

OECD Science, Technology and Industry Working Papers
2003/14

Comparing Labour Productivity Growth in the OECD Area: The Role of Measurement

Nadim Ahmad,
François Lequiller,
Pascal Marianna,
Dirk Pilat,
Paul Schreyer,
Anita Wölfl

https://dx.doi.org/10.1787/126534183836

OECD

Unclassified **DSTI/DOC(2003)14**

Organisation de Coopération et de Développement Economiques
Organisation for Economic Co-operation and Development

12-Dec-2003

English text only

DIRECTORATE FOR SCIENCE, TECHNOLOGY AND INDUSTRY

DSTI/DOC(2003)14
Unclassified

COMPARING LABOUR PRODUCTIVITY GROWTH IN THE OECD AREA: THE ROLE OF MEASUREMENT

Nadim Ahmad, François Lequiller, Pascal Marianna, Dirk Pilat, Paul Schreyer and Anita Wölfl

JT00155838

Document complet disponible sur OLIS dans son format d'origine
Complete document available on OLIS in its original format

English text only

STI Working Paper Series

The Working Paper series of the OECD Directorate for Science, Technology and Industry is designed to make available to a wider readership selected studies prepared by staff in the Directorate or by outside consultants working on OECD projects. The papers included in the series cover a broad range of issues, of both a technical and policy-analytical nature, in the areas of work of the DSTI. The Working Papers are generally available only in their original language – English or French – with a summary in the other.

Comments on the papers are invited, and should be sent to the Directorate for Science, Technology and Industry, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France.

The opinions expressed in these papers are the sole responsibility of the author(s) and do not necessarily reflect those of the OECD or of the governments of its member countries.

---

http://www.oecd.org/sti/working-papers

Copyright OECD, 2003
Applications for permission to reproduce or translate all or part of this material should be made to:
OECD Publications, 2 rue André-Pascal, 75775 Paris, Cedex 16, France.

COMPARING LABOUR PRODUCTIVITY GROWTH IN THE OECD AREA: THE ROLE OF MEASUREMENT

Nadim Ahmad, François Lequiller, Pascal Marianna, Dirk Pilat, Paul Schreyer and Anita Wölfl*

Abstract

This paper examines how measurement problems affect international comparisons of labour productivity. It suggests that these measurement problems do not significantly affect the assessment of aggregate productivity patterns in the OECD area. However, these problems do influence the more detailed assessment of productivity growth, notably the role of specific sectors and demand components in aggregate performance. The paper shows that there are only a few significant problems regarding the comparability of nominal GDP across OECD countries, the most important being the treatment of software investment. In most cases, efforts are underway to reduce the size of these differences. Measurement differences for real GDP are also important, although several of these factors have impacts that work in different directions. Moreover, several of these problems primarily affect the distribution of total GDP across different expenditure categories and across different activities, not necessarily GDP growth. The measurement of labour input is the third major factor affecting labour productivity estimates. A key challenge in deriving comparisons of productivity growth across OECD countries lies in ensuring that estimates of labour input are consistent with the GDP information and that the components of labour input, hours worked and employment are internally consistent. The paper shows that a careful assessment of available data and their underlying concepts is often equally important for developing accurate estimates of labour productivity growth as cross-country differences in the underlying basic data. To address this problem, OECD is currently developing a reference database on productivity at the aggregate level.

* Nadim Ahmad, Dirk Pilat and Anita Wölfl work in the Economic Analysis and Statistics Division of the Directorate for Science, Technology and Industry; François Lequiller is the head of the National Accounts Division in the Statistics Directorate; Pascal Marianna works in the Employment Analysis Policy Division of the Directorate for Employment, Labour and Social Affairs; and Paul Schreyer is head of the Prices and Outreach Division of the Statistics Directorate.

COMPARAISONS DE LA CROISSANCE DE LA PRODUCTIVITE DE LA MAIN-D'ŒUVRE AU NIVEAU DE LA ZONE DE L'OCDE : L'IMPORTANCE DES QUESTIONS DE MESURE

Nadim Ahmad, François Lequiller, Pascal Marianna, Dirk Pilat, Paul Schreyer et Anita Wölfl*

Résumé

Le présent document examine en quoi les problèmes de mesure affectent les comparaisons internationales de la productivité de la main-d'œuvre. Il montre que ces problèmes n'ont pas de répercussions notables sur l'évaluation du profil général d'évolution de la productivité dans la zone de l'OCDE. Par contre, ils ont une influence dès lors qu'on veut pousser plus avant l'analyse de la croissance de la productivité, notamment déterminer la part qui est imputable à un secteur particulier ou à telle ou telle composante de la demande. Les auteurs observent que la comparabilité entre pays de l'OCDE des données sur le PIB nominal soulève peu de problèmes importants et que le principal de ces derniers est lié au traitement des investissements en logiciels. Dans la plupart des cas, des efforts sont déployés pour réduire l'ampleur de ces divergences. Les disparités au niveau de la mesure du PIB réel sont également importantes, même si elles ne jouent pas toutes dans la même direction. Qui plus est, plusieurs d'entre elles se ressentent principalement au niveau de la répartition du PIB total entre les différentes catégories de dépenses et activités, mais pas forcément à celui de la croissance du PIB. La mesure des apports en main-d'œuvre est le troisième grand facteur qui influe sur les estimations de la productivité de la main-d'œuvre. Lorsqu'on veut procéder à des comparaisons de la croissance de la productivité entre pays de l'OCDE, il est essentiel de s'assurer de cohérence des estimations des apports de main-d'œuvre avec les données sur le PIB et de la cohérence interne des composantes des apports de main-d'œuvre, de la durée du travail et de l'emploi. Le document montre qu'un soigneux examen des données disponibles et des concepts sur lesquels elles reposent est souvent aussi important pour l'exactitude des estimations de la croissance de la productivité de la main-d'œuvre que l'étude des écarts entre pays dans les données servant à établir ces estimations. Afin de contribuer à remédier au problème, l'OCDE a entrepris d'élaborer une base de données de référence sur la productivité au niveau global.

* Nadim Ahmad, Dirk Pilat et Anita Wölfl travaillent à la Division des analyses économiques et des statistiques de la Direction de la science, de la technologie et de l'industrie, François Lequiller est le Chef de la Division des comptes nationaux de la Direction des statistiques, Pascal Marianna travaille à la Division d'analyse et de politique d'emploi de la Direction de l'emploi, du travail et des affaires sociales, et Paul Schreyer est le Chef de la Division des prix et des activités d'ouverture de la Direction des statistiques.

TABLE OF CONTENTS

1. Introduction 6
2. Measuring nominal GDP 9
2.1 Issues in comparing nominal GDP 9
2.2 The treatment of expenditure on military equipment 9
2.3 The treatment of financial intermediation services 10
2.4 Underlying statistical systems and the non-observed economy 11
2.5 The measurement of investment in software 12
3. Measuring real GDP 19
3.1 Adjusting for quality change - the role of hedonic price indexes 19
3.2 Measuring real output in services 22
3.3 The choice of index numbers 27
4. Measuring labour input and labour productivity 32
4.1 What measure of labour productivity growth? 32
4.2 Measuring hours worked 33
4.3 Measuring employment 37
4.4 The coherence of employment and hours worked 39
4.5 Adjusting for labour quality 40
5. Concluding remarks 41
REFERENCES 43

Boxes

Box 1. Purchasing power parities in comparisons of labour productivity growth 19
Box 2. A brief comparison of index numbers formulae 29

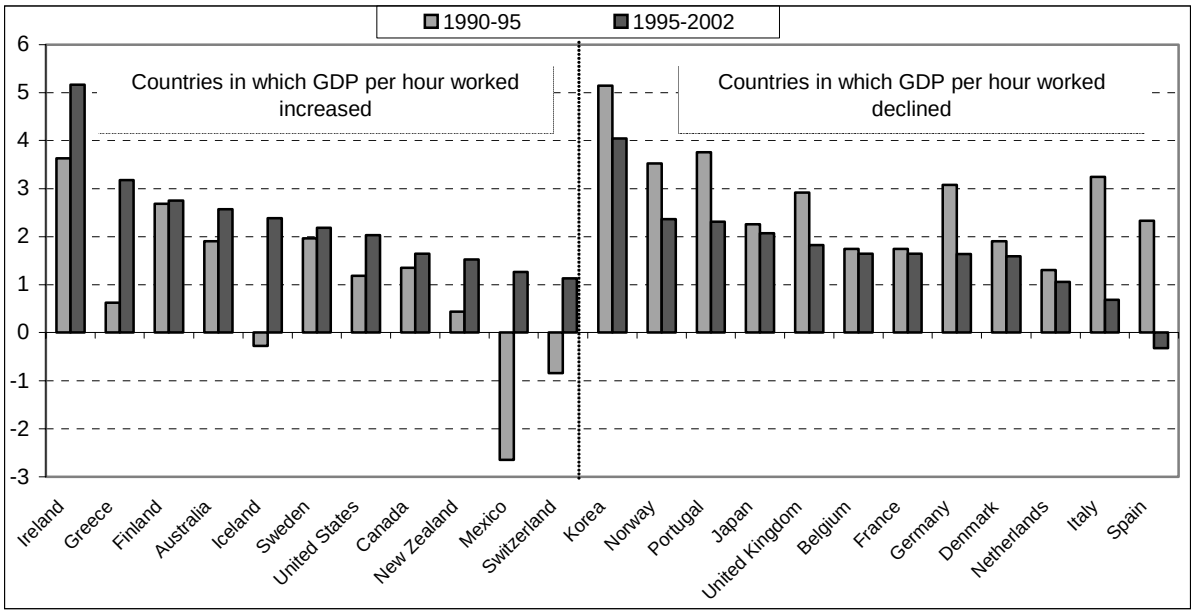
COMPARING LABOUR PRODUCTIVITY GROWTH IN THE OECD AREA: THE ROLE OF MEASUREMENT

Nadim Ahmad, François Lequiller, Pascal Marianna, Dirk Pilat, Paul Schreyer and Anita Wölfl

1. Introduction

Growth and productivity are on the policy agenda in most OECD countries, as countries seek to enhance economic performance, raise standards of living, and address a range of socio-economic challenges. Recent OECD work on growth and the "new economy" pointed to a large diversity in growth and productivity performance in the OECD area and to a range of policies that could enhance growth (OECD, 2001*a*). Work following up on this study demonstrated that some of these differences in growth performance persist until today (OECD, 2003*a*; 2003*b*). Indeed, available estimates of labour productivity growth from the OECD work suggest that the United States has experienced faster labour productivity growth than most large EU countries over the past years (Figure 1). The US experience is especially remarkable since it started from a very high level of labour productivity in the 1990s and since productivity growth accelerated sharply over the 1990s.

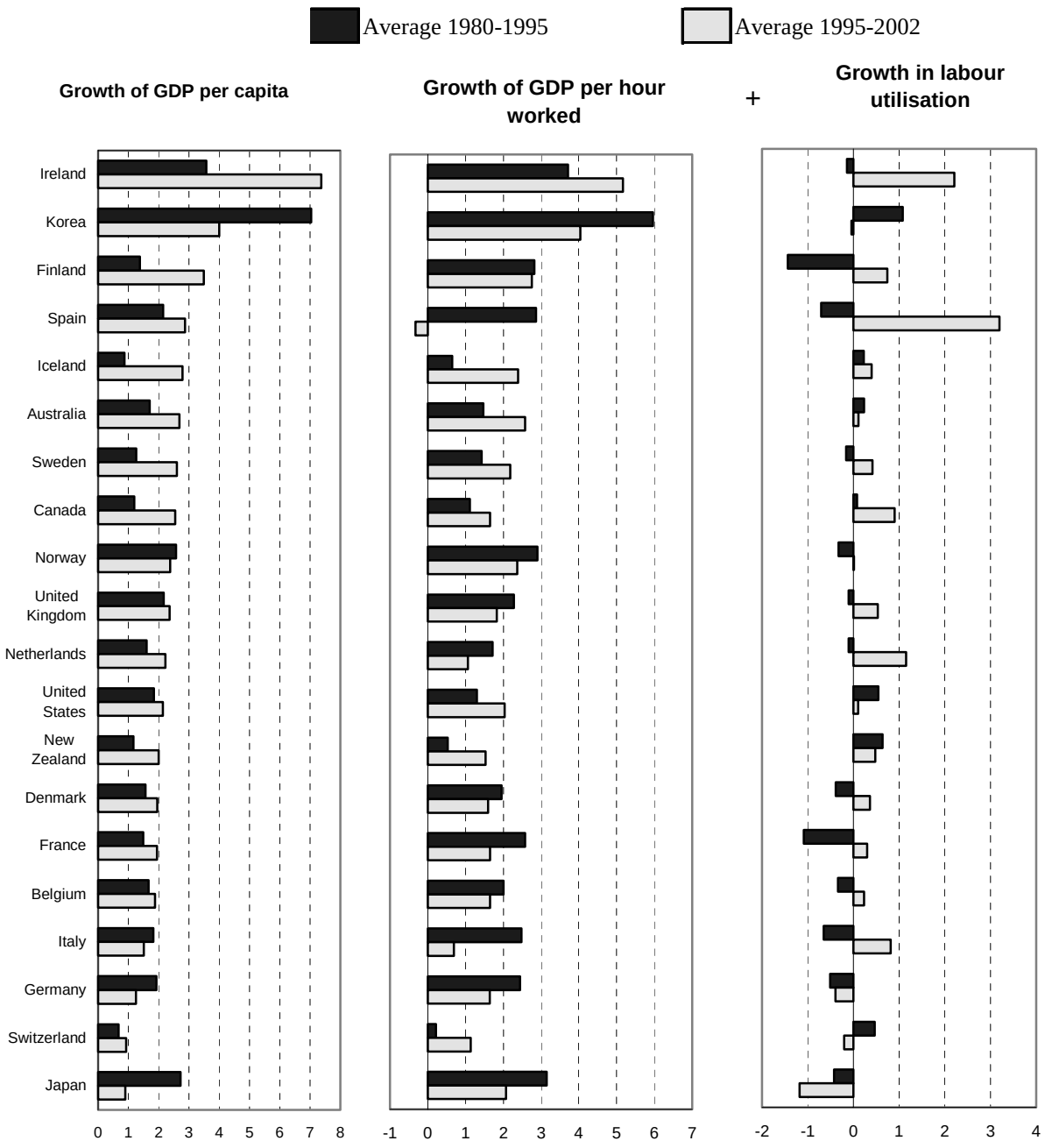
Figure 1. **Growth of GDP per hour worked, 1990-95 and 1995-2002**
Annual average growth rates, in per cent



Source: OECD **estimates** based on the OECD Productivity Database.

Growth of GDP per capita in the United States has been somewhat more rapid in comparative terms, since strong US labour productivity growth was combined with increased labour utilisation over the 1990s, in contrast with several European countries (De Serres, 2003; Figure 2). In addition, US GDP growth has been substantially faster than that in large European countries or Japan, as the US population expanded rapidly over the 1990s.

Figure 2. **Changes in labour utilisation contribute to growth in GDP per capita**
Percentage change at annual rates



Source: OECD, Productivity Database, see De Serres (2003) for estimates adjusted for the business cycle.

The objective of this paper is to examine how measurement problems may potentially affect international comparisons and therefore the validity of cross-country analysis. Substantial progress has been made over the past years in improving the comparability of GDP estimates across countries. This is particularly the case in the European Union where levels of GDP form the basis for contributions to a large part of the EU budget. Nevertheless, as this paper demonstrates, differences in measurement continue to exist, even within the EU. This paper suggests that these measurement problems do not significantly affect the assessment of aggregate productivity patterns in the OECD area, nor do they invalidate the relatively strong productivity

performance of the United States. However, they do influence the more detailed assessment of productivity growth, notably the attribution of labour productivity to specific sectors and demand components.

In examining growth performance, it is important to distinguish two different issues in measuring GDP and productivity, namely *comparability in levels*, and *comparability in growth rates*. In statistical terms, it is more difficult to compare levels across countries than growth rates because levels are more difficult to measure in comparable terms than changes in a single country over time. For example, to compare GDP per capita between countries, data expressed in different currencies must be translated in a common unit. This is usually done using purchasing power parities (PPPs), which are based on the comparison of the prices of a common basket of goods and services between countries. Comparing prices between countries is much more difficult than comparing the change in prices over time in each country. For example, comparing the level of rents between two countries is difficult as it is not easy to find dwellings that are comparable in quality. It is easier to compare the change in the price of rents in each country. For this it is sufficient to compare the price of the rent of the *same* dwelling in two successive years. Of course, the quality of the dwelling may have changed in one year, but the measurement problem is less prevalent.

This argument can be extended to other components of GDP and to GDP itself. Levels of GDP are therefore likely to be less comparable than GDP growth rates, and by extension, productivity growth rates are likely to be more comparable than productivity levels. However, the comparison of productivity levels cannot be ignored, if only because growth and levels are not independent. For example, differences in the methods used to measure the level of ICT investment will have consequences on GDP growth if the growth of ICT investment is significantly different from average GDP growth. This has been the case with investment in software, as explained below.

