31st CEIES Seminar
Are we measuring productivity correctly?

Rome, 12 and 13 October 2006

2007 edition
How to obtain EU publications

Our priced publications are available from EU Bookshop (http://bookshop.europa.eu), where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents. You can obtain their contact details by sending a fax to (352) 29 29-42758.


Luxembourg: Office for Official Publications of the European Communities, 2007

Cat. No. KS-PB-06-002-EN-N

Theme: General and regional statistics
Collection: Methodologies and working papers

© European Communities, 2007
Thursday, 12 October 2006

OPENING SESSION
Welcome to the participants

Ms Martine Van Wouwe, Professor, University of Antwerp, Belgium, Chairperson of the CEIES Subcommittee on Economic and Monetary Statistics

Ms Vittoria Buratta, Director of the Department for Statistical Production and Technical Scientific Coordination (DPTS) of the Italian National Institute of Statistics (ISTAT)

Mr Joachim Recktenwald, Head of Unit ‘National Accounts – Production’, Eurostat

MEASURING PRODUCTIVITY
Chair: Mr. Joachim Lamel, Member of the Board, Austrian Statistical Society

Ms Susanna Mantegazza, Italian National Institute of Statistics (ISTAT) – Italian methodology in measuring productivity

Mr Bernd Görzig, German Institute for Economic Research – Depreciation measurement in the EU countries

OPEN DISCUSSION

MEASURING PRODUCTIVITY (CONT.)
Chair: Mr Joachim Lamel, Member of the Board, Austrian Statistical Society

Mr Bohdan Wyznikiewicz, Gdansk Institute for Market Economics, Poland – Critical assessment of the conventional approach to productivity measurement resulting from the SNA

OPEN DISCUSSION

IMPACT OF HEDONIC PRICE MEASUREMENT ON PRODUCTIVITY
Chair: Ms Lea Bregar, Faculty of Economics, University of Ljubljana, Slovenia

Mr David Fenwick, Office for National Statistics, United Kingdom – The use of hedonic regression techniques for the quality adjustment of computer equipment in consumer prices indices and in the producer prices index: the UK experience

Mr Gavin Wallis, Office for National Statistics, United Kingdom – ICT deflation and productivity measurement

Mr David Wasshausen, Bureau of Economic Analysis, Department of Commerce, United States of America – The role of hedonic methods in measuring real GDP in the United States

OPEN DISCUSSION
Friday – 13 October 2006

MEASUREMENT OF PRODUCTIVITY IN THE NON-BUSINESS SECTOR
Chair: Mr Joachim Recktenwald, Eurostat

Mr Emmanuel Thanassoulis, Aston Business School, Aston University, United Kingdom – Productivity measurement in multi-input multi-output contexts and in the absence of prices

Mr Bob Kuhry, Sociaal en Cultureel Planbureau, The Netherlands – Production, productivity and cost price of public services in the Netherlands

Mr Finn R. Forsund, Professor of Economics, University of Oslo, Norway – Productivity of tax offices in Norway

OPEN DISCUSSION

Mr. Alain Gallais, OECD / Mr. F. Malherbe, Eurostat – Results of the London Workshop on the ‘Improvement of non-market output in education and health’

OPEN DISCUSSION

HUMAN CAPITAL, KNOWLEDGE AND THEIR IMPACT ON MEASURING PRODUCTIVITY
Chair: Botho, Graf Pückler, Germany

Mr Christiaan Stam, Research Fellow at Center of Research in Intellectual Capital, INHOLLAND University and de Baak – Management Centre VNO-NCW, The Netherlands—Knowledge productivity: designing and testing a method for measuring knowledge productivity in order to give direction to knowledge management initiatives

Mr Henry van der Wiel, CPB Netherlands Bureau for Economic Analysis, the Netherlands – Limousines: the way to move forward

Ms Monika Kircher-Kohl, Infineon Technologies Austria AG, Austria – Measuring and driving productivity – a key to success for innovative companies

OPEN DISCUSSION

SUMMING UP

Ms Martine Van Wouwe, Professor, University of Antwerp, Belgium, Chairperson of the CEIES Subcommittee on Economic and Monetary Statistics

REACTION FROM EUROSTAT

Mr Joachim Recktenwald

END OF SEMINAR
Background and aim of the seminar

The Lisbon European Council set an objective for the EU to become ‘the most competitive and dynamic knowledge-based economy in the world’. As a result, changes in productivity have become one of the central indicators for measuring competitiveness and a strategic parameter.

The issue of the economic performance of a branch, a sector, or the entire economy and the direction in which it is moving has considerable relevance for economic policy. This means that overall productivity of the whole economy is a central measure of economic performance. As a rule supranational as well as international and cross-sectoral comparisons are involved. And depending on the results, calls for policy interventions are often made.

