

## OTHER ECONOMIC STATISTICS

### The future of the International Comparison Programme (ICP): Comments by the Australian Bureau of Statistics

#### Introduction

1. The ABS has participated in the 1985, 1990, 1993 and 1996 rounds of the OECD Purchasing Power Parity (PPP) program. For some time, we have had concerns about the program and the broader International Comparison Programme (ICP) of which the PPP program is an important component.
2. The original concerns related to the quality, and hence the credibility, of the outputs from the PPP program, as we do not subscribe to the view that errors in component data will necessarily offset at an aggregate level. Because of these concerns, we have put considerable effort into a rigorous analysis of the 1996 PPP results including updating much of the data analysis presented in the Castles report on the OECD PPP program. Among other things, it has highlighted some serious data mis-reporting by the ABS to the OECD which was not detected during editing; it is probable that Australia is not alone in its mis-reporting. This work has formed the basis for constructive, ongoing communication with the OECD on ways to make improvements in particular areas of the program.
3. These data quality concerns have, in turn, led us to seriously question why users appear to be accepting data that, in our view, contain inaccuracies and are produced on an untimely basis. Why are users not vigorously complaining as they clearly would in relation to national statistics of such a standard? In fact, what are the key public policy uses made of the ICP outputs and how disadvantaged would communities be if the programme did not exist?

#### Some examples of perceived data quality problems

4. Appendix 1 contains a selection of graphs which were prepared by the OECD to facilitate the evaluation of the 1996 PPP results. These are used to illustrate some examples of our data quality concerns. The graphs show, for various "analytical" categories:

. the per capita volume indices (based on national

expenditure data converted using PPPs)

. the per capita value indices (based on national expenditure data converted using exchange rates)

. the comparative price level indices (based on the ratio of the PPP to exchange rates)

The indices cover the 32 countries that participated in the 1996 comparison. They are expressed relative to the OECD Member Countries; Israel, Slovenia, Slovak Republic and the Russian Federation are not included in the OECD total.

5. Graph 5 relates to "Meat". It shows the per capita volume of meat consumption in Australia to be the highest of any of the 32 countries. It is twice that for New Zealand, about 25% higher than that for the USA and over 50% above Canada. These results seem counter-intuitive to us. Further, the per capita meat volume for USA is shown to be less than for France and only a little higher than for Portugal and Italy. These relationships do not seem to be credible.

6. Graph 13 covers "Alcoholic beverages". The per capita alcohol consumption (in volume terms) for the UK is shown as over five times greater than that for Australia. Indeed, the Australian per capita volume of alcohol consumption is shown as only about half of the OECD average. Again, these results seem implausible. In addition, some of the relationships between European countries are difficult to comprehend. For example, the results for Luxembourg and Switzerland seem to be unrealistically high at about three times, and double, the OECD average respectively. On the other hand, the results for Italy, Ireland and Russia seem very low being less than two thirds of the OECD average.

7. Graph 14 shows "Tobacco" prices and consumption. Can the per capita volume of tobacco consumption in Luxembourg be six times greater than the OECD average? Also, are the relativities between other European countries realistic? For example, the consumption of Switzerland and Greece are recorded respectively as more than 50% and 100% higher than that of France.

8. Graph 16 relates to "Clothing including repairs" and Graph 17 to "Footwear including repairs". We find in both graphs the per capita volume comparisons for the USA and Canada difficult to comprehend, with the figures for the USA being more than double and treble those for Canada. In both cases the relative volumes in Canada seem very low.

9. Graph 36 relates to private consumption of "Education".

Comparative price levels across the 32 countries show considerable deviation from the OECD average and it is not surprising that the per capita volume measures make little sense to us. For example, the nearly six fold per capita volume for Australia compared with New Zealand seems implausible.

10. More generally, the graphs highlight the inherent problems involved in undertaking inter-country comparisons:

. The country value data are assumed to be of comparable quality in terms of their comprehensiveness; for example are all countries making appropriate allowances for under-reporting by households of tobacco and alcohol consumption?

. It is presumed that all country value data are dissected into component items in comparable ways. A particular problem we have encountered in the past was the non-comparable classification across countries of alcohol consumed in restaurants: is it part of private consumption expenditure on "alcoholic beverages" or as part of expenditure on "restaurants, cafes and hotels"?

. The inter-country price level indices assume that appropriate judgements have been made as between representativeness (of goods and services purchased within a country) and comparability (of priced items between countries).

11. If any of the above pre-requisites are not present, it is highly likely that the per capita volume comparisons will be dubious.

12. It might be argued that any data quality concerns at the "analytical" category level will tend to even out at higher levels of aggregation. This seems to us to be a brave assumption that will certainly not hold in all cases.

13. It might also be argued that the perceived data quality problems highlighted above for Australia arise because of mis-reporting by the ABS. It is readily admitted that there was significant mis-reporting as highlighted by the detailed analysis of the 1996 results. However, the comparisons are also impacted by mis-reporting by other countries - something over which we have no control.