This paper focuses on comparisons of *labour productivity growth*, leaving comparisons of MFP growth for future work.[1] Moreover, the paper primarily focuses on the comparability of labour productivity growth at the aggregate level; sectoral aspects are only discussed insofar as they affect estimates of aggregate productivity growth. It draws on existing OECD work, carried out across several OECD directorates, notably the Statistics Directorate, the Directorate for Science, Technology and Industry, and the Directorate for Employment, Labour and Social Affairs. The next section discusses a number of key problems in the measurement of nominal GDP, notably the treatment of expenditure on military equipment; the adjustment for financial intermediation services; the non-observed economy; and the measurement of software. The third section discusses issues in measuring real GDP, notably how to separate price and volume changes in areas characterised by rapid technological change, *e.g.* through the use of hedonic price indexes; the measurement of output in services sectors; and the choice of index numbers. The fourth section discusses the measurement of labour input, on the basis of both employment and hours worked. It also examines how to combine different sources in estimating productivity growth for OECD countries. The final section draws some conclusions and provides an assessment of the main measurement problems.

1. Work on MFP growth is also underway at the OECD, *e.g.* in the context of the development of a database on capital services (Schreyer *et al.*, 2003), and of an OECD database on productivity growth.

2. Measuring nominal GDP

2.1 Issues in comparing nominal GDP

Comparability of nominal GDP expressed in national currency units depends to an important degree on the use of a common conceptual framework for measurement. This framework is the 1993 version of the international "System of National Accounts" (SNA 93), which was developed jointly by five international economic organisations (OECD, the United Nations, International Monetary Fund, European Union and World Bank). The European version of the manual is the 1995 European System of Accounts (ESA). Except for two countries (Switzerland and Turkey), all OECD member countries have now adopted the new manual as the basis of their national accounts. The introduction of the new system created substantial problems during the transition period, which have now been overcome.[2] But despite the convergence, some differences still exist between countries in the degree of implementation of the manual, in particular between the United States, Europe and Japan.[3] The examples below highlight those areas that have a significant impact on GDP estimates.[4]

2.2 The treatment of expenditure on military equipment

In 1996, the US National Income and Product Accounts (NIPA) were substantially modified to recognise investment by general government. Prior to 1996, all government expenditure was recorded in the NIPA as current expenditure. The change in classification of these government expenditures on public structures, infrastructures and equipment had a significant indirect impact on GDP, through increases in capital consumption[5], resulting in a systematic increase in the *level* of US GDP at current prices of approximately 2%.

In making this change, the US NIPA accounts came substantially closer to the 1993 SNA, which recognises public investment. However, paradoxically, the extension of the coverage of investment in the NIPA was more extensive than that recommended by the SNA, since it included expenditures on military equipment (aircraft, ships, missiles) that were not considered assets by the SNA.[6] The national accounts in Europe, Canada and Asia strictly follow the SNA in this matter; they only capitalise (and depreciate) military expenditures that can be also used for civilian purposes. This includes some military buildings, airports and transport aircraft, but excludes aircraft, ships or weapons that are exclusively intended for military purposes.

The treatment of military equipment has therefore resulted in a statistical difference in the measurement of GDP. Before 1996, US GDP had to be adjusted upwards for international comparisons; currently, it has to be adjusted downwards. This adjustment is made in the US GDP estimates that are published in the official *Annual National Accounts* of the OECD, but is omitted in the quarterly data published by the US Bureau of

2. A further revision of the SNA has, however, been announced for 2008.

3. The US "National Income and Product Accounts" differ significantly in presentation from the SNA but the US Bureau of Economic Analysis prepares a set of annual accounts for the OECD which are presented in SNA format. The upcoming 2003 comprehensive revision of the NIPA will significantly reduce these presentational differences.

4. Other differences that may affect nominal GDP exist but these are typically less important. For a helpful study see Lal (2003).

5. In the national accounts, the production of government is conventionally estimated as the sum of costs, including the cost of the depreciation of the capital.

6. This issue continues to be debated between national accounts experts; the current principles could be changed in the 2008 revision of the SNA.

Economic Analysis (BEA), and in the OECD's *Quarterly National Accounts*. Fortunately, the size of this adjustment remains limited. In the 1980s, it amounted to around 1.0% of GDP, but in 2001 it was only 0.6% of GDP, following the decrease in military expenditure concurrent with the demise of the Soviet Union.[7] The impact on real GDP growth has tended to be relatively small. Over the past decade, inclusion of this factor in US GDP has implied that growth was 0.03 percentage points per year less than it would have been if the United States had used the same approach as European countries or Japan.[8] Recently planned increases in US military expenditure may reverse this pattern, however.

2.3 The treatment of financial intermediation services

Banking services have historically been a difficult area for statistical measurement as most of these services are not explicitly priced. Some services are explicitly charged, such as the rental of a safe or stock-broking fees. However, the bulk of the income of banks is from interest, dividends, and other property income and is not directly linked to a material service. Such services are "implicitly" charged, essentially through the difference between interest received and interest paid. Thus, in the SNA, this implicit production of banks is estimated using the difference between interest received and paid. This method has been and continues to be debated, but all OECD member countries have estimated this part of bank production, known as "Financial Intermediation Service Indirectly Measured" or "FISIM".

It is relatively straightforward to recognise and estimate FISIM. The key problem is breaking it down between final consumers (households) and intermediate consumers (businesses). Only the first part has an overall impact on GDP. In the United States, Canada and Australia, such a breakdown has been estimated in the national accounts for some time, in accordance with the recommendations in the SNA. In Europe and Japan, despite the recommendation of the SNA 93, the implementation of a breakdown between final and intermediate consumers has been delayed, since there was, until recently, no internationally-agreed statistical method of allocation between final and intermediate consumers. This led to concerns, particularly in Europe, that the change in the SNA in this area would affect the comparability of GDP levels, which are increasingly used as the basis of European Union countries' contributions to the EU budget.

Recently, agreement was reached in Europe on a standard method that will be implemented in 2005. Japan is expected to follow. However, in the meantime, significant differences regarding FISIM remain. In the United States, GDP includes an amount of USD 250 billion (2.3% of GDP) for imputed household consumption of financial services[9], whilst in Europe and Japan, there is no imputed household consumption of financial services. Thus a significant adjustment is needed in order to correctly compare GDP per capita.[10]

---

7. The impact on the GDP level of capital consumption on military expenditure that has civilian uses ranges from 0.2% of GDP for France and the United Kingdom, to 0.1% for Germany, and 0.04% for Italy.

8. This figure was obtained by calculating GDP growth excluding capital consumption of military expenditures and comparing it with official estimates.

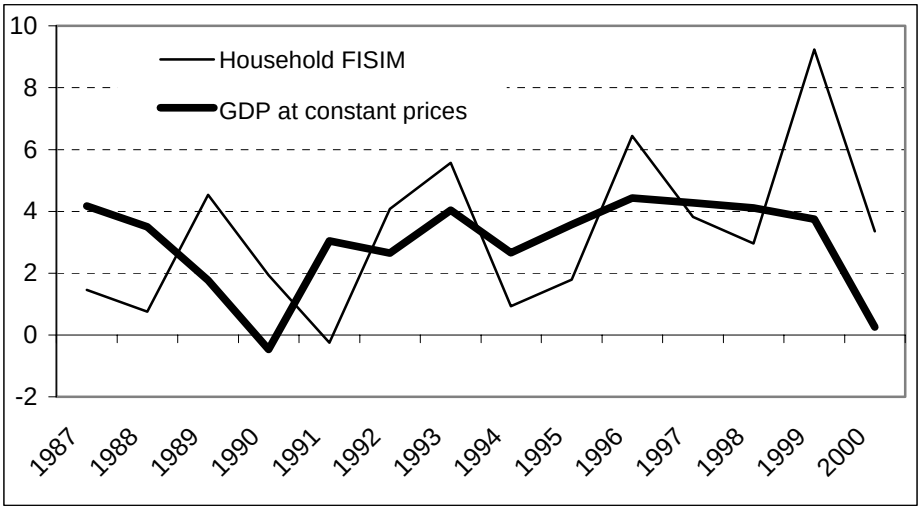
9. To be exhaustive, one should mention that imputed household payments for FISIM are not the only part of FISIM that affect GDP. Imputed government payments and exports also affect GDP (imputed imported FISIM is treated in the NIPA as income and not services). However, these flows are small compared to imputed household FISIM. In 2001, exports were estimated at USD 18 billion, government imputed FISIM at USD 11 billion while household imputed FISIM reached nearly USD 260 billion.

10. The data provided by Canada to the OECD Annual National Accounts are consistent with those for Europe and Japan. However, US data are not adjusted in the same way.

Fortunately, the current difference in levels has only a minor impact on GDP volume growth. To measure the impact of the difference on growth rates one should in theory include estimates of household FISIM within European and Japanese estimates of GDP, bringing them closer to SNA. However, the relevant time series are not readily available for the EU countries and Japan, as the data are still under development. The impact of the United States-Europe/Japan differential can however be measured by adjusting US volume GDP to exclude imputed household consumption of FISIM. Between 1987 and 2001, the difference between official US GDP growth and an estimate excluding imputed household FISIM, is on average significantly less than 0.1% per year. No systematic bias appears between the two GDP estimates over the period (the cumulative difference over 15 years is a mere 0.2%). However, as shown below (Figure 3), over the past two years, volume growth of imputed FISIM has diverged from overall US GDP growth, thus contributing more than average to aggregate GDP growth. If the US national accounts had used the same method as the European countries and Japan, US GDP growth would have been less by 0.1%, both in 2000 and 2001.

In principle, this difference in methodology should be largely reduced in 2005. First, the US Bureau of Economic Analysis has reviewed its method of allocation and will implement a new system in October 2003. This should reduce the difference, from 2.3% of GDP currently, to probably a little over 1% of GDP. Second, the EU member states and Japan have announced that they will implement the allocation of FISIM in their accounts, starting in 2005. Preliminary estimates suggest that European GDP levels would increase by approximately 1.3%, an amount close to the impact in the US. However, there will remain differences in the precise method of compilation.

Figure 3. **Volume growth of household FISIM and GDP, United States, in %**



Source: Bureau of Economic Analysis.

2.4 Underlying statistical systems and the non-observed economy

National accounts typically aim to achieve a maximum coverage of GDP ("exhaustiveness"). This is based, first, on the use of the maximum range of statistical and administrative sources that is available, and, second, on adjustments that are implemented on top of this to take into account the activities that are not observed through these statistics or administrative sources. The less exhaustive the statistical or

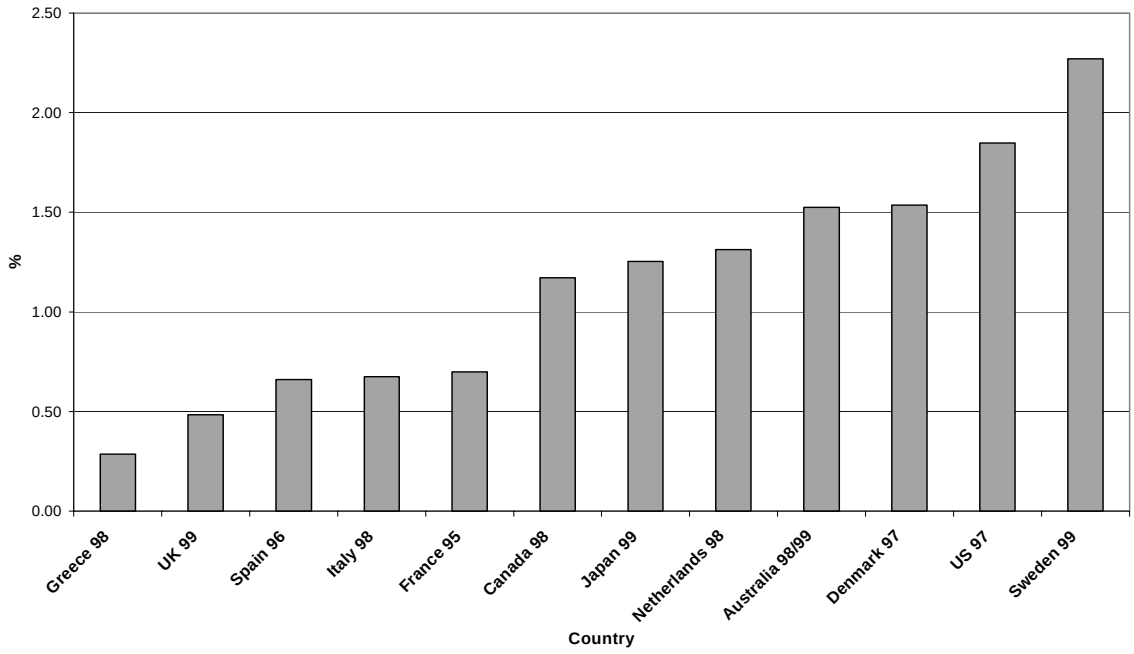
administrative system the more it is necessary to make additional adjustments for the "non-observed economy".[11]

Among OECD countries, Italy, which leads methodological discussions in this area, makes an adjustment of 15% for the non-observed economy. Poland makes an adjustment of around 7%, Hungary 16%, and Belgium 4%. However, even in Europe, there is currently no available systematic synthesis of the adjustment made by each country. Moreover even if this were available, it would still be difficult to compare the reliability of GDP estimates across countries as the adjustments depend on the quality of the underlying statistical system, which differs from one country to another. That said considerable efforts are being made to improve the comparability and exhaustiveness of estimates, in particular by Eurostat. The OECD has accompanied this effort by publishing a *Handbook on Measuring the Non-Observed Economy*. Thanks to all these efforts, the impact on growth rates of differences in statistical methodologies regarding these types of adjustments is likely to be quite small.

2.5 The measurement of investment in software

Another major issue in the comparability of GDP concerns the measurement of software. The 1993 SNA recommended that purchases of software (and any own-account production) should be treated as investment as long as the acquisition satisfied conventional asset requirements. This change added about 1% to GDP in most OECD economies in the mid-1990s. However, the range of the adjustments to GDP differs substantially across OECD countries (Figure 4).

Figure 4. **Total investment in software, percentage of GDP**



Source: OECD, see Ahmad (2003).

11. A complete discussion on the non-observed economy issue in GDP is presented in the November 2002 issue (No. 5) of the OECD's *Statistics Brief* series. This study also covers the question of the measurement of illegal activities, which is currently, in practice, excluded from GDP. However, discussions are in progress among European countries to include these activities in the future.

An OECD Task Force set up in October 2001 confirmed that differences in estimation procedures contributed significantly to the differences in software capitalisation rates, and a set of recommendations describing a harmonised method for estimating software were formulated (see Lequiller, *et al*, 2003; Ahmad, 2003).[12] Most of these recommendations were approved at the OECD 2002 National Accounts Expert meeting and should therefore be implemented by countries. However, this will only happen with a delay of several years. Differences in software and GDP measures will therefore persist for some time. The key measurement issues concerning software investment are briefly discussed below.

Supply versus demand

In practice, National Statistical Offices (NSOs) use one of two distinct methods to estimate software investment. The first is based on how businesses record investment in practice, using a conventional survey based approach (known as a 'Demand' approach). The second is to measure the total supply of computer services in an economy and estimate the amount of software with asset characteristics (the 'Supply' approach). In practice, Demand-based approaches tend to produce systematically lower estimates of investment than Supply-based methods. This mainly reflects the fact that businesses tend to use very prudent criteria when capitalising software, particularly if there are no tax incentives for doing so. Quite often, large software companies do not capitalise any own-account production of software 'originals' at all. As a result, many NSOs use the Supply approach.

To estimate the significance of the difference between the two methods, countries were asked to provide estimates of software investment based on both the Supply and Demand approaches. Only four countries were able to provide demand-based estimates, namely Australia, Canada, France and the United States, pointing to the dearth of demand information available. For Australia and Canada, respectively, supply estimates were 7 and 4 times as large as the demand-based estimates. For France (despite excluding a large proportion of software supply from the calculations) supply was about ⅓ higher. For the United States, supply estimates of purchased software were over ten times as high as demand estimates (USD 123 billion versus USD 11.8 billion).

Differences in the measurement of software should distinguish between two key components, namely purchased software and own-account software.

Purchased software: The investment ratio

Central to the issue of measurement in a Supply approach is how capital expenditures on computer services (software) should be delineated from intermediate consumption of computer services. In other words, the ratio of capitalised software to total expenditure (by businesses and government) on computer services (the investment ratio) can be regarded as a measure of the propensity of any country's statistical office to capitalise software. 'Computer services' is usually defined within international and national product classification systems.[13] It covers a fairly heterogeneous range of services, some with asset characteristics, *e.g.* customised software services, and some without such characteristics, *e.g.* hardware consultancy. *A priori* one would expect the relative size of these services to total computer services to be fairly similar across economies. By extension, if countries applied the same criteria to determine when to capitalise expenditure, one would expect investment ratios to be fairly similar too. A comparison of these ratios therefore provides an insight into the scale of measurement differences across countries. Figure 5, below,

12. The report is available at http://www.oecd.org/EN/document/0,,EN-document-424-15-no-20-31122-0,00.html

13. For example, CPA72, for the European product classification system (or US SIC 73.7 for the US industrial classification system, or NAICS 541511, 511210, 541512, 518210, 541513, 518210, 541512, 541519, in the latest US classification system).

