailoring charges since the price of shirts in a store at the time were quite expensive.

During this period in New Delhi, our two children attended the Woodstock school in Mussoorie (Queen of the Hills), a hill station a few hours from Delhi at 2000 meters rising fairly steeply from the plain. The school was founded in 1854 to provide a Protestant education for girls, the founders included American missionaries. By 1970 it was coed and quite international with very fine freshly ground peanut butter, a hit with son Alex. Woodstock became a family tradition for many missionary families, with offspring often going back to India to work with non-governmental organizations and the like. The Alters were one such family some of whom I came to know at Penn.

The CSO staff arranged appointments with trade associations, importers, dealers and manufacturers to cover the private sector. Even in 1970 India manufactured a variety of appliances, pumps and machinery often under the protection of import quotas or foreign exchange restrictions. India exported a variety of relatively simple products like fans, electric motors, sewing machines and the like. There were also a number of central government enterprises like Bharat Heavy Electricals, Hindustan Petroleum and Steel Authority of India who were consulted on prices. For construction we sought the help of

the National Building Organization, the Public Works Department and several private architectural firms.

Whether in the private or public sector the responses from our contacts were very cooperative whether or not they could help. The one notable exception was the Directorate General of Technology and Trade (DGTT), a world on to itself (Bhagwati and Srinivasan, 1975). With the power to grant licenses to import technically advanced machinery and to produce specific items in India, many sought their favor to enjoy the rents created by the limited number of permits and licenses. Not surprisingly we were made to wait the better part of an hour to learn that they would not be able to supply the kind of information we wanted. Large enterprises had permanent staff visiting various offices within the DGTT, a common situation under *permit raj*. The chance that small firms would get permits was small indeed.

Most CPIs have a target group for whom prices were being collected, like urban workers in the United States. In India in 1970 there were several consumer price indexes, one for rural agricultural laborers, a second for manual workers in mining and manufacturing and the third for non-manual workers. The price survey for the rural index was run by the NSS, the price survey for the manual workers index was organized by the Department of Labour. The Labour Department was located in Simla, the former summer capital of India under the British, and currently a popular site for Bollywood movie producers, as well as a UNESCO heritage site as a narrow gauge mountain railroad. The price survey for the non-manual workers index was the responsibility of the CSO.

For manual and non-manual workers rents were also collected and I did visit two apartments in Delhi sampled by the Labour Department with their enumerators. The accommodations were minimal but adequate. Later I processed a survey of middle class rents in various cities in India and found sensible coefficients on rents and the characteristics of the rental unit like size and availability of shared or private toilet. In the end my Indian counterparts did not believe the different city coefficients that I estimated and chose to use other estimates, the source for which I know not, to obtain national average rents of India for 1970.

To get some sense of differences in type of outlet and variety of product available I visited the different types of markets surveyed for the three indexes. Both the price survey for manual workers and the price survey for non-manual workers sampled some government shops where basic grains, soap, kerosene, cooking oils and the like were available on a ration basis. These fair price shops sold grains with a fair amount of dirt, pebbles and broken kernels, meaning that the differential in price for say, rice, between the fair price shops and free markets was overstated because quality was higher in the latter. The free-market outlets were somewhat higher quality for the non-manual workers at least in New Delhi. A rural market I visited in Eastern Andhra Pradesh was an eye-opener in the sense that the variety of produce available was really minimal, grains, potatoes, onions and some pulses but little more. Items like cooking oil were dispensed from larger containers into vessels brought by the customer. Clearly many purchases were made at bigger weekly markets or in nearby small towns.

How does one arrive at a national average price for India? The NSS does provide unit values for mainly food items and studies have found that rural prices were about 10 percent lower than urban prices. And food prices vary by size of city. The Labour Bureau sampled many smaller urban areas where there might be a mine, or major industrial complex, like Jamshedpur, that were not sampled in the non-manual survey. In the end factors were estimated from price differentials between the cities over 1 million, those between 100,000 and a million, and the remaining large towns and rural areas. For most equipment goods like automobiles no adjustment would be made since they would be purchased from the larger cities.

The Institute of Developing Economies in Tokyo was interested in a direct comparison of India and Japan and invited Satish Kansal (1971) to undertake the study for consumption for the reference year 1965-66. Kansal generally used India's producers' prices adding estimates for transport and mark-ups to compare with prices from Japan's retail price collection. I had met Kansal in India before he went to Tokyo and met him again when we were in Tokyo. Wilma and I went to Japan to go over consumption specifications with their staff for comparisons with the United States as well as India. We made frequent trips to shops in Tokyo with price collectors checking out specifications. The back and forth between translation and discussion between the Japanese was by the end of the day quite exhausting. Needless to say, we were well taken care of in the evenings, when we each had our own Geisha attending us.

(5) Obtaining US Expenditures and Prices

Kravis had many contacts in the US government from his work on processing the Consumer Expenditure Survey with his colleagues in the Wharton School and his work with other agencies. While this made us welcome at the BLS and the Bureau of Economic Analysis (BEA) it did not necessarily mean that these agencies would do the work. In the case of the BEA they already published expenditures at a fairly disaggregated level so the framework was in place. The only price we paid for the data was listening to Bob Parker tell us how difficult it was to fill in the detailed ICP basic headings and the amount of extra work it meant for his staff. However, Bob was a real professional and when he said he could do it, it got done, however unenthusiastically.

With respect to US prices, there are two sources that should be mentioned by way of background. In 1967 a report on the family budgets of city workers referring to 1966 price collection for 39 urban centers was published. (BLS 1967). The publication provided prices for selected food items available from the Department of Agriculture, excluding canned, frozen or processed food (other than bread and corn flakes), a total of 27 items. In addition prices for 200 non-food items were collected and all were published for all US urban and for each city. In spring 1967, prices were used to find the cost of budgets for moderate, and lower and higher expenditure levels for families of four persons. (BLS, 1968) Basically this was the last hurrah for this type of estimate by the BLS for a generation. There was an experimental study of national average retail prices for about 200 non-food items with a reference date of fall 1971 based on collection in 39 urban centers. This was not issued as a Bulletin but distributed as a non-priced document. (BLS, 1972)

These spatial price studies by the BLS were not given priority (and consequently budget) by the Advisory Committee representing the general public, business and labor. It was said that labor unions did not really want their rank and file to know whether it was more or less expensive to live in the locality of their membership. Similarly business would prefer to give perks in a paternal way to staff they wanted to retain or newly hire in a particular location. In addition to the BLS several private firms and non-profit associations also sold or made available estimates of cost of living in various urban centers in the United States. In terms of the ICP and the BLS it meant that our source of US consumer prices should be BLS CPI files. The backing off of the BLS from spatial price statistics circa 1970 may also have influenced what the BLS felt free to supply to the ICP.

The BLS staff with whom we dealt were highly professional. Our contact was Janet Norwood who was in charge of price statistics at the time and who was to become Commissioner of the BLS from 1979-91. The BLS was not willing to estimate national average prices for all of consumption and left that to us. They remained open to collecting some additional prices (it turned out to be about 200 items) for purpose of creating coverage with other countries, but only in one or a few cities, namely Chicago, New York or Philadelphia (Los Angeles and Detroit prices were also occasionally available). Further their concerns about confidentiality were an important consideration, much more so than we had imagined.

On a regular basis we worked with Winifred Stone, Mary Lou Drake and William Berry all of whom knew the specifications very well. They were an invaluable source of information to Alicia Civitella, and both Drake and Berry carried out technical assistance missions to ICP countries. Price experts like Stone were very skeptical of the way we broadened the ICP specifications to include BLS items as well as European, Colombian or Indian items for which the BLS would have preferred to create new specifications. When the ICP would use a specification like local beer without reference to a brand it made BLS experts uneasy to say the least. Our argument was that comparing US beer brands in the United States and say, Nairobi, Kenya, where US brands sold at a premium, also presents problems.

In the end Winifred Stone supplied us price sheets with all indicators of outlet removed. Typically a price sheet covered eight outlets that we averaged in one, three or five cities depending on the item. I remember showing one of these sheets for an Alligator brand rain coat, a fairly upscale item at the time, to one of our Hungarian colleagues, who was amazed at the amount of price variation by outlet. Another surprise for our Eastern European colleagues familiar with one common price was the outlet Sam Goody's, which at the time was the largest record store in the country. LP recordings were color labeled where each color had a price like \$3.99, \$4.99 and so on.

As discussed with respect to rents, the BLS also provided us the individual responses to their rent survey in 1966-67 with about 39,000 observations. I continued to interact with the price experts after the 1978 revision of the CPI which involved a complete change in the way prices were collected and recorded.

(6) Equipment Goods

We also interacted with BLS in obtaining price quotes for equipment goods. Some information was available from the wholesale price index for equipment goods and some from the import price index. Kravis had independently been involved with moving the import price index for machinery from unit values to specification pricing. The main problem of using the value of imports divided by the number of units to obtain unit values is that often the value figures embraced a wide variety of qualities or range of capacities. Kravis recommended that BLS move towards representing the prices of important import groups with fairly tight specifications for a few items as a better method. This change was being developed during the Phase I period and provided another source of prices for producer's durables.

In the case of Kenya and Colombia, most machinery items were imported and their prices could be estimated by importers. As discussed India produced a variety of equipment goods and also estimated hedonic regressions for a number of items. However, the many licenses and permits required in India drove a significant wedge between world prices and Indian prices leading to a high PPP. Matching with Hungary was difficult because brands were different. The prices of the United Kingdom and the EU countries, often with adjustments provided from the coefficients on hedonic regressions for other countries, were modified to more closely match Hungarian specifications. Japan was finally persuaded to compare machinery prices but only after Kravis took a number of industry representatives to a very fine restaurant, the meeting orchestrated by our Japanese counterparts. At that luncheon the industrialists were assured that the ICP comparisons were not for the purpose of showing that Japan's machinery exports were being dumped on foreign markets at prices lower than in Japan.

The comparison of equipment goods was carried out in New York so I am less familiar with the details. One issue illustrates the limits of computing capacity at the time, in this case at the UNSO where FORTRAN was not supported. As a consequence the CPD program at Penn could not be used in New York. Instead the geometric mean of the price ratios of machinery items within basic headings was taken as the PPP for the heading between each pair of countries. The cost of employing this approach was to not use some third country bridges between countries with probably little effect on the overall results.

(7) Construction

Most countries have Departments of Public Works that undertake road building and some building of structures. Actual construction may be done by the private sector on the basis of bids for projects in which case there are records of costs and specifications. Costs of actual projects have the disadvantage that specifics of location within a country, special features of the specifications, and the nature of the site can greatly affect costs. Further, because actual projects typically are done over several years, their costs are out of date and updating a specific project by a general construction price index can be problematic.

An alternative approach is to ask professional bidders to estimate the costs of construction of a particular project. This was done in several countries using architects or

quantity surveyors familiar with building in a variety of settings. In general costs were sought for more specifications because it was judged that any errors would tend to offset each other. Quantity surveyors were used in the United Kingdom and Kenya. Samvit Dhar became important in this work because he had worked with building manuals in the United States. This made it possible to adjust US cost estimates to make them comparable to the submitted costs of other countries that departed from specifications.³⁹ For the United States we employed a private architect in Philadelphia to look over our shoulder regarding residential and non-residential buildings. For other construction we sought the help of agencies of the US government in estimating various specifications. One generalization emerging from the experience of Phase I is that there great variation in residential housing, whereas industrial, commercial, and institutional buildings were more homogeneous across countries. The EU countries preferred to cost thoroughly fewer more detailed specifications using architectural firms, an approach the European Union has continued through 2011.⁴⁰ They also accommodated the Global Office and supplied estimates for a few countries according to the 2005 approach.

The 2005 and 2011 ICP rounds have continued to experiment with different approaches to construction. Because of the high cost of hiring experts for many countries, an approach called *Basket of Construction Components* was used in 2005. A construction component might be the price of an installed footing that included labor, materials and the rental value of any machinery used. Some 22 components were chosen to be priced in all countries, with some idea of the weight that each component would have in a given type of construction, residential, non-residential or civil engineering. In general it was often necessary to hire construction specialists to estimate some of the components as well as the weights. In 2011 the estimates basically used an input approach with some failed attempts to consider markups and labor productivity. In the end the 2005 and 2011 results were broadly comparable largely because most of the same inputs were priced in both years. Plans for 2017 have not been finalized but it is fair to guess that heavy reliance will be placed on input prices, not a very satisfactory outcome.

³⁹ We commonly used the Boeckh Building Valuation Manual and the Chicago Building Cost Manual in Phases I-III of the ICP.

⁴⁰ Parallel with the EU approach, H.Kinston of the Julio Vargas Foundation in Brazil had proposed a similar approach for the ECIEL comparisons.

B. Experimenting with Aggregation Methods on Preliminary Data: 1972-4

(1) Below the Basic Heading Level

The CPD method was described above in the equation in A.(2) and was used in the 1970 ICP in a somewhat altered form. We chose to weight each price inversely to the number of items a country supplied for each basic heading for the reason that the United States generally tended to have more prices in every heading. Use of frequency weighted CPD we argued would make sure each country had the same influence within each basic heading. In fact comparisons of unweighted and weighted CPD revealed very little difference over the 129 basic headings for which comparisons were made.⁴¹

The discussion of CPD thus far has referred to estimation of country PPPs for a basic heading. An alternative to CPD is the method favored by the European Union, the EKS or GEKS method. In the ten country case of the 1970 ICP suppose we calculated all possible binary indexes for a basic heading, United States with Colombia, United States with India, United States with Japan.....and Colombia with United States, Colombia with Colombia (=1), Colombia with India, etc. The 10 x 10 matrix of binary indexes would be symmetrical with the entry United States/Colombia being the reciprocal of Colombia/United States. However, the entries are only transitive if the same items enter into all binaries, the case in the Latin American comparisons, otherwise [Colombia/United States / India/United States] \neq [Colombia/India]. The method used to make the matrix transitive was called *EKS* for Éltető and Köves (1964) and Szulc (1964), which is the geometric mean of all the direct and indirect Fishers for each country. The authors make clear that the method traces to Corrado Gini (1931) and more recently Gini has been given credit and the method is now referred to as *GEKS*.⁴²

In general the differences between the CPD and GEKS estimates become smaller the more entries are in the price matrix becoming zero when the matrix is full. At Penn we pushed for CPD on the grounds that it used all the prices submitted by the countries where at least two countries price the same item. GEKS by contrast might have Colombia

⁴¹ We also used double weighted CPD for basic headings for which we knew the importance to be attached to each price, most notably for house rents and automobiles. We also used the same hedonic regression results for automobiles in consumption and transport equipment in producer's durables. Double weighted CPD was also used to impute basic heading PPPs for instances where a country could supply no suitable prices for a basic heading. In this case we would perform a CPD on the PPPs for the basic headings within a larger expenditure group, for example pork might be missing for a country within the larger aggregate of meats. In this example, we would have basic heading weights for each PPP and could impute a PPP for pork for the country for which it was not available.

⁴² As usually written the GEKS calculation takes the n th root of the product of all the direct and indirect binaries involving each country. For example India would have 10 direct binaries with all the countries including India, and 10 indirect through the 9 other countries and itself. Gini (1931) in fact used a least squares regression to produce a transitive matrix and this is a useful form because allows one to introduce additional variables in the estimation (Aten and Heston, 2009).

submitting 10 of 20 prices for vegetables and India submitting 12 of 20 with let us say only 8 prices for the same item enter the Colombia-India binary, leaving 6 prices unused in the binary computation.

This became a point of difference between the approach in the European Union and that at Penn with both sides being fairly stubborn. CPD was criticized because it appeared to fill holes and estimate prices for items that were not important for the country missing prices for those items. Our defense was that CPD is used to estimate country coefficients, not to estimate missing prices of say, country A. The prices of other countries for the missing items of country A were used as a bridge in the CPD equation and are representative for those countries. To our knowledge no one has demonstrated there is any bias introduced in the estimation of the country coefficients by use of the CPD.

In the end we thought we were at the cutting edge in the use of hedonics and in using all the price data that countries provided for the country by item price matrix for each basic heading. We also felt that our statistician colleagues were slow in the use of regression techniques in the EU comparisons, especially since others in the European Union were using them. In retrospect there were a lot of misunderstandings on both sides and not enough appreciation on our side about the detailed binary comparisons that were being carried out among the European countries. What evolved in the European Union were binary comparisons between each pair of countries at the basic heading level that attempted to weight the importance of each item.⁴³ Beginning with the 2005 ICP some blending of the approaches took place in the sense that countries and regions tried to implement the distinction between representative and available products, albeit with little success. And research suggested that CPD may have some advantages over GEKS.⁴⁴

⁴³ Each EU country chooses from the list of items to be priced in each basic heading those that are *representative and* those that are *available* but not typical, with the remainder being those that are not readily available. Items that are representative in country A and are representative or available in country B would be the binary entry for country A. For B the entry would also include only its representative items with representative and available items from country B. Only if both countries had identical lists of representative items would the entries be the same. Arguments were also used that if GEKS/CPD were used to estimate basic heading PPPs then GEKS/G-K should be used to aggregate the basic headings, a doubtful symmetry in my view.

⁴⁴ See World Bank (2008). A sad aspect of life and the renaissance of CPD is that Bob Summers (1922-2012) did not really appreciate its increased use in his last years. The EU procedure is part of the EU regulations so they use their version of GEKS, which involves meetings of price experts using more resources than have been available for such detailed work in Africa, Asia, or Latin America.

(2) Aggregating the Basic Heading PPPS

The misunderstandings in the discussions of alternative ways to estimate basic heading PPPs were small compared to the heat involved arguments over methods of aggregation above the basic heading level. First, the GEKS procedure can be used to aggregate above the basic heading. In this case the Paasche and Laspeyeres indexes are computed for each pair of countries using their expenditure weights and the Fisher index is calculated. A GEKS procedure is applied to the matrix of Fisher indexes to obtain an aggregate PPP for say GDP. Smaller aggregates like consumption, investment and the like can each be obtained from separate GEKS calculations. The sum of sub aggregates will not equal the total whether it be for GDP or consumption. This non-additive character of GEKS is discussed further below.⁴⁵

At the time we were considering multilateral aggregation methods there were three other techniques that were known to us from practical applications: (1) the Van Yzeren method that was used to compare consumption levels in the ECSC prior to formation of the European Union; (2) the Walsh method that averaged the expenditure shares of all the countries that were used as the basic heading weights in the comparison, the approach used in Latin America; and (3) and the binary star method used by the CMEA and the OECD.⁴⁶ A method suggested by R.G. Geary (1958) was used in an application by the FAO who were interested in an aggregate measure of purchasing power over agricultural output in different countries. S. H. Khamis (1967) proved the existence of a solution to the Geary system and Khamis (1972) and Prasada Rao (1971) established the necessary and sufficient conditions for the existence of a solution. Khamis and Rao both spent time at FAO where both prices and quantities were available to make applications of the Geary system. In the ICP case most of the basic heading quantities are indirect, obtained by dividing expenditures by PPPs. These indirect quantities are also called *nominal* quantities. Summers adapted the Geary system to the case where only PPPs and expenditures were available, a very important contribution to the Phase I report and ICP methodology.

The Geary-Khamis (G-K) system estimates an international price that is the average across the countries of the ratio for each country of the ppp_{ij}/PPP_j for each basic heading. The indirect quantities for each basic heading are valued at these international prices. The indirect quantity is the expenditure for a basic heading, E_{ij} divided by the PPP_j . In the usual G-K, the indirect quantities are used as the weights in estimating the international prices, meaning that larger and richer countries have more influence on the

⁴⁵ As part of his dissertation at the Indian Statistical Institute Prasada Rao (1972) demonstrated that CPD weighted by expenditure shares was another possible aggregation method that was very similar to GEKS in terms of results, and was also non-additive.

⁴⁶ We reported two variations on the star method, one using as inputs the basic heading parities from a binary comparison with the United States with each country. The second used the basic heading parities resulting from the CPD that introduced prices from other countries into the calculations.

average.⁴⁷ Since PPP_j is what all the aggregation methods seek to estimate, it has also to be estimated. In the G-K system it is the total of the national currency aggregate GDP divided by the total when valued at international prices. So the international prices for each basic heading and the PPP of GDP for each country need to be determined simultaneously. A program was developed called METHODS because it calculated results for GEKS, Van Yzeren, Walsh, two binary variants with the United States, and exchange rates as well as the G-K. METHODS would eventually provide output of the multilateral tables at the detailed 152 basic heading level and at the 36 summary heading level.

(2) To Obtain More Support You Need to Show Results

By the end of 1971 preliminary expenditure distributions were available for all the countries, consumption prices had been submitted and processed. The EU countries had pretty much completed all of their surveys, while construction, equipment goods and government remained to be completed for half the remaining countries. The European Union planned to complete a comparison for all their countries including Belgium (Luxembourg included for this purpose) and the Netherlands, about which more below. The group at Penn had support through to the end of 1972 even though about eight additional countries had been invited to participate in Phase II of the ICP with 1973 as reference year. Drafting of a report on Phase I had begun without a firm commitment on whether the World Bank, the UNSD or perhaps the University of Pennsylvania Press would publish the book. In short, the Penn participation and the ICP itself could end in 1972, or further support could be generated to maintain the group at Penn, or a unit could be supported in the World Bank, the expressed preference of Kravis.

The major contact at the World Bank after 1971 was John Edelman who was at the OECD with Kravis and who was one of the associates with Beckerman in Gilbert and Associates (1958). Hollis Chenery was Chief Economist at the time and was supportive of the ICP though any funding would go through the research committee where there were competing projects. Funding from other countries, USAID, and other potential donors was undertaken by Edelman and staff. During the 1972 Kravis spoke a number of times with Edelman. In a January conversation Irv said we would have all the Phase I data by June 1972. In February Irv said we could provide illustrative results by April for food, beverages and tobacco that would involve enough summary categories to give an impression of the effect of different methods of aggregation. Edelman made it clear that the World Bank would need to see some results before it could make decisions on the future of ICP support in the 1972-73 fiscal year beginning in June. It was agreed to hold a meeting in April, which would allow the World Bank enough time to reach a decision on whether to create an ICP unit in house and/or the level of support that they could provide, if any, to Penn after the end of the fiscal year.

⁴⁷ The G-K system could use other weights, for example, the share of each country in the nominal quantities for each basic heading, as illustrated by Rao (1972).

Early 1972 was a very busy time getting ready for the April meeting. The burden fell mostly on Penn because we were doing the computations on consumption data, the late night duty falling on Sultan Ahmad and Alicia Civitella. In Phase I we used brute force to solve the G-K system inverting a large, for the time, matrix, much to the sleep deprivation of Sultan who wrote the program, METHODS, and did the calculations and Alicia who helped and also brought the sandwiches and orange juice.⁴⁸ As the meeting drew nearer, Kravis became nervous and saw to it that we could get access to the extra memory all day and night on the Friday a week before the meeting. Results came forth at 10 pm that Friday night and Kravis drove in from the suburbs to have them for the weekend so we could decide by Monday if the meeting would have to be postponed. Happily the results justified a meeting and the preliminary results were made available to the World Bank by special delivery three days prior to the April 21 meeting in Washington. Kravis had suggested that a number of consultants be invited but in the end only Nancy Ruggles attended as a consultant. Peter de Janosi of the Ford Foundation was an observer, nine World Bank staff were there for all or part of the meeting. Kenessey and the three of us from Penn rounded out those attending.

There were many comments, much of it focused on the different methods and the perceived merits of one or the other. The one feature of the GK system that seemed to appeal to the World Bank attendees was that the results were additive across rows and down columns. Nancy thought that the additive feature of the G-K was a bit of a red herring because other methods could achieve additivity including the Walsh approach, which was her preference. There is a price to pay for additivity, namely the method does not allow substitution as the international prices change because the nominal quantities in the G-K system are fixed. In Diewert's term, the G-K is not a *superlative* index, nor is GEKS, but the latter is constructed from superlative indexes/. At the end of this long meeting Nancy Ruggles in effect said the important thing is not which method of aggregation is used, but rather that the price collection and expenditure distributions were actually completed, permitting substantive discussions of results. I still treasure her comment.

A World Bank decision reported at the meeting is that there would be no ICP unit in the World Bank in 1972-73 and probably not the following year either. In terms of Penn support the World Bank would welcome a proposal to the Research Committee and would seek other support to maintain the same level of our staff until the end of the 1973-4 fiscal year. In the end this support was mobilized. It was assumed that Penn involvement would end in 1974 at which time the Phase I report would be published, the Phase II countries would have submitted their data, and the framework for the Phase III 1985 benchmark would have been established.

⁴⁸ The computing capacity at the time had 256k which was not enough to solve the size of the G-K equations with ten countries and 153 headings. However, after midnight Sultan could obtain an additional 256k which was enough memory to solve the system. Later it was agreed that the G-K was much easier to solve using an iterative approach that Frank Orlando developed. It took less storage and generally converged in under ten iterations.

(3) A Digression on *Men of Affairs*

Kravis would often justify the G-K as our preferred aggregation method on the grounds that *Men of Affairs* wanted results that add up. At that time *Men of Affairs* were pillars of the community, business and government, not, as more recently, well-known people whose behavior the media found to be sufficiently below accepted norms to deserve publicity. Bob (Larry Summers was still in high school) and I did not know any men of affairs at the time but Irv did, so we accepted his judgment. Our second argument for G-K was that it was in the spirit of national accounts that within a country would add up production of each geographic entity to arrive at a total. Similarly we would add the production of each country involved in a G-K to obtain world GDP. So in comparing country GDPs why shouldn't each country have a weight according to its economic size? This logic also led us to use *super-country* weights so that our ten countries could in fact represent the world GDP. In contrast the GEKS or related aggregation methods would give the same importance to Luxembourg as Germany. An appeal of equal weighted indexes is that they capture what a typical person in a country faces in terms of relative prices, which for many purposes like post adjustment, is useful.

Another way of highlighting the above difference is to imagine that the GDP of country X in a benchmark comparison were twice its size. Would that affect the results? For GEKS, Van Yzeren, Walsh and similar methods the answer is *No*. But for G-K the answer is *Yes* because now country X would have a greater weight in determining international prices. And by tilting international prices towards those of country X, the GDPs of other countries would also change. At the time we thought this was a more important property of an aggregation method than a method that was invariant to the size of each country.

The most frequent criticism of our use of the G-K was that the *Gerschenkron Effect* led to the overstatement of the incomes of poor countries. Gerschenkron had shown that using pre-WWI weights to value the growth of industrial production in the Soviet Union greatly overstated the growth compared to using pre-WWII weights because their rapid industrial growth was weighted by the high relative prices of industrial goods circa 1914. The situation in comparing countries is analogous. Valuing a poor country where services are cheaper at the prices of a rich country will make the poor country look richer and vice versa. In Phase I in particular, G-K international prices even with super-country weights, were tilted towards relative prices in the United States. This would tend to raise the incomes of countries like Kenya and India. This is a legitimate criticism and readers were provided with the results of other methods, exposing the dirty linen so to speak. Again this was the price of additivity, but the cost has continually declined as the ICP has included more countries. For example average international prices were closer to Italy in Phase II and to Greece in Phase III to the point that in 2011 if anything international prices are tilted towards poorer countries, which would lead to overstatement of incomes of the high income countries. Further additivity can be achieved by using different weights to

obtain the international prices in the G-K system, for example each country having the same weight.⁴⁹

(4) Organization of Phase II and Completion of Phase I: SOEC

SOEC or, as it now commonly called, *Eurostat*, was initially spread between offices in Brussels and Luxembourg from 1953 to 1967, at which time negotiations began for the United Kingdom, Ireland, Norway and Denmark to join the original six EU members. Also in 1967 the three Communities - Agriculture, Atomic Energy, and Coal and Steel - merged and Eurostat was relocated to Luxembourg. Vittorio Paretti was a strong force as an administrator of Directorate A in Eurostat, within which Hugo Krijnse-Locker ran the section on Intersectoral Relations that included price surveys. While most of Eurostat did move to Luxembourg, some sections of Paretti's staff remained in Brussels because they provided direct support to Finance and other departments of the European Commission.⁵⁰ In Luxembourg staff were scattered in different buildings including the Staar Hotel that had been partially bombed but still had a large meeting room and several floors of offices.

In addition to Krijnse-Locker and Paretti another important contact during Phase I was Phillippe Goybet. One of the early decisions facing Paretti was that all six EU members had in fact participated in Phase I in anticipation of a separate EU report. Both Belgium and the Netherlands were ahead of some other Phase I countries in terms of completed surveys and could have been included in Phase I. However, Paretti had two reservations. First, it would take away from the separate EU publication if all the countries were also in a UN or other publication. (The status of Luxembourg, which shared currency with Belgium except for color and country of issue, is unclear at this point, since it did not participate, in the ICP until the 1980 comparison.) The second reservation was that Eurostat staff thought the Netherlands prices were too low. In any event Belgium and the Netherlands did participate in the 1973 comparisons with the same Dutch prices.

The Directorate Generals of the Institutes of National Statistics (DGINS) began to hold annual meetings with Eurostat staff. In 1972 the Dutch Presidency submitted a document to the DGINS that anticipated what Eurostat would codify 25 years later. Namely that participation in the surveys would have a binding legal basis; the maintenance of secrecy of confidential data; a requirement that the European Commission present a three year program every year; and decisions on budgets should be separate from decisions on programs. While these were general principles for all

⁴⁹ Another additive system developed by Doris Ikle and rendered more understandable by Yuri Dikhanov of the World Bank was used in the regional comparison of Africa in 2005. This system estimates international prices as the harmonic mean of the ratio of ppp_{ij} / PPP_j .

⁵⁰ Political considerations also weighed in split of location of Paretti's sections, namely there was fear that Brussels would form a separate statistical service. Paretti, himself, bought an 11th century fixer-upper castle on the Belgium border in the village of Septfontaines allowing him relative easy access to Brussels or Luxembourg. We had the pleasure of staying in the Chateau de Septfontaines when it had been restored and it was spectacular, with a turret for each of their two children and medieval furnishings to complement the main rooms of the structure.

statistical programs they became central to the PPP program. (De Michelis and Chantrain 2003, p.66)

(5) The Rest of the ICP World

The World Bank and UNSD jointly determined the list of whom the UNSD should invite to participate in the 1973 comparisons. The thinking was that about eight to ten countries were to be invited to yield four to six actual participants. Countries like Morocco, Ghana and Ivory Coast were considered but not actually invited for reasons like statistical capacity or non-existence of national accounts. Brazil and Venezuela were contacted in Latin America as well as Costa Rico in Central America. In the end talks and visits with Venezuela over two years ended with some price submissions but in the end no participation. Iran was invited and eventually came on board, while no Arabic speaking country was invited, mainly because of concerns over national accounts and prices. At that time, Egypt and many other Arab countries required high level approval to make prices collected for the CPI available outside their offices. This was not for confidentiality but rather for political concerns about which I can offer no insight. Singapore was asked and declined, but Korea, Malaysia, the Philippines accepted bringing the total to 16 with Belgium and the Netherlands. The plan was the additional six countries would collect prices with a 1970 or 1973 reference year and back cast basic heading parities to 1970. The original ten countries would update their prices to 1973, typically by using the CPI price movement for 36 summary headings of expenditures.

Drafts of the Phase I report were written and distributed during 1973-74 as prices and other worksheet data were submitted on government and capital formation. When we had the full data set the various binary comparisons with the United States were being finalized, subject to possible revisions of the control totals for GDP and its main aggregates. We also began trying different weights, such as per capita or giving each country the same weight in the G-K calculation, both without super-countries. The per capita weights were more subject to the Gerschenkron effect than the national accounts weighting. We had been using the United States as the reference country in all our multilateral calculations knowing that the results should be base-country invariant. To illustrate this we used Germany as the reference country in one of the runs in early 1974 just before we were going to send out drafts of the report to the Advisory Board and others. This run was carried out in what Bob termed *Black February*.