Castles and Ryten reports

14. The Castles report<sup>1</sup> on the OECD PPP program and the report by Ryten<sup>2</sup> on the remainder of the ICP are both useful studies which highlight data quality problems and major resourcing deficiencies. Castles identifies particular problem areas and proposes specific methodological and conceptual improvements while Ryten emphasises the need for procedural improvements centred around documentation, training, analysis, etc.

15. Both reports describe needs for PPPs and PPP-adjusted GDP data and describe broad applications and categories of users. The reports do not, however, provide conclusive answers to the fundamental questions posed in paragraph 3 above.

16. Both Castles and Ryten conclude that the program is grossly under resourced. However, we question whether the injection of an additional, say, US \$1 million would really transform it. In any case, has the ICP been objectively assessed as the top international statistical priority for the allocation of additional funding, or are there other more critical statistical gaps or deficiencies?

#### Justification for the ICP

17. In order to determine whether the continuation of the ICP is justified, the ABS proposes that consideration be given to undertaking further analysis of the fundamental objectives of the programme and the main uses that are made of the programme outputs, particularly in relation to public policy. Specifically, the primary objective of the ICP should also be clarified. Is it to enable comparisons of price levels between countries, or to support the deflation of expenditure values to derive volume levels?

18. A rigorous cost benefit analysis should assess the total direct and indirect costs of conducting the programme, including the opportunity cost associated with the use of scarce specialist staff resources by a large number of national statistical agencies who are under other international pressures such as those imposed by the IMF Special Data Dissemination Standards.

19. If PPPs are sufficiently inaccurate, long-term average exchange rates could potentially give more reliable results. Ryten (para. 4.1.2) quotes from a paper by Summers and Heston "...but it is at least possible that available ICP estimates of the PPPs are of such poor quality

---

<sup>1</sup>Review of the OECD-Eurostat PPP Program, (STD/PPP (97)5, OECD, Paris, 1997), by Ian Castles.

<sup>2</sup>Report on evaluation of the International Comparison Programme, (E/CN.3/1999/8, United Nations, New York, 1999), by Jacob Ryten.

that they are less accurate estimates of the correct PPPs than exchange rates". This would certainly be a much lower cost approach.

20. As part of the cost-benefit analysis, we propose further study of the consequences of using long-term exchange rates as proxies for PPPs, and the extent of the distortion of the resultant volume estimates. The long interval between the reference period for the PPPs and the date of release (the 1985 ICP results were published in 1994), coupled with disparate domestic inflation rates, could be expected to greatly limit the relevance of the PPPs for contemporary studies.

#### Options for the future of the ICP

21. On the basis of the results of the above analysis, an informed choice about the future of the ICP could be made from the following options:

Option A: Discontinue the entire ICP (i.e. the OECD PPP program and the balance of the ICP).

Option B: Work on implementing major quality improvements to the OECD PPP program outputs through the development of forward work plans incorporating the accepted recommendations from the Castles report. Defer work on the balance of the ICP until the PPP program is assessed as being robust and attention can be focused on reinstating a revamped exercise for the balance of the ICP in about five years' time, drawing on the experience gained from the work undertaken to improve the PPP program.

Option C: As well as improving the OECD PPP program (as per the first part of Option B), develop work plans to improve the quality and timeliness of the outputs from the non-OECD components of the ICP. The plans should reflect the results of a careful evaluation of the recommendations of the Ryten report.

22. The ABS does not consider the continuation of the status quo to be an option.

23. If either Option B or Option C is accepted, then there is a further range of issues which should be addressed before strategic work plans can be devised to facilitate the necessary data quality and timeliness improvements. These are described below as Issues 1 - 6.

Issue 1: What are the essential statistical prerequisites for a successful ICP?

24. The main outputs from the ICP include the purchasing power parities, comparative price level indexes and GDP volume indexes (converted to a common currency using PPPs) in total and on a per capita basis.

25. The data which national statistical agencies are required to provide in order to compile these outputs are national accounts expenditure data by commodity (which the agencies need to reclassify to "basic heading" commodities) and baskets of product prices (usually specifically collected for the ICP exercise).

26. The country national accounts expenditure data by commodity are:

used for internal commodity weighting to compile aggregated PPPs from the product prices; and

deflated by component PPPs to produce comparable country GDP volume estimates.

27. It is a prerequisite for the generation of accurate outputs that are truly comparable between countries, that each national statistical agency must provide:

(i) High quality national accounts data;

(a) in terms of the precision of the classification of expenditure by basic heading commodity, which affects the internal PPP weighting, and

(b) in terms of levels of the broad GDP aggregates. Examples include the quality of adjustments for understated recorded expenditure (e.g. on alcohol and tobacco) and the cash economy, and the extent of adherence to SNA standards (e.g. in categorisation of expenditure between capital and consumption). This impacts on the derivation of the volume measures.