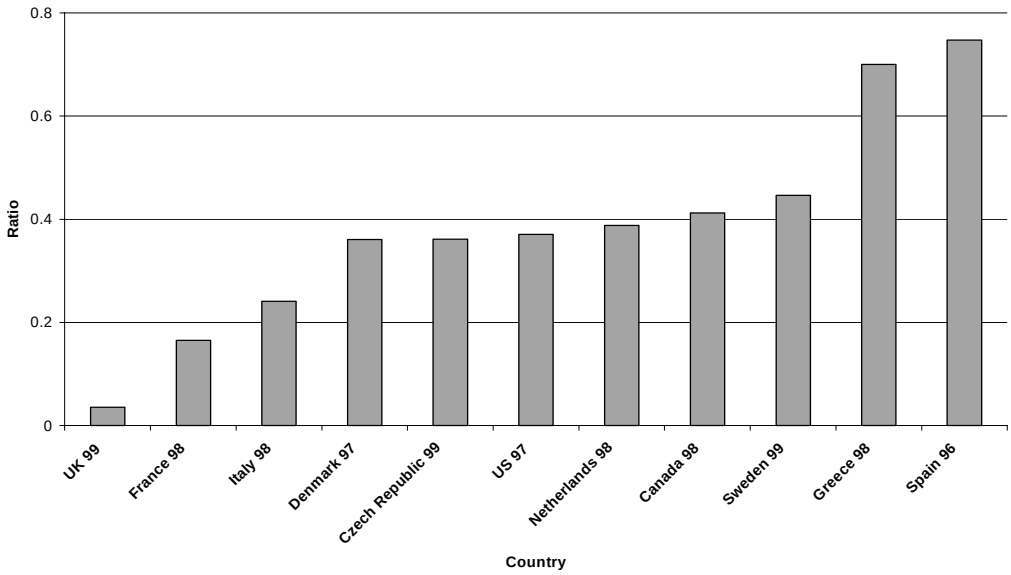
shows the investment ratio for purchased software. It confirms that the range of investment ratios is substantial,[14] with ratios of under 0.04 in the United Kingdom and over 0.7 in Spain.

Measurement, rather than economic factors underpins these differences, reflecting the fact that countries use different criteria to determine whether particular categories of computer services should be capitalised or not. For example, not all countries capitalised software reproductions purchased separately, or the – often ill-defined – category of "software consultancy" services. France, for example, capitalises no computer consultancy services although most other countries capitalise varying proportions, often depending on the nomenclature and detail available in their classification systems. Other differences also impair comparability. For example, not all countries adopt – more or less comprehensive – adjustments to correct for double-counting when using the Supply-method, *e.g.* to correct for: sales of software destined for resale; sales of software based on subcontracts; sales of software to software manufacturers; and sales of software to computer hardware and other machinery and equipment manufacturers. But even amongst those countries that make such adjustments, estimation differences persist.

The OECD/Eurostat Task Force developed a harmonised procedure for estimating software using a supply approach based on recommendations on what types of software should be capitalised, and when, including the nature of adjustments used to correct for any double counting. The expectation is that these recommendations will result in more comparable estimates of software investment ratios across countries. It is also possible to estimate the potential impact of harmonisation on country estimates of software investment by applying a harmonised ratio. Figure 6 below presents two scenarios, the first using an investment ratio of 0.4 (roughly the average of investment ratios across countries, and using criteria broadly in line with the OECD recommendations) and the second using an investment ratio of 0.04 (the UK investment ratio). It shows that the potential for change in software investment levels is substantial, about +/- 1½% of GDP in the UK and Sweden respectively, depending on the investment ratio used.

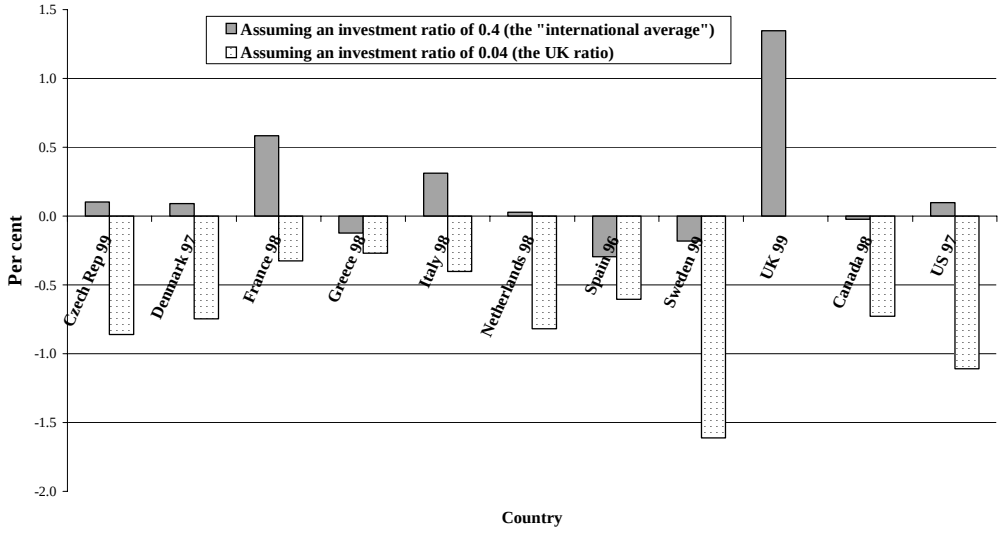
14. Part of the reason for differences in ratios is that definitions for computer services are not exactly equivalent for all countries. This cannot explain the differences between EU countries, however, since the definition used within the EU is exactly the same. It is interesting to note that at the more detailed level, (4 and 6-digit product classification) differences are larger. For example, for a given expense of 100 monetary units on similar (detailed) types of software services, the United States will capitalise 100% whilst France will capitalise only 50%.

Figure 5. **Investment ratios for purchased software**



OECD, see Ahmad (2003).

Figure 6. **Estimated impact on GDP (percentage) using different assumptions for investment ratios for purchased software**



Source: OECD, see Ahmad (2003).

All other things being equal, one would expect to see similar changes to GDP levels. However, this depends on how the change is implemented. In some countries the total levels of investment on all products derived from demand-based surveys (excluding own-account production) are considered to be fairly robust, but the allocation to specific asset types, using Supply approaches, less so. In this regard an increase in *purchased* software investment may result in a decrease in investment in other asset categories. The overall effect may be no change, or a negligible change, in total investment and GDP. However, changes to *estimates of own-account production* of software will usually result in an increase in GDP. For example, when the United Kingdom and Canada first included estimates of software investment they discovered that some software had already been previously recorded by businesses as investment but

allocated to different asset types in official statistics. Therefore the overall change to GDP in both countries was less than the level of software capitalisation. That said, despite the possibility that GDP estimates may only show marginal changes, these changes do matter for productivity measurement, and particularly for estimating the contribution of ICT.

Own-account production

Every country estimates own-account software using an input-method. Japan is the only exception, since it does not yet include own-account software in its national accounts. In principle own-account software is supposed to include all production costs, and intermediate or primary income, such as wages and salaries. All countries include wages and salaries as being the key determinant but many do not include intermediate costs, *e.g.* the Netherlands and Australia. But even where adjustments are made differences exist. For example, Denmark applies a mark-up factor of 1.47 to estimate other input costs, relative to wages and salaries, whereas the United States uses a factor of 1.02. The estimation process for wages and salaries also differs. For example, the definition of an individual engaged in own-account production differs across countries (some definitions account for a wider pool of workers). In addition, the adjustment made to correct for the amount of time these selected individuals spend on other production activities, not own-account production, also differs across countries. Australia, Denmark, Finland, the Netherlands and Sweden make no adjustment, while Canada, France and the United States make an adjustment of 50% for the share of time spent on own-account software production.

Some countries (*e.g.* the United States and Canada) also do not capitalise own-account production of original software designed for reproduction (*e.g.* MS Excel). However, in line with the OECD Task Force recommendations, both countries are in the process of doing so. Table 1 below compares official estimates of own-account software against estimates based on the OECD Task Force recommendations.[15]

Table 1. **Own account software – as a % of GDP**

	Year	Original data	Revised data	Difference
Denmark	1997	0.7	0.4	-0.3
Finland	1995	0.4	0.4	0.0
France	1998	0.3	0.8	0.5
Greece	1998	0.0	0.2	0.2
Italy	1998	0.2	0.4	0.2
Netherlands	1998	0.4	0.9	0.5
Spain	1996	0.0	0.3	0.3
Sweden	1999	0.5	1.0	0.5
United Kingdom	1999	0.4	0.8	0.4
Australia	1998/99	0.5	0.7	0.2
Japan	1999		0.6	0.6
Canada	1998	0.4	1.0	0.6
United States	1992	0.6	0.9	0.3

Source: OECD, see Ahmad (2003).

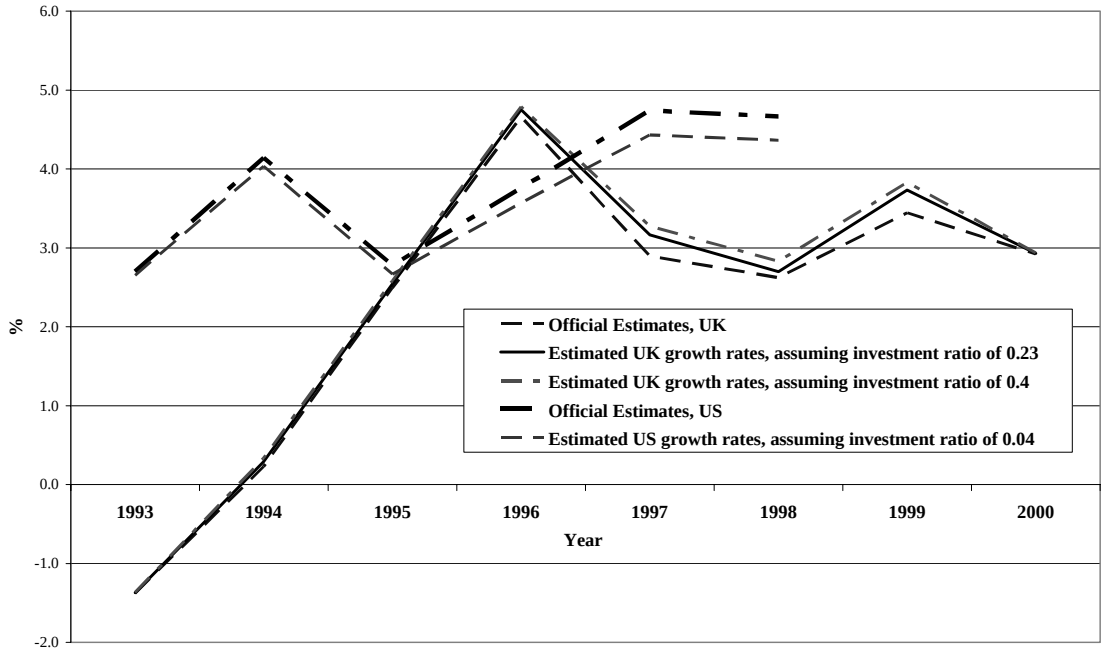
These revised estimates are, however, particularly sensitive to the quality of the employment measures used. In this case, these are data on the Industrial Standard Classification of Occupations 88, ISCO 213, computer professionals. There are believed to be significant international comparability problems with these data, and the "harmonised" revised estimates shown in Table 1 should thus be considered for illustration only.

15. For more information, see Ahmad (2003), *"Measuring Investment in Software"*, STI Working Paper 2003-6.

The impact on growth rates

The impacts of differences in software investment on GDP growth can be substantial. Figure 7 below shows the estimated impact on growth in the United States, assuming an investment ratio of 0.04 (which is the one currently used in the United Kingdom), and for the United Kingdom, assuming investment ratios of 0.23 and 0.4 respectively. In both cases, the OECD Task Force procedure for own-account production is applied, and a number of assumptions[16] are necessarily used. Table 2 shows the size of the changes in each individual year.

Figure 7. **Sensitivity of GDP growth rates to different investment ratios for purchased software**



Source: OECD, see Ahmad (2003).

Table 2. **Estimated changes to GDP growth rates (constant prices), compared to official estimates**
Percentage of GDP

	1993	1994	1995	1996	1997	1998	1999	2000
Change in UK GDP growth rate assuming an investment ratio of 0.23	0.0	0.1	0.1	0.1	0.3	0.1	0.3	0.0
Change in UK GDP growth rate assuming an investment ratio of 0.4	0.0	0.1	0.1	0.1	0.4	0.2	0.4	0.0
Change in US GDP growth rate assuming an investment ratio of 0.04	0.0	-0.1	-0.1	-0.2	-0.3	-0.3		

Source: OECD, see Ahmad (2003).

16. UK information based on ONS supply-use tables 1992-2000. Revised own-account estimates are projected from 1999 using growth in ISCO 213 figures. Purchased software estimates are deflated using the US price index for purchased software adjusted for changes in the US/UK exchange rate. US figures have been calculated using BEA statistics on customised and purchased software to estimate total supply of computer services in non-input-output years, assuming that the investment ratio for purchased software in 1997 is stable throughout the period of time shown above. Own-account estimates deflated using an average of the US price index for purchased software and a price index that rises by 5% per annum (to approximate wages and salaries, without any productivity assumptions/adjustments).

Although in some years GDP growth rates would change by over +/-0.25% of GDP, the trend of GDP growth remains largely unaffected. Moreover, the change in growth due to the different assumptions used is unlikely to be as large from 2000 onwards, since expenditure on computer services since 1999 is likely to have stabilised (indeed for the United Kingdom, estimates of GDP growth would be largely unaffected in 2000). This is because the rapid growth in computer services is unlikely to continue (at least at the same pace), and also because much of the growth in 1999 was likely due to exceptionally high expenditure on software to address the Y2K problem.

Equally, some of the estimated change to GDP growth rates reflects software investment by government. Changes of software expenditure from current or intermediate consumption to investment in these circumstances would overstate the possible sizes of the change to growth rates shown above, because of the way in which government output in current prices is calculated in the national accounts. In these circumstances GDP would only increase/decrease by the imputed capital consumption of the reclassified investment.

The impact on other countries is not assessed in this paper; however, similar results are likely to occur depending on the size of the investment ratio in each country and the harmonised approach used, so changes of between +/-0.25% of GDP would be expected. For example, the Netherlands (with a similar investment ratio and (harmonised) own-account and purchased software (as a percentage of GDP) would be expected to have similar reductions in growth rates to those exhibited by the United States if the UK investment ratio is applied (assuming that growth in software demand is similar in both the United States and the Netherlands). In the same way, for France, increases in GDP growth rates would be expected if an average investment ratio of 0.4 were applied.[17] In all cases the impact on growth rates is likely to be smaller post 1999/2000, assuming that the growth of software expenditure has stabilised at about the same rate as the economy generally. However, if the slowdown in computer expenditure since 1999 (Y2K expenditure) was precipitous, resulting in a negative contribution to overall GDP growth, the direction of change can be expected to reverse.

17. This is similar to the results shown in Lequiller 2001, "The New Economy and the Measurement of GDP Growth", Graph 8. Changes to French GDP growth rates are shown to increase from 0.1% in 1995 to 0.2% in 1998, by applying the US investment ratio to French software estimation. Equally US GDP growth rates are shown to fall by 0.05% and 0.2% respectively over the same period, using the French investment ratio.

3. Measuring real GDP

While the measurement of nominal GDP gives rise to some problems, measurement becomes more complicated when price and quality changes have to be accounted for. This section examines the estimation of real GDP over time. Three issues are key in this context, namely the adjustment for quality changes, the measurement of output in services and the choice of index numbers (see also Box 1).

Box 1. **Purchasing power parities in comparisons of labour productivity growth**

Purchasing Power Parities (PPPs) for GDP are spatial price indices that serve to compare levels of real GDP and its components internationally. As they are associated with *level* comparisons, there is normally no need to invoke PPPs for purposes of comparisons of the *growth* of GDP and productivity. An index of GDP growth of a particular country relative to another one is simply constructed by dividing the two indices of constant price national GDP. Such a comparison would thus be based on *constant national prices* because national deflators are implicitly used in the comparison.

However, in principle, it is also possible to construct an index of relative output growth by using a time series of PPPs, and applying them to one country's current-price GDP. The resulting GDP level, expressed in current international prices, can then be related to the GDP level of another country to form an index of relative GDP growth between two countries. Such a comparison would then be based on *current international prices*.

Conceptually, there is no reason that the two methods should yield the same results. In a world of perfect data, the first method would reflect national price weights and their shifts over time, whereas the second method would reflect international price weights and their shifts over time. The two methods are based on different weighting schemes and indeed, empirical differences can be sizable as recently shown by Callow (2003). One might argue that the comparison of the two methods should in itself be interesting because it reveals effects of different weighting schemes. This is true in a world of complete and high-quality statistics. Practically, however, PPPs are based on a smaller sample of prices and on less detailed weights than the national price indices. Even though for certain products, the PPP price relatives are based on more comparable products than the national series, it is generally the case that for purposes of comparing relative output and productivity *growth*, the comparison based on constant national prices is to be preferred. PPPs should be used when output and productivity *levels* are the object of comparison across countries.

3.1 Adjusting for quality change - the role of hedonic price indexes

A widely discussed issue[18] at the height of the 'new economy' debate was the international comparability of rates of economic growth, given that the United States and some European countries compute price indices for information and communication technology (ICT) products with very different statistical methodologies and results. Because a different price index translates into a different measure of volume growth, the question has been posed whether some, or all, of the measured growth differences between countries are a statistical illusion rather than reality.