The METHODS program showed a summary sheet at the start of a very large printout giving the per capita GDP of the countries with the United States = 100 using the G-K. Kravis and Summers looked at the summary result when Germany was the reference, and not only was it different than when the United States was the reference, but the results looked very unstable. They asked to run it again and with other countries and in no case could they get the same result. We were to include the G-K results in the draft we were distributing the following week and Kravis was upset enough to simply go with the binary results. Summers found that unacceptable. Finally I looked at the detailed international prices and the problem immediately became clear, the net trade balance basic heading had a very large negative international price when Germany was the reference country.

Our mistake was to forget that for the G-K to converge, you could have negative values but to guarantee convergence all values had to be zero or greater. Our fix on this was to run the G-K on the non-negative headings. The PPP from that run was used to value the net foreign balance and net expenditures of residents abroad. Change in inventories was rolled into other investment where any negative entries would be offset by the size of other investment. This treatment did produce results that were reference country independent and allowed us to produce multilateral tables into the draft report that was distributed as promised. Further, we came to the conclusion that this was the better way to value the net foreign balance and net expenditures of residents abroad than to use the exchange rate, about which not all are in agreement. We ended Black February in a better mood with a hint of spring in the air.

One feature of the draft report that provided a very rough check on the results was to extrapolate the binary results from Gilbert and Kravis for 1950 by the growth rate of GDP per capita for France, Germany, Italy and the United Kingdom to 1970. When compared with the benchmark binary results for 1970, the extrapolations understated the benchmark 1970 by 8 or 9 percent for all but Italy where the understatement was 17 percent.

There was no precedent at the World Bank for publishing an ICP type of study, but after seeing the draft, they were willing to undertake its publication. Although finished in mid-1974, the book only came out in 1975 because it was still an era when tables had to be typeset and proofread, a time consuming process in a number intensive publication. We were quite involved with completing the item lists and expenditure distributions for the four new countries participating in Phase II, so were not concerned with publication delays. In trying to move the ICP away from Penn, one argument Kravis would use is that dissemination of the work by a university is inherently slower than at international institutions.

(6) A Digression on the 1967 Comparison

Binary comparisons of prices were carried out between the United States and Hungary, India, Japan, Kenya and the United Kingdom for 1967. The European Union chose not to take part in the 1967 exercise and Colombia did not have the necessary data. India initially said they would not participate but eventually supplied the data. These comparisons were sent to the countries as sort of a first look at whether the results seemed sensible. I do not remember nor can I find documentation on whether the binary comparisons were as detailed as 1970. The results are compared with the 1970 binaries in Table 1 below.

Table 1 Binary Quantity Indexes PC and Paasche/Laspeyres Spreads, 1967- 70

Country	Binary Fishers		Paasche/Laspeyres Spreads	
	1967	1970	1967	1970
Kenya	5.7	5.9	2.10	2.15
India	5.9	6.1	1.86	1.98
Hungary	35.0	40.0	1.59	1.48
Japan	46.0	61.0	1.32	1.25
United Kingdom	61.0	63.0	1.39	1.18
United States	100.0	100.0	1.00	1.00

The results are fairly close between 1967 and 1970, not too surprising given that some extrapolations were probably (a guess) carried out for prices and expenditure details. The fact that all the countries grew relative to the United States is likely as the US real per capita growth rate was under 2 percent for these three years. The Paasche/Laspeyres spreads, the ratio of own versus US expenditure shares, behave as expected. My view is that more attention should have been given to the 1967 work, but we were rushed as it was and the multilateral comparisons were the novelty of the report.

Chapter 3: Responses to the Phase I Report

The Phase I report explicitly warned readers of the inherent limitations on the type of international comparisons that we were reporting. It is only after hearing the views of others that other shortcomings of the final product become apparent. Some of the criticisms of the ICP even 40 years later do not take account of the fact that the results can be no better than the underlying national accounts of the countries and the quality of the price comparisons. We did our best on guiding countries on the matching of items and undertaking price surveys. However, we were bound to accept the national accounts that countries submitted to the UNSD, even if they often seemed improbable compared with other countries. We could and did raise questions with the countries but in the end they had the last say on their national accounts. So our results might be questionable in some areas but that was not necessarily an ICP problem but are often the fault of the national accounts which in turn pose problems for any other type of economic comparison between countries.

It should be mentioned that Hungary reacted to the initial 1973 draft of Phase I with surprise, in particular that the results made them look as affluent as it did. Lengthy communications ensued between Budapest, New York and Philadelphia. We had no prior view on where Hungary sat on the scale of the ten countries and I suppose that the Hungarian statisticians did. The exchange of correspondence clarified many of the differences and one point in particular is worth stressing. Hungary called attention to some of the summary headings like personal care that looked unusually large. We pointed out that often the expenditure surveys had substantial errors because “not otherwise classified” expenditures were often lumped into the wrong categories. This is a general problem with the detail of expenditures sought in the ICP. When countries do not actually have the detailed data, it can show up as a large residual under “other expenditures” or be assigned to a heading that may or may not be appropriate leading to apparent anomalies.

C. Response of Reviewers, Scholars and the Media

(1) Book Reviews

The Phase I report was not an easy read for any reviewer who was not already immersed in the subject or even someone coming back to it like me.⁵¹ One of the first reviews was by Peter Hill, then at the University of East Anglia. Hill noted in his *Economic Journal* review that the ICP project “is perhaps the largest ever undertaken in the history of economics.” (p.161). Hill went on to the OECD as head the National Accounts and Statistics Division in the Department of Economics and Statistics and become pivotal in having the OECD take part in the 1980 ICP benchmark. Under the stewardship of Hill, EU and OECD participation became very important for subsequent ICP rounds. And after

⁵⁰ D. J. Daly in his review in the *Canadian Journal of Economics* wrote, ‘The methods are complex and there are about 125 tables of results.’ He went on to say the book was important and suggested the general reader may only want to read chapters 1, 13, and 15.

he retired from OECD Hill became an active consultant to the World Bank writing a number of chapters in the manual developed for the 2005 ICP.

Perhaps the most searching comment on the Phase I report came in the form of a personal communication before publication to Kravis from Colin Clark, who had read closely the preliminary draft. Clark was very supportive of the ICP in general but quite critical of our handling of non-priced services. The gist of Clark's argument was that he doubted our estimates for services in our separate aggregations of PPPs for services and for commodities. Priced services are in personal care, hotels, restaurants, repairs and the like. Non-priced services are concentrated in education, government and to some extent health, especially hospitals. During the methodological discussions for Phase I it was decided to consider both direct and indirect estimates of volumes for non-priced service. Direct quantities might be hospital bed days, or number of teachers. Indirect volumes might be derived by dividing expenditures on educational staff by relative wages of teachers in each country. We considered educational staff by level of education of teachers at different levels of teaching to narrow the range of human capital.

But in the end we made the equal productivity assumption at whatever degree of disaggregation we could obtain data. For example, we broke down the government expenditures on staff into unskilled blue collar, skilled white collar staff, and higher level civil servants and made wage comparisons that included all benefits. Salary comparisons then became the PPPs, which basically assumed that productivity was the same for a truck driver for example, no matter what type of truck he drove. Two decades later a comparison of government clerks would assume the same productivity whether or not the clerk had a PC. Few believed this at the time but there was little basis for making adjustments by country.

(2) The Media

The book was covered in the *New York Times* by Leonard Silk, in his column in *Newsweek* by Paul Samuelson, entitled 'US; Still the Richest', in the *Economist*, in *Commerce in India* and other publications. The two points that caught the attention of most writers was that the United States was still number one, and that the gap between rich and poor had apparently shrunk. The reason that the position of the United States was emphasized was that Japan, West Germany, and other European countries had been experiencing an "economic miracle" from 1950-1970. In contrast the growth of US GDP per capita over the same period was only half that of countries experiencing the rapid growth associated with reconstruction. So when journalists found a study demonstrating that the United States was number 1 it was a good antidote to the news on national growth.

Often Sweden and Switzerland were mentioned as countries that were edging very close to the United States in per capita GDP but they did not participate in the ICP until 1980. When talking about the results of Phase I Summers would say that the cost of our ten indexes of per capita GDP was \$100,000 each, actually more since the United States was a 100 by definition. He would then go on to explain all the other details behind the ten numbers and why that was such a rich set of source materials for economic analysis.

Summers also tried to exploit the overhead represented by the ten countries and extend the number of countries by developing a simple estimating equation based upon the relationship between PPP and exchange rate values of GDP per capita with the help of Sultan Ahmad. A draft of this paper was presented at the Econometric Society meetings in 1974. In this paper both Sweden and Switzerland were well below the United States. When this paper was eventually published (Kravis, Summers, and Heston, 1978a) Sultan Ahmad's assistance was acknowledged but not as an author, which I do not think was fair...the price of being a graduate student.

A second point often made in the media was that on a per capita GDP basis countries like India and Kenya were 3.00 and 2.04 percent of the United States at exchange rates but 7.1 and 5.7 percent when converted at PPPs. In fact all the nine countries rose relative to the United States. This coincided with the accepted view of the media and the traveler that circa 1970 most countries were cheaper than the United States and so their real income was higher than at exchange rates. However, we were frequently asked, are you saying people in Kenya and India are better off than we thought? At which point we would say no, there has been no overnight change in infant mortality, malnutrition, infectious diseases, or illiteracy. Rather PPP based comparisons are an alternative metric that we believe gives more reliable measures of relative quantities enjoyed across countries.

Our arguments were not totally convincing, not least to international institutions that often received funding because the differences in incomes per capita across countries is larger at exchange rates than at PPPs. But between 1970 and 1975 the international financial system went through the unpegging of the dollar to gold by the United States and subsequent adjustments of exchange rates by many countries. So of the 34 countries in the 1975 ICP six in Europe had become more expensive than the United States, quite different than Phase I. As exchange rates began to move freely so too did their GDP conversions from year to year whereas conversions at PPPs were changed very little. To insulate their estimates the World Bank used an average of exchange rates over several years and for difficult countries where the appropriate exchange rate was not clear, country desks were asked to provide a rate. This fix was sensible enough but it only made clearer that the more appropriate way to compare volumes across countries was to use PPPs to make conversions of GDP totals or per capita.

D. Response of the International Organizations

To quote Mohandas Gandhi, 'First they ignore you, then they laugh at you then they fight you, then you win.'

I had a coffee mug with this quotation that always consoled me after UN or World Bank meetings typically ended with retention of the status quo. For example, the Committee on Contributions at the United Nations began considering whether the ICP results could improve their methodology, which was framed in 1950. That methodology used exchange rates to convert GDP per capita numbers for most countries and then used a sliding scale to determine assessments. Very low income countries in unusual

circumstances, like a civil conflict, received special treatment. Soon after Phase I was completed the UNSD was asked to write papers on how PPP results could be used by the Committee. Over the next 20 years I, among others, wrote papers for the UNSD for the Committee to consider at their annual meeting. We pointed out that if PPP conversions were used the assessment scale would need to be reworked to obtain the desired total of contributions. The reason is that with the United States as reference, the total GDP of the members at PPPs would be larger than at exchange rates. However, the real issue was that the relationship between exchange rates and PPPs varies across countries so that some countries would find their assessment rising and others falling. Once the exchange rate system was in place, any change became a political issue. In looking at the decisions of the 75th meeting of the Committee in June, 2015 no lip service is paid to PPPs and only exchange rates are considered for purposes of country contributions to the budget of the United Nations.

(1) The World Bank

In the 1970s the World Bank was making concessionary loans to member countries based upon their per capita GDP group. The per capita GDP was based on *the Atlas method* which was essentially national currency GDP divided by the exchange rate estimated by World Bank staff. The DDG carried out a study examining whether there were countries changing positions in the per capita GDP ladder or the lending group. In practice country rankings changed a fair amount but not so for country lending groups. Thus nothing in the exchange rate compared to PPP conversion differences suggested a policy need to change their lending metric.

While supporting the ICP in words and resources there was little support for changing the Atlas method, despite several meetings devoted to the subject in the 1975-85 period. Two reasons always mentioned for reluctance to use the ICP numbers were that the results were not timely and that country coverage was spotty among those countries that were eligible for concessionary loans. These were good enough reasons, though my memories are that the discussions tended to have a tone that the PPPs were inadequate. Put another way, I do not recall any meeting that concluded PPPs were a better measure of volumes than exchange rates. By maintaining the World Bank Atlas link to both exchange rates and lending criteria, we felt the DDG was encouraging use of the Atlas for substantive research.

At Penn we initially pushed the value of PPP over exchange rate conversions if you wanted to compare volumes between countries. However, we emphasized that for a number of purposes exchange rates were preferable, like measures of international financial obligations. And we came to the conclusion that it was better if ICP results were not tied to the terms of concessionary loans or to contributions to an international organization as with the United Nations. We did not want countries to feel there was a financial consequence to the prices that were submitted to the ICP that would lead them, for example, to choose high end outlets so their PPPs would be higher and their incomes lower. The goal of such manipulation would be to end up in a lower income group for purposes of borrowing or contributions. This has happened over the history of the ICP and one or two glaring instances will be mentioned later on. (Pic 12)

The ICP contact in the World Bank is now called the *Development Data Group* (DDG) under the Vice President for Research, typically an academic economist on leave for two or more years. The many persons holding this position have supported the ICP, but it has not been a priority even when Larry Summers held the position. Most organizations as data intensive as the World Bank have a Chief Statistician, but that has not been the case under the Vice President for Research or the DDG, despite recommendations to create such a post. The Director of the DGD has usually been a career employee from inside the World Bank-IMF group, an exception is its present director, Haishan Fu, who was at the Human Development Report and most recently ESCAP as Director of Statistics.

(2) The European Union

Any response of the European Union was through the individuals with whom we worked and their response was favorable but I think all were surprised by the comprehensiveness of the Phase I report. It was a much longer report than typical of governments and certainly more academic in terms of providing detailed descriptions of methods and of introducing new types of analysis. For example the introduction of similarity indexes of prices and quantities between all pairs of countries and relating the indexes to income differences was not typically the stuff of statistics departments of international organizations.

Our differences remained with respect to use of GEKS versus CPD at the basic heading level, and over use of country weights in the G-K approach versus an unweighted system like GEKS or Girardi. In their report on the EU comparisons for 1970, Paretti, Krijnse-Locker and Goybet (1974) introduced a European currency unit that was an average of five countries. This practice has many advantages that were not obvious to us at the time. In addition to *Men of Affairs*, the World Bank and the United Nations and others were so committed to the use of the US dollar in their budgets and their mindset that alternatives were not on the table.

To illustrate the point we compare the results for 1970 and 1975 with (a) the United States as the reference country, namely rows 2-4 in Table 1 below with (b) rows 5-7 in the Table where the reference is the geometric mean of eight countries. The relative position of the countries remains the same whatever the reference country(s). However, row 4 where the United States is the reference all countries are growing relative to the United States. Where as in row 7 the major event is the decline in the United States relative to all eight countries in Table 1. Again either representation is equivalent but I think the OECD has made the right decision to present both versions, one using the United States as reference country and one using an OECD average.

Table 1: Using a Single Reference Country or an Average

Country/ Row/ Year	United States	France	West Germany	Japan	United Kingdom	Italy	Hungary	Columbia
(2) 1970	100.0	75.0	74.7	61.5	60.3	45.8	40.3	15.9
(3) 1975	100.0	81.9	83.0	68.4	63.9	53.8	49.6	22.4
(4) 1975/70 %	100.0	109.2	111.1	111.2	106.0	117.5	123.1	140.9
(5) 1970	189.2	141.9	141.3	116.3	114.1	86.6	76.2	30.1
(6) 1975	165.5	135.6	137.4	113.2	105.8	89.0	82.1	37.1
(7) 1975/70 %	87.5	95.5	97.2	97.3	92.7	102.8	107.7	123.3

(3) The International Monetary Fund

We had little contact with the IMF Department of Statistics during the first 20 years of the ICP. IMF sent representatives to meetings at the World Bank on the ICP and one or two of their staff became directors of DDG. When their representatives spoke it was generally to argue for the relative version of the law of one price and generally against their need to know the absolute levels of PPPs for their analytic work. Their assumption was that the starting point for changes in PPPs was a long run equilibrium. Kravis taught and carried out much of his research in international trade and came to the conclusion that PPPs could deviate from their long run equilibrium and *pari passu* from exchange rates.

This led Kravis to choose *exchange rate deviation index* as the term for exchange rate / PPP that was used in the three reports that were drafted at Penn. In retrospect this was not propitious choice of terms for many, especially the IMF, because it could easily be taken as meaning there was something wrong with exchange rates. When the exchange rate deviation index was plotted against per capita GDP the relationship was downward sloping and convex. Users did not readily adopt exchange rate deviation index into their discourse, but its inverse, PPP / exchange rate in percent, termed the *national price level* or just *price level* was quickly adopted. We also provided the price level in the report but gave it less emphasis. The notion of a price level resonates with travelers who often speak of countries as being cheap or expensive with respect to say the United States. Most textbook treatments now use graphs showing that price levels rise with per capita GDP.

In Phase 1 the price level ranged from 29 percent to 100 percent of the United States. This was not surprising but it led us to think this might be the usual result with the highest per capita income country having the highest price level. With hindsight this was probably a naïve view given the demise of the Bretton Woods system between 1968 and 1973. Prices were rising rapidly in the United States in the early 70s and the dollar was

not tied to gold after August 1971, both sources of disruption. As countries were deciding whether or not to float their currencies, stability of price levels with respect to the United States was hardly a sure thing. Even by the Phase II comparison for 1973 France and Germany with lower per capita GDPs had price levels above 100 percent of the United States.

In spring 1980 the new Director of the UNSD, Svein Nordbotten, had drafted a request for funds to support the ICP operations in New York to J. de Larosiere, the Managing Director of the IMF. The request was signed by J. Ripert then Under-Secretary-General for International Economic and Social Affairs. After polite comments on the importance of the ICP work, the response dated May 19, 1980 from Mr. Larosiere wrote, "I am advised that it is unlikely that in the Fund, we would make more than marginal and sporadic use of the data in our current work for many years". The letter appropriately pointed out that the ICP results had a publication lag and there were no plans for annual estimates, but we still took the Fund response, rightly or wrongly, as unnecessarily negative.

(4) The Regional Banks and Economic Commissions of the UN System

The groups discussed in this section did not really respond to the Phase I report but are still relevant to the ICP because they were a potential source of technical support, if not immediately, possibly in the future. They often sent consultants to countries that were ICP participants and where feasible we would try to take advantage of their expertise.

The African Development Bank (AfDB) and Asian Development Bank (ADB) were usually consulted with respect to which countries in their regions would be good candidates to participate in the ICP. Their main activities in the 1970s were to administrate projects in member countries and to the extent the technical assistance to their countries concerned national accounts and price statistics there was room for cooperation with the ICP. The AfDB and ADB were formed in 1964 and 1966 and so were more likely to follow the World Bank with regard to projects like the ICP than to make any independent judgments.

The Regional Economic and Social Commissions of Asia (ESCAP), Europe (ECE) and Latin America (ECLAC) were older being established around 1947 (Africa and Western Asia were much later). We have mentioned the ECLAC studies covering the major countries in the region before the ICP. Likewise, ECE undertook a number of PPP studies, usually binary comparisons often between a western and eastern member like France and Poland. The ECE also undertook production side PPP comparisons. ESCAP in Bangkok had not taken part in any PPP comparisons before its members began participating in the ICP. We worked less with the ECE than ECLAC, partly because our European contacts were heavily involved with the ECE. The Ruggles worked with ECLAC and with Jorge Salazar at the Brookings Institute on Latin American multilateral comparisons. These factors plus proximity led us to be more involved with the work of ECLAC.

Chapter 4: Completion of the Phase II Report

The Phase II report did not include Zoltan Kenessey as an author because he left the UNSD during 1975-6. The transition included a number of months in which he was not in the office though officially with the United Nations until he took the position of Senior Economist with the Board of Governors of the Federal Reserve System in 1976. There he worked on capacity utilization measures and publication of the Industrial Production Index. He retired from the Federal Reserve in 1991 and became Director of the International Statistical Institute in The Hague until 1997, a year before his death.

His replacement at the UNSD was William Murray a career international civil servant with no special knowledge of international comparisons or PPPs. When I took leave from Penn in 1978-79 to work at the UNSD, Bill became my immediate supervisor and valuable source of information on how to get things done in a large bureaucracy. Murray was involved with how the Phase II report would be distributed and felt that it did not justify a separate book as Kravis planned. Murray thought that adding six to the original ten countries, requiring back and forward casting of benchmarks was not much of an addition. Kravis argued that Phase II would maintain the momentum of the ICP and permit comparisons of two benchmark comparisons over time. As was often the case, Kravis won. In retrospect Murray may have been right, in any event my treatment of Phase II will be fairly brief.

A. Measures of Imprecision

Most users of the ICP focus on the index of per capita GDP with the United States = 100. We reported this index with three significant digits knowing that that the results were not free of errors. The sources of error include the expenditures, the prices and the aggregation method. In our attempt to get a handle on errors in the country PPP estimates we used a Monte Carlo type approach based on CPD residual error for the 128 basic headings where PPPs were directly estimated. The basic heading PPPs were perturbed 299 times and after each round a new G-K aggregation was estimated and error estimates were generated from this distribution.⁵² The mean of this distribution was very close to original G-K estimate. The confidence intervals at the 95 percent level were often substantial, for example, Kenya's estimate in 1973 was 6.33 percent of the United States per capita with lower and upper limits of 5.73 and 6.93 percent.

These estimates are over only 128 basic headings where CPDs were estimated, not 152, so the limits may be larger on this account. And more importantly, the confidence intervals suppose no errors in the expenditures, so the estimates provide only the minimum size of the interval for each country. The lowest income group tends to have wider confidence intervals in percentage terms but the relationship with income per capita is not that strong. After the 1975 ICP no attempt has been made to continue error estimates in the ICP reports. However, Deaton (2012) and Hajargash and Rao (2016)

⁵² Today, the number of perturbations would be much larger because of the relative ease of computing compared to the 1970s. Sultan Ahmad did the programming and the code was labeled BANG because this was the period when East Pakistan became Bangladesh.

have provided frameworks for estimating improved standard errors for PPPs and real product.

B. Putting Together the 1970 and 1973 ICP Estimates

In the end 16 countries participated in Phase II with Poland and Venezuela probable until the window closed on the data submissions in 1974. From the European Union, Belgium and the Netherlands were added simply using the revised expenditure and price data for 1970 that were included in the EU report of 1974. All of the EU countries and the other Phase I countries provided updated expenditure data and aggregated price indexes for 36 summary headings for 1973 based on a worksheet developed for this purpose. The basic heading expenditure distributions for 1970 were used to distribute summary expenditures to detailed headings. Similarly it was assumed that the price increase for basic heading PPPs within a summary heading was the same as that for the summary heading.

In a parallel fashion, the additional four countries in greater Asia - Iran, Korea, Malaysia and the Philippines - collected detailed price and expenditure data for 1973 that were backdated to 1970 on the basis of summary heading data for 1970. Exceptions were Korea that provided prices for both years and Iran that collected most prices for 1970 and additional prices for 1973. In all cases, some extrapolations were required from the 36 level to the detailed expenditure and PPP level. In addition headings using hedonic regressions like automobiles and rents were redone to incorporate the new countries.

The question naturally arises as to whether it made sense to go to the basic heading level versus simply aggregating at the level of 36 summary headings. A test was carried out using both levels of aggregation. Differences were under 1 percent for all but three countries. For Korea, Malaysia and the Philippines, all of which provided detailed price data for 1973, the differences were more substantial, between 2.5 and 5.0 percent all in the same direction. We argued that including the 113 more detailed basic heading level data for these three countries moved us closer to the “truth” without increasing differences for the other 13 countries.

Our conclusion from this exercise is that retaining detail improves the comparisons, which we think raises a more general point. There is often a temptation in empirical work across countries to reduce the number of basic headings to the lowest common denominator. However, there is no evidence that finding ways to fill out a tableau of data in even ad hoc ways will be worse than aggregating at a less detailed level. And there is a clear advantage to using more detailed data when available rather than to consolidate it into a higher level of aggregation. This issue has arisen more recently in the context of the planned 2017 ICP where it is likely that there will be more detail in some regions compared to others.

C. Consistency Over Time of ICP Estimates

With two fairly independent benchmarks it seemed worth examining how close you could come to predicting 1973 by extrapolating the 1970 benchmark. One approach was to apply the growth of domestic absorption between 1970 and 1973 to the 1970 domestic absorption. The reason for breaking growth of GDP into growth of domestic absorption and growth of the net foreign balance is that export and import deflators are often weak. GDP deflators and GDP growth rates are affected by errors in the import and export price indexes. In what we called the *international price method* the import and export price indexes of the United States, the reference country, were used to obtain the constant price net foreign balance to add to the domestic absorption to get the estimated 1973 GDP. India was not included in this exercise because there were no export and import price indexes.

We also explored a *net deflation approach* which was to deflate a positive net balance by its national import price index and a negative balance by the national export price index. The 1973 estimates using the net deflation approach were further away from the actual 1973 benchmark than the international price method for all but Belgium. This led to us to adopt the international price approach in extending the estimates for the 15 countries to cover the period 1965-75. The actual benchmark values of 1970 and 1973 were used along with binary estimates for six countries, including India, for 1967. We published this series with United States as 100 in each year, based on current year prices.

D. Geary International Prices

A question about the G-K that was raised by a number of users of the Phase I report was the exact nature of the international prices. When the FAO used the G-K they had expenditures, prices and quantities so that an international price would be in a currency unit like so many dollars for a ton of wheat. When we estimated the G-K the inputs were the direct and indirect PPPs or price levels with respect to the United States, and the expenditures for the 153 basic headings. The international price for a heading was a weight to value the nominal quantities in each basic heading. The weighted average of the international prices over all basic headings was defined as 1.0 in the way we estimated the G-K.

The question raised was whether these international prices were invariant as to the reference country. The answer is they are not invariant to the reference country. However, when the reference country changes so do the nominal quantities because all of the basic heading price levels will be relative to the new reference country. This means the new nominal quantities will be valued at the new international prices and the comparison will remain the same regardless of the reference country.

Another question raised was about the way we presented the detailed PPPs for the net foreign balance in both the Phase I and II reports. For all countries except the United States we used the exchange rate. For the United States we presented the international price that was the result of estimating the G-K over just domestic absorption. It is basically the price level of the United States compared to all the countries. It is used

to value the nominal value of the net foreign balance converted at exchange rates, which is then added to the real domestic absorption to obtain real GDP. The US price level is related to the asymmetrical role of the US dollar in the international monetary system, a good reason to not use the dollar as the reference currency while acknowledging the plus side like familiarity and that international institutions tend to favor reporting dollar values.

Chapter 5: Signing Off - Phase III Report

When it was agreed the 1975 ICP report for 34 countries would be published again by the Johns Hopkins University Press, the Penn group were very, very ready to get out of ICP production, none of us more than Kravis. And the United Nations, the World Bank, the European Union and other international organizations were not unhappy that the ICP was leaving a university setting, with its academic niceties and seeming disregard for deadlines. We did regard the publication of Phase III as our last hurrah and wanted to do more than produce a repeat of Phases I and II, which partly accounts for the long gestation period, the publication only coming out in 1982.

Several factors contributed to the delay including my stay at the UNSD on leave during 1978-79. I was essentially transferring our software to the UNSD and taking the role of Kenessey in the sense of meeting ICP visitors and supervising staff. Murray left all the technical questions to me while he handled bureaucratic details for which I was most grateful. The Japanese government provided in kind ICP assistance in the form of three young staff each for two or three years over the period 1978 to 1985. The first was Shigeru Kawasaki who had spent a year of high school in California and was well trained in statistics. He was of great assistance to me while at the UNSD and was to have a very successful career in the Japanese Statistical Office. Shigeru was followed by Fukui and Watanabe, who were both excellent. I did some of my supervision of the three from Penn though I got to know Fukui fairly well and to enjoy his subtle sense of humor. I was able to meet both Fukui and Kawasaki on one visit to Tokyo in the late 1990s when Shigeru was Director of the Bureau of Statistics and Fukui was a deputy director in charge of price statistics. Bettina and I were also entertained at the Kawasaki home meeting with his wife who I had known in New York, and to meet their daughter and son.

I was on a short-term consultant's contract and not given high priority with respect to small things, like office space. When I finally got my office after several weeks at the UNSD it was very nice indeed, on a high floor in the Secretariat looking over the East river and up river to Roosevelt Island. It was a two window office reflecting my ambiguous status as a fixed term consultant. Kenessey and Murray had three window offices and the Director, Simon Goldberg, had a corner office with four windows. My first tour at the UNSD was quite pleasant in many ways, security was still easy and the work rewarding except that I did very little towards the publication of Phase III.⁵³

Another factor was that Kravis had a backlog of projects to complete including those with Robert Lipsey who generally spent a day a week at Penn working with Irv. Further the three of us were all involved with papers on services (Inman, 1985 and Kravis, Summers and Heston, 1983) that were relevant to the ICP but made only an indirect contribution to writing up Phase III. Finally there was the *hundred-country-paper* (Kravis,

⁵³ A curious practice at the UNSD was that most communication was by letters or telegram, all typically approved by my immediate supervisor and signed by the Director, all of which took time. This could be bi-passed by directly telephoning abroad for which there was no apparent restriction at least not on my phone.

et.al, 1978a), which captured our interest at the time and which Summers and I became more and more involved after 1980.

One application based on the hundred-country paper and on the first two rounds of the ICP was Ahluwalia, Carter and Chenery (1979) that initiated the counts of those in poverty in poorer countries. The paper was based on thirty-six countries for which there were *Kravis factors* and some information on country income distributions. The Kravis factor was the inverse of the price level, namely the exchange rate deviation index, which was used to estimate dollar denominated per capita income and consumption values for 1970. The Ahluwalia paper was part of the research at the Bank on economic growth and income inequality during the 1970s, much of it stimulated by the development literature of the 1950s and 1960s especially the work of Simon Kuznets (1955 and 1963).

The *Kuznets curve* was based on the observed downward relationship between measures on income inequality like the Gini coefficient or the share of the lowest quintile in the income distribution and per capita income. Ahluwalia, Carter and Chenery employed the Kuznets framework using PPP conversions which were only available from Clark or short-cut procedures before the ICP. They also anchored their poverty line concept on the definition that the Indian planning commission used when they established their 2nd Five Year Plan goal of reducing the number in poverty based on a rupee definition of the Indian poverty line.⁵⁴ The hundred-country paper and ICP provided the basis for converting the Indian line to a dollar measure across other developing countries. With variations this use of PPPs and country income distributions has become the basis for future poverty lines of the World Bank including the well-known *dollar a day* line developed when Martin Ravillion was heading the poverty group. This application of the ICP and hundred-country paper to poverty was important in raising the visibility of the ICP among academics, researchers and policy makers.

A. Services in the ICP

In all three ICP reports we presented three aggregations that we thought were of analytic interest: *services and commodities*, *tradables and non-tradables*, and *priced and non-priced services*. The services-commodity distinction was fuzzy but still produced consistent results, namely that the PPP for services rose and for commodities fell as incomes of countries rose. Commodities often have a component of labor in their price, for example purchases of cut up fruits and vegetables at a supermarket, while many services have an element of commodities, such as take-out or doggie-bags at a restaurant. In the ICP reports we simply assigned basic headings to commodities and services but in some other work we attempted to break up some basic headings into part services and part commodities on the basis of input-output tables.