Four plenary sessions are planned for the two-day seminar, each of which will try to articulate the view of both producers and users from different contexts.

The aims of this seminar are:

— to give an overview of the state of the art in measurement of productivity worldwide,
— to identify major challenges for statistics with a special focus especially on the adjusted Lisbon goals and gaps in the European Statistical System,
— to discuss and to describe its impact on the current System of National Accounts,
— to brainstorm alternative solutions, including possibilities as well as limitations.

What is CEIES?

CEIES stands for Comité consultatif européen de l’information statistique dans les domaines économique et social; in English: ‘The European Advisory Committee on Statistical Information in the Economic and Social Spheres’. Its task is to assist the Council and the Commission in the coordination of the objectives of the Community’s statistical information policy, taking into account user requirements and the costs borne by the information producers.

The committee was set up by Council Decision 91/116/EEC of 25 February 1991. The original decision was amended by Council Decision 97/255/EC of 19 April 1997 taking into account the accession of Austria, Finland and Sweden.

CEIES is chaired by the Commissioner responsible for Eurostat. The vice-chairman is Ms. Margit Epler from Austria. CEIES is composed of two private members per Member State, three members from the European Commission, the Chairman of the Committee on Monetary, Financial and Balance of Payments Statistics (CMFB) and the Presidents or Directors-General of the National Statistical Institutes of the Member States.
Programme

http://epp.eurostat.ec.europa.eu/
e-mail : estat-ceies@ec.europa.eu
The European Advisory Committee
on Statistical Information in the Economic and Social Spheres
Secretariat:  Eurostat, Unit A-2
Fax (352) 4301-32629

Organisers:  
*The CEIES subcommittee on Economic and Monetary Statistics:*
Prof. M. Van Wouwe, chairperson; Prof. F.P. Ruane, Mr M. Weale,  
Dr U. Oschisching, Mr R. Camilleri, Mr R.B. Larsen,  
Dr B. Wyznikiewicz, Prof. P. Garibaldi, Prof. J. Komornik,  
Dr J.R. Lopez, D. J. De Lucio;  
with the help of Ms M. Epler, CEIES vice-president;  
*From Eurostat: Mr J. Recktenwald*

CEIES Secretariat:  
Mr G. Wächter, Ms M.-P. Scheidhauer, Ms S. Blair
CONTENTS

The proceedings are a collection of papers prepared by the speakers in advance of the seminar. They do not include the open discussions nor the discussions in the round table/panel session.

The papers presented and published herein only represent the views of their authors and do not necessarily reflect an official position of their institutions or organisations.

OPENING SESSION
Ms Martine Van Wouwe, Professor, University of Antwerp, Belgium, Chairperson of the CEIES Subcommittee on Economic and Monetary Statistics ......................................................... 11
Mr Joachim Recktenwald, Head of Unit ‘National Accounts – Production’, Eurostat ........................................ 13

MEASURING PRODUCTIVITY
Ms Susanna Mantegazza, Italian National Institute of Statistics (ISTAT) – Italian methodology in measuring productivity ..................................................................................................................... 19
Mr Bernd Görzig, German Institute for Economic Research – Depreciation measurement in the EU countries .................................................................................................................. 35
Mr Bohdan Wyznikiewicz, Gdansk Institute for Market Economics, Poland – Critical assessment of the conventional approach to productivity measurement resulting from the SNA ............................................ 50

IMPACT OF HEDONIC PRICE MEASUREMENT ON PRODUCTIVITY
Mr David Fenwick, Office for National Statistics, United Kingdom – The use of hedonic regression techniques for the quality adjustment of computer equipment in consumer prices indices and in the producer prices index: the UK experience .............................................................................................................. 63
Mr Gavin Wallis, Office for National Statistics, United Kingdom – ICT deflation and productivity measurement .................................................................................................................. 84
Mr David Wasshausen, Bureau of Economic Analysis, Department of Commerce, United States of America – The role of hedonic methods in measuring real GDP in the United States ....................... 97

MEASUREMENT OF PRODUCTIVITY IN THE NON-BUSINESS SECTOR
Mr Emmanuel Thanassoulis, Aston Business School, Aston University, United Kingdom – Productivity measurement in multi-input multi-output contexts and in the absence of prices .................. 115
Mr Bob Kuhry, Sociaal en Cultureel Planbureau, the Netherlands – Production, productivity and cost price of public services in the Netherlands ........................................................................ 131
Mr Finn R. Førsund, Professor of Economics, Department of Economics, University of Oslo – Productivity of tax offices in Norway .................................................................................. 149
Mr Alain Gallais, OECD – Results of the London Workshop on the ‘Improvement of non-market output in education and health’ ................................................................. 170
HUMAN CAPITAL, KNOWLEDGE AND THEIR IMPACT ON MEASURING PRODUCTIVITY