(ii) Baskets of price data which are both representative of national transactions and comparable across countries and priced to constant quality. These data form the basic inputs into the calculation of the PPPs.

28. It is particularly important that there is consistency in the classification of the prices and the expenditure values as any significant mismatch of price and value data can result in serious distortions in the resultant aggregate PPPs and volume measures.

29. Because an inherent feature of spatial measures is that errors in the data for one country impact on the results for other countries,

variations in:

the quality of individual countries' basic national accounts aggregates;

the precision with which the expenditure data are classified to basic heading commodities; and

the rigour of the pricing exercise,

can permeate the ICP outputs.

30. It is clear from both the Castles and the Ryten reports that none of these statistical pre-requisites is currently being achieved.

Issue 2: How credible are the outputs from the ICP and what is the extent of user commitment?

31. There are significant problems with the quality and timeliness of the programme's outputs as highlighted by Castles and Ryten.

32. However, the lack of vigorous complaint by users about being delivered statistical tables which include what seems to us to be patently wrong data, and on an untimely basis, must raise questions as to the users' commitment and the extent of their deprivation if the outputs ceased to exist.

33. Castles (para. 2.16 - 2.18) and Ryten (para. 1.3.3) identify a range of uses of PPP-based data by international organisations. However, Ryten states that "PPPs are still in search for well-defined domestic uses" (para. 7.1.1).

34. The Australian experience is that there does not appear to be much local interest in the data. For example, when the 1996 OECD PPP results were finally released in September 1998, local users were mute. They seemed unperturbed by the fact that there had been significant delays in releasing the data. Also, despite the fact that Australia was in the middle of a federal election campaign being fought on economic issues, and that the 1996 data indicated a significant shift in Australia's relative economic performance, no political parties or commentators appeared to react to the new information.

Issue 3: What are the respective roles and responsibilities of national statistical agencies and international organisations and how can the necessary commitments be obtained?

35. Ryten (Annex I, A framework for a critical review) comments on the diffidence of national statistical agencies about the results of the ICP. This perceived diffidence is likely to be associated with the lack of interest in the data by national users (see Issue 2 above) and a lack of clarity concerning the agencies' responsibilities.

36. Castles and Ryten have presented different views about the responsibilities of national agencies. This is a particularly important issue requiring clarification. We believe that ambiguities about the present arrangements may lead to some of the necessary statistical functions not being adequately performed.

37. Castles states that "Responsibility for the estimates should be accepted by OECD/Eurostat, and should not be shared with statistical agencies in Member countries" (page 3).

38. On the other hand, Ryten concludes that responsibility for the results should be shared. The main argument presented is that national statistical agencies are more likely to cooperate if they feel they are accountable for the quality of the output (para. 5.3.1).

39. The ABS does not agree with the concept of shared responsibility for the very reason that Ryten espouses in a different context. He states "In the case of PPPs, the NSOs are in no position to guarantee anything other than the accuracy of their own basic data" (para. 7.1.1). He goes on to say "the quality of A can only be assessed if taken together with B. If B does a poor job, the excellence of A's work is immaterial".

40. We think the last quotation gets to the heart of the issue.

41. The ABS view is that responsibility for the outputs from the programme cannot be shared with member country statistical agencies as individual country results are dependent on the inputs from other countries. Drawing an analogy with respondents to a national statistical collection, individual businesses have a responsibility to provide accurate data - the inputs - but are in no way accountable for the accuracy of the aggregates compiled by the statistical agency - the outputs - which incorporate inputs from a range of respondents over which they clearly have no control. In addition, individual respondents do not have control over methodology, editing, imputation, analysis, etc. While an individual business is expected to be responsible in data reporting, the statistical agency still has full responsibility for the quality control of inputs through processes of editing, querying and analysis. If the agency publishes inaccurate statistics because a respondent mis-reported the agency is fully responsible for the inaccurate statistics.

42. Clearly there is some uncertainty about the respective roles and responsibilities of national statistical agencies and international organisations. We consider it is critically important that those roles and responsibilities be fully articulated and formally agreed by all parties. Certainly national agencies have a major role to play and need to commit to that role and be held fully accountable for undertaking a defined range of statistical functions to a high standard. However, sole responsibility for the quality of the outputs must rest with the international organisations. A successful ICP would have to be conducted as a proper statistical collection with the international organisations taking prime responsibility. Any notion of shared responsibility for the output can only lead to a diffusion of accountability and diminution of commitment.

Issue 4: What should be the economic scope of the ICP?

43. The issue of whether the scope of the ICP exercise should target GDP or be restricted to a subset of GDP on conceptual or practical grounds should be resolved.

44. Castles proposes that the focus in the short term should be on the "comparison-friendly" categories of final consumption expenditure in order that visible gains in data quality and credibility can be made (third last of his Key Conclusions and Recommendations, page 3).

45. He proposes that the collection of prices of inputs for non-market services should cease, and the possibility of using price parities relating to the output of marketed services as proxies should be investigated (second last of his Key Conclusions and Recommendations, page 3). Ryten agrees with these proposals on practical grounds (II.2.22).