⁵⁴ India was the first developing country to establish poverty reduction targets using operational measures like the share of the population or numbers of people to raise above the poverty line. Estimates of the poverty line were made in 1962 as part of the 3rd Five Year Plan and were based on rupee budget that would guarantee 2250 calories a day with a small allowance for other expenditures. The Indian Community Development Program was associated with the poverty goal and both were loosely parts of Lyndon Johnson's *War on Poverty*, not an accident.

The tradable-non-tradable distinction simply added construction to services to obtain non-tradables, hardly a firm breakdown as construction workers are mobile as are many overseas contractors. And tradables are not necessarily items that a country in fact trades. Not surprisingly the relationship of tradables and non-tradables to per capita GDP is parallel to that of commodities and services.

Priced and non-priced services were described in Phase I as other services and equal productivity services. The latter compared wages for a specified type of labor, like skilled manual, as an input price used to obtain PPPs. In correspondence Colin Clark in April 1980 recalled his previous observations that there were large differences between “currency-conversion factors for services valued on price and those valued on labour input”, for the United States and low income countries. Clark wrote further that priced services contained both commodity components as well as labor but that it was doubtful if input-output tables had enough detail to estimate the proportions. He carried out an exercise assuming commodities formed one third of priced services to see how this would impact of the Phase II results for India, Korea, Japan and the United States. The total effect would be to push down the estimates for non-priced services in India and Korea and up for Japan and the United States. In his illustration the total effect would be to lower India about 3 percent compared to the United States in 1978, that is 3 percent of 7.4 percent.

We did a number of experiments on alternative ways to value general government, health and education in preparing the Phase III report. For example, we studied the scores of foreign nationals taking the examination of the Board of Foreign Medical Examiners, a Philadelphia based group who administered an examination across many countries. This exam replaced the pre-WW II practice of certifying foreign nationals to practice in the United States based upon the rating of medical schools around the world. This experiment was an attempt to rank countries by their quality of medical training. We reported but did not use these results because of the issues posed by the exam being used by some foreign medical schools as their terminal exam, and the fact that the exam was in English, so the playing field was not really even.

In the medical area, the main set of non-priced services were those of hospitals. It was possible to obtain direct measures of bed-days per 1000 persons in many countries but it was thought quality of service was highly variable across countries. Robert Barro at one meeting noted that medical services in Kenya were very inexpensive and asked if that is where he should go for his brain surgery, a rather telling way to make the point that it is hard to hold quality constant. Direct pricing of a hospital bed-day was not simple to obtain because of the mix of payment arrangements in different countries, nor would such a price necessarily hold quality constant. We did try to control for quality by asking data on the capital, buildings and equipment per employee, in education, government and health, but we only obtained data for a small number of lower income countries, more for higher income. As a consequence whatever adjustments that were made involved estimates for countries without data on relationships between income and medical capital, a bit of a tail chase.

Measures like time with patients was also considered as a quality measure but this became a slippery slope. A doctor in a rural clinic in India may see 100 patients a day making rapid diagnoses and dispensing antibiotics and other common pharmaceuticals. Whereas doctors in the United States might see 10 or 20 a day. Does that make the US doctor more productive than the Indian doctor? Many would say no, the Indian doctor is more productive because her immediate impact on the health of her patients is greater.

Primary and secondary levels of education have the input side as well as possible measures of output in terms of total students at each educational level, number of graduates and the like. Eurostat and the OECD now use a quality adjusted measure of total students where the quality measure is based on the performance of students in PISA (Programme of International Student Assessment) tests carried out by the OECD. Some 72 countries took part in PISA in 2015. However, for most developing countries, there are a number of issues including enrollment versus attendance (by both students and teachers). Often education budgets are based on enrollment only. Where countries report test results, the sample is unclear because there is little information on whether the schools are representative or are mainly urban and private. As a consequence there is no easy way to apply the PISA country results without more knowledge of the sampling frame.

The World Bank took considerable interest in our efforts and some of their staff wrote memos including Paul Isenman (1979) who questioned the application of our techniques to Africa. Isenman noted that poor countries in Southern Africa commanded higher skill differentials than India where there was more human capital. He also argued that capital was often complementary with labor and so it was important to take account of this as we attempted to do in Phase III. In 1976 Robin Marris, whose brother Stephen had been one of the Associates with Gilbert, left Cambridge to head the Economics Department at the University of Maryland. The Bank asked Robin to provide a critique of comparison resistant services and other aspects of our ICP reports. Marris (1979) experimented with excluding all medical and government employees and measuring education by the number of pupils versus number of teachers adjusted for education. He found these changes had no appreciable effect on the resulting GDP per capita regardless of country or per capita income. These results were based upon the 1973 data of Phase II. We made a major effort in Phase III to improve our estimates for these comparison resistant services, and we think our estimates were more plausible than Phases I and II.

B. The Contributions of in-kind Aid in the form of Advisors

We have mentioned the in-kind contribution of Japan to the UNSD in terms of young professionals who brought statistical and computational skills to the ICP. The ICP also received in-kind contributions from the UK Ministry of Overseas Development in the form of recently retired statisticians who advised participating countries. The contributions of Angus (Harry) Fell in Sri Lanka and Robert Oswald in Syria, Kenya, Jamaica and Malawi were particularly welcome. The two were very different, Oswald a quiet person by nature, but very effective in getting the statistical offices to provide the necessary expenditures, prices and other worksheets in a timely fashion.

Angus was more outgoing, almost flamboyant but equally effective. I had first met him in the Bombay Yacht Club in 1960. At that time there was prohibition in Bombay so Wilma and I were using up our monthly permit of alcohol (it being allowed because it was the custom in our country) at the bar where we met Harry. He had just been doing preliminary tabulations of the Malaysian census on his Kurta calculator.⁵⁵ We would meet again in Hong Kong in the mid-1980s where he was consulting with the statistical office on how to value the investment flow when land was reclaimed.⁵⁶ Later Finn Forsyth was made available to the UNSD to assist in Phase IV.

This in-kind assistance was especially valuable because 18 new countries were being added in the Phase III comparisons, ten from poorer countries. It would have been nearly impossible for the UNSD and the Penn group to adequately cover these additional countries without the in-kind assistance. Similarly Fukui and Watanabe who followed Kawasaki were instrumental in moving software from Penn to New York.

C. Augmented Binary Comparisons

It was planned to have binary comparisons of each country with a regional link country and with the United States. However, as the price data came in some binary comparisons were very thin in that only one item matched in a basic heading and in some cases there were basic headings where both countries submitted prices for several items for a basic heading but none matched. So the basis for traditional binary comparisons of all 561 pairs of countries for 1975 was fairly weak. In 1979 I had returned to Penn from my year at UNSD, and Hughes Picard from the French National Institute of Statistics and Economic Studies (INSEE) had been hired by UNSD to direct ICP operations and to help organize the 1980 comparisons. Picard was a first class statistician and had ICP type experience representing France at many EU meetings. Picard suggested a way to overcome this lack of matching prices for a number of basic headings in Phase III.

Picard's idea was to introduce binary comparisons that were based not just on the prices each country provided, but also on some CPD prices that could be calculated from the CPD estimates carried out for the multilateral comparisons. If country A had a price for an item and country B did not, then we would add the CPD price for that item for

⁵⁵ The Kurta calculator was made in Lichtenstein as the company switched after WWII from watch making to producing cylindrical precision calculators for engineers, surveyors or contractors in the field. It was great for tabulating census data at the site of collection from hand written forms submitted each evening. I bought a Kurta in the early 1960s, which I occasionally bring out to demonstrate to grandchildren how tough life was in our time.

⁵⁶ I met Harry Fell for the third time in 1984 in Hong Kong when we were both visiting the Statistical Office, perhaps 15 years after his retirement from the UK CSO. Harry was there to consult on a national accounts issue and I to discuss Hong Kong's participation in the 1980 and 1985 ICP comparisons. In a bar at the end of the day Harry asked me if we were still using the "Mighty Genghis Khan" method at which at first I drew a blank. It finally occurred to me that Harry was asking about something Bob and I had done in the latest version of PWT which was to do a Geary-Khamis aggregation over both space and time which we had termed the mighty G-K. I said Harry do you mean when I have been retired as long as you I will still be worrying about index number problems? He just smiled, which turned out to be a prophetic smile.

country B. And if only country B had a price for an item, we would fill in the corresponding CPD price for country A. When this was done for all possible binaries, then the available price tableau for the 34 countries would be augmented. The resulting binary comparisons we termed *augmented binaries*.

The rationale then was to correct for the sparse matching of available prices for each binary comparison from the original price submissions of the Phase III countries. Earlier I have argued that there is not necessarily bias in the use of CPD estimated prices because they are neutral depending only on the prices from other countries. We did present all possible binaries for GDP, consumption, investment and government, but did not present detailed binary comparisons except with the United States.

Our conclusion was that the traditional and augmented binary comparisons were highly correlated with each other and with the multilateral comparisons. However, there is a systematic relationship between binary comparisons and multilateral comparisons, namely binary comparisons for lower income countries were systematically lower than multilateral results. The reason is fairly clear. In binary comparisons with the United States each country has equal weight, whereas in multilateral comparisons as we carried them out, the United States had much more weight. The larger weight of US prices would raise per capita GDP of lower income countries compared to a binary where both sets of prices have equal weight.

It is fair to say that augmented binaries did not make a big splash and to my knowledge they have not been used since Phase III. However, it is worth pointing out that augmented binaries have some of the characteristics of the method employed in the European Union. The European Union uses all prices of representative items in country A so long as country B considers the item representative or commonly available. Further for non-EU countries augmenting binaries has some possible research interest. For example, in applying the GEKS approach at the basic heading level and then aggregating the Fishers by GEKS or a weighted GEKS.

D. Analysis of the 1975 Results

Several simple analytic exercises were carried out in presenting the Phase III results as the larger number of countries and wider range of economies allowed more scope for inquiry than in the previous rounds. A schema employed throughout the report was to group the countries into six income groups and to present variables by country and group average. For example, Table 1 in Chapter 10 of the Phase III report (Kravis, et.al. 1982) showed that the average price level for the lowest income group for services was 16 percent and rose to 33 percent in group II, 41 percent in group IV that included Italy and Spain, and leveled off to near 98 percent for group V with the United States as 100 in group VI. Parallel tables examine price indexes by summary expenditure groups; for tradables and non-tradables; and for similarity indexes.

(1) Bridge Country Binaries

One exercise involved comparing original binary comparisons with the United States with the same indirect binaries through bridge countries in each region. For example we used both India and Malaysia as star or bridge countries in the Asia region. We compared the implied binary comparison of each Asian country with the United States using India as the bridge and then Malaysia as the bridge with the direct binary with the United States. The results appeared very sensitive to the reference country. When India was the reference, the indirect quantity per capita relative to the United States was 24.9 percent compared to the direct binary of 19.5 percent, a difference over 20 percent. Most differences were closer to 5 or 10 percent in Europe, under 15 percent in Africa and Latin America. These results were at the GDP level.

When the indirect binaries were carried out at a more detailed level the results were usually closer to the original binary (37 of 50 observations). This point is important because in the 2017 ICP the plan in early 2016 was to combine both new and extrapolated prices. Our results suggest that the more detailed level, perhaps 40 summary headings would be preferable to aggregations at higher levels like consumption. The other calculation was to see how different were the results using the two different reference countries for each region at the GDP or more detailed level. Countries had differed less from direct binaries for 22 of the 34 countries consistent with the above results.

(2) Demand Analysis

In the earlier reports we had run individual regressions of real quantities on price levels and per capita income with mostly the expected negative elasticities on price (28 of 30 summary headings) and all positive on income. These results were gratifying but hardly unexpected. Usually the outlier headings were residuals where the quantities may be large because the expenditures have not been allocated to the appropriate headings, for example “other expenditures”. With 34 countries slightly more ambitious explorations of demand were possible.

Summers took the lead exploring the possibilities of first moving from single equations to estimating generalized least squares. With 34 countries only a limited number of expenditure headings could be used. We tried 4 and 7 large headings like food, clothing, shelter and the like and the results were as expected. The next step was to try estimating a simple demand model, namely the LES or linear expenditure system, over consumption. An interpretation of the system in the context of countries is that there is a minimum subsistence bundle and a representative consumer in each country. Equations are estimated simultaneously for food, clothing, shelter and other so that total expenditures sum to income and all countries (representative consumers) can purchase the minimum bundle.

The results allowed us to tell a story and show that the results compared to the comparable G-K estimates using super-country and equal country weights. The LES is built up from expenditure shares, not totals of consumption, so it is more comparable to equal country weights. We were glad to have carried out these exercises and they were duly noted by others but did not become a mainstream ICP activity. One standard set of

indexes that we published were picked up by others were price and output similarity measures. Let me discuss this with respect to price similarity indexes.

A price matrix was created for each country where the basic heading price level was expressed relative to the average for each country weighted by the real expenditure of each heading. A correlation matrix was run for all pairs of countries that we termed a price similarity matrix and that we used for other analysis. It was structured the way one might expect, namely the closer were countries together in income the higher was their price similarity index. Theil (1985) in a review article on the Phase III Report was supportive of the ICP and detailed empirical work producing the basic heading parities. However, he was critical of Chapter 9 both for its measure of similarity and the use of the linear expenditure system as a demand model. Theil thought he had both an alternative similarity measure and better demand system. Other reviewers of the 1975 ICP were more sympathetic to our analysis and the underlying price parities and expenditure detail that were provided.⁵⁷

E. Estimates for non-benchmark years

As with the earlier reports we compared the benchmark estimates for 1975 for those ICP countries in 1970 that had adequate deflators for domestic absorption. The reason to use the sum of domestic investment, consumption and government is because it is conceptually similar to the benchmark estimates provided in the first phases of the ICP. That is we aggregated over domestic absorption and then added on the net foreign balance at the average price level of domestic absorption over all countries, basically its international price. A conceptually better way would be to further extrapolate PPPs or price levels for summary headings to a later benchmark year and then aggregate the new data to domestic absorption. (This was not done in Phase III but did become the approach in the PWT).

We also argued that this method of extrapolation at the level of domestic absorption was preferable to GDP extrapolations because export and import price indexes are often subject to large errors that become embodied in GDP deflators. The principal results show the extrapolations in current prices of the 1970 benchmark to 1975 for 15 countries (Hungary did not have the necessary deflators). The ratio of the extrapolation to the 1975 benchmark was between 98 and 102 percent for three countries counting the United States; the ratios were less than 98 percent for eleven countries and

⁵⁷ Armstrong (March, 1983 *Economic Journal*) was kind enough to refer back to Peter Hill's laudatory review in 1976 of the Phase I report and conclude, "The task of dealing with more collaborators and processing more data has in no way detracted from the quality of the work and, of course, the results are all the more powerful and interesting." Paul Samuelson (1985) in commenting on the 1984 Nobel Laureate in Economics, Richard Stone, mentions the contribution of the Stone-Geary linear expenditure system and mentions favorably its use in the 1975 ICP. In the interest of transparency, Summers and Samuelson were brothers but it was only later that they acknowledged this in print. Both Bob and his younger brother took their mother's maiden name when they pursued economics as they did not want to trade on the acclaim their older brother had already received.

Iran was 114 percent above the United States. So there was a definite pattern that the extrapolations underestimated incomes compared to the 1975 benchmark on average with less understatement for the richer countries. We tried to find a story to tell that would explain these results, but none seemed satisfactory.

We tried a similar exercise extrapolating from 1970 and 1975 back to 1950 for the seven European countries and the United States. The Netherlands was the only country differing by over 10 percent when either the 1970 or 1975 benchmark was used. And as is common, the closer are two years together the less difference between extrapolations and benchmarks. That is extrapolations to 1950 were on average closer moving from 1970 than from 1975. Eurostat had begun building time series for their countries after the 1980 ICP (Krijnse-Locker and Faeber, 1984)

F. A Digression on 1980

During late 1979 and early 1981 Simon Goldberg retired and Svein Nordbotten from Norway replaced him as Director of the UNSO. I returned to Penn from the UNSD in the summer 1979 and Hugues Picard was hired to be Chief of the ICP section. Picard was on leave from INSEE where he had worked on many issues including consumer prices and had participated as a French representative in many Eurostat meetings. By 1980 we had distributed the Phase III report to all the countries for their reactions that were shared with the UNSD and Penn. In addition Picard was responsible for implementing the 1980 ICP and we would often meet to discuss issues particularly those concerning countries with which we had worked or countries we had visited with respect to their participation in the 1980 round. Another development concerned Eurostat. In 1977 the internal candidate favored to succeed Jaques Mayer as Director General of Eurostat was Paretti but an outsider, Aage Dornonville de la Cour, was chosen. De la Cour tried to have Paretti appointed Deputy Director General, but that was opposed even by the Italian cabinet, but he did put him in charge of three Directorates in 1979. This ended unhappily with Paretti and others leaving Eurostat in 1980.

(1) China's Purchasing Power

A US economic delegation went to China in October 1979. It included authorities on China Dwight Perkins and Robert Dernberger as well as Lawrence Klein, who would receive the Nobel Prize in 1980, and Kravis from Penn. Kravis had flown as a logistics officer on flights to China during WWII and had a personal as well as professional interest in visiting China as it began to open up.⁵⁸ At the time the Chinese yuan was pegged at 1.94 per US dollar, and no one knew how close China's PPP would be to the official exchange rate. Armed with ICP item specifications and worksheets Irv spent much of his

⁵⁸ Irv was with logistics for the well-known *Flying Tigers* commanded by General Claire Chennault, who was a retired US Army Air Force officer who worked in China as a military advisor to Chiang Kai-shek since 1937. Irv kept an autographed photo of Chennault in his office and was always happy to answer questions about his experiences with the Flying Tigers, whose distinguished airplane nose art of a shark baring its awesome teeth was a well-known icon for boys growing up in World War II, like myself.

time with guide and translator in China's shops for consumer goods; or in offices obtaining national accounts, education and health information, and talking to experts on machinery, construction, transportation and government.⁵⁹

Kravis had to wait a few months to obtain some of this information from China but was able to include a report as an appendix to the main report. The reference year chosen was 1975 and Chinese prices were stable over the years to 1979 for most items, so it was not difficult to move prices backward. Kravis had taken specifications for items that the United States had priced for 1975 in anticipation of doing a China-United States binary comparison. He circulated that report to a number of economists who worked on China and did some revisions that appeared the following year (Kravis, 1981). Generally multilateral comparisons will provide higher estimates for lower income countries compared to binary comparisons. Adjusting for this factor, Kravis was able to compare the 1975 Phase III estimates for other low income countries with China.

Most critics regarded the Kravis numbers for China as too high. In any event there was great interest in the estimates especially at the World Bank where they generated much discussion. Further, the previous estimates for China by Hollister (1958) were out of date and they were from the production side whereas the Kravis estimates allowed comparisons with comparable countries. The fact that Irv's estimates were based on direct price comparisons reinforced the importance of PPP based comparisons. By this time Sultan Ahmad (1983) had joined the World Bank and carried out a price collection experiment in China with a 1980 reference year. It is fair to say that Kravis's Chinese exercise was a catalyst for research on both the Chinese economy and PPP comparisons.

(2) The Hill Report

Peter Hill had been hired as a consultant by the ECE, Eurostat and the UNSD to make recommendations regarding Phase IV. Hill was to address aggregation methods and to examine different ways of linking the regions in the 1980 ICP. Around 60 countries were expected to participate in Phase IV from the different regions and a method, most likely different than used in Phases I-III, was needed. Hill (1982) in the published version of the report incorporated the comments on his draft from a meeting in London, October 16-17, 1980 (Hill, Haeder from ECE, Nordbotten and Picard from the UNSD, and Clarke, la Cour and Krijnse-Locker from Eurostat attending). At that meeting it was accepted that the method of aggregation to be used was the G-K though the Eurostat was to review the

⁵⁹ We were amused by Irv's experience buying towels. As was usual in this period in shops a crowd would gather to see what a foreigner was buying. Our specification for towels called for the weight of towel per square meter, about which the shopkeeper had no idea, so Kravis asked him to weigh and measure it. The crowd was amused but not the shopkeeper because Irv did not buy the towel. As a further aside, it was at the time it was common in Kenya for price collectors to actually buy such an item to assure it was a transactions price, and later the goods were donated to charity.

matter, perhaps because Paretti, the most important decision maker, had left a vacuum when he resigned from Eurostat earlier in the year.

Hill was in favor of a weighted aggregation method, a view we were most happy to support. A major difference arose due to the practical problem that the European Union completes and published its comparisons within two years of the completion of its price collections. Thus for 1980 Eurostat published its comparisons by 1982, whereas the other regions would not complete their data collection for several more years. Eurostat did not want their published results for the European Union to be changed by the addition of non-EU members because they were used for administrative purposes such as the allocation of the structural funds. There might have been more room for negotiation on this issue of *fixity* if all countries were on the same time schedule. Kravis (1986, p.8) continued to argue against fixity even after it had been become institutionalized in EU regulations.

(3) Other Meetings

Eurostat made a commitment to comparisons for seven French speaking countries in Africa for 1980 and held a meeting on February 26-29 in Luxembourg with country representatives and Picard from the UNSD. Eight English speaking countries from Africa also participated, typically coordinated through visits from Picard or seconded former staff of the UK Statistical Service. There was a concerted effort to cover the major regions of each country including both rural and urban areas especially among the Francophone countries.

The Japanese government provided funding for a meeting of the ESCAP countries in October 1980 in Bangalore, India. ESCAP, ESCWA, the ADB, and several aid organizations were represented. (Pic 44) The Indian Central Statistical Organization, the Karnataka State Statistical Organization, ESCAP and I (in place of Picard who was occupied with other meetings) set the agenda. Twenty countries were represented from Lebanon to Fiji. Unfortunately only seven of these countries were to participate in the 1980 comparison as part of Asia (Australia, Korea and Japan were through the OECD), one less than Phase III, Syria dropping out. The meeting served the purpose of familiarizing many statisticians in the region with the results of the Phase III comparison and with the data requirements and methods of the ICP. Paul McCarthy was the representative from Australia and was later to be involved with ICP work at the OECD. After leaving the OECD he became involved with the 2005 ICP as a member of the Technical Advisory Group (TAG), and would chair the TAG during the 2011 ICP.

Bellagio again lent its facilities to the ICP in December 1980 for a session on the treatment of services. I had undertaken much of the research and writing of the chapter on services in the Phase III report and drafted an agenda. Country representatives were there from Brazil, France, Hungary, India, Japan, Poland and the United Kingdom. Peter Hill, Richard and Nancy Ruggles, the Penn group, as well as experts from Eurostat, ECE, ECIEL, the IADB and the UNSD, were also in attendance. Alternatives to the Phase III methods were discussed for general government, medical and health services. An attempt was made to obtain cost estimates for different types of hospitals but too few countries supplied the necessary data so the fallback was made to salary comparisons

for different medical occupations again using the assumption of equal productivity for the same job description. For general government and education salary comparisons were also made by job description with allowance for years of schooling.

(4) The Latin American Region

A meeting was held in Rio-de-Janeiro in January 1979 at the Getulio Vargas Foundation, which at that time produced price indexes for Brazil while the central statistical organization, IBGE, did national accounts, censuses and the like. Preliminary meetings took place in November 1978 in Washington, Philadelphia and New York between representatives of the Getulio Vargas Foundation, ECIEL, ECLAC, the UNSD, the World Bank and the IADB who sponsored the Rio-de-Janeiro meeting. The purpose of the United States and Rio-de-Janeiro meetings was to find a way to integrate the 1979 PPP comparison for Latin America with the ICP Phase IV comparison for 1980. The meeting was organized by ECIEL with Dinesh and Janos De-Sousa from the Getulio Vargas Foundation and Jorge Salazar-Carrilla from ECIEL. I was at the UNSD at the time and attended the sessions with Carl Otto, who had replaced Murray. Proximity of the meetings to the Copacabana beach was a not unpleasant aspect of this mission.

The price collection was organized in a similar manner to earlier exercises in Latin America, namely urban and with all countries supplying prices for all items. Seven Central American and Caribbean countries were included with the nine from South America so it was a bit of a stretch to expect the items chosen to be representative for all the countries. Because Portugal and Spain were taking part in the Eurostat comparisons in 1980, consideration was given to linking through these countries. However, that did not prove feasible.

Frank Orlando was the representative of the IADB in Rio-de-Janeiro and had written his dissertation at Indiana on index methods and had convinced us that the iterative approach was a much more intuitive way to estimate the G-K system. It was faster to compute G-K iteratively and easier to explain the steps to those unfamiliar with the system. When all elements of the PPP and expenditure matrices are positive, it typically took five to six iterations to obtain results differing at the fifth decimal place.

The father of Laura Kingston, a staff member at the Getulio Vargas Foundation, was a civil engineer who suggested a method of handling construction that was similar to that adopted by Eurostat. This was the bill of quantities approach that broke down construction projects into component materials and tasks like land preparation, pouring foundation, footings, external shell, and the like. Both the Kingston approach and the Eurostat approach required professional input of architects, civil engineers or quantity surveyors to produce their cost estimates per square meter. The approach used at Penn also employed professionals to estimate unit costs of whole buildings, though in fact when Samvit Dhar relied on building cost manuals, he in effect built up his estimates from costs of components.

(5) The 1975 EU Report

Earlier we mentioned that the EU comparison for member countries was not published officially but in the French journal, *Analyses*, under the names of Paretti, Goybet and Krijnse-Locker. The 1975 EU report (Eurostat, 1978) covered eight of the nine member countries, the three accession countries, Denmark, Ireland and the United Kingdom, and five of the six original members, Luxembourg still absent. The major differences between the superlative aggregation indexes, like GEKS or Van Yzeren, and the G-K is that the latter is additive at the cost of ignoring the substitution effect between basic headings. In estimating the 1975 results Eurostat adopted still another approach, the Gerardi method (1982 and Eurostat, 1978, pp. 26-35).

Greece was an EU associate in the 1960s and gained accession in 1975 and membership in 1981. Gerardi was an Italian staff member at Eurostat and well trained in mathematics and statistics as was his Greek successor Avdoulous. Gerardi strongly felt that in judging the PPP between any pair of countries, the importance of either country should be equal or in his terminology, *equal-distant*. This was a position of Drechsler (1973) who Gerardi cites as an influence on his ideas. Gerardi was also influenced by meetings with van Yzeren in 1977 and the two planned to write a joint paper, though that did not happen (Balk, 2008, pp. 47-8). Balk (2008, p.255) terms the method Gerardi-Van Yzeren because both characterized their approach as that of *tourists* or *immigrants* looking at different ways to judge the cost of living in two places.

One important aspect of Gerardi's presentation is that it was in *purchasing power standards* (PPS) meaning that no single country was the reference country but rather all eight EU countries were the reference. There are several ways to compute this, including taking each country as the reference and then averaging the results. A more general method is illustrated in Table 2 where we have used the 1975 exchange rates to the US dollar and PPPs for women's shoes from the Phase III report for illustration for the EU countries.

Table 2

Country	XR to US \$ 1975	XR/Mean XR on PPS basis (1) / 5.421 (PPS)	Women's shoes PPP / US \$ (PPP)	Women's shoes PPP / PPS (3) / (2)	Women's shoes PL PPS (4) / 5.997
	(1)	(2)	(3)	(4)	(5)
Belgium	36.78	6.785	37.4	5.512	0.922
Denmark	5.746	1.060	5.2	4.905	0.821
France	4.286	0.791	8.25	10.434	1.746
Germany	2.46	0.454	2.53	5.575	0.933
Ireland	0.45	0.083	0.754	9.082	1.519
Italy	652.85	120.441	620	5.148	0.861
Netherlands	2.53	0.467	2.77	5.935	0.993
United Kingdom	0.45	0.083	0.31	3.734	0.625
GEOMETRIC MEAN	5.421			5.977	1.000

Column (1) provides the exchange rate to the US dollar for the EU countries and its geometric mean. Column (2) expresses the dollar exchange rates as PPS by dividing each exchange rate by the geometric mean of column (1). We next take the PPPs per dollar for women's shoes in column (3) and convert these to PPS dividing by column (2). Finally we normalize these by dividing by the geometric mean of column (4). The United States is convenient but unnecessary in the example and we could have eliminated two of the five columns.⁶⁰ The Penn group resisted expressing results as the average of all countries mainly because as discussed in respect of Table 1, the UNSD and the World Bank expressed all their tables in US dollars. However, it clearly made great sense for the European Union because it was not politic to privilege the currency of any one member country. At first glance one might ask why not make it even easier and average the PPP of shoes directly across the eight countries and leave out the exchange rates

⁶⁰ That is we could have begun with country shoe prices for women in national currencies and divided by the geometric mean of the exchange rates and then obtained column (5).

altogether? The answer is that if one of the countries does not have a PPP for a basic heading, not usually an EU issue, then it becomes less straight forward.⁶¹

From the point of view of presentation Gerardi's approach was more general than we had used in our ICP publications. In this regard, both the 2005 and 2011 ICPs have presented their price levels and PPP converted expenditures at both US and world prices. Turning to the Gerardi additive method, it can be described as follows. For each basic heading the PPP and price level was expressed relative to all the member countries. The nominal quantities for Girardi are the national currency expenditures converted at the geometric mean of exchange rates of all countries. The nominal quantities were then valued at each countries PPP for a basic heading relative to the average for that heading. PPP converted values for each basic heading can then be added to whatever aggregate desired and aggregates can also be summed just as in the G-K system.

The Gerardi method does not take account of the difference in relative prices across basic headings. Looking at Column (5) of Table 2 the price level for each basic heading is 1.0 so the Gerardi-Van Yzeren system allows no substitution between basic headings arising from relative price variation. In contrast, the Van Yzeren method does allow such substitution and is not additive. The European Union adopted the GEKS aggregation method in the subsequent comparisons of their member countries. We did not include the Gerardi in our comparison of methods in the Phase III report mainly because it was unknown to us till meetings in Luxembourg in the late 1970s. We did comment on Gerardi's paper in print, which sets out our differences (Kravis, Heston and Summers, 1982a). In retrospect we were rather stubborn in our views and in discussions with our Eurostat colleagues at this time, fundamentally because we were strongly in favor of weighting countries by their economic size, whereas the European Union was for equal weighting. Along with Peter Hill (1982) we thought that the results of binary comparisons were given too much importance by many of our critics without justification when the number of countries was three or more. Despite the high decibel count at some of our meetings, there was always mutual respect and cooperation and many a pleasant dinner together with our colleagues at Eurostat.

⁶¹ Consider estimating basic heading PPPs when the reference country is missing prices for all items in a heading. Then in the CPD estimates the next reference country will be taken alphabetically and the coefficients estimated for the remaining countries. Those coefficients will be off by a constant from the coefficients that are in line with other basic headings. In the 2011 ICP this situation arose in a number of regions. When prices are converted at the average of exchange rates of all countries, then the problem does not arise. I am indebted to Sergey Sergeev for educating me on this point.

Chapter 6: Estimation of GDP of Non-benchmark Countries and Sectoral Output

At the same time that Phase III was being published and Phase IV was being put together, two derivative research activities were being developed. The first was the extension of the expenditure estimates of GDP to countries that had not participated in ICP benchmark comparisons but had adequate national accounts data. Mention has already been made of the hundred-country-paper published in the *Economic Journal* that was based on the 16 countries of Phase II and covered one year, 1970. This was followed by further releases that incorporated the 1975 benchmark and added more years, more countries and some disaggregation of GDP that became the PWT. The second derivative set of studies centered at the Groningen Growth and Development Center (GGDC) at the University of Groningen established by Angus Maddison after he took a Professorship there in 1978.