Mr Christiaan Stam, Research Fellow at Center of Applied Research in Intellectual Capital, INHOLLAND University and De Baak – Management Centre VNO-NCW, the Netherlands – Knowledge productivity: designing and testing a method for measuring knowledge productivity in order to give direction to knowledge management initiatives ................................................................. 177

Ms Monika Kircher-Kohl, Infineon Technologies Austria AG – Measuring and driving productivity – a key to success for innovative companies .................................................................................................................. 216

SUMMING UP

Ms Martine Van Wouwe, Professor, University of Antwerp, Belgium, Chairperson of the CEIES Subcommittee on Economic and Monetary Statistics .................................................................................................. 219

REACTION FROM EUROSTAT

Mr Joachim Recktenwald .................................................................................................................. 225

LIST OF PARTICIPANTS ..................................................................................................................... 229
Opening session
OPENING SESSION

Martine VAN WOUWE
Professor, Antwerp University, Belgium
Chairperson, CEIES Subcommittee on Economic and Monetary Statistics

First of all I would like to thank ISTAT for hosting the 31st CEIES seminar ‘Are we measuring productivity correctly?’

The Lisbon European Council set an objective for the EU to become ‘the most competitive and dynamic knowledge-based economy in the world’. As a result, changes in productivity have become one of the central indicators for measuring competitiveness and a strategic parameter.

An essential factor determining the growth potential of an economy is the productivity of the production factors deployed. The level and direction of productivity movements are therefore among the most important economic indicators. They reflect the capacity of the economy to implement technological progress and give shape to structural changes in the economy.

Given the great significance attached to productivity indicators, it is to be expected that there should be certainty about how the associated statistics are established.

For these reasons the CEIES subcommittee has decided to put this subject on its agenda and to organise a seminar on measuring productivity.

In the first session we look at the new developments in measuring productivity.

Today, it is no longer sufficient to look at labour productivity alone in order to assess growth and structural change. It should be borne in mind that labour productivity is also determined by capital intensity, i.e., the capital investment associated with jobs and ‘knowledge’ deployed. Even the present reference parameter, i.e. labour input per employee or per hour, must be questioned. Therefore harmonizing definitions, concepts and methodologies should be evaluated to develop reasonably comparable data. The most widely recognised measure of productivity is related to labour. A broader and theoretically more pertinent concept incorporates the effect of other factors of productivity and is called the multifactor productivity (also called total factor productivity).

In the second session the impact of hedonic price measurement on productivity is discussed. Changes in quality are not captured or are inadequately captured. Although the hedonic method is increasingly used, there are still ‘significant differences’ with regard to the goods in question and in the methodological fine-tuning. In addition, the transition to hedonic-based price measurement does not automatically increase the international comparability of the results. The much discussed problem is that real parameters – such as GDP growth and productivity – are not or are only partially comparable due to different deflation methods. This problem will only be solved when there is comprehensive
harmonisation of the hedonic method and how it is applied. Therefore deflation in nominal production values should become a special research topic.

The third session will develop the very important problem of the measurement of the output of the non-business sector, i.e. the public sector. The value added of public services represents between 15% and 20% of GDP. As many countries use different methodologies, the international comparability of data representing the public sector is not ensured.

In the fourth session we look at human capital and knowledge and their impact on measuring productivity. Research should focus on the individual production factors – labour, capital and knowledge – and how they are combined. In particular the production factor ‘knowledge’ requires close examination. Recent growth theories accept the fact that education affects productivity levels: education creates human capital directly through knowledge accumulation, but also indirectly through successful research activities. Education seems to be important more via its effects on research rather than directly as a factor input. There is more empirical evidence of influence of higher education in the growth process: countries with higher average years of education tend to grow faster. So, research needs to be carried out to determine the extent to which ‘knowledge’ can be measured or rendered measurable and integrated in a productivity rate.

These are the topics which will be treated in the seminar during the next two days.
If I had to list five macro-economic indicators that have been at the centre of the economic policy debate in the European Union over the last decade, productivity would certainly be one of them. Indeed, productivity is at the heart of what is called the Lisbon process. The issue is competitiveness, and competitiveness is directly related to productivity. You are well aware of the discussion on the productivity surge in the United States and on the reasons why Europe is lagging behind. There has been a heated debate on the importance of information and communication technology for the economic performance on both sides of the Atlantic.

No doubt, productivity is high on the policy agenda.

Hence, there is every good reason for statisticians to ask ‘Are we measuring productivity correctly?’

Productivity is a tricky thing. In general terms, the concept is simple: It is the ratio of the output volume to the volume of inputs used in the production process. Behind that, however, there are a number of most difficult statistical measurement issues.

Let me briefly mention some of them.