46. Castles queries whether the pricing of capital goods (construction and equipment) should continue because of practical difficulties (last of his Key Conclusions and Recommendations, page 3). Ryten proposes sacrificing the pricing of capital in the short term because of the high costs and because it detracts from the effort put into pricing household consumption items in some national agencies.

47. The ABS agrees with Castles and Ryten.

Issue 5: How can complex conceptual issues be addressed?

48. Castles recognises a range of complex conceptual issues that need to be addressed (e.g. in the areas of housing, health, education and collective government services). Ryten does not place the same emphasis on these issues. For example, in Annex I under "Current Beliefs and

Counter Arguments" Ryten states: "comparing the remotest and least developed country with any one of the most developed is essentially not different from comparing real output between two points sufficiently distant from each other on a time scale. In the same fashion as we do the latter by chaining time neighbours until the two extremes are tied to the same chain, so a number of techniques make it possible to chain space neighbours. It may well be that for any link there are more options in space than in time but that hardly constitutes an essential difference".

49. Whilst there is a certain underlying logic to this statement, we consider it may have the unfortunate effect of playing down the unique complexities associated with compiling spatial indexes and serve to understate the difficulties of the exercise.

50. The prime purpose of temporal indexes tends to be to analyse changes for very recent periods and to predict the near future (e.g. inflation studies). While it is true that, over time, long-term series are built up through chaining processes, there is far less dependence placed by users on long-term analysis. That is, most users are not highly dependent on the quality or effectiveness of the accumulated impact of all the chaining processes that have occurred over an extended period.

51. On the other hand, the static spatial index compilation exercise, which produces direct comparisons across countries at different extremes of economic wealth, places an extremely high degree of dependence on the cumulative chaining processes across a large number of countries.

52. Presumably this international chaining is undertaken because there is interest in directly comparing the richest and poorest countries. In the absence of such interest, there would be no reason for such chaining and direct comparisons could be restricted to blocks of similar countries with highly comparable economies.

Issue 6: What are the roles of national accountants and prices statisticians in the programme?

53. Ryten asks "has there been any thought (nationally and internationally) to ensure that national accountants are permanently involved with the estimation of PPPs; with the scope of the ICP; and with national responses to new initiatives taken in the framework of the programme?" (Annex I, section entitled "A meeting on PPPs and its possible agenda").

54. Once the earlier issues are resolved, it is expected that it will become clear that there are important roles for both prices statisticians and national accountants in the conduct of the ICP.

55. The bulk of the data gathering work relates to prices (consumer and producer) and there are many prices conceptual issues to be considered.

56. However, it is also particularly important that the national accounts aggregates that feed into the exercise are accurate and correctly classified (see Issue 1, above). Also, there are emerging national accounts conceptual and measurement issues (e.g. the development of physical quantity indicators for the output of non-market industries) that clearly require input from specialist national accountants.

57. In our view, until such time as the ICP is seen as both a prices and national accounts statistics issue, there is little hope for significant improvement in the quality of the programme outputs.

## Conclusion

58. Clearly there are serious quality problems with the ICP programme, as confirmed by the Castles and Ryten reports. They will be expensive to fix.

59. Before embarking on a quality improvement program, there should be some analysis of whether the benefits justify the costs of the ICP programme. Specifically there should be analysis of the real uses made of ICP data, both internationally and nationally. The ABS has considerable doubt whether the uses of the ICP statistics justify the expense, particularly given the quality problems that exist with ICP statistics. The objectives of the programme may need to be curtailed to something that is more realistic. A better and cheaper alternative may be to use long term average exchange rates as proxies for PPPs

60. This paper proposes that consideration be given to undertaking a cost-benefit study. However, there may be sufficient prima-facie evidence that the use of the PPP and ICP outputs does not justify the overall cost of the programs. In this case, the ICP should be discontinued and the research effort concentrated on the empirical analysis of the adequacy of using long-term average exchange rates as proxies for PPPs.

61. It is suggested that the Commission discuss the following issues.

1. Is there sufficient justification of the usefulness of the ICP programme to warrant the cost benefit analysis proposed in the paper, with a report to the United Nations Statistical Commission and other key stakeholders?

2. If so, who should take responsibility for arranging the analysis?
3. If not, should a study of the adequacy of using long-term average exchange rates be commissioned?
4. Should option A, B or C (see paragraph 21 above) for the future of the ICP be adopted?
5. If option B or C is adopted, then views on the 6 issues identified above (in paragraphs 24 to 57) are sought.