The main challenge for statisticians is to accurately account for the quality changes in these high-technology goods, for example computers. The quality adjustments required are not necessarily calculated in the same way across countries. Consequently, between 1995 and 1999, the US price index of office accounting and photocopying equipment (which includes computers) dropped by more than 20% annually, compared with 13% in the United Kingdom and – at that time - a mere 7% in Germany.[19]

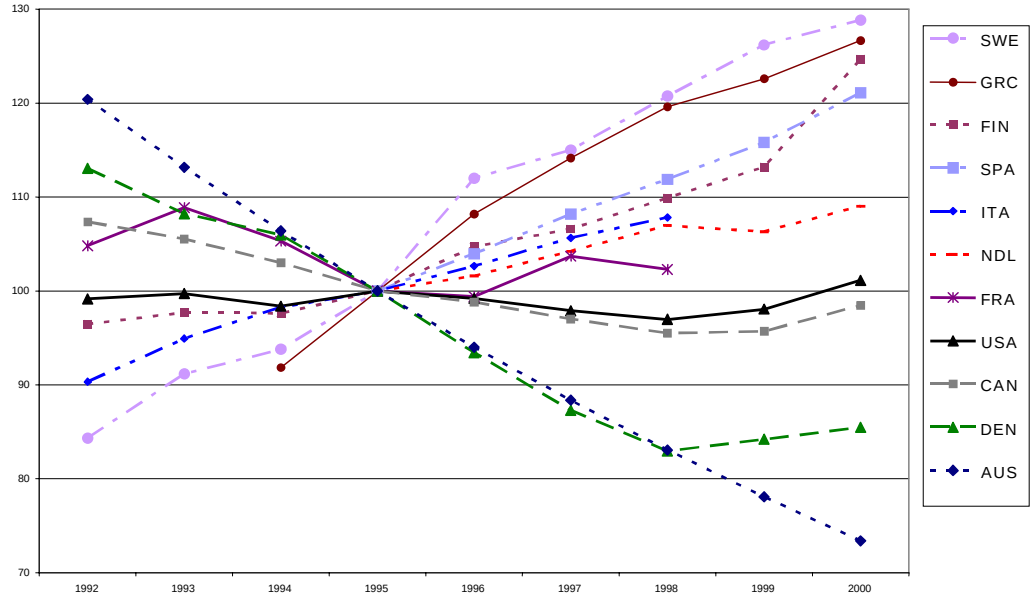
18. See "America's Hedonism Leaves Germany Cold"; *Financial Times*, 4 September 2000; "Apples and Oranges"; *Lehman Brothers Global Weekly Economic Monitor*, September 2000; *Monthly Report of the Bundesbank*, August 2000; "The New Economy Has Arrived in Germany – But No One Has Noticed Yet"; *Deutsche Bank Global Market Research*, 8 September 2000; Wadhwani, Sushil, "Monetary Challenges in a New Economy"; Speech delivered to the HSBC Global Investment Seminar, October 2000.

19. Germany has recently introduced hedonic methods for IT products. The first publication of these data took place in 2002.

Because computers are internationally traded, there are reasons to believe that their price changes should be similar between countries.

Another illustration of the difference in deflators is shown in Figure 8, which shows deflators for software investment. For example, the price index for Australia fell about 30% between 1995 and 2000, while the price index for Sweden rose by about the same amount. This largely reflects the dearth of price information available in this area, and the wide range of alternative proxy deflators that are used to represent price trends for software investment. These include price indices for general inflation, the price index for office machinery, the US price index for pre-recorded software, as well as methods based on inputs such as labour (see Ahmad, 2003).

Figure 8. **Investment in software, price indices in recent years, 1995 = 100**



Source: OECD, see Ahmad (2003).

Thus, at least part of the differences in measured price declines appear to be due to methodological differences. The natural question is: how would GDP growth in Germany, the United Kingdom, or any other country turn out if US methods were applied? Clearly, if the US price index is applied to Italy's or the United Kingdom's investment expenditure, their investment volume will show more rapid growth, as will the volume measures of the computer industry's output. Consequently, one would observe a different pattern of industry contributions to growth in GDP volumes and the role of ICT investment. This is important and policy-relevant information. But is it possible to make a clear-cut statement on how such a transposition of price indices carries over into the single most important measure of growth, economy-wide GDP? Unfortunately not. There are three reasons for this result, often overlooked in the debate.

A first and important distinction in the assessment of the effects of hedonic deflators is whether the product under consideration is used as an intermediate or a final product. Consider a typical intermediate product, say semiconductors, and suppose that they are exclusively sold to other domestic industries, *i.e.* there are no exports. Next, suppose that a statistical office adjusts downwards its deflator for semiconductors. The measured rate of growth of volume gross output of the semiconductor industry will rise, and so will measured real value-added of the semiconductor industry.[20] But at the same time, the measure of real

20. It should be noted that the rate by which the volume measure of value added changes is not the same as the one by which the volume measure of gross output changes.

intermediate input will also rise for other industries, namely the ones that buy semiconductors. The combined measure of real value-added of these industries will decline by the same amount that the semiconductor industry's real value added measure has increased. The economy-wide effect is zero - what has changed are the measured contributions to growth by particular industries.[21] The semiconductor industry will now feature a larger contribution than before whereas its downstream clients come out with a lower contribution to volume GDP growth than before. Such a shift in the measured contribution is, of course, an important change because it may change analysts' assessment of the sources of growth. But it also shows that one cannot readily jump to conclusions about the macro-economic effects of the choice of price indices.

If, on the other hand, a new deflator is used for a product that is mainly delivered to final demand, volume measures of aggregate final demand and GDP will be affected. This is, for example, the case for personal computers, which are rarely recorded as intermediate consumption. Suppose that computers are entirely final products, and suppose that their price index is changed from an annual decline of 5% to 15%. It would now seem straightforward to calculate the effect on measured volume GDP growth as the share of personal computers in total investment or private consumption times the 10 percentage point shift. Thus, if personal computers account for 2% of private consumption expenditure, the measurement effect on total consumption is 0.02x0.1=0.002 or 0.2% per year. This calculation is only valid, however, if all personal computers have been produced domestically, which is rather unrealistic for a large number of OECD countries. This gives rise to the second qualification regarding the impact of hedonic price indices, namely the role of imports.

Suppose that the volume measure of investment and private consumption rises as a consequence of adjusting the deflator for computers. If parts or all of these products are imported, one also has to adjust the price index for imports and the measured rate of volume imports will go up. Since imports enter the GDP calculation with a negative sign, this will partly or entirely offset the positive measurement effect from the other components. Thus, the use of a different deflator for certain products will almost certainly change the measured contributions of individual demand components to macro-economic growth. However, if the products under consideration are imported, these effects will be partly offsetting.

There is yet another possibility if imported products are used as intermediate inputs. Semiconductors, mentioned above, are a good example. Adjusting (downwards) the price index for imports, and consequently upwards the volume index for imports leads to a *fall* in the measured rate of volume GDP growth, that is not counter-balanced by an increase in measured volume growth of investment or private consumption.[22] In this case, the absence of hedonic deflators in a country's national accounts will imply an *overstatement* of real GDP growth (assuming that hedonic deflators represent the preferred measure). Note that this statement holds only if no other price index is changed at the same time. If the correction of the input price index leads also to a correction of the output price index, one faces again a situation of offsetting effects with no or very little impact on measured GDP growth.

To obtain an order of magnitude for the impact of price adjustments of ICT products on the rate of change of volume GDP, Schreyer (2002) and Lequiller (2001) have simulated some of the effects. Overall, the results point to only modest effects. Schreyer (2002) computed a 'multiplier' that translated shifts in ICT

21. This presupposes that the volume of value added is calculated using double-deflation, which is not always the case. Some countries estimate the volume of value added using single indicator methods, by assuming that output volumes move in line with value-added volumes. Where this occurs and taking the semiconductor example above, the volume of GDP will rise as a result of the change in deflator.

22. If imported semiconductors are added to inventories, rather than directly used as intermediate inputs, the rise in measured volume of inventories would offset the rise in measured volume of imports, and the effect on measured real GDP would again cancel out.

prices into effects on measured GDP growth. He found, for example, that if the price indices of the office machinery and computer industry in Germany had been adjusted downward by 10 percentage points per year, this would have resulted in a hypothetical upward adjustment of German GDP volume growth by 0.09 percentage points per year.

A more complete comparison between Germany and the United States has been undertaken by the Deutsche Bundesbank (2001). They combined impacts of differing methods of price measurement and measurement of software investment and concluded that "…if US methods and computation approaches are applied to Germany, the difference in growth between the two countries in the second half of the nineties is around 0.1 percentage point lower annually." About half of this growth difference is attributed to the effects of hedonic price measurement of IT goods. Lequiller (2001) examined the issue of a downward adjustment of computer and software price deflators and their effects on measured GDP growth in France and found a modest effect of 0.04 percentage points per year for the annual rate of growth between 1995 and 1998. Landefeld and Grimm (2000) reviewed the impact of hedonic price indices on aggregate volume growth in the United States. They found that only a small share of measured growth in the second half of the 1990s was associated with the use of hedonic price indices. Landefeld and Grimm estimated that the quality change in personal computers adds, at most, one quarter of a percentage point to the estimate of annual real GDP growth over the period 1995-99. While higher than the figure for France, this has to be put in proportion to a rate of real GDP growth of 4.15% per year in the United States.

To the extent that a change in price indices affects the volume growth of gross output and value-added by industry, it will also affect the rate of labour productivity growth. If measures of aggregate volume GDP growth change, so will measures of aggregate labour productivity and the growth rate of per-capita income. Given that the quantitative impact of hedonic measures on GDP volume changes is likely to be small, aggregate labour productivity growth measures will only be affected by the same small degree.

3.2 Measuring real output in services

The service sector now accounts for about 70 to 80% of aggregate production and employment of OECD economies and continues to grow. But measuring output and productivity growth in many services is not straightforward. What exactly does a lawyer or economist produce? How can the rapidly-changing pricing schemes of telecommunications providers be compared over time? And how should one measure the 'quantity' of health services provided by public hospitals? These and similar questions arise when statisticians attempt to measure the production of services and the output of service industries and the difficulty of this task is hard to overstate.

Generally, measuring volumes in the national accounts requires that current-price values of flows of goods and services can be divided into volume and price components. Typically, this is more difficult for services than for goods. Characteristics of goods can normally be identified and changes in quantities and qualities are measurable in principle, whereas for services even quantitative changes are often hard to measure, let alone quality change. Omitting qualitative changes does not necessarily mean that volume of services will be underestimated. For example, technical progress may improve the quality of medical services but may also lead to a decline in quality when, for example, self-service in post offices, longer queuing times or increased distances to the nearest post office all of which reflect an increased burden on customers and a lower quality of services.

This section discusses some of these issues against a background of international comparability. At the level of individual industries, two types of measurement issues arise: comparability of current-price measures, and comparability of the price indices used to deflate the latter or, in some cases, the comparability of volume indices. When it comes to the impact of service industry measures on the comparability of aggregate GDP series, two other factors come into play. The first is the extent to which a

particular service industry's output is bought as intermediate input by other domestic producers as opposed to the share of output that is delivered to final demand (private consumption, investment, government and export). The larger the share of the former, the smaller the impact of a measurement error at the industry level on measures of aggregate GDP growth, and vice versa: if a service industry delivers exclusively for final demand purposes, any possible measurement will carry over to measures of total GDP. The second factor is the size of the industry in question, or more precisely, its share in economy-wide value-added. The larger the share, the bigger the potential impact of an incorrect measure of industry-level output. The discussion below focuses on three key sectors, namely financial services, wholesale and retail trade, and health and social services. The question of aggregate impacts is discussed at the end of this section.

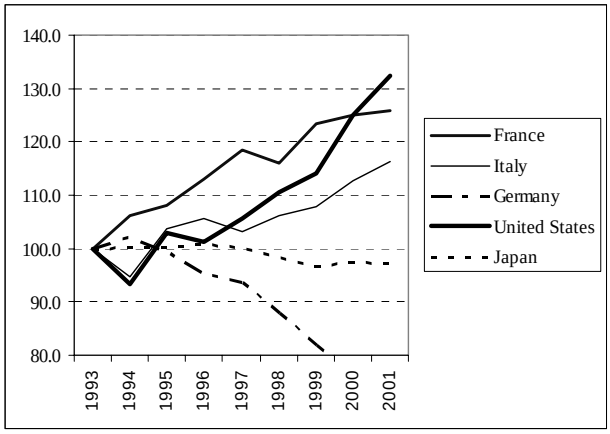
Financial services - banks

One of the differences between national accounts methodologies in OECD countries relates to the treatment of the financial services industry, in particular the non-insurance part of it that was described earlier. It was pointed out that, although the basic approach towards measuring the production of the financial services industry was similar across countries, there were differences in the degree to which financial services were considered intermediate purchases by other industries or final purchases by consumers (and clients abroad). Thus, for countries that treat financial services predominantly or exclusively as an input to business, the methodology of price-volume splits of current-price output will be of little consequence to macro-economic comparisons of growth rates. However, when a sizeable share of financial services is treated as final consumption, the choice of deflation methodology will matter for economy-wide measures of growth and productivity.

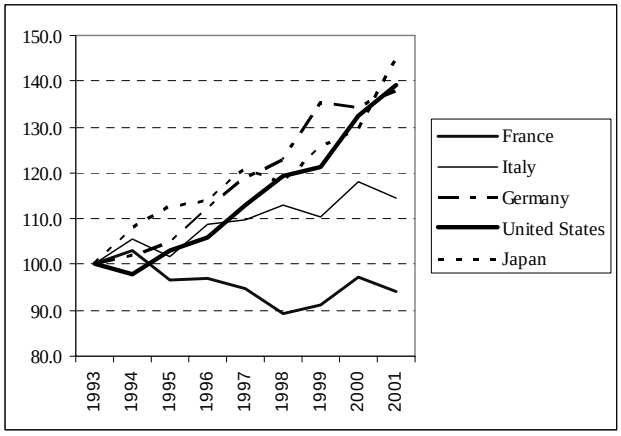
A look at the rates of labour productivity growth of the non-insurance financial services industry across countries reveals large differences that are not easily explained (Figure 9). The US Bureau of Economic Analysis recently (1999) changed its methodology to derive real measures of implicitly priced financial services. Formerly, employee hours worked served as the quantity indicator for the volume changes of implicitly-priced banking services (FISIM). In current practice, a quantity indicator of output developed by the US Bureau of Labor Statistics is used that traces volumes of banking transactions such as number of checks written, or number of transactions on automatic teller machines to better capture the growing volume of transactions. Its introduction led to an upward revision of measured output of the banking industry and to a modest impact on measured GDP: 0.04 percentage points per year over the period 1981-96 (Moulton 2000). In many other countries, the Netherlands being an exception, no such output-based volume indicator exists and the value of financial services in volume terms is derived by applying base-period interest margins to the inflation-adjusted stock of assets and liabilities. The latter method takes no account of quality changes and it is hard to assess whether it tracks the volume of transactions. Measured productivity growth in the banking industry for the United States would then be higher than for other countries if the volume indicator for banking transactions grows faster than the deflated stocks of assets and liabilities that drive volume growth in other countries and if there are no important differences in the deflation methods for directly-priced financial services. However, the data do not seem to bear this out. Measured volumes of value-added productivity in the United States are just about average among the small sample of countries reviewed, and if anything, questions may be raised about the comparability of the data on financial services output within the European Union.

Figure 9. **Financial intermediation industry**
(1993=100)

Implicit deflator of value added



Real value-added per person employed



Source: OECD.

Wholesale and retail trade

Wholesale and retail trade are one of the largest service industries, and on that account alone, their measurement deserves attention. Like banking, they are 'margin' industries for national accounts purposes, meaning that their output is computed as the difference between products bought and products sold rather than as the value of turnover. But their treatment differs from other branches as there is no explicit consumption of retail services in the national accounts. This consumption is included in the values of the goods and services that are sold through retail trade. As such, the volume estimate of retail trade services will not affect the volume growth of GDP, if the latter is obtained from the expenditure side of the national accounts, as is, for example the case in the United States. In such a case, any observed differences in the volume growth of retail services bear on the relative performance of industries, but not necessarily on aggregate productivity measures. This should be kept in mind in the following discussion. To obtain volume measures of output of the distributive industries, margins have to be deflated by an appropriate price index, or alternatively, volume measures of output have to be estimated directly. Such volume and price measures should adequately reflect the services provided by this industry, *i.e.* the provision of goods at a particular location and time, storage and convenience of choice. In statistical practice (*e.g.* Eurostat 2001), it is normally assumed that these services are in proportion to the volume of sales and the latter serve as the real output measure. Or, as in the case of the United State, a sales price index is the key ingredient to deflate retail margins (Lum *et al.* 2001).