A. Origins of the Penn World Table

Summers liked to squeeze as much knowledge as possible from new information leaving as little as possible unused. This was the spirit of the CPD method and the *raison d'être* for developing the PWT. The approach is illustrated from the hundred-country-paper where we moved from the new information, the price level and real GDP of 16 countries in 1970 to estimates for more than 100 countries based on estimating equations. Most of the explanatory power of the regression equations came from regressing \ln PPP converted GDP relative to the United States against the \ln of exchange rate converted GDPs, designated r and n . In our preferred version, we used $\ln n$ and $\ln n^2$ and a measure of openness of economies (exports + imports)/GDP and a measure of price isolation for the period 1963-70. The latter was the mean squared difference between the inflation rate of a country and the world inflation rate for the period 1963-70. With only 14 observations (Hungary and Colombia did not have all the variables) it is difficult for additional variables to obtain much statistical significance when n already explains over 90 percent of the variance in r . The coefficients on the additional variables were of the right sign and equal or greater than their standard error which we took as support for our formulation.

The estimates for the non-benchmark countries were subject to errors of 20 percent or more which are large, but not compared to the errors of using exchange rate conversions versus PPP conversions. We also tried two approaches to estimating non-benchmark countries for 1973. One method was to compute a regression for 1973 as was done for 1970 and use it to estimate the 1973 non-ICP countries. We also used national growth rates of domestic absorption from 1970 to 1973 adding on the net foreign balance for 1973 to obtain the 1973 estimates of GDP. Our arguments for preferring the latter approach seem less convincing than they were at the time.

In Summers, Kravis and Heston (1980) we broke down GDP into consumption, investment, government and the net foreign balance. We used an equation to obtain the real shares based on nominal shares and per capita GDP relative to the United States for the 16 benchmark countries that assured that the sum of the shares equaled domestic

absorption. An important finding of the ICP is that nominal and real shares differ across countries by per capita income, and this is captured in our system of equations. We used 1970 as the reference year and presented a constant price series of GDP per capita from 1950 to 1977 and a current price series adjusted for changes in the terms of trade. We then valued the net foreign balance at the international price of domestic absorption in the constant price series. The methodology employed required that a country have current and constant price national accounts at least at the level of consumption, investment, government and exports and imports. This totaled 119 countries, excluding one benchmark country, Hungary.

Kravis was helpful in subsequent versions of the table but had moved on to other projects more closely related to international trade. He was especially interested in exploring further explanations of national price levels now that there were more degrees of freedom associated with Phase III. This work was done with his colleague, Robert Lipsey (Kravis and Lipsey, 1983). Kravis was always responsive when asked by the United Nations, World Bank or others to advise, participate in meetings or write an article concerned with the ICP (Kravis, 1986), but by the mid-1980s he was engaged in other pursuits. One of his many interests was income distribution and we did use ICP data to examine this issue (Kravis, Heston and Summers, 1984). At the end of the day Bob and I ended up developing the PWT carrying on much of the earlier work with Sultan Ahmad and of course Irv.

In the 1980s we brought out two new versions of the Table, the first based upon both the 1970 and 1975 benchmarks (Summers and Heston, 1984 and 1988). Both of these versions of the PWT were published in the *Review of Income and Wealth* which at the time was edited by Nancy and Richard Ruggles as a labor of love. They also had been among the founders of the International Association of Research on Income and Wealth (IARIW) that published the Review and held regular conferences every two years, typically in Europe where most of member resided. Users wanted digital access to the Tables so we took on the thankless activity of selling tapes of the data between 1984 PWT 3 and 1988 PWT 4. The Ruggles were very supportive of dissemination of the PWT and other data sets and in this regard Richard offered us a deal that we could not resist.

Previously the Review had printed the PWT for all of the countries, variables, and years covered, that is the 30 years in PWT 3. What Richard proposed for PWT 4 was to distribute the Table in the form of a machine readable diskette in each issue, and only provide one year of the printed version. In fact, it took three diskettes for the 121 market economies 17 variables and 26 years covered in PWT 4, which all had to be hand stuffed into sleeves in the back cover of each issue, a task that involved the two daughters of the Ruggles.⁶² Forward looking as this method of data distribution was, it drove librarians up the wall because the diskettes kept disappearing from their journal issues. Hopefully, the culprits shared the contents of the diskettes with others so Richard's vision was partially

⁶² I only learned this at a memorial for Bob in 2012 at the first IARIW conference in the United States in Cambridge in conjunction with the National Bureau of Economic Research summer workshops. After the dinner and speeches Patricia Ruggles reminded me that the task of stuffing the sleeves is well etched in her memory 28 years later.

realized. Fortunately technology moved right along so by the time the next version of the PWT was prepared, a less labor intensive version of distribution was available.

PWT 4 utilized the benchmarks for 1970, 1975 and 1980 as the basis for estimating the reference year per capita figures for 1980. By this time 67 countries had participated at least once in the ICP and we wished to utilize as much of this benchmark data as possible. Sixty countries participated in ICP 1980. The seven that did not had participated in ICP 1975. They were Jamaica, Mexico, Romania, Iran, Syria, Malaysia and Thailand.⁶³ Our approach was to use estimates for consumption, investment and government for 1980 for the 60 countries of the 1980 ICP; and, for the 34 in 1975 and the 16 in 1970, extrapolations to 1980. Bob used an errors-in-variables model to try to assign errors to national growth rates, to benchmark estimates and to their covariance. The purpose was to arrive at weights to generate an estimate for 1980 as the reference year from the total of 110 benchmark observations and extrapolations for each. Bob termed the process *consistentization*, which, once she heard it, Anita Summers said that word was not to be uttered in the house again. Son, Larry, was not against the word itself, but as Chief Economist of the World Bank in 1991 and even before, he did advise his father not to *mess with* national growth rates.

B. The Beginnings of the ICOP Studies at Groningen

Angus Maddison (1926-2010) was well known to Irv because he was at the OECD at the time of the Gilbert-Kravis project. Maddison was Assistant Director of the Economic Development Department at the OECD from 1966-71 and held a number of consultant positions until he moved to the University of Groningen in 1978. Maddison had corresponded with Irv and me about the ICP and followed our work. Angus always had some reservations about (or aversion to) our multilateral approach and our handling of services: the former reflecting a strong preference for binary comparisons, and the latter substantive concerns. His interests were oriented to total production and historical measures of GDP per capita in a comparative perspective. One of the early projects of his GGDC focused on the International Comparisons of Output and Productivity (ICOP) program.

ICOP started out with studies that converted sectoral output in different countries by PPPs to a common unit, beginning with the manufacturing sector. For some sectors like agriculture the FAO had measures that would allow direct quantity measures, while for construction, the ICP estimates could be used. The service sectors are difficult for ICOP and ICP, but ICOP has done a number of service sector studies on retailing and related activities. The advantage of a comparison of the manufacturing sector is that it allows comparison of worker and total factor (with physical and human capital measures) productivity because the data source typically provides value of output, units of output

⁶³ In fact Mexico did collect price and other data for 1980 but chose not to have their results published for reasons that were not explained to me. We felt that the results for Mexico for 1980 were superior to any extrapolation so we used them in PWT 4.

and labor force. Censuses of manufacturing are a standard data source; they vary in frequency, level of detail with respect to items, and coverage of plants or companies.

The conversion factor for national currency values is a PPP derived from unit values, that is, the value of production divided by the number of units produced. The usual criticism of unit values is that often the number of units can cover a wide range of qualities or capacities that may differ across countries. If the item is broadly defined like electric motors, then it may contain pieces ranging from 1 to 5 hp. If the mix is roughly the same across countries, this type of heterogeneity is not a real problem. But if for some reason some countries specialize in low horsepower motors and other countries in high horsepower motors then use of unit values as prices may lead to systematic under or overstatement of PPPs. Studies have verified that the more detailed the classification of a manufacturing census, the closer will unit values be to prices.

A number of Ph.D. students and staff at Groningen who worked at the GGDC carried out productivity studies for particular sectors or countries contributing to the expansion and improvement of ICOP studies since their initial papers circa 1983 and their first data set in 1987. Adam Szirmai was in the Faculty of Economics when Maddison joined Groningen and interacted with the ICOP group though his principal focus is on social development and inequality which he now pursues at the UNU-Merit and the Maastricht Graduate School of Government. Bart van Ark joined the graduate program in 1981 and became a Ph.D. Researcher 1986-87 and spent 1988-90 as a Research Officer in London at the National Institute of Economic and Social Research continuing research on productivity. His dissertation compared productivity performance in ten countries from 1950-1990 becoming an important ICOP study.

We had talked of trying to reconcile the ICP and ICOP estimates and Dirk Pilat spent a semester with us at Penn with that being one item on the agenda. With good faith on both sides, progress was nil. Dirk's dissertation work was oriented to productivity comparisons but at the OECD he soon became involved with the contribution of IT to productivity as did the ICOP. He now heads the Science and Technology Division at the OECD.

Bart van Ark succeeded Maddison as Director of the GGDC, as Angus became more involved with historical studies. As the PWT began to build a historical data set back to 1950, Irv would often joke with Bob that he would next want to go back to the Holy Roman Empire. Angus Maddison of course beat us to it going back to the Roman Empire and earlier. In later chapters we will touch further on the legacies that Maddison and Kravis left in terms of more recent developments of the ICOP and the PWT.

Chapter 7: ICP after Penn - Regionalization and 1980 and 1985 Comparisons

The organization of the 1980 and 1985 ICP benchmarks ushered in the regionalization of the program. ESCAP was the center for Asia and the Pacific comparisons and as mentioned earlier a meeting was held in Bangalore, India, to discuss the time schedule and data demands on the countries. The Economic and Social Commission for Western Asia (ESCWA) talked about participation of its countries but for a variety of reasons did not participate in 1980. Eurostat again organized participation of countries in Africa as part of their technical assistant missions to improve statistics in former colonies. Eurostat hired a number of African nationals in Luxembourg including Michel Mouyelo-Katoula from Cameroon who would become Global Manager of the ICP for the ICP 2011. We have mentioned the ECIEL 1979 and its ECLAC update to 1980 that formed the basis for Central and South America. Eurostat carried out comparisons for its 13 members with Israel pricing the same product list, Austria coordinated the participation of Finland, Hungary, Poland and Yugoslavia, and finally the OECD joined the 1980 ICP retrospectively bringing in Canada, Japan, Norway and the United States.

Picard had coordinated much of the 1980 comparison at the UNSD, but he returned to INSEE in 1982 before all of data were processed. I was an Advisor at the Pakistan Institute of Development Economics in Islamabad from June 1982 to January 1983 and a candidate to replace Picard as Chief of the Price Statistics section at the UNSD. An offer of a two year appointment with the option to continue arrived in December 1982 and I joined the UNSD again in January 1983. By this time the UNSD had moved to DC-2, a just completed building on East 44th Street adjacent to the Secretariat. An office issue arose immediately because my new rank called for three windows and so they arranged for such an office looking East onto the brick wall of an older building. I said I preferred an odd shaped but slightly smaller office with only two windows but very good light and an unobstructed southern view, a battle I easily won because both offices were vacant at the time. By this time the computer capability of the Division was quite adequate and most of my time was devoted to completing and writing up the 1980 comparisons and planning for the 1985 ICP round.

Y. Kurabayashi was Director in this period and my immediate supervisor was Madhu Palekar who primarily worked on national accounts. Samvit Dhar was at the UNSD at this time along with Roshan Traku, both part of my section. My section also included work on methodology of time to time price indexes and the principal researcher on this work was Wolfgang Schubert, who fortunately did not need my advice or supervision.

I also assisted the Committee on Reduction of Military Budgets (UN Department of Disarmament, 1986) that asked for countries to submit a breakdown of expenditures and prices for some representative items for personnel, equipment, installations (airfields, ports and the like), fuels and maintenance. The Committee wanted to convert these national expenditures by PPPs to estimate comparable quantities across the eight countries who responded to the survey, Australia, Austria, Finland, Italy, Norway, Sweden, the United Kingdom and the United States. Of these countries only Italy practiced conscription which posed an interesting problem. As a consequence of

conscription the PPP for compensation for Italy was unusually low compared to the other countries. But also the expenditures on compensation were low so that one got the correct quantity of personnel dividing expenditures by salaries.

If no adjustment were made, there would be a distortion because the relative share in lira for Italy would be less compared to other countries with a volunteer military. The correct way to deal with such distortions would be to raise both the compensation component of Italy's military expenditures (and government and GDP) and to adjust salaries to what Italy would have to pay if there were not conscription. I think the Committee rightly thought this was an effort not justified by the likely difference it would make to the numbers or to their audience. Bettina Aten and Heston (1993) were to use the results of this study to extend the estimates of real military expenditures from these eight countries to a much larger group of 134 countries, quite a stretch.

A. The Integration of the OECD into the 1980 ICP

Picard was involved with the coordination of the UNSD with the work being done in Austria, Luxembourg and Paris. The hard work of item matching and coding for the new countries had been completed before I rejoined the UNSD in 1983. The expenditure distribution and conceptual framework was similar to previous ICP rounds. However there was one knotty problem that arose for one of the important participants, the United States.

1. The Problem Posed by the US CPI Revision in 1978

In 1978 the BLS introduced the revised CPI that changed the price collection protocol among other procedures. It was no longer a simple matter of obtaining price collection forms (with outlets deleted) from the BLS or of obtaining special collections outside of catalog prices. In the field the BLS price collectors would have a checklist for one or more *Entry Level Items* (ELIs) which are basically ICP basic headings. The collector would ask a store manager which were the volume sellers, of say soft drinks, and check off the size, type of container, type of drink (sports, carbonated) as well as location and type of outlet. As long as the same item was the volume seller the next month, the BLS had the price ratio between time t_{+1} and t that could be appropriately weighted to get a time to time index. But only by searching the checklist responses could one determine what item was actually priced.⁶⁴

Because the system was fairly new to BLS staff in 1980 and more so to OECD staff, the method to obtain US consumption prices was a trial and error process that was

⁶⁴ As noted the BLS uses the term Entry Level Item in the same way that the ICP uses basic heading. Within the ELI there are clusters from which ICP items are described on the cluster checklist. The checklist provides a set of characteristics or specification for a particular good or service. One can think of ICP items within a basic heading as a clusters within an ELI and a particular item as a cluster of characteristics.

left to Michael Ward.⁶⁵ Ward (Pic 29) had been hired as consultant to work on OECD participation in the ICP and was given the task of obtaining US consumption prices from the BLS, a task that was also very much in the interest of the UNSD. We undertook this work together in a windowless space that the BLS provided for us to extract prices from the checklist responses of Chicago, New York and Seattle. As I recall we spent three long days going through the files and did come up with a large enough number of item price observations to permit 1980 US participation. Because Canadian and US markets are very homogeneous with respect to specifications including brands and outlets, often US prices were linked through Canada in arriving at OECD parities for some basic headings.

In subsequent ICP rounds the work is done by the BLS essentially matching OECD specifications with items in the CPI files. The digital access was much improved so that the work has become less labor intensive I would guess. One of the ironies of the revised CPI framework is that it appears at first to make it much more difficult for the BLS to make place to place comparisons within the United States compared to the previous CPI. However, Kim Zeischang (Pic 40) of the BLS saw that a hedonic variation on the CPD approach would in fact produce estimates of price levels in each of the CPI centers for each ELI. The variation is that the checklist characteristics are added to the right hand side of the CPD equation so as to tighten the specification.

By going through all of the checklists and pulling out prices along with the set of characteristics of the item within a basic heading, one can estimate the basic heading price level for each CPI center. There are weights to aggregate the basic heading parities to totals like food, clothing or all consumption. Zieschang had three researchers carry out this exercise (Kokoski, Cardiff and Moulton, 1994) with Philadelphia as the reference CPI center. Their work showed that the procedure was feasible but it was not taken up as continuing activity by the BLS in part because budget requests to do a special survey were not approved. The reason to request budget for an additional survey was that some in BLS price statistics believed that spatial PPP studies really required a different sampling frame than was used in the CPI. Since it was highly likely that such an additional funding request would be turned down based on past experience, perhaps it was more that the BLS at the time really did not want to be in the business of producing spatial price indexes.

One thing is clear. Some of CPI items that are quite suitable for time to time indexes are not suitable for place to place comparisons. A memo of understanding was reached in 2003 to allow the BEA to begin to make use of the revised CPI data base to estimate regional price levels to convert state incomes to comparable volumes. The BEA team led by Bettina Aten (Pic 19) began to estimate hedonic CPD equations of the type that Zieschang's group in the BLS estimated a decade earlier. (Aten, 2006 and Aten, Figueroa (Pic 50) and Martin (Pic 32), 2011) This is when anomalies began to show up. For example Bar Mitzvah catering for 50 guests could be used in a single center because

⁶⁵ Michael and I first met when I was invited to give a lecture on the ICP while in Tokyo at the United Nations Statistical Institute where Ward was teaching an eight week session on International Statistics.

the collector would be returning to the same catering service each period. However, across the CPI cities it was a checklist specification that apparently leaves room for interpretation because price differences of several hundred percent could occur across CPI cities. However, the CPI data base is so rich that such items could be dropped leaving more than enough overlap across cities of the remaining goods and services that robust estimates of most basic heading parities could be obtained.

(2) Putting Together the OECD and the 1980 ICP World

The ECE holds a Conference of European Statisticians each year and in 1979 they approved a European Comparison Programme in response to the regionalization of the ICP. The ECE only provides a forum for discussing methods and results and does not make estimates themselves. There were 18 European countries in the 1980 ICP, 12 being the Group I countries that were members (or in line to be) of the European Union. Austria was included with Group I to provide a link between the Group I and Group II countries by means of binary comparisons of Austria with each of the Group II countries, Finland, Hungary, Poland and Yugoslavia.⁶⁶ With strong encouragement from the European Union, the OECD agreed to join the 1980 ICP but data collection for Canada, Japan, Norway and the United States did not begin until 1983. As part of the OECD effort Norway became the 18th European country in the 1980 ICP. The ECE issued its report in 1985 and the ICP report was published in two parts in 1986 and 1987 (UNSD and Eurostat). The ICP world comparison for 1980 had to be assembled after most of the regions had completed their work. The arrangement worked out was that the UNSD received basic heading prices for 20 countries representing all the regions. One aspect of moving to regions was that the classification systems were not standard across regions in 1980 so it was necessary to standardize them in order to compute CPDs for example, causing still more delays. The EU system was adopted because the largest number of countries were already using it.

I was at the UNSD by the time the core price data were received by the UNSD so we proceeded to estimate CPDs across the core countries. The next step was to link these to the remaining countries in each region. For example Kenya and Senegal were the two African core countries. We took the geometric mean of their price levels with respect to the United States from the core CPDs and with respect to Africa from their regional CPDs for each basic heading. If the price level was 40 percent with respect to the United States for a basic heading then the price level of each African country would be multiplied by 0.4. The same was done in each region to provide a full matrix of price levels for all 60 countries for all basic headings.⁶⁷

⁶⁶ The European Free Trade Association was formed in 1960 consisting of Austria, Denmark, Finland, Portugal, Sweden and the United Kingdom. Austria, Denmark, Portugal and the United Kingdom were EU members or candidates by 1980. That left Norway and Finland to be linked to the OECD through Austria.

⁶⁷ In the 2005 ICP a better method was used for the linking of the 18 Ring countries to their respective regions. It was developed by Erwin Diewert and is described in the ICP Manual on the World Bank website.

We next aggregated the price levels and expenditures to GDP using the G-K method and super country weights as in Phases I to III. The total GDP for each region was summed up from the G-K results for each country in a region. Fixity within regions was preserved by allocating the regional total to each country according to the distribution in the regional aggregation. This procedure was used in the 2011 ICP and termed the *Country Aggregation and Redistribution (CAR) method*.

B. The 1985 ICP Round: Plans and Implementation

Plans were developed for the 1985 ICP round in meetings in 1983 and 1984 while I was still at the UNSD. Regionalization of the 1980 ICP made it clear that linking the regions was a major issue that required advanced planning so that it coincided with the price collecting activities of countries. In planning the 1985 round it was proposed that there be a set of 30 core countries that make detailed binary comparisons with partners in different regions. Ideally this would have provided for two to three binaries between each pair of regions. In addition Austria would continue to participate in the EU-OECD comparisons along with the Europe Group II; and Japan would participate fully in the ESCAP regional comparison and the OECD comparison. Any of the core participants from poorer countries would need separate funding for their special price collections.

(1) The Best Laid Plans

Missing the academic environment I chose to return to Penn in the Fall of 1984. Fortunately Lazlo Drechsler (1929-1990) joined the UNSD as Chief of the ICP section from 1985-89. Drechsler had a strong background in PPPs having worked in the Hungarian Statistical Office and having written several papers on methodology. He coined the word *characteristicity* as a criteria for choosing items to match in price comparisons and the phrase *comparison resistant services*. Drechsler and I had both participated in planning for Phase V and he fully understood what was involved in implementation of the ICP. Further Lazlo knew his way around the UNSD because he had done work for the ECE in 1964 where he was involved in developing the 1968 System of National Accounts (SNA) and producing correspondences with the Material Product System (MPS) used in the Soviet Bloc countries at the time. So Lazlo joined the UNSD eminently prepared plus he had a close relationship with Kravis, who continued as an advisor to the UNSD. Unfortunately there were a number of constraints facing the 1985 ICP that were not anticipated.

The first problem was UNSD funding for travel by countries to ICP meetings, for assistance to core country candidates for carrying out extra price surveys and for hire of consultants in technical areas like construction and capital equipment. By 1985 the number of possible core countries had gone from 30 to ten and in the end the only binary that was completed was between Kenya and the United Kingdom. The links of Austria to the Group II countries and the OECD remained along with the link of Japan through the OECD to ESCAP countries.

(2) The Latin America Problem

No countries in South America participated in the 1985 ICP in large part because of their experience with the 1980 round. The prices and expenditures were collected in 1979 by research institutes affiliated with ECIEL and not the statistical agencies in their countries with whom the UNSD communicated. ECLAC did update the basic heading parities and expenditures to 1980 but without the national governments participating. So when the 1980 results were released the countries quite rightly asked questions about how the results were obtained, questions to which in many cases ECLAC was unable to respond. My view is that there was little enthusiasm in the countries to take part in Phase V and without their interest ECLAC was in no position to seek funding.

The European Union did support the participation of seven Caribbean countries and these comparisons were completed. However, despite efforts by Eurostat, linking to the other regions was not carried out so in the report (UNSD and European Union, 1994) there were only 56 countries shown in the tables, not the 64 that participated. The seven Caribbean countries were reported separately; and Nepal, which only reported consumption, was also not in the final tables.

(3) Leadership

Lazlo worked very hard to complete Phase V and organize Phase VI receiving little funding from outside the United Nations and also with little support from the then director of UNSD, William Seltzer. Seltzer, a US demographer, was an internal candidate and we overlapped briefly during my UN days. Bill faced many pressures in UNSD to accomplish more with less and he certainly found the open-ended character of the ICP something that was easy to relegate to a low priority. Then in 1988-89 Drechsler contracted cancer leading him to return to Hungary in 1989. He suffered a painful year in Hungary before and died in November 1990. An equally tragic loss to the ICP was the death of Hugo Krinsje-Locker (1928-1990) from acute Leukemia in September 1990. In 1989 Hugo left Eurostat to join Eurocost, an institution that he had helped establish that would bid for contracts to undertake price surveys. They began with a contract to undertake the African ICP price surveys for Eurostat for the Phase VI comparisons, a project that Michel Mouyelo-Katoula directed. Drechsler's position was not filled, one further step of the UNSD away from operational involvement in the ICP. I think by the 1990s there was an influential group of economic statisticians in administrative positions that were ready to see the ICP fade away. Eurostat contracting practices⁶⁸ came under increasing scrutiny in the late 1980s and 1990s, which impinged on morale. Fortunately this did not affect Eurostat cooperation with the OECD and together they were a substantial force for continuing the ICP, even if all their member countries were not.

⁶⁸ The scrutiny came to a head in September, 2003 when Yves Franchet was transferred to another position, ending his 16 year tenure as Director of Eurostat. It was claimed that a part of some contracts was being set aside in a slush fund with a murky audit trail. The defense of the practice was that Brussels was so slow in processing payments that some slack was needed in order for Eurostat to do its work in a timely fashion. No evidence of personal gain was provided unless you count purchasing membership in riding and volley ball clubs and expensive restaurant meals as personal gain.

(4) Picking up the pieces of the 1985 ICP

The two major difficulties to be faced in putting together the 1985 ICP related to linking the regions and at what level of detail. With Drechler's departure the decision was made to ask Eurostat to carry out the computations, which sounds straightforward until one gets to the details. The Caribbean problem was already mentioned and if we had been consulted, we would have suggested that it would be worth a try to take advantage of the fact that Jamaica took part in both the 1975 and 1985 comparisons, which we did in the PWT. In the PWT we also included Nepal for 1985 at the level of consumption, investment and government using similarity of real and nominal national currency shares in GDP and the similarity of the consumption PPP to that for all of GDP. Certainly Antonis Avdoulous, who replaced Gerardi and did the computations for Eurostat for 1980 and 1985, would not have been expected to make those type of adjustments even if he thought they were sensible.

Linking of Group I and II countries in Europe was at a detailed level but not so for Japan and ESCAP, which was done at the level of GDP, a second best solution compared to using whatever detailed breakdowns are available. Another compromise in the 1985 report was to provide detailed PPPs and expenditures for 53 headings of which only 40 were not included in other aggregates, whereas much more detail was available for many of the regions. Probably this decision was made because that number of headings was least common denominator available for most countries, a common enough practice.

Our position has always been to use as much detail as is available for each country but not to show any detail that is implicit in the aggregation procedure that is used. For example, suppose footwear expenditure and price data is available for men, women and children for many countries but only footwear as a total for the rest. We would keep the detail for all countries, even though the implicit expenditures for some countries will be imputed proportionately or as in similar countries for those who have provided only totals. Our argument is that keeping the detail can only improve the comparisons and has no clear downside.

The Phase V report was published in 1994 with even more of a lag than the Phase I report and about the same time that Eurostat-OECD were releasing their 1993 comparison. Not surprisingly this caused much concern about how the ICP should be carried on in the future. We address a number of these concerns in Chapter 8 that covers some other developments of the 1980s and 1990s including the unhappy life of Phase VI of the ICP.

Chapter 8 Exit and Voice

This chapter takes its title from the Albert Hirschman book *Exit, Voice and Loyalty* (1970) that contrasted reactions to deteriorating quality of service by firms, governments and non-profit institutions. The complaints, or *voice* against the United States Postal Service (USPS) declined when United Parcel Service (UPS) provided an *exit* for both letters and parcels. Further the USPS in fact was able to improve the quality of its service as a result, perhaps by the example of UPS. An historian colleague and I administered a major in International Relations (IR) at Penn for many years with no home department, no office, no budget and a program that lacked quality but had a great deal of student interest and *voice* because students did not want to *exit* to another major. One of our students described the lack of administrative support for IR to his parents who rewarded this *voice* with an endowed chair and other funding to improve the IR major, which it did.

This chapter pursues the exit-voice analogy with respect to the ICP experience at the global level, especially in the late 1980s and 1990s. After 1980 when the OECD joined with the European Union in PPP comparisons there was a significant number of developed economies participating. In addition the CMEA countries continued their comparisons until the breakup of the Soviet Union in 1990. At this time Hungary, Poland and other countries sought to join the European Union while the OECD was expanding its technical assistance in national accounts to many of the Commonwealth of Independent States and Eastern European countries. Even while the scope, quality and frequency of Eurostat comparisons was increasing and the OECD was becoming more active, there were core OECD countries that were questioning the program.

A. The World Bank and the ICP in the 1980s

The ICP is directly under the DDG⁶⁹ which in turn answers to the Vice President for Development and Chief Economist, a position created for Hollis Chenery in 1972 that he occupied until 1982. Jean Baneth (Pic 7), a career World Bank employee directed the DDG from 1980 to 1989. Baneth completed a special international M.A program at Yale during my time there so we were acquainted. He initiated some in-house studies of the implications of using ICP results in an operational way and organized a meeting on the subject in Easton, Maryland, in May 1983 that was mentioned earlier. The setting was a very pleasant inn on the Chesapeake Bay but the meetings themselves were less congenial for Summers and myself.

There was a major meeting in Luxembourg in November 1983 on the completion of Phase IV and the organization of Phase V. A similar meeting was held in Bellagio in September 1984 to insure that the core countries for the 1985 ICP were on schedule to complete their binary comparisons. This was my last meeting as a UN staff member before I returned to Penn. At the meeting Baneth offered me a position at the Bank, which at one time I might have been interested, but not then. He asked me to suggest someone else and I immediately recommended Michael Ward since the position among other

⁶⁹ Data Development Group is presently administered by Haisan Fu (Pic 22). It was formerly known as the Department of International Economics and by other names.

things concerned national accounts, which happily Michael accepted. It became clear at this and other meetings that Baneth would continue to say that the ICP was valuable and to provide some resources for related research projects but was quite set against any practical applications within the World Bank.

As mentioned earlier we were not in favor of country loan status being tied to ICP results because we thought it would create incentives for countries to manipulate their price submissions to make their countries look poorer. However, the position of a country in the GDP per capita scale in the ICP depends on the price submissions for all countries. So when one country chooses to price in high end outlets or choose the most expensive brands as India did in 1985, it can indeed make India look poorer. The 1985 estimate, however, showed that Bangladesh was 10 percent better off per capita than India, an unbelievable, unwanted and embarrassing consequence.⁷⁰

However, Irv in particular thought it was inexcusable for the World Bank to distribute freely its glossy World Bank Atlas based on exchange rates, sometimes averaged or adjusted where there were dual markets. We argued that certainly PPP conversions for benchmark countries or approximations for non-benchmark countries were much closer to the “truth” than exchange rates. When Stanley Fischer became Chief Economist in January 1988 Irv’s pleas found a more sympathetic ear. And by Spring of 1989 Baneth was out of DDG and instead was director of the Geneva Office of the World Bank. His successor, John O’Connor, was from the IMF Statistics Office and not particularly sympathetic to the ICP. He definitely was not friendly to the PWT, which was being put to use in some international publications like the Human Development Report. At some time during John’s period as head of DDG a policy was called to my notice that if available international organizations would only use each other’s estimates and not those of outside sources.

B. The Philadelphia Meeting, August 1990

Twenty-Five Years of the ICP: A Review and Future Plans was the name of the meeting held at Penn with the support mainly of the World Bank and with some logistical support of the UNSD. Kravis served as meeting convener and the venue was the newly completed Executive MBA center of the Wharton School that was next to the Economics Department building so at least for us the arrangement worked well. William Seltzer was Director of the UNSD from 1986 to 1994 and as mentioned earlier had in our view given little priority to the ICP. Seltzer attended the meeting along with two other staff while O’Connor, D.C. Rao, Michael Ward and Sultan Ahmad represented the World Bank.

⁷⁰ India is but one example. China was not in the ICP at this time, but during the decades when its reported annual growth rates were 10 percent or higher and in the United States they were 3 percent or lower China by official figures never gained on the United States. And when China participated in the ICP they also sought to appear poorer than they were. Sometimes countries want to appear better off for national pride I suppose. When Italy revised its national accounts upward by over 10 percent a representative proudly declared they were now above the United Kingdom per capita to which the United Kingdom representative replied that Italy could now contribute more to the EU Social Fund.

Antonis Avdoulos, Pietro Benedetti and Kanti Munnsad of Eurostat attended while David Roberts was the only representative of the OECD. Munnsad, who coordinated the African comparisons of Eurostat, in affect represented Africa.