Australian Bureau of Statistics

January 1999

## Appendix 1

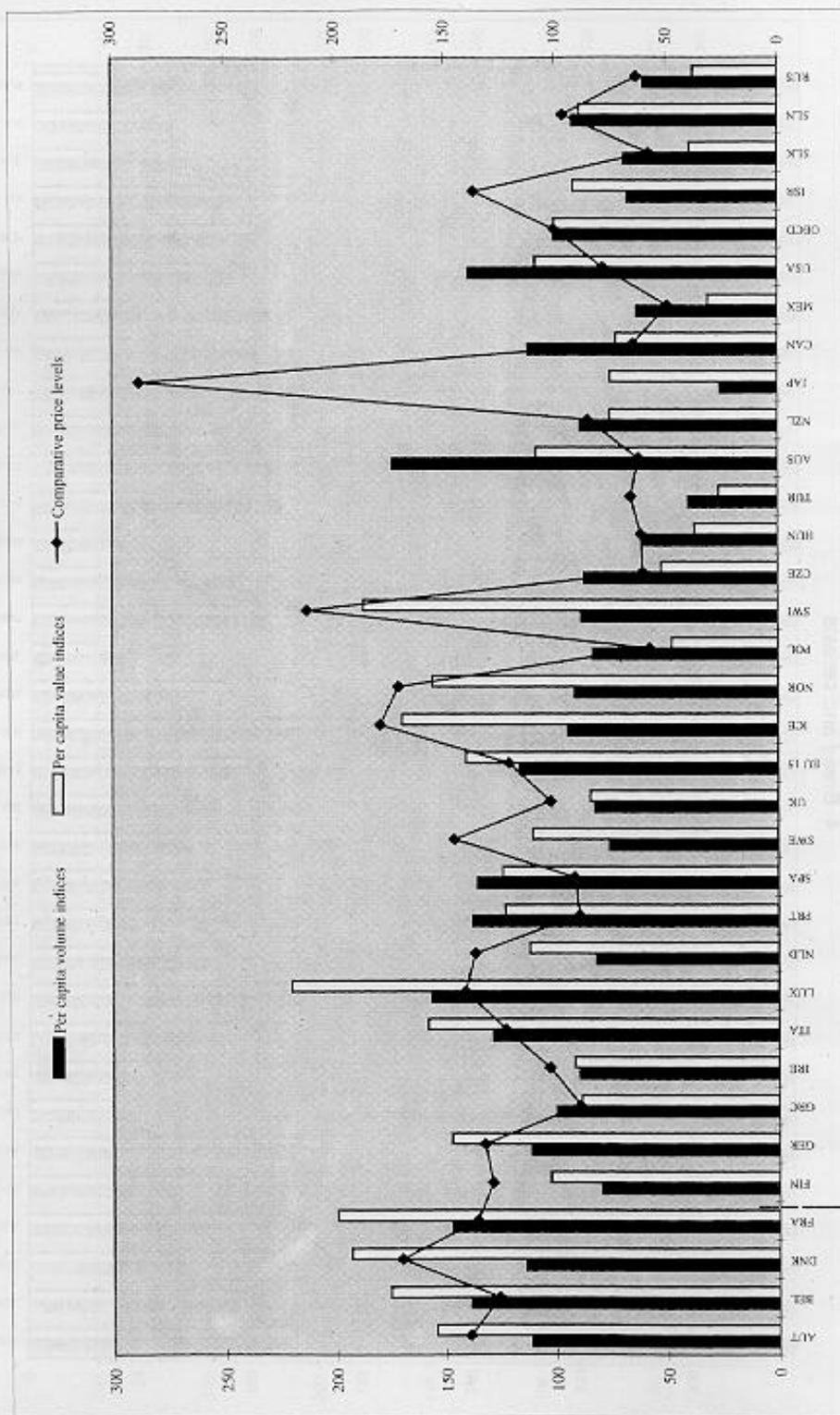
7/27/98

### EUROSTAT-OECD PPP PROGRAMME

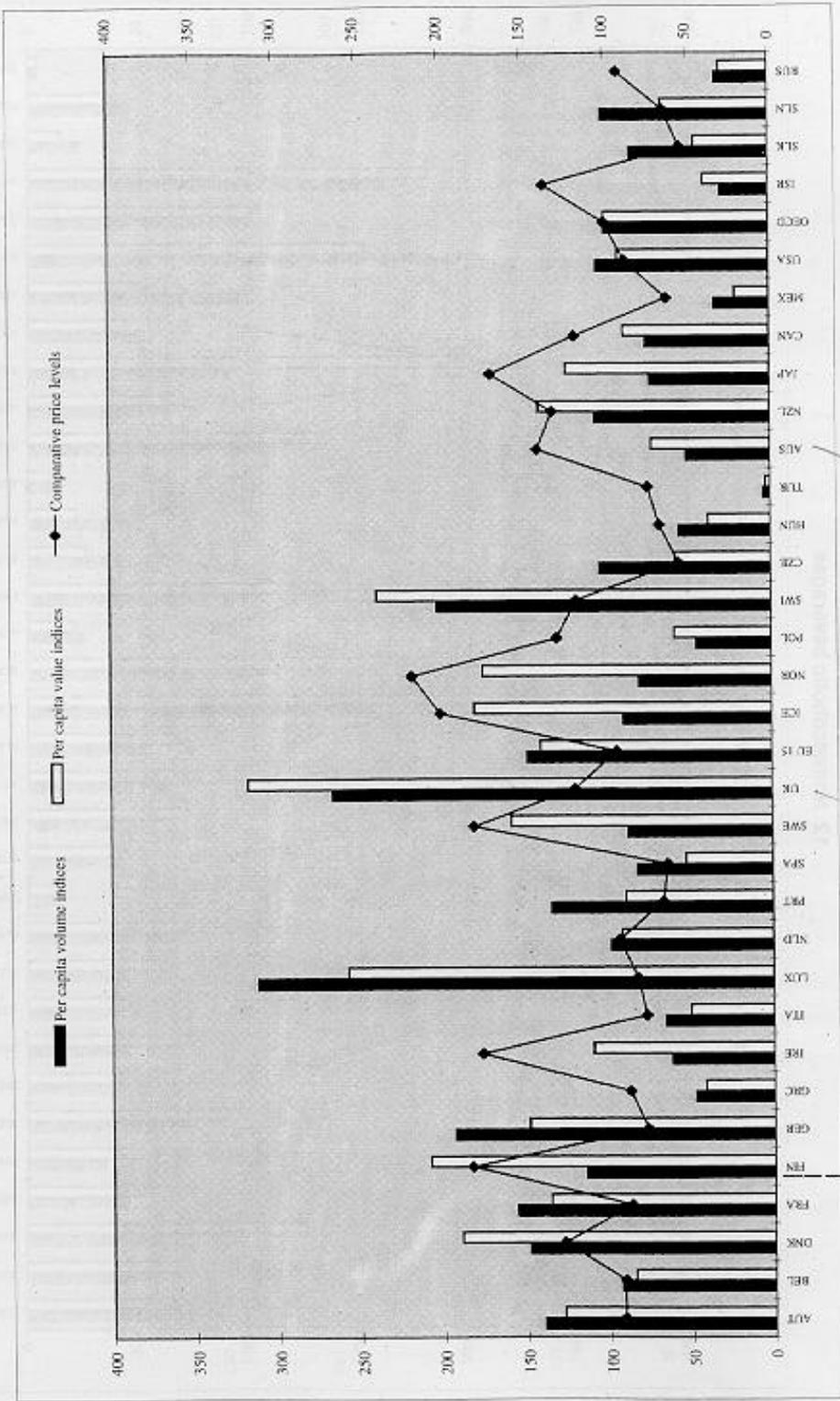
#### GRAPHS TO FACILITATE THE EVALUATION OF THE 1996 BENCHMARK RESULTS FOR ANALYTICAL CATEGORIES

1. The graphs show for each analytical category:
  - the per capita volume indices (based on national expenditure data converted using PPPs);
  - the per capita value indices (based on national expenditure data converted using exchange rates);
  - the comparative price level indices (based on the ratio of the PPP to exchange rate).
2. The indices cover the 32 countries that participated in the 1996 comparison. They are expressed relative to the OECD total, that is the total for the 28 participating OECD Member Countries. Israel, Slovenia, Slovak Republic and the Russian Federation are not included in the OECD total.