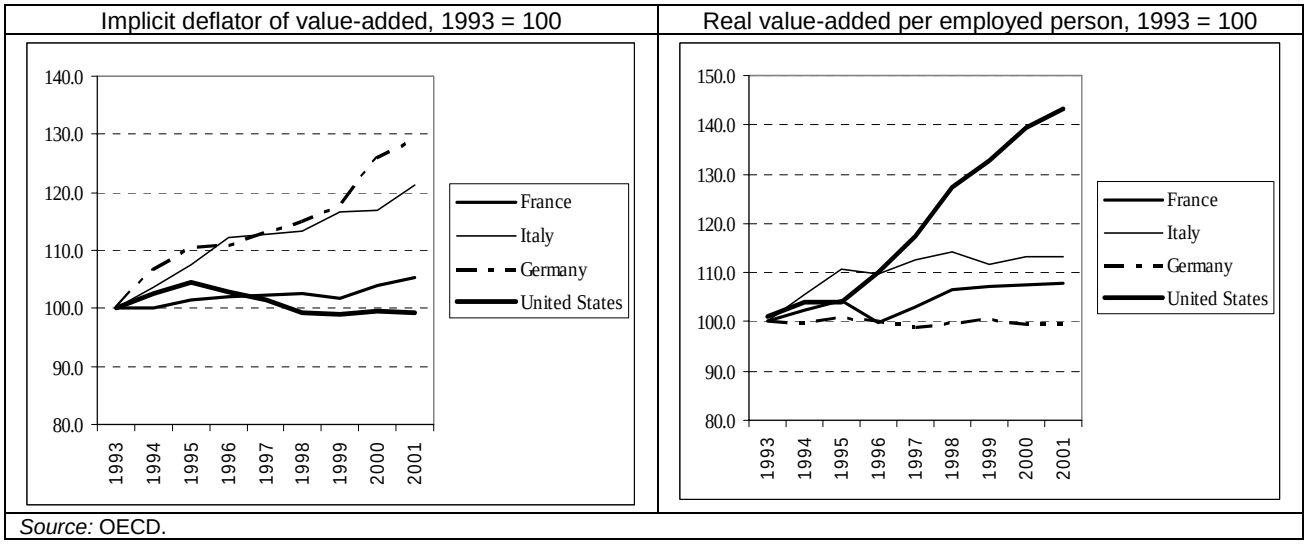
There are two possible weaknesses of such a treatment: *i)* it ignores all changes in the quality of distribution services provided. For example, an extension of opening hours that adds to shoppers' convenience but does not necessarily increase the volume of sales would go unnoticed in output and productivity statistics; *ii)* quality changes in volume measures of sales would affect volume measures of distribution services. If the volume measures of sales (or the sales prices) that serve as proxies for volume measures of distribution services change because of a change in the quality of the goods sold, such a volume movement would be translated into a volume change of distribution services.[23] However, there is little reason to believe that distribution services would have gone up or declined simply because a particular good distributed by the industry is of better or lower quality than a similar item in the preceding

23. This implies that different methodologies between countries to get at the price-volume split of the goods transacted by retailers and wholesalers would lead to a different measure of the volume output of the distribution industry.

period. Thus, if the movements of the United States sales price index that is used to deflate retail margins reflect a good deal of quality adjustment (*e.g.* as a consequence of hedonic techniques used to establish the price changes of some of its components) then the volume of retail services could be mis-measured compared to the volume measure of retail services in other countries.

An examination of real value-added and productivity series in the distribution industry across some OECD countries shows indeed that developments are diverse with much more rapid productivity and value-added growth in the United States than in other countries (Figure 10). It is difficult to say whether such differences are due to differences in statistical methodology. If they reflect methodological differences, these must stem from differences in implicit deflators, given similar methodologies in measuring current-price output. At the same time, the strong performance of the US distribution industry in the 1990s is a well-documented phenomenon, and not only on the basis of national accounts data. In summary then, due to the particular treatment of distribution services in the national accounts, the impact on the comparative growth rates of GDP is limited or nil, but whether the industry-level productivity measure provides for an accurate international comparison of the sector has to remain open.

Figure 10. **Prices and productivity in the wholesale and retail industry**



Source: OECD.

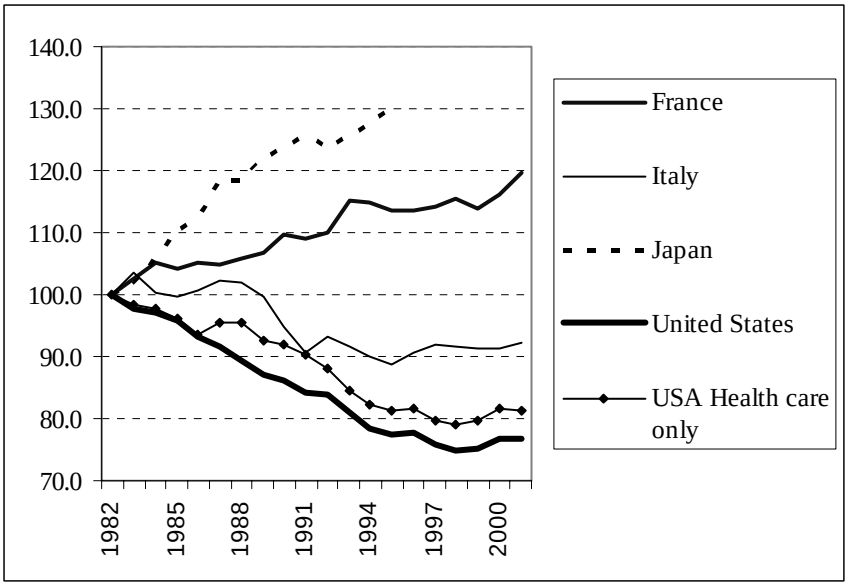
Health and social services

These services account for a large and growing part of the service industry. In many countries, significant parts of this sector operate on a non-market basis, implying that observed prices and fees do not cover costs to any significant extent. Currently, a vast majority of OECD countries measure volumes of health services as the sum of deflated costs. However, such input-based methods insufficiently grasp the quantity and quality of output, thereby mis-measuring – typically understating productivity growth. This point has repeatedly been made for the United States whose productivity index for health and social services declined over the past 20 years (Figure 11). Bosworth and Triplett (2002) indicate that medical care productivity in the United States was understated by a large amount before 1992, the period when a new industry price index replaced the old CPI-based medical care deflators in national accounts. The PPI indexes rose more slowly than the CPI indexes they replaced and led to an uptake in measured productivity after 1995.

Overall, one is left with a picture where measured productivity in the health and social services has been in secular decline in the United States and Japan, slightly increasing in France and steadily increasing in Italy. There is no reason to believe that adding further countries would render the overall picture more

homogenous. Given the nature of the industry and given the difficulty in obtaining appropriate price and volume measures, it would seem that at least some of the cross-country differences in productivity growth are due to differences in measurement. As health care is a service with a large final demand component, any differences in comparability at the industry level carry over to the aggregate level. However, it is very difficult to put a number to the size of these effects. A number of National Statistical Offices have begun to move away from input-based measures for health and social services, and, indeed, government (non-market) output more generally, introducing output based (quantity) indicators where possible, to capture volume change. The United Kingdom for instance now measures over half of government output using output based indicators.

Figure 11. **Real value-added per employed person in the health and social work industry, 1993=100**



Source: OECD.

A "zero-productivity" simulation

Given the difficulties in assessing possible measurement biases empirically, several studies (Slifman and Corrado 1996, Vijselaar 2003, Sharpe *et al.* 2002) proceeded as follows: noting that certain service industries were characterised by prolonged periods of negative measures of productivity growth, they concluded that poor measurement may provide an explanation. Gullickson and Harper (1999, 2002) and Slifman and Corrado (1996) then examined what would happen if productivity growth in these services had not been negative but zero and simulated the impacts on measured overall labour productivity growth.

A similar simulation was recently undertaken by Wölfl (2003), in line with Sharpe *et al.* (2002) for the United States and Canada, and Vijselaar (2003) for the Euro area. Such a thought experiment is primarily intended to show the potential size of the problem and neither suggests that negative productivity growth necessarily implies mis-measurement, nor that the size of the adjustment made is the correct one. But it provides a diagnostic tool to examine key areas for measurement problems. The main merit of the analysis lies in its attempt to trace effects of mis-measurement in one particular industry to the aggregate level. Whether or not measurement biases on the industry-level carry through to total GDP depends on the extent to which an industry delivers to final demand. The analysis thus uses information from input-output to provide weights for such deliveries.

Wölfl's (2003) empirical results suggest relatively strong indirect effects on other industries' productivity measures from the adjustment of output in services such as financial intermediation or renting of

machinery and equipment. Indirect effects tended to be rather small when the adjustment of output measures was in hotels and restaurants, or other social and personal services with a large final demand delivery. For instance, in the case of France and Germany, almost 60% of gross output of financial and business services flows to other industries, while the respective share is only about 20% in the case of trade, hotels and restaurants, and 5% in community, social and personal services.

Overall, the analysis suggests that the principal impact of possible mis-measurement is a shift in the attribution of productivity growth to specific industries, whereas the impact on aggregate productivity growth tended to be moderate if France, the United States and Germany – the countries for which a simulation was carried out – are examples to go by. For example, if rates of output change of all those service industries that exhibit negative rates of labour productivity growth are adjusted upwards, this leads to a 0.08 percentage point change in measured GDP growth for the United States over the period 1990-2000 and 0.19 percentage points for France. While these effects would directly translate into measured labour productivity growth, they are comparatively small.

3.3 The choice of index numbers

GDP is a single measure of economic activity, drawn up as a combination of the prices and quantities of those goods and services that enter the realm of national accounts. For any given year, total GDP at current prices is a simple sum of its components. But for the comparison of GDP growth, this data has to be compared over time and the specific way in which quantities and prices are aggregated and compared with other periods is synonymous with the choice of an index number formula. Many such formulae exist and have been examined in the context of a two-period case.

But when more than two periods are involved, the question arises which period should be taken as a basis. Options include: choosing the first or the last observation as a base (direct comparison), or using the chain principle. The chain principle makes use of the natural order provided by the march of time: for every index for period t, period t-1 provides the base. In a national accounts context,[24] three types of index numbers are encountered in practice.

Fixed-base Laspeyres volume indices

These indices use a base year which typically changes every five years. If the base year is called t_0, and if P_i^t and Q_i^t are individual prices and quantities in year t, the Laspeyres volume index is given by $Q_L^{t/t_0} = \frac{\sum_i P_i^{t0} Q_i^t}{\sum_i P_i^{t0} Q_i^{t0}}$. At present, t_0 corresponds to 1995 in many OECD countries and we could also say that $\sum_i P_i^{95} Q_i^t$ is the GDP or one of its components of year t at 1995 constant prices. The volume change between two years t and t-1 is then given by:

$$\frac{Q_L^{t/t0}}{Q_L^{t-1/t0}} = \sum_i \left(\frac{P_i^{t0} Q_i^{t-1}}{\sum_i P_i^{t0} Q_i^{t-1}} \right) \frac{Q_i^t}{Q_i^{t-1}}$$

24. See Whelan (2000) for a very instructive discussion of chained index number formulae in the national accounts.

Thus, it is a weighted average of the volume changes of its components with weights that are quantities of the period t-1 valued at prices of the base year t_0. "Fixed base" Laspeyres volume indices are currently in use in many European countries (*e.g.* Germany, Finland, Sweden, Spain and Switzerland). A complication arises when it comes to the expenditure components of GDP: private consumption, investment and government expenditure are added up but imports are subtracted to obtain a value for GDP. In terms of growth rates, this means that one component of GDP, imports, enters with a *negative* weight; this will typically complicate assessments about overall possible biases.

Chained Laspeyres volume indices

These use an annually-changing base year. Here, the base year moves continuously and corresponds to the year t-1. Other than that, the Laspeyres formula remains, *i.e.* between any two adjacent years it is always the price vector of the first year that serves as the reference. Thus, volume GDP is expressed in previous years' prices, and the volume index of GDP becomes: $Q_L^{t/t-1} = \frac{\sum_i P_i^{t-1} Q_i^t}{\sum_i P_i^{t-1} Q_i^{t-1}} = \sum_i \left(\frac{P_i^{t-1} Q_i^{t-1}}{P_i^{t-1} Q_i^{t-1}} \right) \frac{Q_i^t}{Q_i^{t-1}}$.

Thus, the year-to-year growth of volume GDP is a weighted average of the volume changes of its components with weights that are quantities of the period t-1 valued at prices of the same period t-1. Annually chained Laspeyres indices are the formula recommended by Eurostat for its Member countries. About half of the EU countries have changed their national accounts to a chained, annually re-based Laspeyres method, and the remainder will change by 2005. The Australian and New Zealand accounts are also based on this index number formula.

Chained Fisher volume indices

Like in the chained Laspeyres case, the base year changes over time in this case. But rather than limiting the price weights to be of period t-1 as in the Laspeyres formula, the Fisher volume index allows for additional price weights to be taken from the period t. This is done by computing a Paasche index $Q_P^{t/t-1} = \frac{\sum_i P_i^t Q_i^t}{\sum_i P_i^t Q_i^{t-1}}$ and forming the Fisher index as the geometric average of the Laspeyres and the Paasche index: $Q_F^{t/t-1} = \left(Q_L^{t/t-1} Q_P^{t/t-1} \right)^{1/2}$. Thus, the Fisher index formula neither gives preference to the price weights of the period t-1 nor to those of period t – it treats them symmetrically.[25] Annually weighted Fisher indices are presently used in the national accounts of the United States and Canada.

International comparability

With these definitions in mind, the question arises whether the choice of index number formula matters for international comparability. No specific effort is made at this point to discuss the various advantages and drawbacks of these and other index number formulae (see Box 2).[26] However, some of the arguments arising in such discussions are linked to comparability and in this sense they will be covered.

25. Note that the Fisher index, unlike its components, the Laspeyres and Paasche indices, cannot be presented as a weighted average of its components. This is sometimes considered a disadvantage. For a more general evaluation that brings forth many of the advantages of the Fisher index number, see Diewert (1992).

26. The System of National Accounts (1993) offers a useful discussion of the various index numbers. For the relative merits of fixed and chain-weighted index number formulae, see also Szulc (1983) and Hill (1988).

Box 2. A brief comparison of index numbers formulae

Fixed base and chained Laspeyres indices: For a comparison of these two formulae, it is instructive to take another look at the weights that enter the "fixed-base" Laspeyres formula introduced earlier in a comparison between two adjacent years t and t-1: these weights take a product's quantity in period t-1 and attach a price weight of period t0: $P_i^{t0}Q_i^{t-1}\Big/\sum_i P_i^{t0}Q_i^{t-1}$. Now assume that the price of product i has fallen between t0 and t-1 and all other prices remain unchanged: $P_i^{t0} > P_i^{t-1}$. Compared to a chained Laspeyres index where quantities of period t-1 are valued at prices of the same period t-1, the "fixed base" case values the quantities of period t-1 at prices of the base year. Given the case at hand, this means that the "fixed base" weights are higher than the weights in the case of a chained Laspeyres index. If the relative price decline continues and the consumers increase the volume of quantities that they buy ($Q_i^t\Big/Q_i^{t-1} > 1$), total GDP growth between t-1 and t will be measured at a higher rate in the "fixed base" case than under the chained index.

This 'substitution bias' will be large[27] if the base period is relatively remote from the reference periods t and t-1 and if the relative price decline has been significant and continuous. Note, however, that the sign of the substitution bias depends on the direction in which the quantities of product i evolve: in a market dominated by demand, a (relative) price decline will tend to lead to an increase in quantities sold, and the "fixed base" case will tend to over-estimate GDP growth between t-1 and t. In a market dominated by producers, a (relative) price decline will tend to reduce the supply of a commodity and thus a decrease in volumes. Consequently, the "fixed base" case would under-estimate GDP growth between t-1 and t. The sign of the "substitution bias" is therefore, to a considerable extent, an empirical matter. *A priori* statements on its sign are only possible under assumptions about consumer or producer behaviour and about the presence or absence of shocks to supply technology or consumer preferences.

Chained Laspeyres and chained Fisher index: How do these two chained index numbers compare? Again it is difficult to make a general statement on purely theoretical grounds. However, if one assumes that markets reflect consumer behaviour and that further there are no shifts in consumers' preferences over time, Samuelson's (1947) and Allen's (1949) theorems apply[28] and it can be shown that the Laspeyres quantity index will exceed the Paasche quantity index: $Q_L^{t/t-1} \geq Q_P^{t/t-1}$. As the Fisher index is the geometric average of the two, it will also be true that the Laspeyres quantity index exceeds the Fisher index.

Bringing the three types of indices together, theory indicates that under typical demand-side conditions (falling (rising) relative prices are associated with rising (falling) quantities) and in the absence of significant shocks to consumer preferences, the following relationship holds:

$$\left(Q_L^{t/t0}\Big/Q_L^{t-1/t0}\right) \geq Q_L^{t/t-1} \geq Q_F^{t/t-1}.$$

The question about international comparability can be formulated as follows: given the same set of prices and quantities for GDP components in two countries, does it matter which index number formula is chosen to compute volume GDP growth? As is often the case, there is no general answer. Rather, it depends on three factors. The first is the degree to which relative prices change over the period considered. This will determine the presence or absence of a substitution bias in conjunction with certain index number formulae (see Box 2). The second factor is the reaction of consumers and producers to such changes in relative prices. If both relative price changes are large and the elasticities of demand and/or supply are high, the choice of the index number formula will matter. In other cases, this may be less of an issue. The third factor concerns the relative role of imports: if imports are large compared to total demand, and if import

27. The "substitution bias" noted here is not one that has been defined in a theoretically 'clean' way – a more rigorously defined bias would be measured as the difference between any particular index number and a theoretically correct (but possibly unobservable) quantity index.