Loh Mongkow represented ESCAP, Pedro Sainz- ECLAC, Michael McPeak and Frank Orlando the Inter-American Development Bank, and Simon Nocera the IMF. Katrina Reut and Michelle Vachris of BLS, who provided US prices to the OECD, and Hermann Haberman from the Office of Management and Budget, which is the closest thing to a national statistical office, also attended. Albert Franz from Austria, the coordinator of the Europe Group II countries in Europe, and Drechsler took an active role in the discussions. There was an agenda involving a review of previous benchmarks, reports from the regions, discussions of methods, and issues in linking the regional results. The ICP handbook draft that Drechsler had initiated was discussed along with the major issue of funding.

A sub-agenda concerned the overall organization of the ICP and what should be the role of the UNSD. Phases IV, V and VI all suffered from the fact that the UNSD was tasked with coordinating the global ICP comparison which required more initiative, funding and management than the UNSD could muster. When the ICP began to be built up from regional comparisons, strong central coordination was needed to insure that there was common coding of items and global ICP basic headings across regions. And if regions chose to have a smaller or larger number of basic headings than the global norm, they would be consistent with the global ICP number. These conditions were only achieved in Phases I-III of the ICP and in the 2011 ICP.

While the meeting endorsed this role for the UNSD, it was not all clear that it would happen given the leadership in the UNSD and the importance attached to the ICP by the UN Department of Economic and Social Affairs. Hermann Haberman, who attended only the last afternoon of the meeting, also expressed a very negative view of the whole program. Since Haberman was to succeed Seltzer as head of the UNSD in 1994, this was not an encouraging omen.

Another tension that became clear at this and many other meetings was between the statistical offices providing the ICP data and the potential users of the results. Most potential users were growth or policy economists that normally were unaware of the participation of their own statistical offices in the ICP. The argument that the ICP could be justified in a country because it enhanced their statistical capacity both in national accounts and price indexes was generally accepted especially for lower income countries. Another plus for the ICP was the growing number of journalists that began to use PPP conversions for consumption or GDP comparisons in their articles lending support for the benchmark comparisons. My overall view of the meeting was that there was reason to be pessimistic about global ICP prospects including its own self-inflicted wounds, for example, poor communication with countries and untimely dissemination of results, and the negative views of some statisticians, who rightly found it a difficult program to administer in a way as to make their constituents happy. And reason to be sanguine about the future of ICP because of the very solid programs of the European Union and the

OECD and the growing use of PPP converted aggregates by the media, textbooks, and researchers.

As noted Drechsler succumbed to cancer just three months after the meeting. Further Irv Kravis was diagnosed with Parkinson's disease in 1990 leading to a somewhat reduced activity level in terms of advising the ICP.⁷¹ For example, Irv had been invited to attend a conference in Nagoya, Japan in November, 1990 but elected not to go and I attended in his place. Kravis had retired from teaching but continued to attend the office at least three days a week until late in 1992. He was designated a Distinguished Fellow of the American Economic Society in 1992 and died at the airport in Philadelphia where he was catching a plane to the Annual Meetings to receive the award. The ICP lost its founding father, Penn and its economics department lost a very good citizen, and all who knew him lost a warm, congenial and effective human being.⁷²

An ICP Handbook had been given high priority by the UNSC as were all technical manuals related to the activities of the UNSD. The ICP Handbook was begun by Hugo Krijnse-Locker and Lazlo Drechsler and after their deaths Michael Ward had brought together a first draft that was distributed to a wide group of PPP experts. I was then asked to revise the Handbook taking into account the comments on the first draft. The Handbook was published in English in 1992 and subsequently in the other five UN languages. As events unfolded little use was made of the Handbook because, except for the ECLAC countries, it was late for the 1990-93 benchmarks and by the time the 2005 benchmark was underway a much more thorough version began to appear online.

⁷¹ Irv and Lilian Kravis had four children, none of whom took up economics, but two did follow Lilian, herself a distinguished pediatric allergist, into the medical profession. In addition to standard medications, recently treatment of Parkinson's has included emphasis group events such as singing or dancing, and generally keeping active. Lilian pushed Irv very hard to keep active, which was certainly in the right direction.

⁷² Irv had a very good sense of humor and knew when to use it, as the following illustrates. Much to his irritation a group of faculty, including myself, had petitioned to remove the then provost for not consulting them on a unilateral change in the rules governing graduate student fellowships. A typical academic tempest in a teapot. As president of the faculty senate, Irv, having cut short a meeting in Luxembourg on the ICP, returned to chair a meeting of the Senate to determine the degree of faculty support for the petition. It was a tense meeting broadly breaking down into a senior group of faculty supporting the provost, quite a distinguished scholar, and a younger group of faculty. At one point a senior medical school department chief called out one of his younger colleagues for supporting the petition at which point Kravis said that we need to talk about the immediate issue as there is not enough time for us to settle within department differences between colleagues. After that the meeting did focus on issue at hand and voted to censure the provost who then resigned.

C. A Digression on Kravis, Summers and Penn

The University of Pennsylvania today is an Ivy League institution within a group informally known by that name as early as 1933. Whereas the Ivy League is now known as an elite group of institutions with high academic standards, this was only partly the case prior to the 1950s. For one thing standards of admission were much lower in earlier periods with strong preferences for athletes, alumni and men. Several schools were all-male into the 1960s though often associated with sister schools who similarly began admitting men. Athletics was a big business with Penn being nationally competitive into the 1950s and *the Game* between Harvard and Yale a major athletic event for other than students and alumni.⁷³ In 1954 the Ivy League became a Division I athletic conference with the agreement that scholarships were not to be awarded to student athletes. This meant that top athletic prospects generally went to other colleges and universities and the Ivy League was no longer competitive in football, the big money sport. From 1948 to 1953 Harold Stassen was President of Penn which he used as a stepping stone for his candidacy for the Republican nomination for President. This reflected the nature of Penn's Board of Trustees all of which slowly changed after 1954.

Penn and Philadelphia had some special features notably a Quaker influence but a history of discrimination as captured by Digby Baltzell in his popularization of the term WASP (White Anglo-Saxon Protestant) in his writings about Philadelphia. Digby was a popular member of the Sociology Department at Penn, who I knew and with whom I played squash from time to time. When I joined Penn in 1962, Digby was incensed that one of his young colleagues was prevented from joining the Philadelphia Cricket Club because he was Jewish. A number of departments were Waspish at the time but not so Economics, which thanks to some open and supportive Quaker faculty in the 1930s, had a number of notable Jewish students and faculty at the time included Kravis, Kuznets, Summers, Levine, Malenbaum, Weintraub, Bloomfield and Larry Klein.⁷⁴ As noted, Milton Gilbert had been a graduate student at Penn, and before going to OECD, had shaped the National income program at BEA with much support from Kuznets.

It was a small social faculty in the 1960s and Wilma and I and other WASPs were a part of weekend dinners and other events all of which slowly faded away in the 1970s

⁷³ Penn played football at Franklin Field which for many years also hosted the professional Eagles of Philadelphia. Lest one think all was athletics, contemporary with construction of Franklin Field the University built a large museum across the street housing major archeological artifacts from Alaska, Sumer and other Middle East sites as well as Central American excavations.

⁷⁴ During the McCarthy witch hunts in the 1950s Dan Thorner an economist specializing in India had been a victim who left Penn in 1952 for India pursuing research, publishing 3 books on Indian Agriculture, teaching in India as well as advising the Planning Commission. In 1962 he joined the *École des hautes études en sciences sociales* in Paris where he finished his career. Klein, who had joined the communist party in the 1940s, and was for this reason was denied tenure at Michigan in 1954 after which he joined Oxford. Kravis was instrumental in recruiting Klein to Penn in 1958 where he was awarded the Nobel prize in 1980.

as the faculty grew larger. I was also half time with a South Asia department that had been created almost single handedly at partition and Indian independence by a well-known Sanskrit scholar, W. Norman Brown.⁷⁵ Norman and his wife Helen, who had spent several years in France during and after WW I, were both Quakers, and their cocktail parties brought together a very diverse group. Norman was well known for his capacity to drink martinis and his great knowledge of baseball.

As I got involved with the ICP, I came to know Bob and Irv much better both socially and in a research environment. Irv, if not a mentor, was a model of the focus needed to see research through to its completion. Irv would be willing to hear out some of the suggestions that Bob and I would offer, but never lost sight of the goal and when it should be met. When Irv and Lilian Kravis traveled with the Summers, Bob and Lilian were quite laid back about time whereas Anita and Irv would display anxiety or even some irritation with the others if they thought they might be late. Opposites attract I suppose.

When the research was going badly as it did in Black February, Irv had a characteristic mannerism. He would slowly take off his glasses and systematically rub his eyes and then slowly return them to his face. If a meeting was moving toward a decision he was against or someone made a proposal that he thought unacceptable he would go through the same ritual. I never knew whether this was Irv's way of stifling an outburst, his way of collecting thoughts, or simply an automatic reaction to an unpleasant development.

Bob and I did our share of university service but nothing like Irv who I do not believe ever turned down a request to serve on or more often chair committees outside the department. He was also instrumental in moving the economics, political science and regional science departments outside of the Wharton school in 1970. When Wharton was founded in 1881 it was the first business school in the United States and included an economics department, a geography department (becoming regional science under Walter Isard), and a political science department. Irv's argument was that the three departments mainly taught students in Arts and Sciences with perhaps 20 percent in Wharton, with undergraduates applying separately to each school. Kravis's position was it made educational sense for the departments to be in Arts and Sciences. Wharton colleagues responded that it would be a costly move for our pocket books, and they were right.

⁷⁵ When I joined Penn there was a great deal of collegiality across schools and departments and members of the Wharton Faculty and the Department of Oriental Studies, for example, often had lunch with each other. George Taylor in the Wharton industry department, who was nationally known in part for negotiating the settlement of the national steel strike of 1959, once was attending a University Museum lecture by Samuel Kramer, who was an eminent Sumerian scholar and an entertaining speaker. At the reception after the talk someone cornered Taylor and asked George why isn't a famous labor specialist like you called "eminent" like your friend, Kramer? Alas, this kind of interaction also faded away during the 1970s as the faculty grew in size.

There are many stories about Bob's way of illustrating points in his lectures that often stuck in the heads of students. He liked to illustrate moving to a limit by likening it to emptying a tube of tooth paste where no matter how much you squeezed there was always an iota left in the tube. One engineering student long after graduation sent him a tube that she claimed had been vacuumed empty. He would also illustrate tradeoffs in economics by saying on the one hand this, and the other that, where he would place one hand under his suit ala Napoleon.

Bob and I had many discussions together with our research assistants and he had two pedagogical tools that he used to get the assistants more involved in the research. One was to ask a new recruit to carry out some calculation and when they discussed it Bob would compare it to the result that another assistant had previously carried out. Not surprisingly this irritated some new recruits but usually they came to appreciate the importance of checking their results. The other ploy was to ask an assistant to carry out a calculation and to sit down with Bob and the results. Bob would not look at the output but rather ask the assistant what they expected to find, and why, which typically required considerable discussion. Only then would they both look at the output together. In these conversations and many other contexts he would often say "I would have thought..." which usually meant that what you previously said was at best partly right or outright wrong.

The three of us had a pretty good sense of when a result was doubtful or simply could not be correct, which Bob referred to as the *sniff test*. In the *Making of Index Numbers*, Irving Fisher (1922, p.353) lists a criterion for the plausibility of an index number formula as *Absence of Freakishness*, which when dealing with a very large number of formulas, as Fisher did, is like Bob's sniff test. The value of emphasizing the plausibility of results has become more important because students tend to think that if a result comes from a computer computation then it cannot be wrong. Whereas data input, output statements and coding are but a few sources of freakishness.

Bob described some of his experiences growing up when he and his brother Paul (Samuelson), who was seven years older, were together. Paul would ask Bob to argue one side of a debating question and after he had lost, they would switch sides, and Bob would lose again. Tough brotherly love. Bob and Anita instilled in their three sons at dinner and other occasions the notion that they needed to be able to argue for their ideas, and the notion that when something is big or small they needed to ask the question, *compared to what?*⁷⁶ The eldest, Larry, was on the debate team in high school, and I got to know both Larry, and the youngest, Johnny, in those years playing tennis. As Larry became known for his economic skills, Bob described being introduced to George Schultz at an economics conference where Schultz said "oh yes, you are Larry's father". In relating this it was clear Bob was proud of Larry but also thought it humorous he was no longer just Paul's brother but had a new distinction. Not many could wear both hats so gracefully.

⁷⁶ Nancy Birdsall gave an endowed lecture at Penn in honor of the Summers', and referred to their sons saying one became a doctor, one a lawyer and the third went into the family business.

Ending this digression on a personal note, my marriage to Wilma (1935-2016) broke up in the late 1980s. By the early 1990s Bettina Aten and I had begun living together and became domestic partners ten years later. We met on the squash courts and I was later to learn after multiple defeats that she had been five times Brazilian women's champion. After completing her MBA at Penn she returned for a Ph.D. in regional science and began an academic career before joining BEA in 2003. While in academia she continued to work on PWT and did all the programming and computing of PWT 6.0 in 2002.

D. The Unhappy Life of the Phase VI Global Comparison

In the planning stage of Phase VI the reference year was to be 1990 but it soon became apparent that only the European Union and the OECD would make that date. (The 1990 comparison was the last of the five year EU-OECD comparisons. The three year comparison cycle started in 1991.) There had been a major effort to develop a core set of prices that would be collected in all the regions. Sultan Ahmad and I were involved with drawing up a core list and codes. However, there had never been an agreement about countries providing their prices to any entity other than the region. The UNSD was supposed to be the central coordinating office, but no initiative was taken to obtain prices directly from the countries or regions as far as I can remember. Perhaps this was because prices were never available for all regions for the same year. So despite building a blueprint for having a large number of item prices for common items collected by countries across all regions, Phase VI essentially became a series of regional comparisons.

1. The Regional Comparisons

The sequential character of the regional benchmarks posed major problems for producing a global comparison. That said, the regional comparisons represented a larger and more balanced group of countries than any previous benchmark, a total of 115. For the first time ESCWA was able to complete comparisons for nine countries, counting Egypt and Palestine. And the ECLAC region was well represented by eight countries from Latin America, Panama and twelve from the Caribbean. Conveniently EU-OECD comparisons were on a three year cycle so their results could be used for either a 1993 or 1996 global comparison.

The African Comparison for 1993

The report for Africa (Eurostat, 1996) provides approximately 200 basic heading deflators and about 50 aggregations to summary headings. The basic headings follow Eurostat's for EU countries with one or two exceptions like own construction. The aggregations are carried out using both the GEKS and G-K methods and the results are provided at the level of 50 aggregations. The authors were Michel Mouyela-Katoula and Kantilal Munnsad.

The ESCAP Exercise

The ESCAP office in Bangkok was the operational center for the 1993 regional ICP. Fourteen countries were full participants in the 1993 benchmark, while Malaysia and

Laos only partially participated. China participated in some city comparisons and India did not take part, I suspect because of results of the 1985 comparison where Bangladesh had a higher per capita income than India. As mentioned earlier India appeared to have provided some unusually high prices in certain basic headings that had the effect of lowering India's 1985 PPP converted GDP. The Japanese government provided financial assistance for the 1993 comparisons and provided a consultant to ESCAP. The World Bank and the ADB both provided support for meetings and travel. And China hosted two meetings, the last to discuss the initial results of the comparisons in Beijing in August, 1997.(Pic 49)

In the initial ESCAP planning, Shanghai was to make a binary comparison with Tokyo and Guangdong province with Hong Kong. The price collection was completed for both cities but neither the representative from China or Japan approved of the results so they were not reported.⁷⁷ The comparison of Hong Kong and Guangdong was completed and included separately as a binary comparison in the report. (ESCAP, 1999, p.35). This comparison is affected by the how well the provincial income accounts reflect the weights of consumption and capital formation. The published price level of Guangdong was 63 percent of Hong Kong in 1993, which is probably too high.⁷⁸ Readers should probably know I was heavily involved in the preparation and writing of the report.

The European Union, the OECD and Associates

Fortunately EU-OECD comparisons were on a three year cycle so whether the reference year was 1993 or 1996 for Phase VI was not of great moment. In fact the only attempt to produce a global result was carried out at the World Bank for 1996. An advantage of a later year was that the OECD was incorporating comparisons with the countries of the former Soviet Union that comprised the Commonwealth of Independent States (CIS). And between Eurostat and the OECD many of the former CMEA countries were also incorporated in the comparisons. All of these countries were adapting to market versus state set prices and to the SNA versus MPS system of national accounts so the experience they were gaining each year was likely to improve the quality of their comparisons within the OECD. A total of 51 countries, including Israel, were covered by Eurostat and the OECD in 1996.

⁷⁷ Because the consultant that Japan provided ESCAP had to return to Japan before all the prices were received, I was asked to undertake most of the calculations and write-up for all the comparisons. As a consequence I had seen the Shanghai-Tokyo binary and thought it should have seen the light of day.

⁷⁸ The estimated price levels of consumption, 58 percent, government, 27 percent and investment, 77 percent show the relative variation that one would suspect. However, the investment share in Guangdong and most of China was regarded as too high by many outside observers. Further, in consumption goods there were some anomalies, for example, reported clothing prices in Guangdong were higher than Hong Kong, the former being an exporter and the latter an importer. Interestingly, China reported relatively high prices for clothing in the 2005 ICP too.

The Economic Commission for Latin America and the Caribbean

The Caribbean was well represented in the 1996 comparisons, in part because of substantial support from the European Union and the World Bank. I attended meetings in Kingston, Jamaica, and Port au Prince, Trinidad, and was impressed with the competence of many of the statisticians. However, because the countries are so small, statistical offices are always spread thin across a number of projects. The only Central American country to participate was Panama, which took part in the Latin American comparison for 1996. With the addition of these 21 countries and those of EU-OECD over 60 percent of the countries had a reference year of 1996.

The Economic and Social Commission for Western Asia

The 1993 regional report (ESCWA, 1997) was a first for the region and a significant joint statistical effort of the participating countries. It helped of course that all participants were Arabic speaking and that the World Bank contributed to support for meetings and publication. As was the case with Africa the aggregations were done with both the G-EKS and G-K methods which produced some anomalies. The countries in Western Asia that have high per capita GDPs have small populations like Qatar and Bahrain, whereas Egypt has by far the largest population followed by Yemen and Syria. In Western Asia the total GDP of these three countries is larger than the more affluent per capita income countries so the international prices are closer to the less affluent countries. As a consequence the differences between the G-K and GEKS aggregations show no consistent pattern.

2. Official Global Report on 1996 ICP: a no-Show

The World Bank staff in DDG made an effort to put together a global ICP table from all of the regional PPPs that had been developed from 1993 to 1997. To do this it was necessary to update PPPs for Africa, Asia and Western Asia from 1993 to 1996. Then the ECLAC countries had to be linked in with other regions. With the cooperation of the BLS, US prices were obtained for most of the consumer items priced in ECLAC countries thereby providing a link at the basic heading level. Next, the basic headings were collapsed to the lowest common denominator, 31 basic headings, which for countries with more detailed aggregations were often summary headings built up from several basic headings. Again, PPP and expenditure detail was lost for some regions by collapsing basic headings with no obvious gain in precision at the global level.

We jointly carried out with the DDG some aggregations with the PPP and expenditure data for all 115 countries at the 31 basic heading level. Our calculations were done by Bettina Aten using code she had written and tested for some years. However, for reasons we could never fully resolve the results obtained by the DDG did not match with our results. In some instances the DDG staff had made new assumptions or updated the input data often with more recent national accounts. In the end, with what we understood to be a data set on which no more revisions would be made, we still did not match. In any event, at the DDG or higher level it was decided to only use their input data for internal use or to supplement the World Development Indicators, a widely used data base provided by the World Bank.

Thanks to easy digital access and the burgeoning growth modelling literature, the PWT was enjoying increased usage not only in academia but also by researchers within the World Bank and the IMF. John O'Connor and others in international organizations were not comfortable with a data set like the PWT being preferred by many for their research needs. Various attempts were made to reconcile the differences between the PWT and the time series of GDP published by the DDG, so as to make the PWT redundant, or at least that is my interpretation. In any event, the fact that the PWT, for the better we thought, broke down GDP into consumption, investment, government and the net foreign balance, was not something DDG was ready to do, no rapprochement occurred. As to the fate of Phase VI, we used the DDG data base in aggregations underlying PWT 6.0, and as was our practice, we provided the DDG generated 115 by 31 matrices of PPPs and expenditures on our website. This 1996 data set has had a large number of downloads between its availability and the distribution of the 2005 ICP results in 2008.

E. The Castles and Ryten Reports of 1997

Personally, I was not acquainted with anyone involved with instigating either of the reports so my account relies on unnamed but, of course, reliable sources. C. Louis Kincannon became the first Chief Statistician of the OECD in September 1992 until he returned to the US Census Bureau as director in 2000. Hermann Habermann was the Director of UNSD from 1994 to 2002, when he returned to Washington to join Kincannon at the Census Bureau. For reasons never clear to me, Habermann had a very negative view of the ICP which he shared with Kincannon. My understanding is that both men would have been happy to see the OECD and the UNSD programs scuttled. To repeat, my interpretation of their antagonism (and that of Seltzer) is that the ICP requires a different type of cooperation between countries than do other statistical activities. There are common statistical practices across countries with respect to censuses and surveys and their analysis but in the end it is a national activity. By contrast the ICP is an international activity requiring cooperation between countries, particularly with respect to price collection. A related point is very well formulated in Ryten (1997, p.29), "In the case of PPPs, there can be no knowledgeable domestic critic, for the simple reason that it is only from an **international perspective** that the data can be critically appraised."

Another reason that the ICP was a difficult sell to country statistical offices was that the staff tended to have strong backgrounds in statistics but less so in economics. We often found that planning ministries, finance ministries, central banks were not familiar with the ICP program nor that their countries were participants. Put another way government economists who might be interested in the ICP results in terms of actual use were not in the same network as those producing the results. And perhaps more important is the point made by Angus Deaton that the results are most used by the international organizations and researchers, so there is good reason for country skepticism about claims for the value of the ICP to the countries.

1. The Castles Report

Within the OECD there were frequent questions by member countries about the quality and value of the PPP program that supported the idea of appointing a consultant, to review and evaluate the EU-OECD program. While Kincannon thought it a good idea, he was hardly alone, just as Habermann could find country support for a review of the ICP for the UNSC. The consultant chosen for the EU-OECD review was Ian Castles (1935-2010), who had served with the Australian Treasury in 1958, became Secretary of the Ministry of Finance in 1979 and was appointed Australian Statistician in 1986. Castles was a very well qualified consultant from a country that had questioned its participation in the OECD program, so the appointment was generally welcomed.

Castles interviewed national statisticians in Australia, the United Kingdom, Canada and the United States, and staff at the World Bank, Eurostat, the UNSD, various departments within the OECD, and Bob Summers⁷⁹. He agreed with many of the common criticisms of the ICP with respect to handling of government services including education and health, construction and machinery. Overall, Castles (1997) concluded that the OECD program should be continued but with more resources and with the aim of improving problem areas. One surprise to those less friendly to the ICP was the strong support within the OECD of staff in departments outside of statistics. In terms of the EU-OECD comparisons the Castles report was supportive.

2. The Ryten Report

The 1995 meeting of the UNSC it was “(a) Agreed on the need to conduct an evaluation of the International Comparison Programme (ICP) to address the reservations held by some countries about ICP implementation and the uses of ICP results, and to seek ways to improve the credibility of ICP data;” and “(b) Appointed a steering committee to supervise the evaluation process.” Subsequently, a document (E/CN.3/1997/3/Add-1) was prepared for the 1997 UNSC setting out the terms of reference for the evaluation drawn up by the UNSD, Eurostat, the World Bank and the IMF.⁸⁰ These organizations financially supported the consultant who was chosen from a short list drawn up by the same organizations. Further they requested that the evaluation take into account the Castles report.

⁷⁹ For reasons forgotten, I was not at Penn the day Castles came to Philadelphia. Bob said he asked good questions and he enjoyed the interview but was not at all sure what Castle’s review would say. I never met Castles but corresponded by email with him and David Henderson, who became active supporters of PPPs with respect to valuing the economic effects of climate change. I had written a paper for a conference on the International Panel on Climate Change that was critical of the use of exchange rates in judging the costs of climate change that Castles and Henderson liked. We kept up a correspondence that I have continued with Henderson, who it should be noted was chief economist of the OECD Economics and Statistics Department from 1984 to 1992 after an academic career at Oxford and University College London.

⁸⁰ The new interest of the IMF is discuss below.

The consultant chosen was Jacob Ryten, who did his graduate work in economic statistics at the London School of Economics and spent most of his career at Statistics Canada where he retired as Deputy Chief Statistician in 1997. Neither Bob nor I met Ryten prior to completion of the report though we had been advised to expect the worse. Indeed the Report was highly critical of recent ICP rounds and of the ICP manual, which in retrospect seem appropriate. His style was often dramatic in the sense that there might be dire consequences of not following his advice, which was often quite specific. And some have suggested that the report was a little self-serving in that it proposed there be a high level global coordinator of the ICP, a position coinciding with his retirement. But his recommendations were much like Castles in that they advocated continuation of the ICP with more resources, more communication with the countries, a manual directed at the operational needs of countries, and strong support for regional coordination. And above all to not let the ICP languish as it had done the previous decade at least.

Needless to say, Bob and I were most pleased with the bottom line of both reports. The Ryten report in particular, must not have been well received by Habermann, particularly its call for UNSD to play a major role in a renewed ICP. Subsequent UNSC sessions beginning in 2000 were annual and most had documents on ways for the new ICP to be organized.⁸¹ I was not privy to all that took place in response to the Ryten report until a major meeting took place at the World Bank in 2002.⁸²

Chapter 9: Finding a Home for the PWT, Part I

By the 1990s and Bob and I began to seek new blood to take over the PWT at the same time the use of our table began to enjoy much wider use. In fact, what had begun for us as an academic exercise had become a data bank that was being sourced by academic and other researchers so we began to feel a responsibility for maintaining the PWT. Bob was near retirement from teaching and quite ready to leave day-to-day maintenance of the PWT to others. Beginning in the 1980s we received research grants

⁸¹ The UNSC has 24 members, five from Africa, four each from Asia, Eastern Europe and Latin America, and seven from Western Europe and others. Members serve four year terms and every two years a new chair, three vice chairs and a rapporteur of the UNSC are chosen. There evolved a group called *Friends of the Chair* of the UNSC, with special interest in the ICP that was important in establishing a framework for the 2003 ICP.

⁸² David Roberts was privy to these events and notes, "There is an important difference between Castles and Ryten. Castles was pro PPPs and when Australian Statistician supported Australia's participation. Ryten when in Statistics Canada was very anti PPPs and did much to hinder Canada's participation. When he was appointed consultant we assumed the ICP would be given the kiss of death. His Damascene conversion confounded many." Subsequently Castles and David Henderson a former Head of Economics at OECD were both very active in promoting use of PPPs in studies on the consequences of climate change. They liked a paper that I had written on the subject and I came to know their views after that and certainly agree with Roberts' evaluation.

from the National Science Foundation (NSF) that supported research assistants, travel and summer stipends for about 30 years.⁸³

The ICOP program at Groningen expanded during the 1980s and 1990s with a number of graduate students doing dissertations related to productivity levels and growth in OECD countries, Asia and Latin America.⁸⁴ Marcel Timmer joined the Groningen faculty in 1988 working closely with Bart van Ark, a protégé of Maddison. Bart began his affiliation with the Conference Board in New York in the 1990s finding another home for productivity studies.

A. Initial Efforts to Pass the Torch

Daniel Nuxoll (Pic 7, 17) wrote his dissertation at Brown University, which had used ICP data to investigate some index number issues. He was teaching at Virginia Tech when we invited him to be included in an NSF grant and to see how well we worked together. He took part in PWT 5.5 in 1993 and we wrote a joint article on the Belassa-Samuelson effect (Heston, Summers and Nuxoll, 1994). However, Dan was not that enthusiastic about life in Blacksburg, Virginia, especially when his prospective wife lived in Washington. So he chose to work with the Comptroller of the Currency in Washington on loan risk and moral hazard, a long way from the PWT.

We had also had informal discussions in the late-1990s with the World Bank and the IMF about their interest but we concluded they did not want to make the necessary commitment at that time. Even so in their presentation to the UNSC in 2000, the World Bank paper said that they would be absorbing us into their work program. I believe this document was written by Yonas Biru, who had taken over from Sultan Ahmad after his retirement the previous year.⁸⁵

⁸³ Bob's son Larry kidded us as NSF awardees we were the Hoyt Wilhelm, a very good knuckleballer who pitched well into his 40s; and George Blanda, who was a quarterback and placekicker for 26 years in professional football. These awards were of course a validation of our work that we warmly welcomed.

⁸⁴ While the Penn economics department once had faculty like Kuznets and Richard Easterlin who were well known for their empirical research, the faculty beginning in the 1970s had begun concentrating on game theory and other specialties well removed from ICP-PWT type concerns. So by the time our research assistants were at the stage of choosing a dissertation topic, any work related to our research was not an attractive choice. And it is doubtful if any new faculty member would want to be part of our team if they did not already have job security. Mention should be made of two of graduate assistants Mark McMullen (Pic 16) and Sean Dougherty (Pic 51), who did not finish the Ph.D. program but of whom we had a high opinion and both have successful careers.

⁸⁵ Biru was quite ambitious and initially received support from the Director of DDG, Shaida Badiie (Pic 47), who herself had only been appointed in 1997. When he joined DDG, the staff included Michael Ward, Sultan Ahmed, and Yuri Dhikanov. Yuri was hired as a consultant in 1997 by way of Penn. Andre Schleifer, his advisor at Harvard, called Summers in spring 1997 and said he had an offer we could not resist, namely to take on Yuri for the summer because he could not get enough of PPPs, not an interest of anyone at Harvard,

We also tried to sell PWT to the National Bureau of Economic Research (NBER) and the then President, Marty Feldstein, suggested involving Robert Feenstra (Pic 24,25) at the University of California at Davis, which we did. Subsequently Feenstra and I organized a cooperative annual workshop beginning in 2004 termed simply *PPP Workshop*. The first session included Feenstra, Bob and I, and Bart van Ark, Marcel Timmer and of the University of Groningen. In addition, Prasada Rao at Queensland in Australia, John Romalis (Pic 32,33) then at Chicago, and Robert Hill at Graz University in Austria were included in our NSF proposals and funding. Marshall Reinsdorf and Bettina Aten of the Bureau of Economic Analysis, and Erwin Diewert (Pic 30,33,40) of the University of British Columbia also attended the first PPP Workshop, which is now in its thirteenth year.

B. China and the International Monetary Fund: Any Port in a Storm

As criticisms rained down for slow implementation, poor linking of the regions in Phases V and VI, lack of an adequate manual, and a weak leadership structure at the UNSD, the global ICP ship appeared to be taking on a lot of water. The European Union and the OECD did a lot of bailing, the World Bank somewhat less, and the ICP ship did not sink in the 1990s, but it needed help. And the first rescue vessek on the horizon flew an IMF flag.