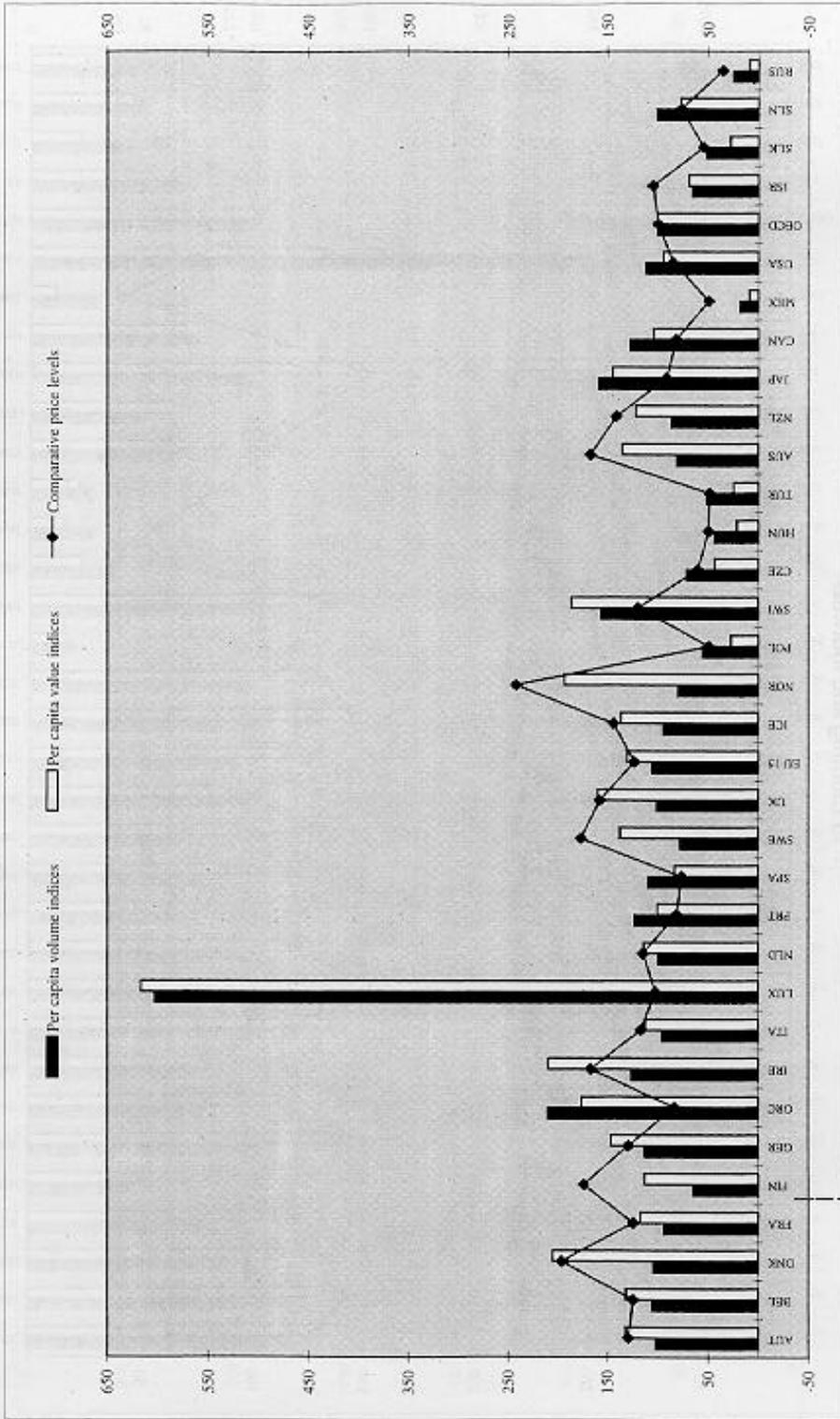
### 5 Meat



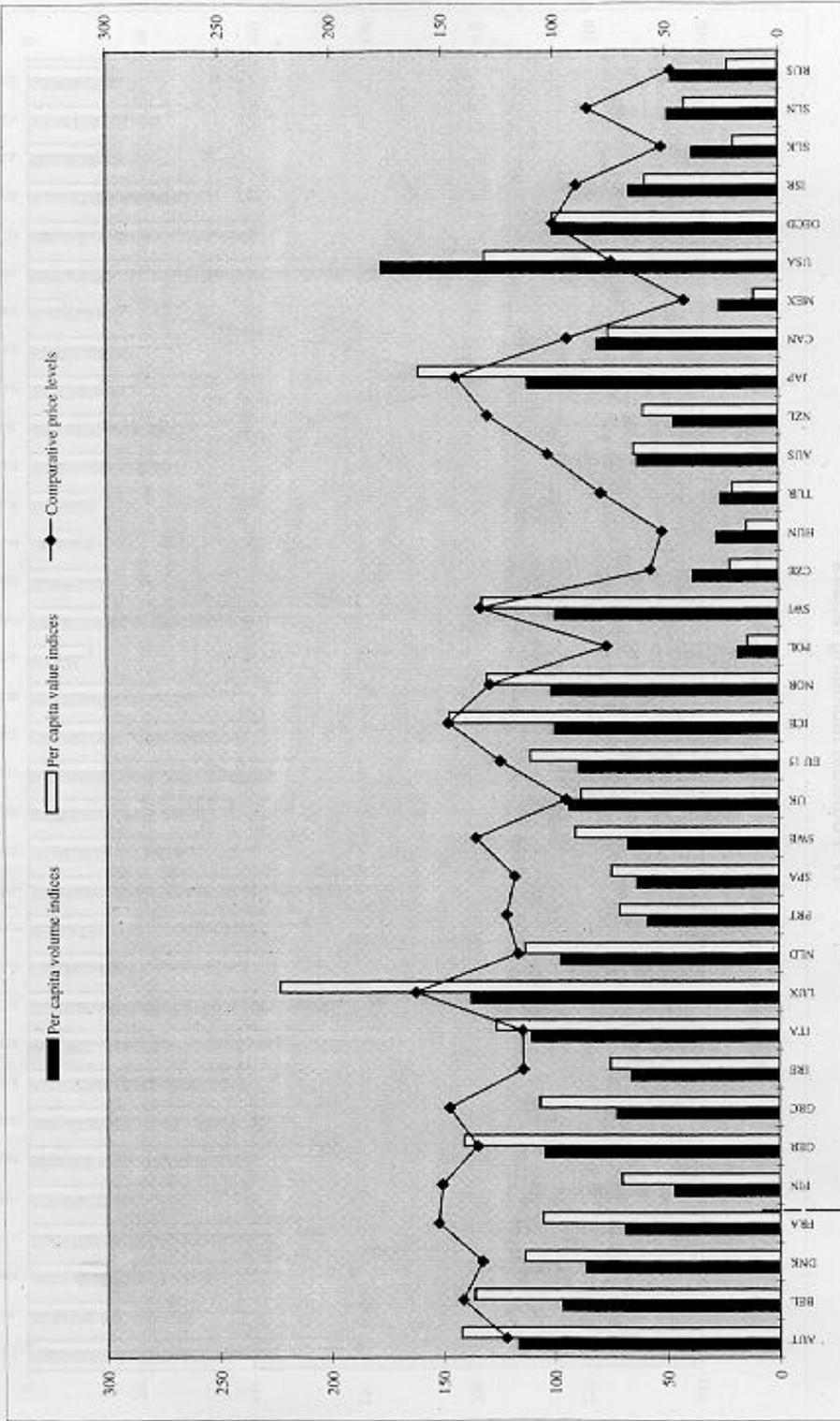
### 13 Alcoholic beverages