28. Diewert (1973) examines the issue more broadly and derives the conditions under which empirical observations are consistent with assumptions about consumer behaviour. In a recent unpublished note, Diewert extends the same idea to the supply side and determines the conditions under which observed price and quantity vectors are consistent with optimising producer behaviour.

prices and volumes behave in the same way as prices and volumes of private consumption, investment and government consumption, then a bias at the level of individual GDP components may be offset among them and have little effect on measured overall GDP.

Empirical evidence

Schreyer (2001) used detailed final expenditure statistics[29] to assess the effects of choosing different index numbers in the presence of significant relative price changes. The latter arose because hedonically-adjusted deflators for ICT products that were used in the study showed rapid declines compared to other goods and services. Table 3 below reports index number effects for five countries. Results were obtained by: *i)* replacing 'traditional' deflators of ICT products by hedonic deflators – this improves comparability and leads to significant shifts in relative prices; *ii)* re-computing the volume change of final demand components with different index numbers. The latter permitted the calculation of index number effects. More specifically, Table 3 shows the impact of the choice of a Fisher quantity index over a Laspeyres quantity index. Because the study used a direct comparison between two non-adjacent years, the effects obtained correspond to a difference between $Q_L^{t/s}$ and $Q_F^{t/s}$. Other empirical studies yield similar results. For example, investigations by the Australian Bureau of Statistics and by the Dutch Central Bureau of Statistics (CBS 1997) show that the effects of replacing a fixed-weight Laspeyres index number formula by a chained Fisher index number formula are fairly modest at the level of total GDP, but significant for individual components (Table 4). Overall, the following conclusions are reached, based on observations covering roughly the 1985-95 period and five OECD countries:

- The base-year weighted Laspeyres index tends to overstate volume growth in subsequent years. The alternative Fisher index whose weights reflect both prices of the current and the comparison period tends to produce lower rates of real growth. This difference in the simulation results between the two index number formulae marks the index number effect shown in the following table.
- Index number effects tend to be more pronounced for individual final expenditure categories than for total GDP. This is due to compensating effects from imports that enter overall GDP computations with a negative sign. The extent of this compensating effect depends also on the weight that imports occupy in total GDP.

With EU countries gradually implementing chained Laspeyres indices and several non-European countries using chained Fisher indices, the question of comparability will be reduced to the differences between chained Fisher and chained Laspeyres indices. The SNA 93, although it recommends the use of chained Fisher indices, considers the chained Laspeyres formula a satisfactory substitute. Generally, these differences are expected to be smaller than between fixed-based and chained indices. As a consequence, these differences in index number formulae should have only minor effects on the comparability of aggregate national accounts data.

29. The effects of the choice of index number formulae in the presence of relative price changes depend also on the level of aggregation at which an index is constructed. Generally, a high level of detail is preferable to a low level of detail.

Table 3. **Effects of using different index number formulae in the presence of rapid relative price changes**
Volume growth, average percentage change at annual rate

Based on:	Private consumption expenditure	Government expenditure	Investment	Exports	Imports	Total final expenditure
			Canada, 1986-92			
Fixed-weight Laspeyres index[1]	2.21	2.85	3.49	5.66	6.57	2.21
Fisher index[2]	2.13	2.73	3.05	5.22	5.95	2.15
Index number effect[3]	-0.08	-0.12	-0.44	-0.45	-0.62	-0.06
			France, 1985-96			
Fixed-weight Laspeyres index[1]	2.25	2.04	2.44	4.86	4.95	2.22
Fisher index[2]	2.18	2.03	2.21	4.71	4.99	2.13
Index number effect[3]	-0.07	-0.01	-0.23	-0.15	0.03	-0.08
			Japan, 1985-94			
Fixed-weight Laspeyres index[1]	3.40	2.46	6.05	4.43	7.66	3.67
Fisher index[2]	3.27	3.59	5.16	2.84	7.25	3.41
Index number effect[3]	-0.13	1.13	-0.88	-1.60	-0.41	-0.26
			Netherlands, 1986-93			
Fixed-weight Laspeyres index[1]	3.18	1.98	2.93	5.87	5.49	3.42
Fisher index[2]	2.88	1.97	2.28	5.46	4.84	3.11
Index number effect[3]	-0.31	-0.01	-0.65	-0.41	-0.65	-0.31
			United States, 1987-93			
Fixed-weight Laspeyres index[1]	2.15	1.03	2.38	8.46	4.55	2.28
Fisher index[2]	2.08	0.99	1.73	8.01	3.99	2.13
Index number effect[3]	-0.08	-0.04	-0.65	-0.45	-0.56	-0.15

1. Volume growth rate of final expenditure component, based on aggregation across products with a Laspeyres-type index with fixed weights from the initial year. The underlying data were either directly available from national input-output tables, or specifically calculated, *e.g.* for the Netherlands whose input-output tables are based on quantity indices with annually-adjusted weights. Hedonically-adjusted price indices were used for ICT products.
2. Volume growth rate of final expenditure component, based on aggregation across products with a Fisher-type index that averages weights from the initial and the final year. As in the previous case, full quality adjustment was simulated for all relevant ICT products.
3. A positive sign indicates that volume growth of the expenditure component has been understated, a negative sign indicates overstatement.
Source: Schreyer (2001).

Table 4. **Volume changes in GDP and its components: difference between results from fixed-weight Laspeyres index* and chained Fisher index**

Percentage points, average annual rates of change

	Australia				Netherlands			
	GDP	Private final consumption	GFCF	Imports	GDP	Private final consumption	GFCF	Imports
1985	-0.1	0.0	-1.6	-0.9	..	..	..	..
1986	-0.3	-0.1	-1.9	-1.0	..	..	..	..
1987	-0.2	-0.1	-0.7	-0.2	0.0	0.0	0.0	-0.2
1988	0.1	-0.1	-0.2	-0.4	-0.8	0.1	-0.1	-0.2
1989	0.1	0.0	0.1	0.0	-0.2	0.0	-0.1	0.1
1990	-0.1	0.0	0.2	0.1	-0.2	0.0	-0.1	-0.2
1991	0.1	0.1	0.8	0.4	-0.1	-0.2	0.0	-0.2
1992	-0.1	0.0	0.5	0.9	0.0	0.0	-0.2	-0.2
1993	0.1	0.2	2.2	1.3	-0.5	-0.1	-0.3	-0.7
1994	-0.3	0.0	2.0	1.5	..	..	..	..

*Price base for Laspeyres index: 1989 for Australia, 1986 for the Netherlands.
Source: Australian Bureau of Statistics (1997); Central Bureau of Statistics (1997).

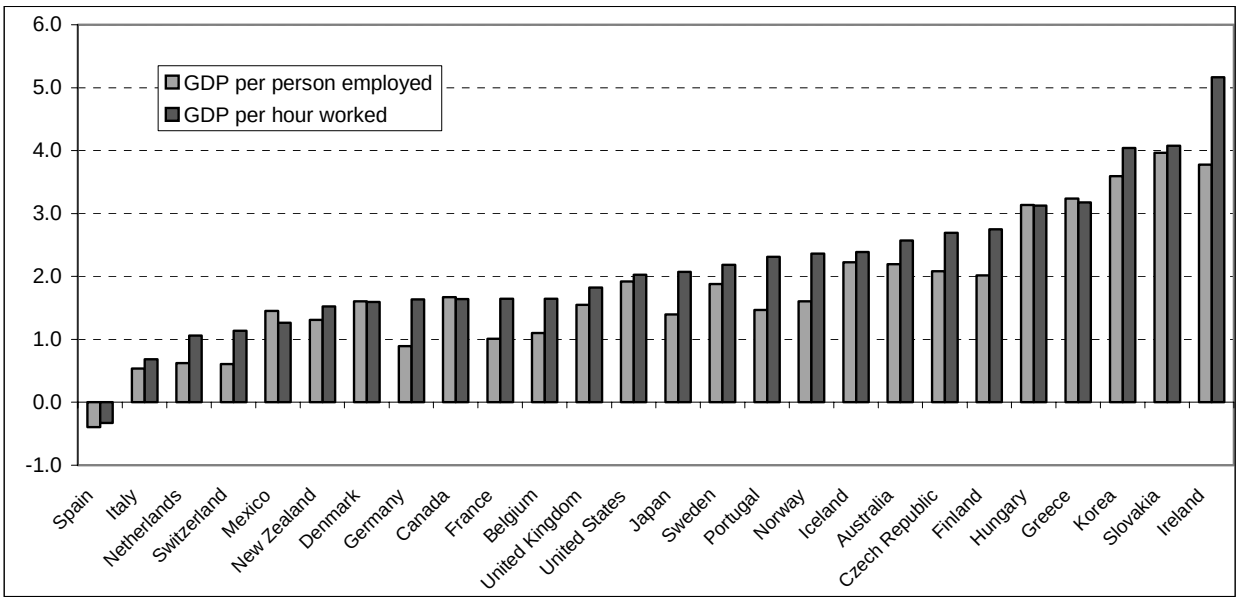
4. Measuring labour input and labour productivity

4.1 What measure of labour productivity growth?

Once measures of real GDP and labour input have been derived, measures of labour productivity growth can simply be derived by relating a quantity index of output (GDP) to a quantity index of labour input. The index for labour input can refer to total employment, total hours worked, or some quality-adjusted measure of labour input. These measures are likely to lead to substantially different estimates of labour productivity growth (Scarpetta *et al.*, 2000; OECD, 2003*a*). In principle, each of these possible measures is a valid measure of labour productivity; the choice of a particular measure depends on the purpose of the analysis and, in many cases, on data availability.

However, while all these alternative measures are valid, some measures provide a more accurate assessment of labour productivity growth than others. For example, adjusting for hours worked is particularly important in a cross-country comparison of productivity. Figure 12 illustrates the impact of adjustment for hours worked on the basis of OECD data. The graph illustrates that in most OECD countries, productivity growth over the 1995-2002 period was much more rapid when adjusted for hours worked. This is mainly due to the decline in statutory hours and an increase in part-time employment in many OECD countries. Ignoring the change in hours worked will lead to substantially different estimates of labour productivity growth and related international comparisons.

Figure 12. **Labour productivity growth in OECD countries, 1995-2002**



Source: OECD, calculations based on the OECD productivity database.

Adjusting these estimates for differences in the composition of labour is more difficult, as the required data are often not available in a standardised way across OECD countries. In particular, adjustments in some countries can rely on much greater detail than in others. OECD adjustments for labour composition have primarily relied on educational achievement, which is only a partial adjustment for composition (OECD, 2003*a*). Other studies, *e.g.* Jorgenson and Yip (2002) have used more detailed adjustments. Some work is currently underway to eventually apply such more detailed adjustments to a wider range of OECD countries, depending on data availability in individual countries.

This section addresses the main measurement problems in deriving estimates of labour input for OECD countries in more detail. It looks in turn at the measurement of hours worked, employment and labour quality.

4.2 Measuring hours worked

Comparing labour productivity, in terms of levels and growth, requires that the units of labour input used for the calculation are compatible and consistent with output measures, that is, GDP measures. Labour input measures should, therefore, be reported according to the SNA production boundary, which covers any work for pay and profit – whether in paid employment or self-employment jobs – for resident producers, that is, in resident units of production. Depending on the analysis, the units of labour input can either be the number of employment head counts — number of jobs, number of workers, number of full-time equivalent jobs or employment - or the total number of hours worked per year.[30] The latter is the preferred measure for productivity analysis as it reflects both work intensity – which is important due to the emergence of part-time work and declines in the average full-time workweek in many countries – and the volume of work engaged per year in the production of goods and services by producers within the geographic boundaries of a country.

Until recently, only estimates of hours actually worked per person employed on average per year[31] were available for a number of countries, as these were regarded as a convenient way to publish figures on hours-per-worker, but also because they separated out the effect of changes in the level of employment. But the pervasive use of these estimates in the calculation of labour productivity measures demands estimates of total hours of work and, therefore, employment estimates that are compatible with hours-per-worker figures. This section addresses the issue of comparability and data consistency over time for hours worked and employment series that are reported by national statistical offices, and their impact on the measurement of labour productivity growth.

As noted above, total hours of work estimates used with output measures should be consistent with SNA production boundaries, which require full coverage of jobs – employee and self-employment jobs – all industries, all establishments and firms irrespective of their sizes, non resident workers in resident units of production minus resident workers in non-resident units of production, armed forces whether or not stationed abroad, in sum all persons engaged in economic activities located within the geographic borders of a country, with the exception of household production workers.

However, data sources traditionally used to report hours of work do not fully comply with the coverage requirements of the SNA. Administrative sources, like social security registers, are limited to persons covered by the relevant regulations. Besides, the scope of employer or enterprise surveys is often limited to non-farm employees, excluding managers and supervisors, public administration employees, employees in personal services, small firms, conscripts and regular armed forces. The remaining activities and occupations are covered using administrative sources, industry censuses or labour force survey results, in particular in the case of self-employed and farm workers. On the other hand, labour force surveys report activity status of resident persons and have a broader coverage of industries, occupations and establishments. However, labour force surveys often make certain exclusions based on age and geography, and have partial or no coverage of collective households and of armed forces. The size of military personal is often obtained separately from administrative sources. It is, therefore, expected that, in general, labour

30. Total hours worked: Total hours worked is the aggregate number of hours actually worked during the period in employee and self-employment jobs.

31. Average hours actually worked per person in employment per year can be interpreted as the average length of a work year of full-year workers.

force survey coverage will differ only marginally from the coverage requirements of the SNA, except in countries with a sizeable amount of cross-border workers.

In practice, countries that produce their own national estimates of annual working time (*e.g.* eight countries out of the 24 OECD countries that provide data on hours worked to the OECD) attempt to come close to National Accounts coverage. To achieve this goal, they use the best available data sources for different categories of workers, industries and components of variation from usual or normal working time (*e.g.* public holidays, annual leave, overtime, absences from work due to illness and to maternity, *etc.*). For example, in many countries actual hours are derived from establishment surveys for production and non supervisory workers in employee jobs and from labour force surveys for self-employed, managers and supervisory workers, farm workers and public sector employment. Hours lost due to sickness are estimated from the number of days not worked from social security registers and/or health surveys. Vacation time is derived from establishment data on paid leave or from the number of days of statutory leave entitlements (assuming all vacation time is taken by workers).

There are at least four reasons that enable the use of labour force survey results to produce estimates of annual hours of work, at least at an aggregate level and for countries not already producing annual hours of work series, and providing that the estimates hold up to scrutiny.

- First, labour force surveys collect data on weekly actual hours worked during the survey reference week. In many cases, they also collect usual or normal hours worked, that are, although subject to reporting bias, the unique source for unpaid overtime hours.

- Second, the coverage of labour force surveys tends to be both broad and relatively close to that of the SNA production boundary.

- Third, labour force surveys are conducted in all OECD countries, so that it is potentially possible to generate estimates of annual hours worked for each OECD country.

- Fourth, in many European countries, labour force surveys are carried out on a continuous basis covering all 52 weeks of the year, which allows direct estimation of annual hours actually worked. Indeed, a methodology for producing estimates of annual working time was developed by Eurostat in the mid-1990s, based on Spring results of the Community Labour Force Survey (CLFS). Data on (statutory) vacation and holidays for this survey were obtained from external sources.[32] The methodology was subsequently extended at the OECD, to take into account second jobs and persons absent from work during the reference week. This method currently serves as the basis for a number of EU countries which do not produce estimates of annual working time and was also used in the OECD growth project (Scarpetta, *et al.*, 2000).

Data quality

Estimates of hours worked also depend on the quality of the underlying data. This issue is difficult to address. It is easier if one considers, instead of the estimates of labour volume itself, estimates of average hours of work per person employed. For this statistic, small variations from the theoretical SNA

32. One might expect some seasonal effects since the CLFS is conducted in the spring and absences from work due to illness, weather, *etc.* may vary seasonally. However, simulations using quarterly national data in a number of countries having quarterly surveys revealed that the results were remarkably robust. The use of external figures for vacation and holidays essentially amounted to assuming that persons took all of their allotted vacation and that any work on public holidays was compensated for by absences at other times, both of which are undoubtedly false (but unavoidable) assumptions under the circumstances.

employment coverage will have little impact. The coverage and quality of an associated employment figure that would be used to generate an estimate of labour volume can then be addressed separately.

In a number of countries there also exists a well-established National Accounts series of annual hours worked, generally based in part on establishment survey data, which can be confronted with a series generated from the national labour force survey. Such data confrontations generally show that the labour force survey estimate of annual hours worked per person employed (total economy) exceeds the NA series, sometimes by as much as 10-15%. The question that immediately arises concerns the origin of this discrepancy. It is easy enough to dismiss household survey data as being subject to reporting error and the vagaries of proxy response (which they generally are), but this is not helpful. Although the labour force survey may be more prone to error than establishment surveys or administrative data sources, it is the only available source for estimates of annual hours in quite a number of countries (see Table 5). Thus identifying the precise origin of any errors is necessary if one is to arrive at an acceptable estimation methodology applicable to a large number of countries.