The IMF disseminates each year the *World Economic Outlook* (WEO) that provides estimates of economic growth in the previous year and estimates for a few following years. Up to spring 1992 the growth rates were aggregated for regions and the world using exchange rate (a three year average) converted GDP as weights for each country. The IMF staff were aware that growth rates in Asia seemed very low given the very high rates of GDP growth that were observed for China since 1978 and were generally confirmed by the media and other observers. Further the very low rates of growth in Europe and the United States were receiving a very large weight in the index for the world. Their conclusion was that the country weights should be the national GDP converted at PPPs, which was initiated in the May 1993 WEO. Did Bob and I agree? Yes, yes, yes! Were Bob and I surprised? After the lukewarm support of the IMF staff to PPPs in the past, we were not just surprised, we were astonished. Was this a game changer? Yes, particularly because the DDG at the World Bank had little choice but to give more prominence to PPPs.

except perhaps Larry Summers. We did so and during that summer he met John O'Connor who offered him a consulting position at the World Bank. Yuri was not interested in supervisory work and Michael Ward was quite involved with national accounts advising, so when Sultan retired in 1999 Yonas became the face of the ICP. In fact when the position of Global Manager of the ICP was advertised in 2002, Ward among others was advised not to apply because of his age, but Yonas did apply though he was not really considered because of his lack of managerial experience. In fact an agricultural statistician, Fred Vogel, who incidentally was older than Michael Ward, was selected. I received the impression that the decision makers at the time wanted few if any voices from the past involved in the 2003 ICP, and it was only through the persuasion of Michael Ward that I was asked to be on the TAG.

Where did the IMF obtain its PPPs? IMF (1993, p. 165) says, “The PPP-based are derived from PPP estimates of GDP from the International Comparison Program (ICP), supplemented by World Bank and IMF estimates for countries not covered by the ICP and for China.”⁸⁶ Bob Summers did inquire further with IMF staff and did not receive any more details about countries other than China. To illustrate the magnitude of the change, the weight of the industrial countries declined from 73.2 to 54.4 percent while the share of Asia went from 7.3 to 17.7 percent when moving from exchange rates to PPPs. Overall with this change WEO world growth estimates rose by over half of one percent.

C. Rise of Empirically Tested Growth Theories

Ironically, we came to the view that the PWT and the ICP were much more complementary than competitive with each other especially in the 1990s when usage of the PWT greatly expanded. What was the source of this expansion? First was the publication of PWT 5 (Summers and Heston, 1991) that covered 1950 through 1988 using data from the 1985 ICP for 56 countries. Only the reference year 1985 was published in the article and diskettes were freely available from the NBER so it was much easier for users to access directly to their computers. Coincidentally, or we like to think because of the availability of the PWT, Robert Barro published in the same journal issue (Barro, 1991) his much cited article on economic growth and convergence of countries.⁸⁷ There followed too many papers testing models of convergence on versions of the PWT so that claims of significance to any particular growth formulation became questionable.⁸⁸ We were not complaining of course, and were duly rewarded by being designated as Distinguished Fellows of the American Economic Association in 1998.

⁸⁶ The estimate for China was criticized in the Wall Street Journal (6/1/1993) where Bob Summers was quoted as saying “It ain’t science with a capital S”. The IMF number increased China’s GDP by 3.5 times its exchange rate number or a price level of 28 percent. It placed China with the third largest output in the world, while the PWT estimate had China above Japan at number two. We re-examined our sources and concluded with Angus Maddison that China’s growth rates were too high producing extrapolations of estimates for 1975 and the 1980s that were in turn too high. In our data appendices to the PWT we provided extensive write-ups of our treatment of China in each revision of the data.

⁸⁷ Both papers were based on presentations at a conference in spring 1990 in Park City, Utah. Kim, Morse and Zingales (2006) tabulated citations of articles published in 41 leading journals between 1970 and June 2006. Of the 146 articles receiving more than 500 citations, Barro’s article was number 33 and PWT 5 was 37 with 1111 and 1070 citations respectively. No articles above these ranks were published after 1990.

⁸⁸ To give some idea of the excesses of academe, (Wall, 1995) published a note with the title, *Cricket versus Baseball as an Engine of Growth*. The term, *Engine of Growth* appeared in a large number of papers using the PWT in the 1990s so the spoof begins with the title. The results for 95 economies showed that cricket playing countries grew 43 percent less and baseball countries 80 percent more from 1960 to 1990 than countries playing neither sport. Policies like subsidies and baseball instruction were called for in countries with no history of the game but getting cricket playing countries to give up the game loomed as a more difficult task.

This development made the PWT competitive with the data sets of the World Bank and other official organizations. According to a study by Simon Johnson and colleagues (Johnson, Larson, Papageorgiou, and Subramanian, (2013)) two thirds of all published articles through 2006, including by staff at international institutions, cite the PWT in their publications on economic growth and related topics as their data source.⁸⁹ Further, wider use of the PWT made acceptance of using PPPs as conversion factors to compare real quantities across countries more common among financial journalists and other media.

The 1990 World Development Report was devoted to poverty and its correlates including education, health and mortality, indebtedness and location. Poverty line estimates were based on the 1975 ICP with 1985 as the reference year. The report came up with a line of \$275 for the extremely poor and \$370 for the poor following the earlier methodology of Ahluwalia, Carter and Chenery (1979). When Martin Ravillion headed the poverty group a *dollar a day* was introduced as the poverty line. The dollar a day line was often misunderstood especially when updated. But it was very successful as a simple easily grasped number that conveyed to those in richer countries a very important world problem.

The United Nations Development Programme introduced another way of looking at world welfare based on their Human Development Index (HDI) that provided a ranking of countries. The initial Human Development Report in 1990 included the HDI that was attributed to Mahbub al Haq in collaboration with Amartya Sen. The HDI was an unweighted average of an index of literacy, life expectancy and per-capita income at PPPs.⁹⁰ The HDI was subject criticism for its lack of weighting and other limitations but it gained quite a bit of publicity and which carried over to the ICP and PWT.

By the 1990s both the ICP and the PWT had led to the use of price level comparisons not just in textbooks on economic development and international trade but also college texts on economic principles. Undergraduate statistics, development and econometric courses might assign papers based upon using PWT data to examine some issue like income inequality across countries, or the demographic correlates with GDP, or the Adam Smith question of why some countries are richer than others. A whole new generation of college students entered the 21st century familiar with the use of price levels and PPPs, provided of course they had some economics. The Big Mac index of the *Economist* also played a role for younger and older readers, about which more below.

⁸⁹ Their article pointed out an important problem in using different versions of PWT with respect to our series on growth of GDP. The method used in PWT 8.0 and beyond takes care of that problem.

⁹⁰ The HDI was not really a new idea as we have seen in our discussion of physical indicators that were used in the 1950s and 1960s by Hungarian researchers and by Beckerman and Bacon. The Overseas Development Council headed by Jim Grant encouraged Morris D. Morris to produce a Physical Quality of Life Index (PQLI) in the 1970s that gained some traction because its underlying data series, child mortality, life expectation at age one and basic literacy were readily available. The most striking finding was how poorly some high per capita GDP oil states did on the PQLI index compared to some poor states of India like Kerala.

Our view was that these developments made the job of selling the ICP easier for the UNSD, the World Bank, Eurostat and the OECD.

When we were ready to introduce new versions of the PWT that would integrate the EU-OECD comparisons with the 1996 benchmark results, an undergraduate research assistant asked me why we did not put the PWT online? There followed a tutorial trying to bring me up to date on creating a website and how users would access it. Bettina Aten created PWT 6.0 in October 2000 and it was available on the web soon after. We began tracking usage of the PWT about the same time that Google and other search engines were making it easier to find us under PWT.⁹¹ The Table below indicates the power of the web in expanding access to a research data set like the PWT. The column on hits is not an indication of real usage but is included to show how rapid was the growth in this period. More meaningful columns in terms of usage are Unique Visitor and Pages downloaded. Even Unique Visitor is probably an overstatement because tracking is by apparatus used so that if I access PWT at home and in an office, my understanding is that it counts as two unique visitors. In any event we were pleasantly surprised by the usage in the years after 1991.

Tracked Annual Use of PWT 2004-09

Year	Unique Visitors	Visits	Pages	Hits	Hits per Visit
(000's)					
2004*	11.1	15.3	70.3	160.7	10.5
2005	92.2	139.0	712.5	2986.6	21.5
2006	117.3	174.8	910.7	3887.1	22.2
2007	129.2	196.1	1054.2	4503.2	23.0
2008	139.1	209.0	1078.3	4508.7	21.6
2009	142.5	215.9	1057.8	4474.8	21.6

* 7 months

Our self-serving view is that most of the non-European OECD statistical offices were unaware of the research uses made of their PWT data even within their own countries. Results of the ICP benchmarks were clearly essential to the existence of the

⁹¹ It turns out that PWT has other meanings. For example, it is the symbol of Pawtucket, Rhode Island, and according to another informant it would connote “poor white trash” in Oklahoma.

PWT which in turn contributed to the recognition of the value of PPP and real volume estimates. There was a feedback to statistical offices and support, if reluctantly, to continue the ICP. At an official level the fact that the IMF came to the decision in about 2000 to use per capita PPP converted GDP as a basis for country contributions was another important factor.

D. Aside on the Big Mac Index

The *Economist* magazine introduced the Big Mac index of Pam Woodall in September 1986 as a semi-humorous examination of the relationship between PPPs and exchange rates.⁹² The Big Mac was available as a standardized item in a comparable physical setting in eighty countries by 1997. There were still local variations like use of halal meat in Islamic countries, but by and large, the Big Mac was a similar combination of services and commodities in all countries. Though the weights are different it represents non-tradable and tradables like GDP and consumption. The *Economist* has maintained the Big Mac index for 30 years now and it has stimulated associated research that has kept the index in the news and in our view contributed to acceptance of the view that it was important to maintain the ICP.⁹³

The Big Mac index inspired many competitors including a blue jeans index that was maintained for a few years in the Netherlands, later a Tall Latte index using Starbucks coffee, an Australian I-pod index and there are probably others. The use of one commodity, like wheat or gold, as a measure of value across time and space has a long history. In the case of the Big Mac early criticism was that consuming a Big Mac in the United States where it was such a common practice was very different than in say, India, where it was it was a sign of status to be served in an air-conditioned, clean restaurant and a well maintained bathroom. It is likely that when McDonalds first entered a country there was a price premium due to their novelty but time and local competition soon drove down such premiums.

Early on comparisons of the Big Mac index were made with the PWT by Pakko and Pollard (1996, p.6). The simple correlation was .85 in 1991. Poorer countries with lower PWT price levels tended to be lower than Big Mac indexes and the opposite for higher income countries with higher PWT price levels. Of more interest was that the Big Mac index was persistently above or below the exchange rate over the years, with the exception of Canada. The Big Mac index appears closer to the PPP of the PWT or the ICP than to the exchange rate which suggests that is a fair approximation of the PPP for consumption.

⁹² In fact Howard Banks (1984, pp. 110-11) wrote an article in *Forbes* with the title, *The Big Mac index*. In that article Banks looked at how many Big Macs a days' wage would buy in various countries, a theme in the work of Orley Ashenfelter and Stepan Jurajda to be discussed below. See Ashenfelter (2012).

⁹³ In the 2005 ICP report (World Bank, 2008, p. 4) the Big Mac is used as an introduction to the concept of purchasing power parity.

Big Mac prices were also compared within the United States by Anthony Landry (2008 and 2011) with consistent results as one moves from neighborhoods within the city of Dallas to other cities in Texas and then to other cities in the United States. Leary (2008) reports the standard deviation of Big Mac observations within Dallas is 0.19, within Texas 0.21, and within the United States, 0.31, where the unit is a dollar. For the world cities that are covered by the *Economist* the standard deviation was 1.43. As expected the dispersion of prices is greater the larger the geographic area. This remains a problem for the Global ICP and for large regions like Africa and Asia.

In another context, Landry (2011) examined the variation of the prices of Big Macs within Manhattan and across neighborhoods of the other boroughs of New York City. The distance from Penn Station in New York was recorded for each location, the average being 2.6 miles in Manhattan and 9.6 miles in the other boroughs. In June 2011 the averages were pennies lower in Manhattan with a standard deviation of .20, about the same as Dallas in 2008. The standard deviation for the other boroughs was 0.34 slightly more than Landry obtained for other US cities in 2008. A plausible explanation is that most consumers drive to their McDonalds, but in New York City most are on foot. The difference in prices across cities in 2008 ranged from \$2.24 in Adel, Georgia, to \$3.84 in Philadelphia, over 50 percent. Within Manhattan in 2011 it was \$3.59 to \$4.24, about 17 percent while in the other boroughs the range was \$3.29 in Brooklyn to \$4.56 in Queens, about 32 percent.

The Big Mac is a standard item sold in a standard physical outlet that shows substantial price variation within large cities. Clearly location is a major price determining factor within an urban center like New York affecting rents on business properties and convenience for customers. Moving from within city to across city comparisons one has to also factor in differences in wages, local business laws regulations, and transport costs of inputs on the supply side and average incomes on the demand side.⁹⁴ Studies on the prices of service items as standardized in production and final product as the Big Mac are useful productivity differences. The continuing work of Ashenfelter and Jurada on their index of McWages, the number of Big Macs that the wages of a McDonalds' server can buy in a day, is being extended from across countries to within countries. When we look into the future of the ICP it may very well be that Big Mac prices will have a role.

As to the PWT it became clear that Feenstra was not ready to take it over alone circa 2006 because of his other commitments and because his main interest with respect to the PWT was from the international trade perspective. In fact, we did work on this problem the next few years and a framework was developed and published (Feenstra, Heston, Timmer, and Deng, 2009). Bart van Ark was actively associated with The Conference Board in New York since 1997 and moved from Groningen to New York in 2007 and is now a vice president and Chief Economist. At various times we discussed possibility of housing the PWT at the Conference Board but in end it did not appear a good fit. So Bob and I were left to search again for a home for the PWT.

⁹⁴ Landry (2008) shows a weak positive relation between personal income and Big Mac prices in different neighborhoods of Dallas.

Chapter 10: Getting the 2005 ICP Aloft and Landed

As discussed in Chapter 8 the 1990s was a good news/bad news period for the ICP with some positive momentum by the end of the decade. The UNSC in 2001 endorsed a new round of the ICP but recommended that implementation be put off for another year until financing of a new round was further spelled out with options as to country coverage being discussed at the next session of the UNSC. The World Bank document presented at the UNSC 2002 session proposed country coverage of 70, 80 and 100 countries with budgets up to \$15 million, all representing a quantum leap in financing. In fact 146 countries participated in ICP 2003 so in contrast to previous benchmarks the number of participating countries exceeded early plans. The reference year was to be 2003 with a meeting in Washington later in March 2002 to get input from Eurostat, the OECD, the IMF and others.

A governance structure was to be established which in the interim would be the Friends of the Chair. It was decided there would be an ICP Executive Board from representatives of some country statistical offices and of the concerned international organizations. A Technical Advisory Group (TAG) was to be established that would advise the Global Office and indirectly the Executive Board on methodological issues and to oversee the writing of a new ICP manual. Finally there was the group of regional coordinators who were hires of international organizations like the AfDB and the ADB, with the exception of Latin America where Canada played an important role. The regional coordinators were in direct contact with the countries and hosted meetings on price collection and national accounts, and special areas like construction, rents, education, government and health. The position of Global Manager was posted in Spring 2002 and Fred Vogel joined the World Bank in early November 2002 on a five year contract. Vogel is a statistician by training and previously was a senior manager in the US Department of Agriculture supervising a large staff, and often directing major field surveys. His work provided him a number of international contacts through country visits, specialized meetings and through the International Statistical Association.

A. The First Two Years: Reality Sets In

When Fred Vogel took charge the two principal staff members with ICP experience were Yuri Dikhanov and Yonas Biru who have been mentioned previously. An immediate issue was to assemble staff even though they could not immediately be hired on a permanent basis. A number of short term contracts were issued and a number of outside consultants were hired, some from the BLS familiar with consumer price specifications. The short term contract applicants were predominantly recent graduates from abroad whose employment at the World Bank would permit them to work in the United States for the immediate future. This was a good thing in that a number of them were well trained but with no ICP experience of course. Fred was justly proud of one of his initial appointments, Nada Hamedeh from Lebanon, who was fluent in at least Arabic, English, French and Spanish. In this latter respect, as well as her affability that was very effective

in personal and larger meetings, Nada reminded me of Krinjse-Locker. Too bad they never met.

Edwin Dean was appointed the Chair of the TAG in 2003 after a long career at BLS where he was head of productivity measurement in the United States as well as comparisons with other countries. Ed was very active in the Committee on Income and Wealth, a research group holding meeting with topics related to measurement of national accounts, wealth, and price adjusted changes over time. We had met a number of times over the years and were familiar with the other's work. It was a considerable surprise to me when I learned in 2004 that Ed had resigned as Chair of the TAG apparently because he understood the TAG would report directly to the Executive Board rather than the Global Office.

Bettina Aten and I had moved to Washington in the fall of 2003. I had retired from teaching at Penn and Bettina had just left academe to take a position at the BEA to start a program in measuring price differences across the United States with the aim of measuring real personal income in different cities and states. (Aten, 2006) Living in Washington made me a convenient candidate to replace Dean, but I am sure there were the same reservations to doing so that existed when I was not initially considered for the TAG. In any event I was asked to chair the TAG and agreed. The main tasks were to consult with the Global Office and to sit in on staff meetings as necessary and plan the agenda with Fred Vogel and run the two day meetings of the TAG twice a year. Sometimes there were meetings with the regional coordinators that were scheduled to take place before or after the TAG meeting, which I often attended.

(1) The Initial TAG Members

The first TAG included Ed Dean as chair, David Roberts from the OECD, Silke Stapel and later Paulus Kojijn from Eurostat, Kim Zieschang from the IMF and Sergey Sergeev who at the time was working at the Statistical Office of Austria and carrying out many of the computations for Eurostat.⁹⁵ I had met him one or two times before but was not really aware of what an asset he would be. The academics included myself, Prasada Rao from the University of Queensland, Brisbane, and Erwin Diewert from the University of British Columbia.

At the time the Global Office was not only to carry out the PPP comparison but also take account of the use of the PPP results by the Poverty group in the Bank, which

⁹⁵ Sergey was a protégé of Yuri Ivanov of CISSTAT in Moscow, the agency that coordinated economic comparisons between the Commonwealth of Independent States. Among the TAG members Sergey was legendary for the speed with which he responded to emails or the distribution of new papers or minutes of the TAG. Any response after 24 hours must mean Sergey was traveling where internet was not readily available or he was sick. His knowledge of the literature was extraordinary and sometimes a bit embarrassing when he would remind me of what Kravis, Heston and Summers had written in one of their volumes.

for the next decade was headed by Martin Ravillion.⁹⁶ A special committee of the TAG was assigned to look into poverty line measurement consisting of Prasada Rao, Angus Deaton, a Global Office member, Dipankor Coondoo for a short time, and myself. At the time Angus was carrying out poverty studies using detailed household expenditure surveys from India and Indonesia with Olivier Dupriez of DDG. When I was asked to recommend to the DDG a replacement for Ed Dean on the TAG, it was natural to suggest Deaton and happily he agreed. In addition to the TAG members and staff from the Global Office, several consultants attended most meetings. One was Peter Hill who was primarily tasked with overseeing the new ICP manual including working with those writing chapters and also drafting a number of chapters himself.

The other was Derek Blades who had recently retired from the OECD where he had been heavily involved in national accounts and the ICP. I had met Derek when he was with the National Accounts branch at the UNSD during my first appointment there in 1979. I had known of his work on estimating the value of subsistence activities in a variety of economies and over the years had always enjoyed his company at various meetings. David Roberts (2015) has written a moving Memorial that I will not try to replicate. Suffice it to say the fun we had arguing, the many meals we enjoyed together, notably with Angus, David and Bettina, the talk of boarding schools in England, and jogs together are a sample of the wonderful memories he left. He always made contributions to TAG meetings that were thoughtful, well-reasoned, and full of good humor.⁹⁷

The 2005 TAG got along very well with each other and with the Global Office staff with one or two exceptions.⁹⁸ The DDG asked the TAG in 2007 to suggest additional names of younger colleagues that would be passed on to the Executive Board as future TAG members. A number of persons were suggested and eventually Robert Hill (Pic 29) and Marcel Timmer (Pic 43) were added. Robert was Peter's son who had studied at the University of British Columbia and did his dissertation with Erwin Diewert on a novel method for linking and aggregating countries using minimum spanning trees. Timmer had

⁹⁶ The most common adjective to describe Martin by the Global Office and most of the TAG was arrogant. He had little use for the ICP group, and usually found reason not to attend meetings to which he was invited, or to show up for a few minutes to say his piece and leave. I attended an annual World Bank meeting in the late 1990s when Deaton was critiquing the methodology of how the Poverty Group was updating their Dollar a Day poverty line. This was my first introduction to how Martin and Angus would talk past each other, or rather how Martin would not acknowledge that he did not really understand PPP estimation and/or that he could be wrong.

⁹⁷ In one interaction Derek introduced a way to modify the raw number of housing units to try to indicate the quality of the flow of housing services from the existing housing stock. In past work the proportion of units of electricity, running water and toilet were averaged to obtain an index of overall quality. Derek presented an impressive argument for age of dwelling as detracting from quality of housing to which Deaton asked whether Windsor Castle should be downgraded because of its age. Everyone had a good laugh including Derek.

⁹⁸ With respect to the staff, Yonas Biru became a problem at a personal level for many of the TAG and all of the staff by 2007. Yuri Dikhanov presented different issues about which more below.

joined the Groningen faculty when Bart van Ark was there and expanded the ICOP program and was at home with international comparisons. These were two very good additions. Unfortunately, it was decided to expand the TAG for the 2011 ICP to include many country representatives, who were often new to the ICP, so to several of us elders from the original TAG felt it lost the easy give and take of *the good old days*.

The original plan to produce a 2003 benchmark soon slipped to 2004 and eventually 2005. Some regions had to start from scratch, for example Africa, which had previously been coordinated out of Luxembourg by Eurocost. The AfDB did hire Michel Katoula-Moyela, who was familiar with the Eurocost comparisons for Africa, as a coordinator, a positive step. And a number of country representatives had worked on previous African ICP exercises, which was also helpful. But the number of participating African countries nearly doubled (from 22 to 40) compared to previous rounds so there was much country learning to be done. And the bureaucracy at the AfDB was not efficient in channeling outside funds to the countries. By contrast, almost all the countries in Latin America and Asia had participated in previous ICP rounds. ESCWA started with few countries and no staff with ICP experience, but an excellent liaison in Hamadeh, who herself was from Lebanon. Further the fact that eventually 146 countries participated meant that the total size of the 2005 ICP was almost a third larger than envisioned. The lack of synchronization across the regions had major implications for how the comparisons were carried out eventually leading to changes in methods of how the regions were linked, hardly a new feature of ICP rounds.

(2) The Coding Issue

The first problem was to agree on the number of basic headings for expenditures and second to draw up the specifications for the items to be priced. The starting point on expenditures was naturally previous ICP rounds requesting expenditure detail for about 150 basic headings. A list of item specifications was supplied to the regional coordinators who in turn asked countries in the region to indicate the items or variations of the items they could price and whether they had items to add to the list. In this way a global list was assembled, applying the Eurostat expenditure codes that were already being used in 51 countries.⁹⁹

While there was a common coding for the expenditure headings in the other regions, the item codes within a basic headings were usually different. So all regions would code an item beginning with its basic heading number but then assign a regional number to an item. So *tomatoes* would always be in the basic heading of *fresh vegetables*, it would not necessarily have the same item number in the Africa as in Asia-Pacific. This was partly due to the fact that Latin America and Asia were ahead of the Global Office in producing a final list of items to be priced. This in turn occurred because funding in each region was on its own time line, with the ECLAC comparison, which was being carried out

⁹⁹ The EU-OECD comparisons included members of one and/or the other organization plus a few countries more loosely affiliated with either organization. Two examples are Israel and Russia with the OECD. Russia provided the binary link between the OECD and the other nine countries in the Commonwealth of Independent States, bringing the total to 51.

with substantial assistance from Statistics Canada, needing to complete its report a good year or more before Africa and West Asia. The ADB also needed to produce a report well before the Global Office received all the country and regional data submissions. Since countries in the regions wanted to preserve the regional results through fixity, the Global Office was left with the problem of linking the regional results, but with no easy way to use the regional prices because of the lack of common item coding across regions.

It was decided to approach this problem in the manner of the 1996 comparison by singling out a set of core or *ring* countries in each region. In practice, the Global Office was not free to simply draw up a list of the best set of countries based on coverage and statistical capacity because being a ring country was totally voluntary. One more administrative problem that was dropped in Fred Vogel's lap.¹⁰⁰ He had some resources to assist poorer countries but in the end the 18 ring countries were a mixed bag. In order to have a common coding system for the ring countries it was necessary to draw up a new ring list starting with the EU-OECD codes and list. Input from the ring countries added a few items, eliminated others and altered some details like packaging and size. But basically it was a Western European list and the ring countries generally provided prices whether it was a representative item or not.¹⁰¹ This led to the conclusion of Deaton and Aten (2016) and Inklaar and Rao (2016) that the price levels for Africa and Asia for household consumption were 20 percent too high, understating their real consumption.

B. Price Validation in the Regions

Validation here assumes that countries have already edited the price observations they have collected to ensure that the average price they submit has not been affected by unusual observations or clerical errors. In the 1970 ICP we had developed sensible

¹⁰⁰ I remember several conversations with Fred when he would speak nostalgically about his time in the Department of Agriculture when he could implement a program effectively because his colleagues were on a payroll that he administered. In DDG he had very little leverage to get countries to do what he asked. If they had a major survey scheduled for example, they could not simply set it aside and undertake an ICP ring country survey that was not part of the original terms of their participation. In the end the ring countries included at least two in each region, Oman and Yemen in Western Asia and Brazil and Chile in Latin America. Africa had five countries, while Asia-Pacific and EU-OECD each had four. Egypt is counted in Africa but it is also in Western Asia for many purposes.

¹⁰¹ Yonas Biru was quite adamant that in the African ring countries, which had priced almost all the items on the ring list, these items were commonly available in their markets. I was taking part in the meetings on the evaluation of the ring country price submissions and was in the camp that did not believe his claim. My sense is that Yonas was at this point taking his grievance out on the staff and as well as in substantive discussions affecting the ICP. His grievance was that when Vogel's five year contract was up, the Executive Board voted to renew it through the completion of ICP 2005. My conjecture is that Yonas was given some sort of assurance that he would replace Fred when his contract was up. In any event he turned his wrath on the Global Office staff, particularly the decision makers, and wrote a quite unpleasant letter to the Executive Board. When he was eventually dismissed, he went through a formal World Bank grievance process and was eventually given a generous settlement that he, of course, thought inadequate. A sad story for all.

ways to detect prices that were outliers because of incorrect units, clerical errors, unclear specifications and the like. Let us call these matching errors. With only ten countries and frequent direct contacts, we may have caught most of these matching errors. We tried to capture outlet variations and geographical differences in prices but with much less success. Let us call these sampling errors. By the 1975 ICP our success in obtaining responses from countries when some prices appeared to have matching errors, was less successful. In the meantime Eurostat was carrying out price validation at frequent meetings of all its countries, clearly a more satisfactory approach than we could afford to undertake at Penn. In the early 1990s Eurostat and the OECD started using the *Quaranta Table*, named after Vincenzo Quaranta of the Italian Statistical Office who developed it for Eurostat as a diagnostic tool. The Quaranta Table basically highlights price observations that appear out of line with either the average GEKS price level of the country for that basic heading or the average price level for other countries.¹⁰² The table comprises a simple spreadsheet and needs no technical support.

(1) The World Bank Price Validation Software

One could well ask why the Global Office needed its own software for price validation when the EU software was available and usable on PCs and laptops with any of several programs like Microsoft Excel. There are two justifications. First, it could be considered as technical assistance to the regional offices providing them with the capability of validating prices from the countries in preparation for meetings of price experts from the countries. Combining technical assistance and the ICP makes the output more attractive to donors. Second, since the regions were using CPD at the basic heading level, it would make sense that in looking for price outliers it should be in the framework of CPD, not GEKS. However, Quaranta tables could have been easily transferred to the countries at a lower cost. Further they are not method-specific, so it was irrelevant whether GEKS or CPD is used or any other method. Even so, the Global Office chose to design tables that are CPD specific.

Yuri Dikhanov was tasked with writing a special program for validation of the the CPD method and he did so incorporating the important features of the Quaranta tables and collapsing them into one Table per basic heading. The first calculation was to estimate a CPD from the average prices submitted by the countries and calculating the residuals from the regression. So the table was a matrix of countries and products with an entry for each cell for which a country submitted a price, which was the CPD residual. The CPD residual combined both deviation of the price from the country basic heading

¹⁰² A very clear explanation with illustrative tables is provided in OECD (2012). In the Quaranta Table all prices are converted to a common currency at exchange rates. In this way a product price can be compared with the average of all prices for the product. Prices > than 1.25 or < 0.75 are considered questionable and any prices > 1.5 or < 0.5 are considered highly questionable. In any case they are referred back to the countries, not discarded. The process is iterative as there is an initial basic heading PPP calculated for each country that must then be recomputed as prices are edited. Because the EU-OECD does editing with groups of countries, further sets of Quaranta Tables are generated as the process takes in more and more countries.

parity as well as the deviation from the geometric average of each product price across countries, a very nice feature.

Yuri left the residuals in logs, not a user friendly version. It would have been much more intuitive to express the numbers naturally as decimals centering on 1.0 rather than in logs centering on 0.0. A point made time and time again much to Yuri's annoyance. But Yuri is stubborn and his Table is easy enough to use sitting with a group at the Bank once you are familiar with it, though whether country representatives in the regional meetings felt the same way I doubt. Potential outliers were flagged by colors so one was immediately alerted to a price that posed questions. Yuri's Table also permitted comparison of the price level of the basic heading for each country with a preliminary aggregation over consumption from an initial CPD run as does the Quaranta Table. And like the Quaranta Table, the Dikhanov Table involves iterations as prices are corrected or deleted.

(2) The *Tool-Pack*

Diagnostic tables to catch errors in price collection or entry are very useful in building understanding of the ICP at the staff level in the countries and regions. At the basic heading level one can see the interdependence of the country prices with the prices in other countries, which is one of the important insights provided by the ICP. Further the Quaranta tables and Dikhanov software can be used alone or embedded in a larger framework. Both require as input the matrix of countries and prices for each basic headings, the basic heading expenditure if preliminary aggregations are estimated, and any information that provides different weights for the country average prices within a basic heading.

Even before Fred became Global Manager the DDG had assigned an Information Technology group working with Yuri to develop a data-input software package labeled *Tool-pack*. It incorporated the checking procedures described above as well as a wide variety of aggregation programs. It would allow user-input of individual price quotes that would be averaged within basic headings, provide measures of dispersion and be aggregated using different multilateral methods. It could in principle also calculate consumer price indexes. In initiating the *Tool-pack* the vision was that it would provide regions and individual countries with computing capability and a ready-made ICP do-it-all calculator. From the beginning when it was first demonstrated to the TAG, and subsequently from regional offices, there was strong criticism of the *Tool-pack*. It was a black box, handed down from the Global Office to statisticians in country offices.

Users had a multitude of options at their fingertips, such as CPD, GEKS, and Tornquist, at the basic heading level and GEKS, Ikle and Geary methods above the basic heading level, but no idea how the underlying computations were carried out. Thus it was doomed to fail as a training device and worse, hindered the promotion of transparency and understanding of the ICP computational procedures. The second criticism was that the inevitable problems that arise from developing software to work in a variety of settings – different versions of multiple operating systems on different

platforms in different languages, meant that any glitch had to be referred back to the Global Office for troubleshooting. When the code is not explicit as to the steps involved, troubleshooting is not an easy matter. Rather than bundling many different options into one package my view is that revealing code through comment statements and the steps along the way provides a better training experience.