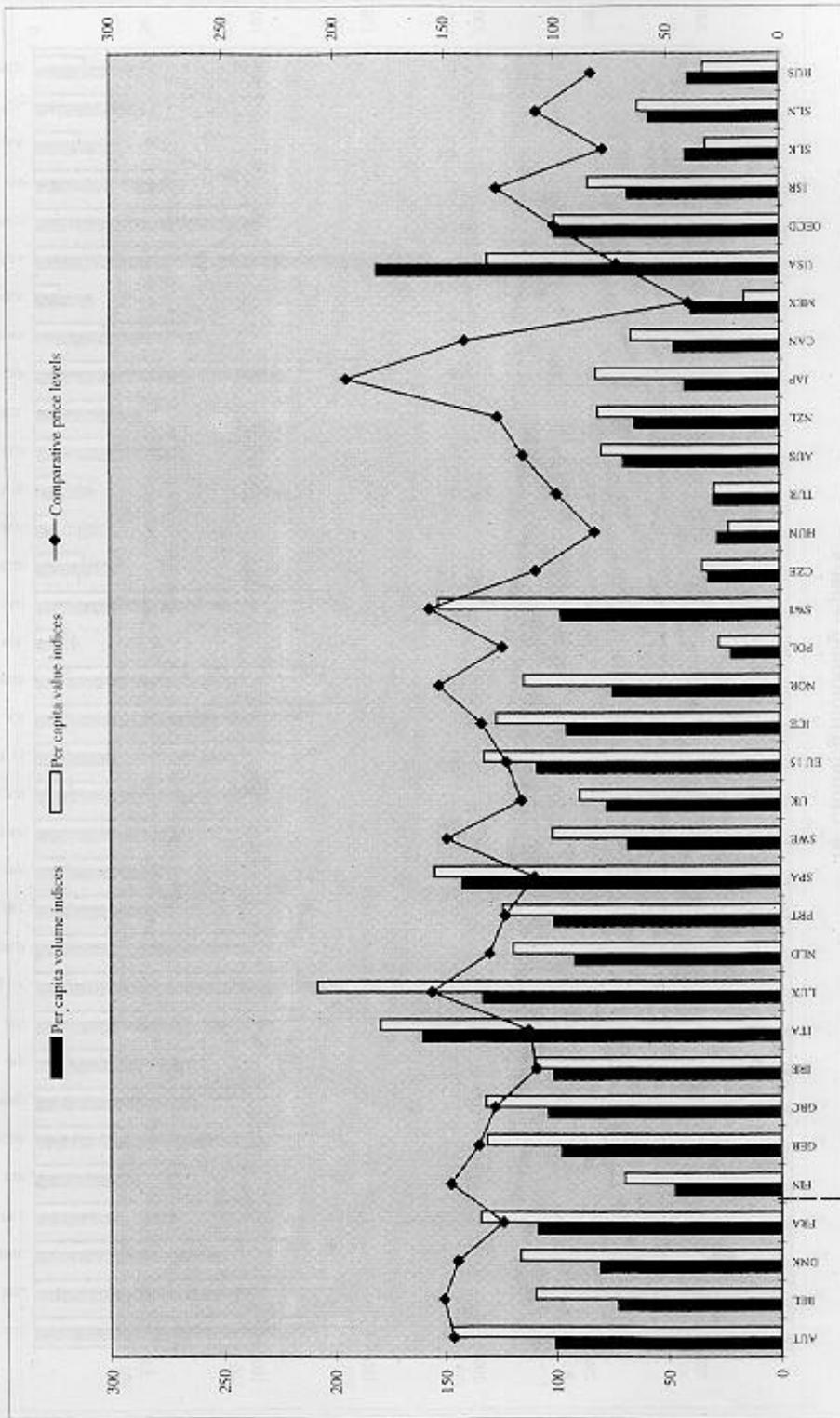
# 14 Tobacco



### 16 Clothing including repairs



### 17 Footwear including repairs



### 36 Education