Table 5. **Main sources for estimates of working time**

	Annual actual working time per worker[a]
Australia	LFS - (NA)
Belgium	CLFS
Canada	LFS (NA)
Czech republic	LFS
Denmark	CLFS
France	Establishment survey (NA)
Finland	LFS
Germany	Administrative
Greece	CLFS
Iceland	LFS
Ireland	CLFS
Italy	CLFS
Japan	Establishment survey
Mexico	LFS
Netherlands	CLFS
New Zealand	LFS
Norway	Establishment survey/LFS (NA)
Portugal	CLFS
Slovak Republic	LFS
Spain	LFS
Sweden	Establishment/LFS (NA)
Switzerland	LFS (NA)
United Kingdom	LFS
United States	Establishment survey (NA) Jobs -

LFS: Labour Force Survey; CLFS: Community Labour Force Survey; NA: National Accounts.
a. Average hours actually worked per worker per year, except for Canada, Iceland, France, Germany, Japan, Switzerland and United States where working time is reported on a per job basis.
For more detailed notes and sources, see Table F. of the Statistical Annex of the *OECD Employment Outlook.*

Concepts of hours worked

The comparability of estimates of working time may also be affected by the use of different concepts of hours worked. There are six main concepts of hours of work that are retained in different sources of labour statistics: actual hours, usual hours, contractual hours, legal hours, paid hours, and hours offered (by employers). Labour force surveys collect weekly actual hours according to the ILO resolution concerning statistics of hours of work adopted by the tenth international conference of labour statisticians (ICLS) in 1962. It includes normal hours of work (*i.e.* legal hours, contractual hours), the number of hours in excess of which any time worked is remunerated at overtime rates, time spent for the preparation of the

workplace, short rest periods at the workplace, minus hours paid but not worked due to annual leave, public holidays, sick leave, maternity leave, *etc.*, meal breaks, and travel time to work. To fully account for all hours worked, actual hours worked must include in addition so-called "unpaid overtime", that is, hours worked over and above the standard working hours of an establishment, (generally) by managers and professionals. Strictly speaking, such hours are not overtime since they are most often considered to be part of the implicit contract of employment. Weekly usual hours of work is the other series collected by labour force surveys, which comprise normal hours during a typical workweek of the year, plus regular overtime worked whether paid or unpaid. On the other side, establishment-based surveys, social security or administrative registers record contractual hours, paid hours and hours offered by employers.

Aggregate comparisons of annual hours of work from different sources, although informative, provide little information about the nature of the estimation problem described earlier. For this, confronting the results from different data sources with regard to the components of annual working time is needed, that is, with regard to information on normal hours of work, overtime, time lost due to absences on annual leave, public holidays, sickness and maternity leave. This is the sort of information that can provide clues for the specification of an estimation methodology that can correct for observed deficiencies in sources.

Results on working time from labour force surveys

It was in this spirit that a preliminary comparison of the main components of working time was carried out in a document prepared for the OECD Working Party on Employment and Unemployment Statistics (WP7) in 1998 (OECD 1998*a*). It should be noted that these results are based on data for only a few OECD countries; they would need to be buttressed by results from a wider group. The main findings were as follows:

- "Standard" hours of work from establishment-based surveys and labour force surveys (LFS) differ by 1 to 3 % for the four countries included in the analysis (France, Germany, Netherlands and Switzerland), with labour force surveys yielding the higher estimates.

- Monthly labour force survey estimates of hours not worked due to holidays, even when adjusted for the irregular occurrence of holidays during the reference week seem to be seriously downward biased (Canada).

- Estimates of hours lost due to illness, work accidents and maternity leave from labour force surveys appear to be underestimated by about 45% to 60% compared with administrative sources (in France, Germany and Switzerland). These seem to be associated largely with a serious under-reporting of part-week absences.

- Labour force surveys also seem to underestimate overtime work (*e.g.* in Germany). However, this is not entirely certain, because some regularly occurring overtime may be included in usual hours of work in labour force surveys.

- Finally, in the aggregate, empirical results from two countries suggest that labour force survey estimates yield figures for annual hours of work that are only somewhat higher than those from establishment surveys. This is due in part to the fact that biases in estimates of components of working time tend to cancel out to a certain extent. Note, however, that such relatively close agreement is not found everywhere.

For the OECD growth project, the annual hours worked estimates that were generated for certain countries from labour force surveys, incorporated two adjustments inspired by the above results: *a)* vacation and holidays hours not worked were obtained from external sources, whenever possible; *b)* observed hours of

absence from work due to illness or maternity were doubled relative to those observed directly in labour force surveys. However, it should be repeated that these conclusions are based on results from only a few countries and would need to be supported by more broadly-based analyses.

A second study (OECD, 1999) subsequently examined the effect of so-called "unpaid overtime" worked by managers and professionals on the estimates of annual working time from labour force surveys relative to those from administrative or establishment survey sources. For the latter sources, it is expected that recorded hours for these occupational groups would be the same as the conventional hours of work of other workers in the same establishment. The impact of this "unpaid overtime" recorded in labour force surveys varied by country from as little as no measurable effect to as much as 40 hours per year, depending on the country. However, the impact for individual sectors of activity was far more significant, with certain sectors showing effects of as large as 200 hours per year. That is, the additional hours worked by managers and professionals, over and above those worked by full-timers in other occupations in the same sector, increased the annual hours figure for the sector by about 200 hours from what it would otherwise have been.

To sum up, some uncertainty remains as regards measures of hours worked in OECD countries, although this uncertainty is greater as regards the level of hours worked than as regards its growth rate. Nevertheless, comprehensive estimates of annual working time currently only exist for a limited number of OECD countries and the quality of series on annual hours worked therefore differs across OECD countries, introducing some uncertainty in measures of labour productivity growth.

4.3 Measuring employment

Concepts and coverage

Several sources of employment data exist that can, in principle, be used for analysis of labour productivity. There are several conceptual reasons for differences in these series. These are closely related to the sources used to collect the data:

- Employment concepts in data sources vary: establishment surveys, administrative registers - social security files, tax and insurance records, business registers – and industry censuses collect data on the number of jobs recorded in resident units of production whereas household surveys and population censuses report data on the number of persons in employment based on resident (non institutional) population, in line with national concepts.

- Changes in survey concepts and design may have an effect on employment estimates over time and have adverse consequences on data consistency and comparability. For example, since 1987, Dutch employment figures are based on a more that 12 hours worked per week concept, labour force survey data for the United States prior and after 1994 are not comparable due to a major redesign of the Current Population Survey questionnaire. Labour force survey results are also revised after the results of each (decennial) population census, which are also responsible for series breaks.

- Measurement units are either the number of jobs or the number of workers, depending on data sources (*e.g.* establishment surveys collect data on the number of jobs whereas household surveys report data on a per worker basis).

- Data exclusion: age exclusion (*e.g.* persons aged 67 years or more in the Danish labour force survey) and geographic exclusion are common in household surveys. On the other hand, employer surveys have limited industry coverage – *e.g.* farm employment, personal services,

public sector – limited worker coverage – non-production workers in goods producing industries and supervisory workers in services – and do not cover small firms (below a certain number of workers), *etc.*

- Survey frequencies – monthly, quarterly, annual survey – and survey reference periods vary between countries and surveys.

The degree of multiple jobholding and the extent of cross border workers make up largely for the difference between data sources. Usually, labour force survey results enable the conversion of the number of workers to jobs as they collect details on workers main jobs and additional jobs – however, there is no distinction possible between second and more jobs.

Country practices

Country practices also differ. In practice, National Statistical offices report annual estimates of employment that are based on best sources available in their respective country, in other words, that are both comprehensive and representative of activities performed over the year. It is therefore common that when establishment surveys or administrative sources (*e.g.* social security registers, employees covered by collective bargaining agreements, *etc.*) are retained as the main source, the figures for the employed population not covered by these sources are taken from labour force surveys or other administrative sources. The establishment survey based series are usually benchmarked to employment levels reported in registration data that are used for sample selection. Therefore, employer-survey based employment estimates are regularly revised with the updating from registration data.

On the other hand, labour force survey estimates are benchmarked to population census results and population projections between two census collections.[33] In practice, labour force survey results are revised retrospectively after each census round (to account for population under- or over-count), but country practices vary regarding adjustment methods applied to revise labour force aggregates. In addition, labour force survey scope is, in many cases, limited to civilian non-institutional population or private households and contains age and geographic exclusion, as noted in the previous section on LFS based hours worked measures. The latter exclusions require adding, *ex post*, excluded categories to derive nationally representative employment estimates.

For all these reasons, official estimates of employed persons are based on labour force surveys only in countries with monthly or quarterly surveys (conducted or not on a continuous basis, that is, covering all weeks of the year). On the other hand, estimates of number of jobs, according to national or domestic concepts, are usually based on establishment surveys, industry censuses and administrative registers. As an illustration, the United States report two sets of monthly employment series derived, on the one hand, from an employer survey – Current Employment Statistics (CES) - and, on the other, from the labour force survey – Current Population Survey (CPS). However, the two series show divergent trends in employment levels, probably explained by an underestimation of CPS-based employment levels (Juhn and Potter, 1999) between two consecutive population censuses, differences most probably attributable to the difficulty in projecting net migration estimates.[34]

33. Population projections are based on population registers; administrative records on the number of births and deaths during the year and on net migration.

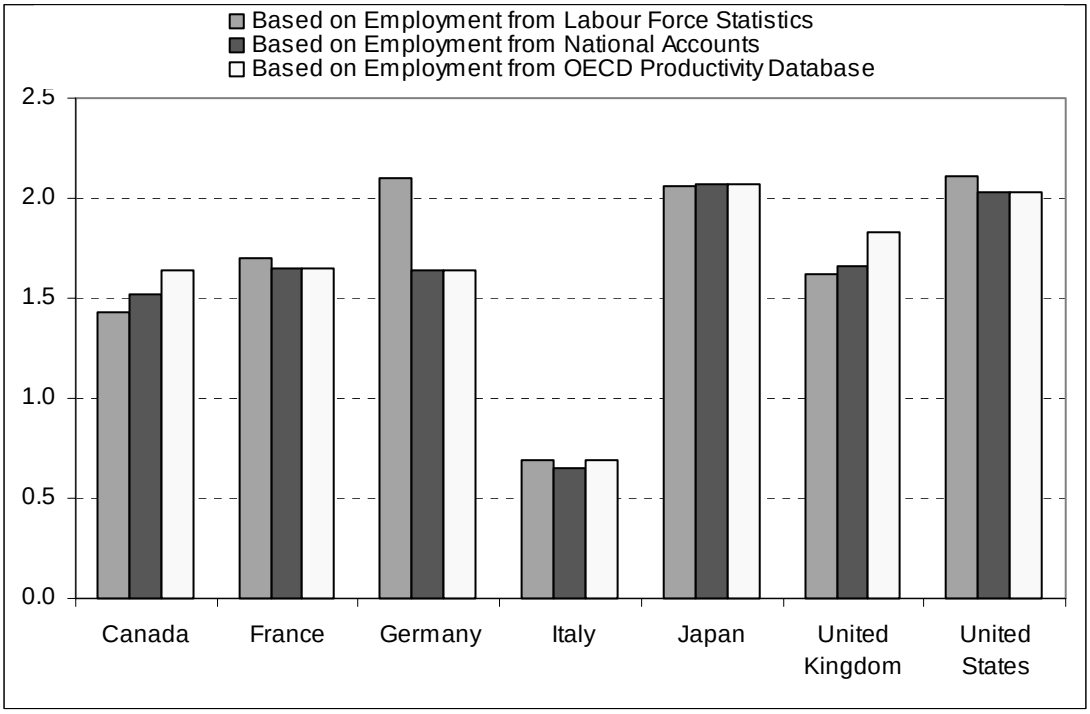
34 . Incidentally, the difference between labour force surveys and enterprise surveys is sometimes exploited as an information source to obtain estimates of the non-observed economy. Of course, this supposes that each survey is in itself correct.

4.4 The coherence of employment and hours worked

Total hours worked can be derived by combining estimates of annual hours worked per person employed with average employment levels over the year that are in accordance to National Accounts production boundaries. Such employment series have not been collected systematically or examined very closely in the past. The quality of the labour productivity measures has relied on the vigilance of the analysts and their understanding of the annual hours and employment series.

In practice, the employment figures generally used for labour productivity calculations at the OECD have been those that are collected for the purpose of OECD economic forecasts, whose nature owes more to historical habit and, to some extent, to national data availability, than to a concern for comparability or adherence to national accounts concepts. Work is currently underway at the OECD to develop a series of estimates of total hours worked that ensure the consistency of employment and annual hours worked over time in calculations of labour productivity. Ensuring such coherence is important for calculations of labour productivity growth (Figure 13).

Figure 13. **Estimates of labour productivity growth using different employment series, 1995-2002**
Annual compound growth rates, in percentage



Source: OECD estimates.

For example, Figure 13 shows that the use of employment estimates from labour force surveys or national accounts will somewhat understate Canada's labour productivity growth over the 1995-2002 compared with the series in the OECD Productivity Database that uses a consistent set of estimates of employment and annual hours worked. These inconsistencies can make a substantial difference in calculations of labour productivity growth; up to 0.5 percentage points in the case of Germany and about 0.2% in the case of Canada and the United Kingdom. For non-G7 economies, these differences are sometimes even larger, depending on the differences between various sources of employment information.

4.5 Adjusting for labour quality

Labour input reflects the time, effort and skills of the work force. While data on hours worked capture the time dimension, they do not reflect the skill dimension. When total hours worked are the simple sum of all hours of all workers, no account is taken of the heterogeneity of labour. For the estimation of productivity changes, the question is whether, over time, the composition of the labour force changes, *i.e.* whether there is an increase or decrease in the average quality of labour input. By most measures, there has been a steady increase in the quality of labour (OECD, 1998*b*). An increase in the average quality of labour implies that a quality-adjusted measure of labour input would rise faster than an unadjusted measure of labour input. Successful quality-adjustment is tantamount to measuring labour in constant-quality units. In the context of productivity measurement, Jorgenson *et al.* (1987), Denison (1985) and the U.S. Bureau of Labor Statistics (BLS, 1993) have tackled this issue.

Measuring constant-quality labour input is interesting from several perspectives. First, it provides a more accurate indication of the contribution of labour to production. One recalls that MFP measures the residual growth in output that cannot be explained by the rate of change in the services of labour, capital and intermediate inputs. When quality-adjusted measures of labour input are used in growth accounting instead of unadjusted hours worked, a larger share of output growth will be attributed to the factor labour instead of the residual factor productivity growth. In other words, substituting quality-adjusted labour input measures for unadjusted measures can better identify the sources of growth, by distinguishing between externalities or spill-overs – captured by the productivity residual – and the effects of investment in human capital. Second, a comparison of an adjusted and unadjusted measure of labour input yields a measure of the corresponding compositional or quality change of labour input. This can usefully be interpreted as one aspect of the formation of human capital, and is thus a step towards measuring the effects of intangible investment.

The theory of the firm stipulates that, under certain conditions (*i.e.* the firm is a price-taker on labour markets and minimises its total costs), labour of a certain type will be hired up to the point where the cost of an additional hour of labour is just equal to the additional revenue that using this labour generates. This equality implies that, for a measure of total labour input, the individual labour inputs of different quality can be weighted using their relative wage rates (or, more precisely, the shares of each type of labour in total labour compensation).

Even when labour input is differentiated only by a simple trait, such as occupation, information requirements are severe: data are needed that distribute the number of total hours worked across different occupations, by individual industry and by individual year. In addition, quantity measures of labour input (*i.e.* hours worked) have to be accompanied by price measures (*i.e.* relative average compensation) to construct weights for aggregation. Such rich data sets are normally both difficult and costly to collect and therefore not readily available in practice.[35]

Even when such data are not available, "implicit differentiation" can provide a useful, albeit incomplete, adjustment for labour quality. Implicit differentiation arises when labour input (*i.e.* total hours worked) is measured by detailed industry without, however, distinguishing between different types of labour within each industry. If the rate of change in hours worked by industry are aggregated to the economy-wide level using each industry's share in total labour compensation as its aggregation weight, these weights will be relatively large for industries that pay above-average wages and relatively small for industries with below-average wages. Assuming that above-average wages reflect an above-average skill composition of the work force, some of the quality change of labour input is implicitly taken into account. Statistics

35. For empirical results see United States Bureau of Labor Statistics (1993), Fosgerau *et al.* (2001) and Scarpetta *et al.* (2000).

Canada's industry-level productivity statistics provide an example of implicit differentiation, since the indices of labour input at the sectoral level are built up from hours-worked data for more detailed industries that are weighted by their shares in total sectoral labour compensation.

Most OECD estimates of labour productivity do not yet incorporate adjustments for labour quality, although some efforts have been made. For example, some work has examined labour composition effects at the industry level using occupational data (OECD, 1998*b*). Other OECD work has calculated labour input as a weighted sum of different groups of workers with different levels of education, each weighted by their relative wage (Scarpetta *et al.*, 2000). Compared with other adjustments available in the literature, this breakdown was rather crude, although it has shed some light on the role of compositional changes in labour input consistently for a range of OECD countries, thereby permitting cross-country comparisons. More OECD work on better measures of labour composition may be carried out in the future.