The Tool-Pack was promoted by the DDG as an important output of the ICP. The push came from Shaida Badiie as Director who felt the Tool -Pack would enhance the image of the DDG, all at a very substantial cost, especially since some of the software had to be outsourced. This enthusiasm was not shared by the Global Manager who had to deal with the problems of trying to scale down the effort. In terms of IT, the Tool-Pack harks back to a view of a computing center that services requests from clients. This contrasts with the view that the final users or assistants with substantive knowledge of the desired output write the code and check the results.

In Phase I of the ICP we were pushing the limits of the Penn computing capacity, whereas by the 2005 ICP most laptops could easily carry out all the computations required by the ICP in much less time. There really is no need to have an intermediate level black box, like the Tool-Pack, above the ultimate user of the ICP computations. Spreadsheets like the Quaranta and Dikhanov Tables are very helpful but should be independent of other methods and computations.

The second criticism is that there are many slips in getting the system to work in Washington and to work in the field because of the way computers are configured. When the code is not explicit as to the steps involved, troubleshooting is not an easy matter. Rather than bundling many different options into one package my view is that revealing code through comment statements and the steps along the way provides a better training experience.¹⁰³ The Tool-Pack was promoted by the DDG as an important output of the ICP. The push came from Shaida Badiie as Director who felt the Tool-Pack would enhance the image of the DDG, all at a very substantial cost. It is not clear that this enthusiasm was shared by the Global Manager and other staff who had to deal with its problems. In terms of IT, the Tool-Pack harks back to a view of a computing center that services requests from clients. This contrasts with the view that the final users or assistants with substantive knowledge of the desired output write the code and check the results. In Phase I of the ICP we were pushing the limits of the Penn computing capacity, whereas by the 2005 ICP most laptops could easily carry out all the computations required by the ICP in much less time. There really is no need to have an intermediate level, like the Tool-Pack, above the ultimate user of the ICP computations. Spreadsheets like the Quaranta and Dikhanov Tables are very helpful but should be independent of other methods and computations.

¹⁰³ The DDG did not support a software like SAS or Stata, which allow coders to document each step in a program. Rather they had programmers in C and C+, or equivalent, that were not easy to follow even for the initiated. Most of the staff worked with Excel, which is clear enough for simple sequences of calculations, but is not always easy for someone new to follow when the number of steps is large.

C. Linking the Regions at the Basic Heading Level in ICP 2005

It became clear by 2004 that the Global Office would not have direct access to the country prices collected in each region for the 2005 ICP. As a consequence the Ring country method was adopted as the way to link the countries at the item level within each basic heading. The TAG was involved in discussing chapters on the ICP Manual that Peter Hill was editing as well as writing a chapter on elementary indexes at the basic heading level. In addition several TAG members were writing papers on the methods that should be employed in moving from ring country prices to linking regions at the basic heading level. The TAG was also tasked with recommending methods for the Regions, except for the EU-OECD countries.

(1) The CPD and GEKS Methods at the Basic Heading Level

A surprising development to me was the degree of support that the CPD approach had among the TAG members, given that Eurostat and the OECD were committed to the GEKS approach. At a practical level the results of both methods produced very similar results so long as the basic matrix of country-item and prices was fairly full. In an attempt to settle on a recommendation to the regions, several experiments were carried out to test the two methods. The starting point was a full matrix of item prices for a basic heading where both methods reduce to a geometric mean that were taken as “truth”. CPD and GEKS estimates differ when more and more holes are introduced into the original price matrix. While the differences were not great, the CPD estimates were consistently the same or closer to “truth” than the GEKS estimates.

This led the TAG to recommend the CPD to the regions but certainly not discouraging the use of GEKS. It should be noted that in Latin America all of the basic headings contained full price matrices so choice of method did not arise. The CPD is slightly easier to compute than the GEKS but even by 2006 this was not an issue in any of the regions. This discussion covers most of the basic headings in consumption except housing services, public education and hospitals. In government it includes the various occupational and skill categories under personnel expenditures. Construction and machinery were handled differently.

(2) Using the Ring Country CPDs to Link the Regions

When the ring countries submitted their prices a validation exercise was carried out that raised a number of questions. We have already discussed the flags raised by a number of items for which African countries provided prices whose common availability seemed doubtful. Another much discussed issue was whether the submissions of some ring countries, like Sri Lanka, appeared especially doubtful. We included in the validation process the relative position of countries for each basic heading from the ring country CPDs compared to their position in the regional CPDs. In general these comparisons were somewhat erratic in Asia and Africa and not that encouraging in the OECD, mainly with respect to Japan. Consideration was given to dropping some ring countries, or giving their prices less weight, but in the end only some outliers for which the ring country provided no response to questions, were dropped.

The Global Office was really between a rock and a hard place in making these decisions because the ring countries agreed to do extra work so it would have seemed impolitic to not use their prices. During the ring country validation experts were sent to select Asian countries including China and Malaysia, a ring country. For China the question was whether the areas where China priced, fourteen large cities and their hinterland, really represented the whole country, and whether the outlets chosen to collect prices were typical. These issues are discussed below.

CPDs were estimated providing country price levels for the 18 ring countries for each basic heading where the United Kingdom was the reference country. Even here the question was raised by Sergeev as to whether the CPD should not be weighted so that each region received the same weight. As it was actually estimated the five African countries (six counting Egypt) were receiving more weight in the CPD than the two countries each in Latin America and Western Asia. I agreed with Sergeev on this point but we did not carry the day.

Diewert (2004) had written a paper on how the ring country price levels for each basic heading might be used to provide price levels for all countries in each region with a common reference country. The paper also outlined how the regional results could then be aggregated to produce global results at various levels including GDP. Diewert presented his paper at a TAG meeting where the light finally went on for me about what Erwin was proposing at the basic heading level. Taking the two country case of Oman and Jordan in Western Asia the idea was to first take the geometric mean of their coefficients from the ring CPD, call it PL_{Ring} with the United Kingdom as reference. Then compute the geometric mean of their price levels from the West Asian CPD, call it PL_{WA} . Each country price level from the West Asian CPD is then multiplied by (PL_{Ring}/PL_{WA}) to express their price levels with the United Kingdom as the reference.¹⁰⁴ Thus there was for each basic heading a price level for each country with a common reference that provided the necessary data to link the regions at higher level of aggregation. In practice the price levels were converted to the United States as the reference country using the UK/US price level at the GDP level from the EU-OECD regional results to make the conversion.

(3) Aggregating the Regions in the 2005 ICP

Diewert (2004) also proposed possible ways to aggregate the regions in 2005. The essential idea was to estimate a PPP for all the countries in the region for each basic heading. Basic heading expenditures at exchange rates were multiplied by the price level of each country (as a decimal) and summed up over all the countries in a region, the real expenditures. The nominal expenditures were simply the sum of basic heading expenditures at exchange rates for all the countries in a region. The nominal expenditures divided by the real was then the basic heading price level (as a decimal) of the region.

¹⁰⁴ Both Peter Hill and I had faced this kind of linking issue in the OECD in the case of Peter and the 1980 ICP in my case. And both had made the same type linking calculations but in a more ad hoc way than what Diewert proposed. Our joint reaction to Erwin's method was 'why didn't we think of that?'

Then a GEKS was computed for the six regions at different levels of aggregation. The TAG (*mea culpa*) endorsed this proposal.

Even before publication Sergey explained that the technique was not reference country invariant, to which I, and eventually Erwin and others agreed. However, it was late in the day and the differences appeared small so the final report went ahead with the original reference countries. However, I came to understand that this was not the only short-coming of this method of linking the regions. In particular interactions of individual countries in different regions did not enter into the aggregation procedure, which I saw as a limitation. For example, comparisons of India or China with other large countries like Brazil, Japan or the United States used no price level information involving these pairs of countries. In any event, the 2011 ICP was not constrained by the need for ring or core countries because all countries agreed to submit their prices of core products to the Global Office. This allowed for the use of the CAR method as described with respect to the 1980 ICP that still preserved fixity in each region. In fact the CAR method could have been applied to the 2005 data and is shown in PWT 7.0.

(4) China's 2005 Participation

In the 1993 ICP China agreed to bilateral comparisons between Shanghai and Tokyo and Beijing and Hong Kong. They took their participation seriously enough to host an ESCAP workshop in Beijing in 1997 to discuss the preliminary results. The Shanghai-Tokyo binary comparison did not satisfy either Chinese or Japanese officials and was not approved for release. However, the Hong Kong binary with Beijing did go forward and was published as an Appendix in the final ESCAP report (UNESCAP, 1999). For the 2005 ICP China agreed to participate again on a limited basis collecting prices in 14 cities and their hinterlands. This limited geographic coverage was finally accepted, after considerable discussion, by the other 22 countries in the Asia-Pacific region at a meeting at the ADB and formed the basis of the regional report (ADB, 2008).

The Global Office also accepted this arrangement based heavily on a principal components exercise carried out by Yuri Dikhanov. Prasada Rao, some members of the TAG, and I were not convinced by this exercise and questioned the acceptance of the 14 city survey as representative of all of China. Our dissents were not effective for the ICP but we did significantly alter the PPPs for household consumption downward by 20 percent in PWT 7.0 released in 2011. The basis for this large adjustment is in Part C of the *Description of PWT 7.0* under PWT 7.0 in the Groningen Growth and Development website (<http://www.rug.nl/research/ggdc/data/pwt/>). The treatment of China's prices and growth rates (Part D) runs to 20 pretty dense pages that I will not reproduce here.

However, it is worth making one point clear. The hinterland of the 14 cities in some cases incorporated rural areas but all included suburbs. It was found that for many items the hinterland prices were higher than the city prices. This is not particularly surprising since the pricing was mainly of shop items with very few if any services and no rental information. Our criticism of the 14 city survey was not that it failed to capture rural-urban

differences in prices so much as how well were other regions of China represented? My documentation relied on the work of other scholars and on internal Chinese data, like cost of living allowances in different prefectures. The evidence all suggested that the prices collected for the 14 cities were higher than in other areas of China. Outside observers generally concluded that China's Statistical Office did a very thorough job of collecting the prices in terms of obtaining the targeted number of items and matching specifications. However, in matching specifications it was felt that the outlets used were often high end and likely reflected well off rather than average Chinese consumers. If true this would mean the 14 city price level for China would be too high compared to countries pricing in more typical outlets, the case in most other countries in Asia and Pacific.¹⁰⁵

Having said that Chinese National Bureau of Statistics (NBS) did a good technical job of collecting a large number of price quotes in 2005 there remained a noticeable use of more expensive outlets and brands for any given specification. In the United States, for example, the BLS makes point of purchase surveys to find the type of outlet in which consumers buy various items. This is not commonly done in the ICP surveys in Africa and Asia, and was certainly not the practice in China. In fact, the ADB and other countries in the 2005 ICP asked that the Global Office send pricing experts to China to go around to the outlets with price collectors to determine what was collected for various specifications. It was on this visit that the experts agreed that the outlets were not necessarily typical for non-food items and within an outlet, where there were choices for a given item, the better brands were priced. As was found in the 1993 ESCAP ICP, Chinese textile prices were higher than those in Hong Kong markets even though many of these clothing items were produced in China and exported to Hong Kong.

Even before ICP 2005 was initiated, the NBS sent staff to many countries to learn about the ICP, though for most of the delegation this was a foreign visit focused on tourism as much as substance. Not an uncommon practice with congressional delegations in the United States, with parallels in other countries. Paris and the OECD was a favorite spot especially during the years in which an attempt was made to link China to the OECD comparisons. David Roberts and Derek Blades eventually threw up their hands on this effort as it became clear that the NBS was not going to supply matching prices for items for which mutual agreement had been reached. Prasada Rao in Brisbane provided ICP lectures for NSB groups after their visit to Canberra, and before they went to the Gold Coast. I spoke to different NSB groups in New York, Philadelphia and Washington and at least two other countries, all with no apparent impact on how the NSB viewed China's role in the ICP. It seems clear to me that since the 1980s China wanted to claim a rate of economic growth without historical precedent while their per capita output compared to other countries remained low. The NSB expressing the views of higher levels of government expressed reservations about the results of the 2005 round, arguing that their economic level was too high. It would be

¹⁰⁵ In an OECD-China price comparison exercise David Roberts and team found that the NSO collected prices from up market outlets and in these outlets priced the higher priced items that matched the specifications. In the selection of representative items it was clear that items bought by the urban middle class were chosen.

nice to say that it was easier to work with the NSB in the 2011 ICP, but as will sadly become clear below, Global Office and TAG relations with China became even more difficult.

D. The Final Results of the 2005 ICP

The Global Office faced a great deal of pressure to complete the 2005 Report but there were time lags because the Executive Board, the Regional Offices and in some cases individual countries were given a chance to review a draft report. The IMF very much wanted the results in order to implement PPP based quotas for member countries in 2008. As a consequence there was not as much time to analyze the results in the final report as there should have been. However, a major supplement to the 2008 report for the 2005 ICP was published later, namely *Measuring the Real Size of the World Economy* (World Bank, 2013) edited by Fred Vogel with some early assistance of Prasada Rao. The volume runs to nearly 600 pages volume with chapters by the usual suspects.

(1) The Form of Presentation

The Tables of the 2005 report were notable in one respect dear to my heart. While the United States was the reference country, results were also expressed with the world average as the reference per capita and total. Botswana's per capita GDP for 2005 is reported as 28.9 percent of the United States and 134.4 percent of the world average (based on the 146 participating countries). The United States dollar remained the reference currency for both the ICP benchmarks of 2005 and 2011. I had recommended to Irv and Bob that we also express the per capita GDP for each country relative to both the United States and a world average in the 1975 ICP report.¹⁰⁶ Irv was not sympathetic nor did I get any support for doing this in the 1980 ICP report. The 2005 report also calculates a world price level of 81 percent of the United States and a world per capita GDP as 7291 US dollars at PPPs and 8976 US dollars at exchange rates.¹⁰⁷ The report also provides the necessary data to calculate the price level of each country relative to the world average, which is also on my wish list for future reports.

My argument for also expressing country price levels relative to the average is because the United States occupies a very asymmetrical position in world financial markets because of the willingness of individuals and institutions to hold dollars as a safe currency and asset. The amount of US currency held abroad exceeds that held in the United States, especially for \$100 bills that are convenient for illegal transactions and which the Federal Reserve in 2016 quit printing for this reason. The total is well over half a trillion dollars. In addition to central bank holdings, many bank accounts held abroad are denominated in dollars, and many transactions are specified in dollars like those in

¹⁰⁶ This could be the simple average of the 34 countries or preferably the average based on the PPP converted GDPs of the super-countries.

¹⁰⁷ In the early versions of PWT we felt constrained by space and also felt our users were more than capable of doing the computations themselves. It should be noted if a poorer country like Pakistan were the reference currency the word price level with respect to Pakistan would be greater than 1.0.

oil, the so called petro-dollars. This can produce significant fluctuations in the value of the dollar abroad for no reasons fundamental to the United States economy. This affects ICP PPP results in different benchmarks as reflected by the price level in the United States relative to the world average, which has fluctuated between 140 and 120 percent of the world average between adjacent benchmarks. If price levels are expressed relative to the United States these US price level fluctuations show up as price level fluctuations in other countries where nothing has necessarily changed.

In TAG meetings and in discussions with the staff I pushed for explicitly including exports and imports as well as the net foreign balance in the tables of national currency expenditures. The advantage of explicitly showing exports and imports is that users are often interested in the total volume of trade relative to GDP, and there is no good reason not to make this information available. Further, even though it could be easily calculated by users I pushed for explicitly showing domestic absorption, the sum of consumption, domestic investment and government. One reason for this is that as we have shown earlier, for purposes of extrapolation, domestic absorption is more reliable than GDP because of relatively low quality of export and import price deflators in many countries. Again time was short so this did not happen in the 2005 report but did make the 2011 publication.

(2) The Ikle Method of Aggregation Used in Africa

Doris Ikle did her graduate work at New York University, Columbia, and Johns Hopkins. Her initial research was directed at index numbers and she published an article proposing a new aggregation formula (Ikle, 1972). Bob Summers was one of the referees on the paper and found it tough reading and asked for few substantive but many language revisions. Yuri Dikhanov became interested in the Ikle index in the 1990s and several staff at the DDG also liked the idea once Yuri had made it more intelligible. The index turns out to be additive, a trait favored by the DDG at the time, and is a variation of the G-K index as described by Balk (2008, p. 247) and Cuthbert (2000).¹⁰⁸ Using the G-K framework in the form used at Penn, the relative price of each country was multiplied by each countries' nominal quantity, namely $Q_i = \text{Expenditures on basic heading at exchange rates} / \text{Price level of basic heading}$. In Ikle the relative price of each country is multiplied by the harmonic mean of the Q_i s.

This definition of the international price provides additive results that are not subject to the bias associated with the Gershenkron effect. The Africa region wished to have an additive result in 2005 and chose, with Yuri's counsel, Ikle. In my view additivity and equal country weighting within the G-K can as easily be obtained by simply taking an arithmetic unweighted average of the relative prices of each country, an alternative noted

¹⁰⁸ Jim Cuthbert and his wife, Margaret, had at one time worked for the OECD as consultants and made several contributions to the ICP literature. Because they mainly worked on consultancies and lived in Scotland I only met them a few times, a loss for me because I only appreciated their work somewhat later. Balk, like Dikhanov, helped make Ikle more visible proving the existence of a solution for the system. The method is sometimes referred to as the Ikle-Balk-Dikhanov index.

by Peter Hill (1982) as well as Balk and Cuthbert. Equal country weighting in G-K using an arithmetic or harmonic mean produces very similar results.

(3) Did the 2005 ICP Results Produce Surprises?

The answer to this question clearly depends on whom you ask and how much they believed previous ICP benchmarks and the World Bank's World Development Indicators. The major tendencies of previous ICPs were clearly evident in the results, namely that country price levels rose with per capita GDPs whether converted at exchange rates or PPPs. I thought that the price levels of many African countries were too high and even more so for India and China given their recent high rates of economic growth. However, many on the Executive Board and Global Office were quite ready to ignore previous ICP benchmarks and accept the 2005 ICP as "truth". This was nowhere more evident than in the World Development Indicators of the World Bank. Quoting Deaton and Heston (2010).

"For example, the 2007 version of the World Development Indicators (WDI), World Bank (2007), lists 2005 per capita GDP for China as \$6,757 and for India as \$3,452, both in current international dollars. The 2008 version, World Bank (2008a), which includes the new ICP data, gives, for the same year, and the same concept \$4,088 for China and \$2,222 for India. For comparison, GDP per capita at market exchange rates is \$1,721 for China and \$797 for India."

Outside the World Bank there were many comments on the 2005 ICP of a political nature that implied that the new results were making the poor countries look poorer so as to keep the World Bank in business. Others like myself thought that more time was needed for users to become familiar with the new results to better judge the 2005 ICP. In the meantime a familiar ICP problem arose, namely that before moving ahead with a new round it was necessary to have the blessing of the UNSC and to set up governance and secure funding. While this was being sorted out many of the Global Office consultants were let go or transferred to other tasks, Vogel remained as a consultant, and Nada Hamadeh was caretaker of the ICP. Dikhanov was temporarily transferred to other work and Yonas Biru was in limbo.

Sometime in 2009 I resumed chairing an *Interim Tag* until a new governing structure was in place. I made it clear my term as chair would end with the appointment of a new TAG and Global Manager, a position advertised in 2009. From many applicants the selection committee chose three candidates to interview: Michel Mouyelo-Katoula (Pic 20), formerly of Eurocost in Luxembourg and during the 2005 ICP manager of the staff at the AfDB; Marshall Reinsdorf (Pic 43) of the BEA and formerly at the BLS, economist; and Silke Stapel, member of the TAG and manager of the Eurostat comparisons. In the end, Michel was chosen as the new Global Manager starting in 2010.¹⁰⁹

¹⁰⁹ Reinsdorf was not really seriously considered because Silke and Michel had so much more ICP experience. The choice between the two was moderately contentious for several reasons. First, Michel carried some unjustified guilt by association because he had been at Eurocost when their contracts with Eurostat were being investigated. Further Yonas and Michel did not work well together during the 2005 ICP, not in small part because Michel had

Chapter 11: Reflections on the ICP and the PWT

My relationship with the ICP is approaching 50 years and not surprisingly some days, months and years have been better than others. I have been involved with other research during this period, much of it in India and Pakistan on subjects as diverse as sacred cows, camel transport, corruption and historical estimates of national income in India. Beginning with the ICP at age 33 and still dabbling at age 82 is a life that would not appeal to all and not surprisingly a number of family, friends and colleagues have asked me why. I have played a lot of tennis and still play squash and since moving to Albuquerque, pickleball (look it up) so I have always had athletic and other interests. However, like my Scottish statistician friend Angus Fell, I still find the ICP intellectually interesting and almost all involved, congenial.

A. Transferring the PWT to Groningen

One of the reasons I remained involved with international comparisons was that the PWT was still being supported enough to keep a skeleton staff at Penn to maintain it while all the while I was trying to pass it on. By this time Bob was not really taking part in the PWT, so it fell on me to explore possibilities. At one stage Bettina Aten and I tried to sell it to the Philadelphia Federal Reserve, which had an ongoing relationship with the Economics Department at Penn in terms of joint workshops and seminars. There was some enthusiasm there, just as there was at the Board of Governors, but in the end no follow through. One new hire in Economics at Penn had some interest, but not surprisingly her career agenda did not allow her enough time to take over the PWT. We maintained our annual international workshop where I would report my lack of progress in finding a new home.

By 2010 the problem of moving the PWT took care of itself in that Marcel Timmer by then was secure in his position at Groningen and his younger colleague, Robert Inklaar, was well on the way. At our 2010 workshop Feenstra, Timmer and Inklaar offered to take over the PWT making some very important improvements, namely explicitly estimating export and import PPPs and developing constant price series that were not sensitive to each new benchmark, as was the case in the Penn version. I mulled over their offer for a full two seconds before exclaiming YES! All thought at this stage that the PWT was generic enough that the name would be maintained. We produced a final version at Penn, version 7.0, that incorporated the 2005 ICP with our modifications in 2012. Feenstra, Timmer and Inklaar brought out PWT 8.0 in July 2013 also based on the

a higher salary than Yonas, a point the latter took personally. I had gotten on well with Silke and we had dinner the night before her interview when she described her vision as Global Manager. Silke can come on strong and I am sure it set off some alarms the next day because her vision involved a large budget for meetings and much emulation of Eurostat (a red flag to a bull for many non-Europeans). Yuri made it clear that he would not stay in DDG if Silke was chosen. In the end Michel was a good choice because he was a very effective manager.

2005 and previous ICP results and thoughtfully compared the differences with PWT 7.0. (Feenstra, Inklaar and Timmer, 2015)¹¹⁰

It was a transition that fit well into the work of the GGDC because it had already become the archive for the Maddison Project to quantify economic history of the world more so after 1800 but also pre-1800 stretching back to BCE. The GGDC is the center of the Maddison Project arranging meetings and circulating papers to a loose...I am a member...group of researchers who try to reach a consensus on how to update Angus's work using more updated rounds of the ICP. The present proposal jointly authored by Bart van Ark is to use the 2011 ICP, which I also support. In addition to their continuing work on productivity, another major project at GGDC headed by Marcel Timmer is to use world input-output relationships in order to trace supply chains so that that the import, export and factor content of products can be estimated. This exciting work does take time of Timmer, particularly, and Inklaar but bless them, they are young.

The transition of the PWT went smoothly, at least from my standpoint with Inklaar now answering questions of users. I do not miss the questions, especially those that are clearly answered in the documentation, or those by paid consultants who want you to do their work. Bob was always trying to find a way to prevent access to the PWT until a user had read the documentation or had passed some kind of test, but we were not able to find a way to pull this off. Bob passed away in spring 2012 just before his 90th birthday, so he did not get to see the PWT truly become generic and his vision realized. Anita and Larry Summers did not want a memorial service at Penn, especially since many of his friends had passed. The IARIW that published the *Review of Income and Wealth* where early versions of the PWT were published, was holding their meeting in Cambridge that summer. The organizers agreed to honor Bob at a dinner meeting during their week long session which was convenient for both Anita and Larry, who also gave a brief talk. I shared some of the history of the PWT and its evolution and my memories of how Bob, who was in charge of public relations and marketing, and I, as production manager, worked together. We both discussed substantive issues at length and rarely ended in disagreement. Our personalities were quite different, Bob more up front and action oriented and I was more laid back, but we both enjoyed our conversations and shared many a good laugh, often at ourselves.

B. The 2011 ICP

I was a member of the TAG during the 2011 benchmark and a member of sub-committees, including the Computation Task Force tasked with monitoring the Global Office estimates. I also worked with Paul Konijn (Pic 23) on housing, and on an ad hoc group dealing with China. As mentioned earlier the TAG was much larger in the 2011 round with members from a number of governments, Italy, Norway, South Africa and the United States for example. Luigi Biggeri (Pic 20) from Italy was both an academic from Florence as well a former member of the Italian Statistical Office and made many useful

¹¹⁰ The GGDC maintains the PWT website and archives all previous versions of the PWT at Penn including documentation. Their latest version is PWT 9.0, which incorporates the 2011 ICP results, was issued in June 2016.

contributions. The Regional Coordinators also attended the TAG meetings because both were scheduled in the same week. The number attending from the Global Office was a bit larger, and included the regularly attending consultants, Derek Blades and David Roberts (Peter Hill no longer attended). All those attending had useful things to say but the large size of the group made it a bit unwieldy, especially for those seniors who were not so good at remembering names or catching all that was said by the speakers in a larger room. (Pic 20)

There were two notable changes in the EU-OECD representation. Silke Stapel had moved from Eurostat to a more senior position in the European Commission. Silke was replaced by Paulus Konijn. Both were very competent, and Paul in my view was more willing to compromise in the sense of not making the best the enemy of the good. On the OECD side Roberts reached retirement age, and his long-time colleague Francette Koechlin (Pic 20), was his official replacement, though Francette had also attended 2005 ICP meetings. In addition Kim Zeischang was promoted within IMF and so shared representation with Mick Silver (Pic 21). Mick was known to many from his work on the CPI in the United Kingdom while teaching at the University of Cardiff so he was a welcome addition at the technical and geniality level.

The reason for the Computation Task Force arose out of the documentation, or rather lack thereof, for the 2005 ICP. For example, when researchers began experimenting with the data set of basic heading expenditures and price levels made available in the summer of 2008 after publication of the 2005 report they found they could not replicate what was published. This was not the only case where more documentation of 2005 would have been valuable. Michel was strong on transparency so three groups were to replicate Yuri's computations for 2011, Sergey and independently Robert Hill in Austria, and Bettina Aten and I in Washington, with Paulus and Francette also on the task force. This did not always work smoothly, but in the end we all breathed more deeply when Sergey and Yuri's results matched. The rough spots were how the regional and Global Office results were put together.

The initial computations in the Global Office were to compute CPDs from the Global Core List prices for the 156 countries supplying them by summer, 2013.¹¹¹ All got the same results, hurray! One option would be to simply aggregate these basic heading price levels and expenditures by GEKS to get a world total and use the CAR approach to provide regional GDP totals. The regional results would then be used to allocate regional total GDP to each of the countries. However, this very simple and straightforward approach that I advocated was not what was actually done

At a meeting of the TAG at which I must have been asleep, a more round about approach was adopted in words, but without the actual steps written symbolically so that a competent programmer could follow the logic without room for interpretation. In this approach use would be made of the country price levels for each region for each basic

¹¹¹ The 2011 reports there were 199 countries net after excluding countries participating in more than one region. There were 22 countries in the Caribbean and 21 in the Pacific islands that were linked on later, the Pacific islands only at the level of household consumption.

heading based on the Global Core and regional prices. The argument was that these regional price levels would be superior to those from the core prices alone because they were what the regions used in their comparisons. My view remains that the purpose of this exercise was only to link the regions and to obtain regional totals of GDP in a common currency. It is not obvious to me that linking the core and regional price levels to obtain a new set of price levels for the CAR aggregation would be an improvement on the simple approach. This is especially so, given the detailed steps involved in the linking.

The three validation groups tended to interpret the guidelines differently giving rise to voluminous emails between Sergey and Yuri that the Hill and Aten groups found difficult to follow. One issue was the fact that the PPPs of each country in the regions were submitted with reference to a regional reference country like Nigeria in Africa. So the steps included moving from reference currencies to a global reference currency. As Sergey pointed out it would have been much better to eliminate these steps and begin with all price levels and expenditures expressed in dollars as the reference currency where the reference would be the United States or all countries, the preference of Sergey and me. Agreement was finally reached, but it took time and patience.

(1) Digression on Regions

Calling the EU-OECD a region is increasingly a misnomer. It does have a European (and Commonwealth of Independent States) core but it includes Japan, Korea, Australia, New Zealand, Canada, Mexico, Russia and the United States in the 2005 comparison and Chile in 2011. And it removes important countries from North and South America and from Asia. So anyone wanting strictly geographical regions needs to construct such entities themselves bearing in mind there will be some anomalies adding the results of Korea and Japan to the results of the Asian region. In 2011 there were also two single countries, Iran and Georgia, whose results are not strictly comparable with those of the other countries in their regions, and some notable regional countries excluded, namely Argentina, because official price statistics at the time were politically manipulated.

These issues are likely to grow in the future as geo-political isolation for some individual countries increases and the OECD includes more countries. The way the EU and OECD countries are put together is by blocs that are a geo-political mixture corresponding to no United Nations or World Bank region. Many of the arguments for regarding the regional comparisons as more homogeneous are belied by the OECD. Africa and Asia are also quite heterogeneous. These comments may have less relevance as the Global ICP moves to a more flexible format as discussed in section C below.

(2) The 2011 Results and Poverty Fracas

The 2011 Report made clear that there were a number of changes in methods between 2005 and 2011 that made it important for users to exercise caution in interpreting changes between the 2005 and 2011 ICPs. For example just applying the CAR method to the 2005 inputs increased the per capita GDP of China by 12 percent and that of India by 8 percent, removing tens of millions from poverty. Global Office warnings were not

typically heeded, perhaps because the headlines were so attractive. Most comments focused on world income distribution, more equal in 2011, and the very large reductions in the world poverty count. A typical media story was to embrace the reduction in poverty around the world while there were other stories saying it was too good to be true, and the World Bank was tinkering with the numbers. In the 2005 ICP the World Bank was accused of making the poverty count too high so they would still have an important mission to improve the world income distribution. Hard to please everyone.

Within countries like India, where the left and right have been squabbling over trends in inequality since Independence in 1947, the left reacted to the 2011 ICP Report by saying the poverty count was grossly understated. The then Congress coalition took the results as evidence that their economic policies were raising incomes of all including the very poor. India has its own poverty count based upon periodic Commissions, the most recent being chaired by C. Rangarajan, a Ph.D. in Finance and Economics from Penn in the 1960s, who had been Governor of the Reserve Bank of India. That report issued in 2014 showed an increase of a hundred million in poverty compared to the previous Commission. Interestingly this was accomplished by adding to the poverty line a higher expenditure share for other goods and services more comparable to better off countries, because that was a goal suitable for India's economy.

The World Bank's poverty group faced a problem of whether or how to integrate the 2011 ICP as it updated its poverty count based upon the 2005 ICP, (Ravallion, Shaohua, and Sangraula, 2009). When the World Bank seriously addressed this issue Ravallion had taken a faculty position in the Economics Department at Georgetown University. In the view of the Global Office and poverty types like Deaton, Rao and me, ICP 2011 should be used as the basis for any future poverty line. A meeting in January 2015 of both the Global Office and poverty group plus observers from non-profits and government aid agencies as well as Aten, Deaton and I held useful discussions.¹¹² The final outcome was for the poverty group to begin implementing a new poverty line of \$1.91 a day. This line was based on the results of the 2011 ICP and underlies the World Bank estimates of the total number in poverty by country, region and world.