5. Concluding remarks

The discussion above suggests that several factors affect the comparability of estimates of labour productivity growth across OECD countries. The first concerns nominal GDP. There are several significant problems regarding the comparability of nominal GDP across OECD countries. Cross-country differences in the measurement of military expenditure and FISIM are quite small and account for less than 0.1% of differences in growth rates between the main OECD regions. Differences in the measurement of software are somewhat more important, primarily for those countries where software investment is currently underestimated to a significant degree. Improved measures of software investment may increase GDP growth in some countries (*e.g.* Japan and the United Kingdom) by up to 0.2 - 0.3% over the second half of the 1990s. In the case of FISIM and software, efforts are underway to reduce the size of these differences, which would improve international comparability in the medium term. In all cases, these measurement problems are more prevalent when comparing GDP levels than for comparisons of GDP growth rates. Other areas may matter too though, particularly those components of GDP that are calculated via imputation, for example imputed-rent, a significant component of GDP. Further examination of these areas may be helpful.

Measurement differences for real GDP are also important, although several of these factors have impacts that work in different directions. Moreover, several of the measurement problems primarily affect the distribution of total GDP across different expenditure categories and across different activities, not necessarily GDP growth. For example, improving the output measurement of service sector deliveries to intermediate demand will primarily affect the contribution of these services to aggregate growth at the cost of other sectors; it will not necessarily affect total GDP growth. No definite assessment of the size of the differences due to mis-measurement of real GDP can therefore be provided, although some best practices can be derived. For example, the use of hedonic price measures to reflect rapid quality change will simultaneously require the use of chain-weighted indices. Using hedonic prices in combination with fixed weights will lead to an overstatement of GDP growth and reduced international comparability. Efforts in this area to spread the use of hedonic price measures across countries, to move towards chain-weighted indices and to improve output measurement in services should help address some of the key problems involved.

The measurement of labour input is the third major factor that determines comparability of labour productivity estimates. The choice of employment series for productivity analysis can have a substantial impact on the resulting estimates, as can the estimate of hours worked. A key challenge in deriving comparisons of productivity growth across OECD countries lies in ensuring that estimates of labour input are consistent with the GDP information and that the components of labour input, hours worked and employment are internally consistent. The combination of inconsistent series of employment and annual hours worked can lead to biased estimates of labour productivity growth.

Differences in estimates of labour productivity growth across studies can also be due to the timing of estimates. Since productivity growth fluctuates considerably, inclusion of one extra year can lead to a different assessment of productivity, in particular over relatively short periods (see Annex Table). For example, US labour productivity growth since 1995 amounted to 2.1% annually if calculated over the 1995-2000 period, to 1.9% annually over the 1995-2001 period, and to 2.0% annually over the 1995-2002 period.

An important factor in this context is also revisions by statistical offices, either of real GDP growth or of the associated measures of labour input. These variations arise because the comprehensive data sets that are needed to construct these estimates are not available rapidly. For example, aggregated business accounts for statistical use are only available several years after the year under review. Statistical offices have therefore implemented methods based on more rapid indicators to obtain more timely statistics, *e.g.* on GDP or employment. This system will by necessity lead to revisions. For example, GDP growth rates for major OECD economies over the 1995-99 period have been substantially revised from the beginning of 2000, when the "new economy" bubble was at its peak, to the beginning of 2003. Estimates of average annual GDP growth for the United States changed only little, from 4.07% in April 2000 to 4.10% in January 2003, for Japan from 1.06 to 1.2% over the same period, for Canada from 3.23 to 3.82% and for the four major European economies from 1.89 to 2.05%. The timing of estimates used in different studies is therefore an important factor that needs to be taken into account when comparing their results.

This finding leads to another conclusion that can be drawn from this paper: a careful assessment of available data and their underlying concepts is often equally important for developing accurate estimates of labour productivity growth as cross-country differences in the underlying basic data. Almost all OECD countries now follow the SNA 1993, all have labour force surveys that follow standardised survey practices and all are – though at different speeds – moving towards better practices in measuring real GDP. Some of the most important differences between various studies of labour productivity growth are linked to data choices, *e.g.* the combination of employment, labour and GDP series used in particular studies. To address this problem, OECD is currently developing a reference database on productivity at the aggregate level, which intends to resolve the problem of data consistency.

Finally, much work remains to be done at the national level. Investment in statistics is important and it is hoped that the present document is useful in identifying priority areas where the pay-off of such investment is largest.

REFERENCES

Ahmad, N. (2003), "Measuring Investment in Software", *STI Working Paper 2003-6*, OECD, Paris.

Allen, R. G. D. (1949), "The Economic Theory of Index Numbers", *Economica N.S.* 16, 197-203.

Bassanini, A. and S. Scarpetta (2001), "Does Human Capital Matter for Growth in OECD Countries? Evidence from Pooled Mean-Group Estimates", *OECD Economics Department Working Paper No. 289*, Paris.

Bosworth, B. and J. Triplett (2002), "'Baumol's Disease' Has Been Cured: IT and Multifactor Productivity in US Service Industries"; Paper presented at the New York Federal Reserve Bank conference: Economic Data: New Needs for the 21st Century.

Bureau of Labor Statistics (1993), *Labour Composition and U.S. Productivity Growth, 1948-90*; US Government Printing Office.

Callow, Julien (2003); "Getting Europe to Work"; *Credit Suisse First Boston Fixed Income Research*; 10 April.

Central Bureau of Statistics (1997), "The Use of Chain Indices in the Netherlands"; paper presented at the OECD-UNECE-Eurostat Meeting of National Accounts Experts, Paris.

Denison, E. (1985), *Trends in American Economic Growth 1929-82*; The Brookings Institution, Washington DC.

De Serres, A. (2003), "Structural Policies and Growth: A Non-technical Overview", *Economics Department Working Paper No. 355*, OECD, Paris.

Deutsche Bundesbank (2001), "Problems of International Comparisons of Growth - A Supplementary Analysis"; Appendix to *Monthly Report*, May.

Diewert, Erwin W. (1973), "Afriat and Revealed Preference Theory"; *The Review of Economic Studies* 40, 419-425.

Diewert, Erwin W. (1992), "Fisher Ideal Output, Input and Productivity Indexes Revisited"; Journal of Productivity Analysis, Vol. 3, No 3, 211-248.

Eurostat (2001); *Handbook on Price and Volume Measures in National Accounts*, Luxembourg.

Gullickson, W. and M.J. Harper (1999), "Possible Measurement Bias in Aggregate Productivity Growth", *Monthly Labor Review*, February 1999.

Gullickson, W. and M.J. Harper (2002), "Bias in Aggregate Productivity Trends Revisited", *Monthly Labor Review*, March 2002.

Hill, Peter (1988); "Recent Developments in Index Number Theory and Practice", *OECD Economic Studies* No 10, Paris.

Jorgenson, D.W., F. Gallop and B. Fraumeni (1987), *Productivity and U.S. Economic Growth*, Harvard University Press, Cambridge.

Jorgenson, D.W. and E. Yip (2001), "Whatever Happened to Productivity Growth", in E. Dean, M.J. Harper and C. Hulten (eds.), *New Developments in Productivity Analysis*, University of Chicago Press, Chicago, pp. 205-246.

Juhn, Chinhui, and Simon Potter (1999), "Explaining the Recent Divergence in Payroll and Household Employment Growth", Current Issues in Economics and Finance, December, Vol. 5, No. 16, Federal Reserve Bank of New York.

Lal, K. (2003), "Measurement of Output, Value Added, GDP in Canada and the United States: Similarities and Differences", Statistics Canada, *mimeo*.

Landefeld, Steven J. and Bruce T. Grimm (2000); "A Note on the Impact of Hedonics and Computers on Real GDP"; *Survey of Current Business*, December.

Landefeld, Steven J. and Robert P. Parker (1997); "BEA's Chain Indexes, Time Series, and Measures of Long-Term Economic Growth"; *Survey of Current Business*, May.

Lequiller, François (2001); "La nouvelle économie et la mesure de la croissance du PIB"; *Série des documents du travail de la Direction des Etudes et Synthèses Economiques G2001/01*; INSEE.

Lequiller, F., Ahmad, N., S. Varjonen, W. Cave and K.H. Ahn (2003), "Report of the OECD Task Force on Software Measurement in the National Accounts", *Statistics Directorate Working Paper 2003/1*, OECD, Paris.

Lum, Sherlene K.S., Brian C. Moyer and Robert E. Yuskavage (2001); "Improved Estimates of Gross Product by Industry 1947-98"; *Survey of Current Business*.

Moulton, Brent (2000); "Measurement of Banking Services in the U.S. National Income and Product Accounts: Recent Changes and Outstanding Issues"; document available at http://www.bea.doc.gov.

OECD (1998*a*), "Annual Hours of Work: Definitional and Comparability Issues", Working Party on Employment and Unemployment Statistics, DEELSA/ELSA/WP7(98)2.

OECD (1998*b*), "Labour Input and Labour Composition Measures by Industry: An Exploratory Analysis for the G7 Countries", DSTI/EAS/IND/SWP(98)1, Paris.

OECD (1999), "Hours of Work - Comparability and Polarisation Issues", Working Party on Employment and Unemployment Statistics, DEELSA/ELSA/WP7(99)8.

OECD (2001*a*), *The New Economy: Beyond the Hype*, Paris.

OECD (2001*b*), *Measuring Productivity - OECD Manual*, Paris.

OECD (2003*a*), *The Sources of Economic Growth in OECD Countries*, Paris.

OECD (2003b), *ICT and Economic Growth – Evidence from OECD countries, Industries and Firms*, Paris, forthcoming.

Samuelson, Paul (1947); *Foundations of Economic Analysis*, Cambridge Mass.

Scarpetta, S., A. Bassanini, D. Pilat and P. Schreyer (2000), "Economic Growth in the OECD Area: Recent Trends at the Aggregate and Sectoral Levels", *OECD Economics Department Working Paper No. 248*, Paris.

Schreyer, Paul (2001); "Information and Communication Technology and the Measurement of Real Output, Final Demand and Productivity"; *Economics of Innovation and New Technology*, Vol. 10. 2001.

Schreyer, Paul (2002); "Computer price indices and international growth and productivity comparisons"; *Review of Income and Wealth*.

Schreyer, P. and D. Pilat (2001), "Measuring Productivity", *OECD Economic Studies*, No. 33, 2001/2, OECD, Paris, pp. 127-170.

Schreyer, P., P.E. Bignon and J. Dupont (2003), "OECD Capital Services Estimates: Methodology and A First Set of Results", *OECD Statistical Working Papers*, forthcoming.

Sharpe, Andrew., Someshwar Rao and Jianmin Tang (2002), "Perspectives on Negative Productivity Growth in Service Sector Industries in Canada and the United States", Paper presented at *Workshop on Service Sector Productivity*, May, Brookings Institution, Washington, DC.

Slifman, Larry and Carol Corrado (1996), "Decomposition of Productivity and Unit Costs", *Occasional Staff Studies*, OSS-1, Federal Reserve Board, Washington, D.C.

Szulc, Bohdan J. (1983), "Linking Price Index Numbers", in: Diewert, Erwin and Claude Montmarquette (eds.), *Price Level Measurement*, Statistics Canada.

Vijselaar, F.W. (2003), "ICT and Productivity Growth in the Euro Area: Sectoral and Aggregate Perspectives", forthcoming in the proceedings of the IVIE Workshop on *Growth, Capital Stock and New Technologies*, to be published by the BBVA Foundation.

Whelan, Karl (2000); *A Guide to the Use of Chain Aggregate NIPA Data*; Federal Reserve Board; and United States Bureau of Economic Analysis.

Wölfl, Anita (2003); "Productivity Growth in Service Industries: an Assessment of Recent Patterns and the Role of Measurement"; *STI Working Papers 2003-6, OECD*, Paris.

Annex Table: Growth of GDP per hour worked, 1970-2002, OECD estimates based on OECD productivity database

(total economy, annual compound growth rates, in per cent)

	1970-80	1980-90	1990-2000	1990-2001	1990-2002	1990-95	1995-2000	1995-2001	1995-2002	1995	1996	1997	1998	1999	2000	2001	2002
Australia	1.7	1.3	2.0	2.2	2.2	1.7	2.4	2.6	2.5	1.5	3.5	3.3	3.5	1.5	-0.1	4.0	2.0
Austria	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Belgium	4.1	2.6	2.2	1.8	1.7	1.8	2.6	1.9	1.6	-3.3	2.6	1.8	-0.1	5.4	3.3	-1.8	0.2
Canada	2.1	0.9	1.5	1.5	1.5	1.2	1.8	1.9	1.8	1.6	-0.3	2.8	1.5	2.4	2.8	2.0	1.7
Czech Republic	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	1.9	2.8	2.7	4.1	4.0	-0.1	0.0	2.0	3.7	7.5	2.0
Denmark	3.3	1.6	1.8	1.7	1.8	2.0	1.5	1.3	1.7	4.0	1.2	0.9	1.1	-0.4	4.8	0.3	3.8
Finland	3.1	2.8	2.8	2.7	2.7	2.4	3.2	2.9	2.9	1.4	1.5	5.0	3.7	-0.1	6.0	1.4	2.5
France	3.8	3.0	2.0	2.0	2.0	1.8	2.2	2.1	2.1	2.3	0.5	1.8	2.6	1.6	4.4	1.9	1.6
Germany	3.9	1.7	2.7	2.6	2.5	3.4	1.9	1.8	1.7	2.5	2.3	2.0	1.3	1.5	2.2	1.3	1.3
Greece	n.a.	1.8	1.5	1.7	1.8	0.3	2.9	3.1	3.1	1.6	1.8	5.1	-0.6	2.4	5.7	3.9	3.6
Hungary	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	2.7	3.1	3.1	3.2	1.5	4.1	3.2	0.7	4.0	5.2	3.2
Iceland	5.4	1.2	1.2	1.3	1.5	0.1	2.5	2.6	2.6	-3.7	3.5	5.8	2.6	-2.6	3.0	3.2	2.8
Ireland	4.9	3.8	4.6	4.5	4.7	3.8	5.6	5.3	5.4	4.8	3.8	9.7	2.6	6.5	5.2	4.0	6.0
Italy	4.3	2.3	2.0	1.8	1.6	2.7	1.0	1.0	0.7	3.6	0.5	1.4	0.9	0.8	1.6	0.6	-0.9
Japan	4.5	3.3	2.5	2.4	2.4	2.8	2.1	2.0	2.1	2.6	2.5	2.3	0.8	2.7	2.3	1.7	2.2
Korea	n.a.	6.2	5.0	4.8	4.8	5.6	4.2	3.9	4.1	5.3	5.2	4.5	1.2	4.3	5.8	2.2	5.1
Luxembourg	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Mexico	n.a.	n.a.	0.1	0.2	-0.1	-2.6	2.2	2.0	1.3	-8.6	-1.1	-0.7	4.1	-0.3	9.0	1.0	-2.8
Netherlands	4.5	1.8	1.4	1.5	1.2	1.4	1.4	1.6	1.1	1.6	0.4	1.2	2.6	2.6	0.0	2.8	-1.9
New Zealand	0.3	0.8	0.9	0.9	1.0	0.3	1.6	1.5	1.5	-0.6	0.2	1.9	0.9	2.6	2.3	1.0	1.3
Norway	4.9	2.5	3.0	2.9	2.9	3.5	2.4	2.4	2.4	3.4	3.6	2.7	0.2	1.7	3.8	2.3	2.3
Poland	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Portugal	n.a.	n.a.	3.0	2.8	2.6	2.9	3.1	2.7	2.3	2.9	3.1	4.6	2.8	1.1	4.2	0.3	0.3
Slovakia	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	3.9	3.6	4.1	3.7	0.7	4.9	4.8	5.3	3.7	2.1	7.1
Spain	4.5	3.6	1.0	0.9	0.9	2.2	-0.3	-0.4	-0.3	0.3	0.1	0.5	-0.9	-0.3	-1.1	-1.0	0.6
Sweden	2.5	1.2	2.0	1.9	2.0	1.7	2.3	2.0	2.2	1.7	1.6	3.5	1.9	2.0	2.6	0.6	3.1
Switzerland	2.0	1.0	0.4	0.5	0.4	-0.5	1.4	1.4	1.1	1.1	0.8	1.8	0.9	0.1	3.2	1.4	-0.3
Turkey	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
United Kingdom	2.8	2.0	2.4	2.3	2.2	2.7	2.1	1.9	1.9	1.6	1.8	1.8	2.4	1.8	2.7	1.2	1.5
United States	1.8	1.3	1.6	1.5	1.6	1.2	2.1	1.9	2.0	0.4	2.3	1.6	1.9	2.1	2.4	1.1	2.9
EU-11*	3.6	2.3	2.1	2.0	1.9	2.6	1.6	1.5	1.4	2.0	1.2	1.7	1.5	1.4	2.2	0.9	0.9
EU-13**	n.a.	n.a.	2.1	2.0	1.9	2.6	1.7	1.6	1.5	2.0	1.3	1.9	1.4	1.4	2.4	1.0	0.9

Notes: Estimates are available as follows (first year in parenthesis) - Korea (1980); Greece (1983); Portugal (1986), Mexico (1991), Hungary (1992), Czech Republic (1993) and Slovakia (1994). Hours worked estimates for Austria, Luxembourg, Poland and Turkey are not available.

* EU-11: Excluding Austria, Greece, Luxembourg and Portugal. EU averages are based on GDP in 1995 PPP USD.

* EU-13: Excluding Austria and Luxembourg. EU averages are based on GDP in 1995 PPP USD.

Source: OECD Productivity Database.