The outcome of this review of the World Bank poverty estimation has been a reorganization of the work and an increase in the transparency of the calculations. One criticism of the use of the ICP for estimation of the poverty line is that many of the items and basic headings are not necessarily relevant for the very poor, for example air fares and dishwashers. By agreeing to base poverty estimates on PWT 2011 the poverty group is in a position to estimate their own consumption PPPs using the core price list, eliminating questionable items or basic headings from the list. As will be discussed below, it is not clear this will be an option in the future versions of the ICP.

Another improvement that I think should be made in poverty line estimates emerges from the work of Deaton and Dupriez (2009) and Deaton (2010). They point out

¹¹² USAID under the guidance of Steve O'Connell (Swarthmore and previously Penn) took a great deal of interest in the poverty results based on the 2011 ICP. O'Connell's colleague, Don Sillers, undertook a number of simulations using the 2011 results that helped inform the January meeting.

there is an interdependence between the poverty line and the number in poverty that has not been taken into account in previous World Bank work, including the current \$1.91 poverty line. Previously an unweighted average of the national poverty lines of the ten poorest countries were converted at household consumption PPPs from the ICP. As a starting point many of these national poverty lines were higher than countries not in the bottom ten, like India, where the numbers in poverty are in the 100s of millions. Second the average would be more convincing if weighted by the number in poverty in each country. Third the per capita GDPs from the ICP used to determine the poorest countries are subject to errors of 20 percent or more, so picking a bottom ten countries introduces further uncertainty around the initial line.

A better procedure would start with a larger sample of poverty countries based upon total numbers in poverty from previous World Bank estimates and ICP per capita incomes. An initial poverty line would then be estimated by taking an average of national PPP converted poverty lines weighted by number in poverty of the larger sample. From this initial poverty line new estimates of the numbers in poverty can be calculated and used to re-weight the national poverty lines to begin a new iteration until there is convergence. In my experience the convergence requires just a few iterations. More importantly, taking account of the interdependence between the poverty line and counts appears to make a difference.

(3) China Redux

The economic media picked up many stories about the size of the Chinese economy emphasizing that it had the world's second largest economy in the 2005 ICP, though only roughly half the economic size of the United States. However, some China enthusiasts thought that the 2005 ICP estimates were low and claimed China's economy would soon be larger than that of the United States. For whatever reasons the Chinese did not welcome that prospect and unfortunately this affected their participation in the 2011 ICP in a number of ways. First, they obtained a position for one of the NBS staff on the Executive Board of the 2011 ICP with a serious impact on the timely completion of the 2011 benchmark.

Second, Chinese participation within Asia created serious frictions with both the ADB staff and the Global Office. This took the form of asking to see the prices of other countries before making Chinese prices available and other fairly overt attempts to manipulate the results. During validation the ADB did an analysis of the increase in prices between 2005 and 2011 in the prices submitted to the ADB (41 percent) and the Chinese CPI (22 percent), a large difference.¹¹³ As part of participation in the 2011 ICP each country had agreed that the prices and expenditures submitted to the regional or Global Office would be validated after which further analysis would be carried out at the regional and global levels. The NSB provided 856 products from the regional list of which 383 were from the global core list. In addition China priced 200 global core items,

¹¹³ In fairness, the ADB listed other countries whose prices required further scrutiny including Hong Kong, Indonesia and India.

not priced in any country in the Asia-Pacific region, and which would have been of no use in the regional comparison. It will be recalled that global core items were the basis of linking the regions in the ICP but it had been agreed that validation of the global core items was first to be done in each region, so there was no mechanism for China's additional core prices to be validated.

After the NSB had submitted data and their prices were validated at the ADB they continued to raise time consuming issues with respect to how Chinese data were being used. China was the only country to raise such issues, and they were the only country to submit a set of revised prices apparently to see whether this would change the estimates. In fact, the set of revised prices were by no means all higher than before and their use would not have basically changed the results. Rather the issue was that if one country was allowed to do this how could one refuse a similar request by other countries.

Third, the main Chinese concern was not with their economic position within Asia and the Pacific but rather their position with respect to the United States and several other countries. This led the NBS at a late stage to demand that they be treated as a single country and not be included in the ADB comparisons. It is not clear that treating China as a single country would necessarily achieve the result that the Chinese wished, but this was considered a non-starter by the Global Office and the TAG because countries had agreed to abide by the accepted multilateral methods. The last event that pushed everyone involved to the brink occurred when the preliminary 2011 results were reviewed by the TAG in September 2013 with explicit agreement that the results would not be distributed, an agreement violated by the NBS.¹¹⁴

Martine Durand was OECD Chief Statistician and as Chair of the Executive Board had to deal with the Chinese demand to be treated as a single country. The Global Office was geared up to publish the results in 2014 but the eventual compromise pushed publication to 2015, an unfortunate outcome. What was agreed by the Executive Board was that the NBS would put their concerns in writing and there would be a meeting at the Bank November 20-22, 2013 including an NBS delegation, ADB and Global Office and EU-OECD staff, a representative from India, and a task force of the TAG, including Aten, Deaton, and me. I knew two of the NBS representatives from previous meetings and I felt embarrassed for their unprofessional presentations because I knew they were capable of so much more. They claimed to provide new evidence that was all clippings from the media. They talked about housing but all their evidence was based upon property prices not the flow of rental services. Angus at one point asked the Chinese why they were wasting our time, a common sentiment among the non-NBS representatives. And perhaps among the NBS delegates too, as their discussion points were presented in a repetitive but unconvincing fashion.

The result was that Angus Deaton, Alan Heston, Paul McCarthy, Prasada Rao, and Fred Vogel sent an open letter to the Executive Board explaining why the Chinese

¹¹⁴ At one point Deaton suggested that the Chinese be asked the result they wanted and that would be included in the report. This solution would have removed the pretense that China was abiding by the ICP protocol like the other countries.

demand was technically not feasible and totally at odds with the agreed terms of participation in the ICP. Fred Vogel wrote an initial draft and Paul McCarthy added his input to produce an excellent version that ran to seven pages after adding a few TAG suggestions. The final letter to the Executive Board answered NSB concerns about the preliminary results of the 2011 ICP where China was included with the ADB group. I have been told that there was some support within the Bank and the Executive Board to go along with treating China as a single country prior to this letter. After all the drama the initial results with China in Asia were published with each Table having an end note saying in part “The National Bureau of Statistics does not recognize these results as official statistics.” My own take is that the published China estimates for 2011 are an improvement over 2005. I would like to report that China’s cooperation with the ICP will go more smoothly in the future. However, as of this writing it is not clear that China will follow other countries in Asia and provide newly collected or updated 2017 prices extrapolated from 2011.

C. Moving the ICP Towards Annual Updates

After publication of the 2011 ICP in 2014, an evaluation of the ICP took place under the auspices of the Friends of the Chair and was presented to 47th session of the United Nations Statistical Commission. This section will take up three major recommendations from the evaluation report that were accepted by the UNSC: permanent place of the ICP in the international statistical system, governance structure of the project, and moving to a rolling benchmark system.

(1) You’ve Come a Long Way Baby

The UNSC “supported the recommendation of the Friends of the Chair group...that the International Comparison Programme become a permanent element of the global statistical programme and that it be conducted at more frequent intervals”. Having begun as a research program in 1968 at the UNSD and Penn, moved to a regional basis in the 1980 ICP, survived major reviews under less than friendly circumstances in the 1990s, the ICP is now as *permanent* as such words convey. A few hundred people have been involved at the global and regional offices over this period and several thousand in country offices in the role of national accounts and price statisticians as well as price collectors. It has been a major international economic project, perhaps the largest to the present time.

It is a project that demands cooperation among countries in matching of specifications of goods and services, not a typical activity in statistical offices outside of international trade statistics. The most gratifying part of the ICP for me over the past 15 years has been the group of younger government statisticians and researchers who have taken on this work with interest and often enthusiasm. It is that generation who will be carrying out the recommendations of the UNSC and will be modifying existing methods to produce more frequent ICP comparisons using less resources than in the past.

(2) Rolling Benchmarks

The UNSC also “agreed with the proposal to adopt a rolling benchmark approach of surveys to be spread over a three-year cycle starting in 2017...”. The rolling benchmark approach was an OECD initiative originally suggested by Katerina Reut (BLS) and Barbara Slater (Statistics Canada). They argued that countries would find participation easier if data collection was continuous. At the time Eurostat collected consumer prices for the five year benchmarks (t) between mid t-1 and mid t+1 with machinery and equipment and construction priced every year (paid for by Eurostat). So it was relatively easy for the European Union to adopt the approach. The OECD adopted the approach at the same time but as it was the countries that paid for the pricing of capital goods, so these were priced only once every three years. House rents and government compensation were collected every year by Eurostat and every three years by the OECD. Since 1991, EU comparisons have been annual comparisons and EU-OECD comparisons have been three yearly comparisons.

The UNSC envisions the remaining ICP countries undertaking a similar exercise with the initial reference year being 2017. The 2017 results will serve as an update of the global 2011 benchmark and will be based on surveys undertaken in the regions at different survey dates since 2011. All regions will have done price surveys for 2017 for at least some portion of the basic headings. The UNSC recognizes it is easy to recommend but implementing rolling benchmarks involves a degree of coordination of survey timing between regions other than EU-OECD that is not necessarily feasible. As a practical matter this means that the countries and regions will need to extrapolate many prices or basic heading price levels to 2017 so as to provide the Global Office with the inputs necessary to produce a global aggregation for 2017.

One area of research the Global Office is testing is whether non-official prices can be productively employed, for example, as time to time extrapolators. One reason for this is that country price indexes are often not very detailed, certainly not at the basic heading level. Sources for non-official prices can include special collections contracted to firms who do this on a regular basis, scanner data recording quantities and prices by bar code for each transaction, and prices obtained by web scraping of websites for supermarkets and the like.¹¹⁵ Some of these techniques are quite feasible in Latin America, Western Asia and Asia and the Pacific, less so in Africa. Another question is whether fixity outside the European Union and the OECD is to be maintained every year or some more flexible approach will be adopted, like having preliminary annual global results followed by more final results with fixity when each region is satisfied with their results. The World Bank background paper recommended that results be provided with and without fixity, an idea I strongly support because I believe it adds information beyond that published in the present report, and because of the increasingly non-contiguous character of some regions.

I would also suggest that the World Bank provide for users, on request, a table of all possible binary price and quantity matrices for consumption and domestic absorption

¹¹⁵ In the early ICPs we regularly used Sears’ catalogues, both because BLS also used them, and because for appliances like refrigerators, Sears was the largest seller in the United States.

and possibly GDP. Researchers can of course do this on their own if their request for the basic heading parities and expenditures is approved. But I would argue that non-researchers would also be interested in such binary results within and across regions. This would involve over 100,000 numbers but they are not being published, only provided on request, and these days would not even be qualified as *big data*.

(3) Governance

The governance structure would consist of a Governing Board taking the place of the Executive Committee, an Interagency Coordinating Group and a Technical Advisory Task Force. The aim of these changes would be to increase efficiency and offer more flexibility to the structure. As I read the UNSC report, membership in these bodies would be a balance of countries, regions, experience and expertise. I believe the Technical Advisory Task Force will not be a continuing body like the TAG, but rather a rotating group that would be chosen to examine particular issues as they arise and report to the Global Office and other groups as appropriate. The Technical Advisory Task Force should be able to function with smaller and probably fewer meetings, which should reduce costs, and operate in a more focused way. An Interagency Coordinating Group would be efficient in the sense these agencies already meet on a fairly regular basis so there could be fewer special meetings because many regular attendees are likely to also be familiar with the ICP. Given a certain unwieldy governance structure of the 2011 ICP the prospects for the future appear an improvement.

Before leaving the UNSC report, I must say I was surprised that there was virtually no mention of PPP comparisons within countries. At one time Eurostat recommended but did not fund country surveys of regional differences in prices. The EU countries have periodically made studies of differences in prices in major metropolitan areas for a sample of items that are used to adjust to national prices those prices actually surveyed for the estimation of PPPs. I have never thought this was a very satisfactory adjustment for larger countries but at least an effort is made. The BEA with the leadership of Bettina Aten, does produce metropolitan area and state price parities annually on the basis of the time to time price data collected by the BLS and the Census Bureau. But if the truth be told these are not used for the purpose of adjusting US prices provided to the OECD for PPP purposes. Luigi Biggeri with the support of Prasada Rao has been instrumental in pushing regional comparisons in Italy and most recently at a conference in Nanchang, China. (Pic 50)

D. In Good Hands

This chapter on reflections includes commentary on the immediate future of the ICP, reflecting the difficulty of getting me to shut up on a topic that continues to hold my interest. There are a lot of challenges ahead for the Global ICP but I am optimistic that new approaches will be developed that will improve its quality while using less resources. The source of this Panglossian view of the ICP future is the entry of younger better trained and motivated staff into the global, country, and regional offices. In some ways I am surprised that the same expenditure side approach to GDP has been maintained as long as it has, given the continued criticism of GDP. Perhaps the Groningen group will in the

future bring out industry side productivity measures that parallel expenditure side PPP estimates in a consistent way. Hopefully the EU-OECD approaches to better measures of health and educational output can be extended to other regions. Add to the wish list improvements in measuring output of government and to better estimate the flow of owner occupied and rental housing. These almost 50 years association with *The Project* has been a rewarding experience, especially the friends I have made, and in too many cases lost.

References

- Ahluwalia, Montek S., Nicholas G. Carter and Hollis B. Chenery (1979) "Growth and Poverty in Developing Countries," *Journal of Development Economics* 6.
- Ahmad, Sultan (1983), "International Comparison of Chinese Prices, 1980" Comparative Analysis and Data Division, World Bank, December.
- Ark, Bart van (1993), "The ICOP Approach—Its Implications and Applicability," in A. Szirmai, B. van Ark, and D. Pilat editors, *Explaining Economic Growth; Essays in Honour of Angus Maddison*, North Holland, Amsterdam.
- Ashenfelter, Orely (2012), "Comparing Real Wage Rates," *American Economic Review*, April, pp. 1-28.
- Asian Development Bank (2008), *Purchasing Power Parity: 2005 International Comparison Program in Asia and the Pacific*, Economics and Research Department, Manila.
- Aten, Bettina (2006) "Interarea Price Levels: An Experimental Methodology," *Monthly Labor Review*, BLS, September.
- Aten, Bettina and Alan Heston (2009) "Chaining Methods for International Real Product and Purchasing Power Comparisons: Issues and Alternatives," Ch 10, in *Purchasing Power Parities of Currencies: Recent Developments in Methods and Application*, D.S. Prasada Rao editor, Edward Elgar Publishing Ltd. March, 2009.
- Aten, Bettina and Eric Figueroa and Troy Martin (2011) "Notes on Estimating the Multi-Year Regional Price Parities by 16 Expenditure Categories: 2005-2009," *Survey of Current Business* Bureau of Economic Analysis, May.
- Aten, Bettina and Figueroa (2014) "Real Personal Income and Regional Price Parities for States and Metropolitan Areas, 2008–2012," *Survey of Current Business* Bureau of Economic Analysis, June.
- Balk, Bert M. (2008) *Price and Quantity Index Numbers: models for measuring aggregate change and difference*, (Cambridge University Press: New York.
- Barro, Robert (1991) "Economic Growth in a Cross Section of Countries," *Quarterly Journal of Economics*, May.
- Beckerman, Wilfred (1966), *International Comparisons of Real Income*, Paris: OECD Development Center.
- Beckerman, Wilfred and R. Bacon, (1966) "International Comparisons for Income Levels: A suggested New Measure," *Economic Journal*, 76 September.

- Bhagwati, Jagdish and T.N. Srinivasan (1975), *Foreign Trade Regimes and Economic Development*, India. Columbia University Press, New York.
- Braithwaite S.N. (1968) 'Real Income Levels in Latin America' *Review of Income and Wealth*, June.
- Bureau of Labor Statistics (1967), *City Worker's Family Budget: Pricing, Procedures, Specifications and Average Prices*, Bulletin No. 1570-3.
- Bureau of Labor Statistics (1968), *Three Standards of Living for an urban family of four persons*, Spring 1967 Bulletin No. 1570-5
- Bureau of Labor Statistics (1973), *Average Retail Prices of Selected Commodities and Services*, Fall 1971
- Castles, Ian (1997), "Review of the OECD-Eurostat PPP Programme" OECD document STD/PPP(97)5.
- Clark, Colin (1940), *The Conditions of Economic Progress*, London, Macmillan.
- Clark, Colin (1957), *The Conditions of Economic Progress* 2nd edition, New York, St Martins Press.
- Clark, Colin (1976, "Economic Development in Communist China," *Journal of Political Economy*, vol 64, No. 2.
- Clark, Colin (1979), 'Productivity in the Service Industries' in *Reflections on Economic Development and Social Change: Essays in Honour of Professor V.K.R.V. Rao*, Bombay, Allied Press.
- Committee on Scholarly Communication with the People's Republic of China (CSCPRC, 1980), *Report of the CSCPRC Economics Delegation to the People's Republic of China*, National Academy of Sciences, Washington.
- Cuthbert, James R. (2000) "Theoretical and practical issues in purchasing power parities illustrated with reference to the 1993 Organization for Economic Co-operation and Development Data, *Journal of the Royal Statistical Society*, 163. Part 3, March.
- Cuthbert, James R. (2003), "On the Variance/Covariance Structure of the Log Fisher Index, and Implications for Aggregation Techniques," *Review of Income and Wealth*, 49, No. 1, March.
- Deaton, Angus and Olivier Dupriez (2009), "Purchasing power parity exchange rates for the global poor,"
- Deaton, Angus (2010), "Price indexes, inequality, and the measurement of world poverty" *American Economic Review*, Proceedings, May.
- Deaton, Angus and Alan Heston (2010a), Understanding PPPs and PPP-based national accounts, with, *Journal of Macroeconomics*, October.

Deaton, Angus (2012) “Calibrating Measurement Uncertainty in Purchasing Power Parity Exchange Rates”, Research Program in Development Studies, mimeographed, Princeton.

De Michelis, Alberto and Alain Chantraine, *Memoirs of Eurostat: Fifty years serving Europe*, Luxembourg, 2003.

Diewert, W.E. (2004), ‘On the Stochastic Approach to Linking Regions in the ICP,’ Discussion Paper 04-16, Department of Economics. U. of British Columbia.

Diewert, W.E. (2013), “Methods of Aggregation Above the Basic Heading Level: Between Regions,” in *Measuring the Real Size of the World Economy*, World Bank

Dikhanov, Yuri (1997), ‘Sensitivity of PPP Based Income Estimates to Aggregation Procedures,’ Data Development Group, World Bank.

Drechsler, Lazlo (1973), “Weighting of Index Numbers in International Comparisons,” *Review of Income and Wealth*. March.

Economic Research Institute (1963), “Analysis of Price Comparisons with in Japan and the United States,” Economic Bulletin No. 13, Economic Planning Agency, Japan.

Economic Commission for Europe (1985) “International Comparison of Gross Domestic Product in Europe 1980”, *Conference of European Statisticians Statistical Standards and Studies*, No. 37, United Nations, New York.

Economic and Social Commission for Asia and the Pacific (1999), *ESCAP Comparisons of Real Gross Domestic Product and Purchasing Power Parities, 1993*, United Nations

Economic and Social Commission for Western Asia and the World Bank (1997), *Purchasing Power Parities; Volume and Price Level Comparison for the Middle East. 1993*

Eltető, O., and P. Köves, 1964, “On a problem of index number computation relating to international comparison,” *Statisztikai Szemle* 42, 507–18.

Eurostat (1978), *Comparison in Real Values of the Aggregates of ESA 1975*, Luxembourg, January.

Eurostat-OECD (2012), *Methodological Manual on Purchasing Power Parities*, Paris.

Feenstra, Robert, Alan Heston, Marcel P. Timmer and Haiyan Deng (2009). “Estimating Real Production and Expenditures Across Nations: A Proposal for Improving the Penn World Tables”, *Review of Economics and Statistics*, February.

Feenstra, Robert, Robert Inklaar, Marcel Timmer (2015), [The Next Generation of the Penn World Table](#)” *American Economic Review*, 105(10), 3150-3182

Eurostat (1996), *Comparison of Price Levels and Economic Aggregates 1993: The Results for 22 African Countries*, prepared by Michel Mouyelo-Katoula and Kantilal Munnsad, Belgium

Fogel, Robert William, Enid M. Fogel, Mark Guglielmo and Nathaniel Grotte (2013), *Political Arithmetic: Simon Kuznets and the Empirical Tradition in Economics*, Chicago.

Geary, R. C. (1958), "A Note on Comparisons of Exchange Rates and Purchasing Power Between Countries," *Journal of the Royal Statistical Society* 2.

Gilbert, Milton and Irving B. Kravis (1954), *An International Comparison of National Products and the Purchasing Power of Currency*, OEEC, Paris.

Gilbert, Milton and Associates (1958), *Comparative National Product and Price Levels*, OEEC, Paris.

Gillingham, Robert (1983) "Measuring the Cost of Shelter for Homeowners Theoretical and Empirical Considerations," *Review of Economics and Statistics* May.

Gini, Corrado (1931), « On the Circular Test of Index Numbers, » *Metron*, 2, 3–24.

Gerardi, Dino (1982), « Selected Problems of Inter-Country Comparisons on the Basis of the Experience of the EEC », *Review of Income and Wealth*,

Hajargasht, Gholamreza and D.S. Prasada Rao (2016) « Computation of Standard errors for Purchasing Power Parity (PPP) Exchange Rates from the International Comparison Program (ICP) » Paper for MIT Workshop on PPPs, May, 19.

Heston, Alan (1973), "A Comparison of Some Short-Cut Methods of Estimating Real Product per Capita," *Review of Income and Wealth*, March, pp. 79-104.

Heston, Alan and Robert Summers (1988), "What Has Been Learned About Prices and Quantities From International Comparisons:", *American Economic Review*, May.

Heston, Alan, R. Summers and D.Nuxoll (1994) "The Differential Productivity Hypothesis and Purchasing Power Parities: Some New Evidence," *Review of International Economics*, October.

Heston, Alan and Robert Lipsey ed. (1999), *International and Interarea Comparisons of Prices, Output and Productivity*, Studies in Income and Wealth, No. 61, University of Chicago Press.

Heston, Alan and Bettina Aten (2005) "Regional Output Differences in International Perspective", in *Spatial Inequality and Development*, edited by Ravi Kanbur and Anthony J. Venables, UNU-WIDER Studies in Development Economics, Oxford University Press, New York. 2005

Hill, T.P. (1982). *Multilateral Measurements of Purchasing Power and Real GDP*, Eurostat, Luxembourg.

Hollister, William Wallace (1958), *China's Gross National Product and Social Accounts. 1950-57*, Glencoe, IL, Free Press.

Ikle, D.M. (1972) "A New Approach to the Index Number Problem," *Quarterly Journal of Economics* 86, 188-211.

Inman, Robert (1985), editor, *Managing the Service Economy: Prospects and Problems*, Essays commissioned for the inaugural conference of the Fishman-Davidson Center, Wharton School, U. of Pennsylvania, Cambridge University Press, New York.

International Monetary Fund (1993) *Survey*, May.

Ivanov, Youri and I. Rhyzov (1978) 'A new stage in the activities of the Council of Mutual Economic Assistance in the field of international comparisons of national product', *Review of Income and Wealth*. Pp177-84.

Johnson, Simon H., William Larson, Chris Papageorgiou and Arvind Subramanian. (2013) Is Newer Better? Penn World Tables Revisions and Their Impact on Growth "Estimates," *Journal of Monetary Economics* Vol. 60, No. 2 (2013): 255–274.

Kansal. Satish M. (1971), "A Comparison of the Cost of Living in India and Japan," *Journal of Developing Economies*, June.

Khamis, S.H. (1970), "Properties and Conditions for the Existence of a New Type of Index Number," *Sankhya*.

Kim, E, Gan, Adair Morse and Luigi Zingales (2006), "What has Mattered to Economics Since 1970," *Journal of Economic Perspectives*, Fall.

Kokoski, Mary, Pat Cardiff and Brent Moulton (1994), "Interarea price Indices for goods and services: An hedonic approach using CPI data," Working Paper no. 256. Washington, BLS, July.

Kravis, I.B, Z. Kenessey, Alan Heston and Robert Summers (1975), *A System of International Comparisons of Gross Product and Purchasing Power*, Baltimore: Johns Hopkins University Press. (XI, 294).

Kravis, I. B., and Alan Heston and Robert Summers (1978a), "Real GDP Per Capita for More Than One Hundred Countries," *Economic Journal*.

Kravis, I.B, Alan Heston and Robert Summers (1978b) *International Comparisons of Real Product and Purchasing Power*, Baltimore: Johns Hopkins University Press, (IX, 264).

Kravis, I.B. (1981). "An Approximation of the Relative Real per Capita GDP of the People's Republic of China," *Journal of Comparative Economics*. 5. 60-78.

Kravis, I.B, Alan Heston and Robert Summers (1982) *World Product and Income: International Comparisons of Real GDP*, Johns Hopkins University Press, May 1982.

Kravis, I.B, Alan Heston and Robert Summers (1982a) Comments on: "Selected Problems of Intercountry Comparisons on The Basis of The Experience of The EEC", *Review of Income and Wealth*, Comment on Gerardi.

Kravis, I. B., R. Summers and A. Heston (1983a), "The Share of Services in Economic Growth," in F.G. Adams and B. Hickman (eds.), *Global Econometrics: Essays in Honor of Lawrence R. Klein*, M.I.T. Press.

Kravis, I. B. and Robert Lipsey (1983), "Towards an Explanation of National Price Levels," *Princeton University Studies in International Finance*, No. 52.

Kravis, I. B., R. Summers and A. Heston (1984) "Changes in World Income Distribution," *Journal of Policy Modeling*, May: pp. 237-6.

Kravis, I.B. (1986), "The Three Faces of the ICP," *Research Observer*, 1, No. 1.

Krinjse Locker, H. and H.D, Faerber (1984) "Space and Time Comparisons of Purchasing Power Parities and Real Values," *Review of Income and Wealth*, 30.

Kuznets, S., (1955) "Economic growth and income inequality" *American Economic Review* 49, March, 1-28.

Kuznets, S., (1963), Quantitative: aspects of the economic growth of nations, VIII: Distribution of income by size, *Economic Development and Cultural Change* 11. no. 2, Jan., pt. 2, 1-80

Landry, Anthony (2008) "The Big Mac: A Gblal to Local Look at Pricing", Economic Letter, Federa; Reserve Bank of Dallas Vol. 3, No. 9, September.

Maddison, Maddison (2004), "Macromeasurement Before and After Colin Clark," *Australian Economic History Review*, March.

Mizoguchi, Toshiyuki (1968), "A Comparison of Level of Consumption of Urban Housholds in Japan and on Mainland China," Technical Paper #4, Institute of Economic Research, Hitotsubashi University, Tokyo.

Noda, Tsutomu, (1966), "Comparisons of Level of Living in Japan and Selected Asian Countries," Economic Research Institute, Economic Planning Agency, Tokyo.

OECD (2012), [Eurostat-OECD methodological manual on purchasing power parities, Paris.](#)

Paige, D, and G. Bombach (1959), *A Comparison of National Output and Productivity of the United Kingdom and the United States*, Paris: OEEC.

Pakko, Michael R. and Patricia Pollard (1996), "For Here or To Go? Purchasing Power Parity and the Big Mac," *Review*, Federal Reserve Bank of St. Louis, January/February.

Pakko, Michael R. and Patricia Pollard (2003), "Burgernomics: A Big Mac Guide Purchasing Power Parities," *Review*, Federal Reserve Bank of St. Louis, November/December.

Ravallion, M, C. Shaohua, and P. Sangraula, (2009) "Dollar a Day Revisited" *The World Bank Economic Review*, 23(2):163-84.

Roberts, David (2025) "In Memoriam: Derek Blades," *Review of Income and Wealth*, March.

Ren, Rouen (1997) *China's Economic Performance in an International Perspective*, OECD Development Centre, Paris.

Ryten, Jacob (1999) "Evaluation of the International Development Programme," Item 3 (e) UN Statistical Commission Thirtieth session, 1-5 March.

Salazar-Carrilla, Jorge (1978), Prices and Purchasing Power Parities in Latin America, 1960-72, An ECIEL Study.

Samuelson, Paul (1985), "The 1984 Nobel Prize in Economics," *Science*, January.

Summers, Robert (1973), "International Price Comparisons Based Upon Incomplete Data," *Review of Income and Wealth*, March.

Summers, R., I. B. Kravis, and A. Heston (1980), "International Comparison of Real Product and its Composition: 1950-77" *Review of Income and Wealth*, March.

Summers, R. and Alan Heston (1988), "A New Set of International Comparisons of Real Product and Price Levels: Estimates for 130 Countries, 1950-1985", *Review of Income and Wealth*, March.

Summers, Robert and Alan Heston (1991) "Penn World Table (Mark 5): an Expanded Set of International Comparisons, 1950-88", (with R. Summers) *Quarterly Journal of Economics*, May.

Syzilagyi, Gyorg (1966) 'An intercountry comparison of the national income of planned economies,' *Review of Income and Wealth*.

Szulc, Bodan, (1964), "Indices for multiregional comparisons," *Przeglad Statystyczny*, 3, 239-54.

Theil, Henri (1983), "World Product and Income: A Review Article," *Journal of Political Economy*, Vol. 91, No.31.

United Nations Statistical Commission and Economic Commission for Europe (1969), *Comparison of Levels of Consumption*, New York.

United Nations Statistical Office and the European Community (1986-7), *World Comparisons of Purchasing Power and Real Product for 1980: Summary Results for 60 Countries*. Published 1986, and Part two: Detailed Results for 60 countries published in 1987 by the United Nations.

United Nations Department for Disarmament Affairs (1986), *Reduction of Military Budgets; Construction of military price indexes and purchasing power parities for comparison of military expenditures*, A/40/421, New York.

United Nations Statistical Division, (1992), *Handbook of the United Nations International Comparison Programme*, New York.

United Nations Statistical Division and the European Community (1994), *World Comparisons of Real Gross Domestic Product and Purchasing Power, 1985: Phase V or the International Comparison Programme*, ST/ESA/STAT/SER.F/64.

United Nation Economic and Social Commission for Asia and the Pacific (1999), *ESCAP Comparisons of Real Gross Domestic Product and Purchasing Power Parities, 1993*, Bangkok.

Wall, Howard J. (1995), « Cricket versus Baseball as an Engine of Growth » *Royal Economic Society Newsletter*, July, pp2-3.

World Bank (1990), *World Development Report, Poverty*, Washington, D.C.

World Bank (2008a), *World Development Indicators*, Washington, D.C.

World Bank (2008b), *Global Purchasing Power Parities and Real Expenditures*, Washington, D.C.

World Bank (2013), *Measuring the Real Size of the World Economy*, edited by Frederick A. Vogel, Washington, D.C.

World Bank (2015), *Purchasing Power Parities and the Real Size of the World Economies*, Washington, D.C.

Yotopoulos, Pan A. and Jenn-Yih Lin (1993), "Purchasing Power Parities for Taiwan: The Basic Data for 1985 and International Comparisons", *Journal of Economic Development*, Korea, No.1

Yzeren, J. van (1956) 'Three methods of comparing the purchasing power of currencies', *Statistical Studies*, Netherlands Central Bureau of Statistics, No 7, December