Take the volume of output. Its measurement is anything but straightforward. In the case of market production, appropriate methods of deflating current price aggregates are needed to obtain the volume of output. For non-market activities, such as health and education, there are no output prices. Hence, the challenge is to measure the volume of output directly. On top of this, across all activities the changing quality of goods and services needs to be taken into account.

Now look at the labour input. It is obvious that it is preferable to measure it in terms of actual hours worked rather than by a simple headcount of the persons employed. But how best capture hours worked, in particular unpaid overtime? And what about the quality of human capital?

There is also the general conceptual issue of single factor productivity vs multi-factor productivity. What are the pros and cons? And finally: Should the estimation of multi-factor productivity be the business of statistical institutes?

A long list of most interesting questions!
Therefore, I am grateful to CEIES for organising this seminar, and I am grateful to ISTAT for hosting it.

I am looking forward to the presentations of our invited speakers and to lively discussions with all of you.
Measuring Productivity
ITALIAN METHODOLOGY IN MEASURING PRODUCTIVITY

Cecilia JONA-LASINIO
Massimiliano IOMMI
Susanna MANTEGAZZA

Italian National Institute of Statistics – ISTAT
Directorate of National Accounts

Executive summary

The purpose of this document is to illustrate the methodology adopted at the Italian Institute of Statistics to calculate total factor productivity and to show some results for the Italian economy over the period 1992-2003. We evaluate to what extent the Italian output growth can be attributed to technological factors (TFP growth) and how much should be assigned to the growth of inputs.

In particular, we want to assess the role of ICT capital accumulation in stimulating economic growth. We do this by means of both aggregate and industry growth accounting models. In order to isolate the underlying sources of gains in productivity and to a better understanding of the forces driving economic growth it is necessary to identify the contribution of each industry to aggregate productivity growth. In fact, positive productivity growth in one sector can be offset by negative productivity growth in another.

It must be noticed that the results and the methodology reported in this paper will be subject to some modifications in the near future. In fact, all the data used in this paper are going to be revised as a result of the ongoing benchmark revision of national accounts that will be concluded by December 2006. Besides new data sources and estimation methods, the revised national accounts will reflect two major conceptual innovations: the adoption of the supply and use framework and the introduction of chain indexes to measure changes in volume and price. These innovations will entail some changes in the methodology currently adopted to estimate TFP growth.

The structure of the paper is as follows. The methodology is illustrated in section 2, while section 3 is devoted to data description. The empirical results are shown in section 4 and some conclusions are drawn in section 5.

1 Istat working group on productivity: Susanna Mantegazza, Antonella Baldassarini, Massimo Costanzo, Nadia Di Veroli, Laura Esposito, Massimiliano Iommi, Cecilia Jona-Lasinio, Livia Mastrantonio.
1. Methodological issues

1.1. The growth accounting framework

Economic growth can be broken down into components associated with changes in factor inputs and the Solow residual that reflects technological progress and other elements (Barro, 1998). Growth accounting is basically a matter of measuring this residual, or total factor productivity (TFP) growth, and using the result to divide the growth of real output into an input component and a productivity component. This technique allows to establish how much of an observed rate of change of an industry’s output can be explained by the rate of change of combined inputs. In this paper we adopt the standard growth accounting model based on the seminal work by Tinbergen (1942) and Solow (1957) and its development by Griliches and Jorgenson (1967).

The model relies on several simplifying assumptions that are not necessarily met in practice, but provide a reasonable approximation to many markets. In particular, it is assumed that perfect competition holds in all markets; that production technology (both at aggregate and at the industry level) is characterized by constant return to scale, Hicks-neutral technological change and absence of adjustment costs to change the level of any input (so that there are no quasi-fixed inputs and firms are always in a long-run equilibrium). If these assumption holds, some non-observable variables such inputs marginal product and the corresponding output elasticity can be measured by inputs prices and their income shares. Moreover since these shares sum to one, the capital share can be measured as a residual.

The analysis has been done both at the economy-wide and at the industry level. The two levels of analysis have been linked through a decomposition of aggregate TFP growth into the contributions from TFP growth in each industry.

1.1.1 Sources of growth at the aggregate level

Identifying the sources of growth at the economy-wide level requires a framework that describes the relation between aggregate output and primary inputs. There are two alternative representations of aggregate economic activity, one in terms of value-added and another one in terms of deliveries to final demand. In a closed economy the two representations coincide, but in an open economy they originate a different measure of output, a different definition of the primary inputs and, consequently, a different growth accounting equation. The difference between the two representations is due to imported intermediate inputs. In fact, the value of final demand is equal to aggregate value-added plus the value of imported intermediate inputs, while, with respect to inputs definition, in the deliveries to final demand model intermediate inputs are primary factors of production like capital and labour.

We adopted the deliveries to final demand model because it allows a more general representation of aggregate economy given that it requires only constant returns to scale sectoral production functions

---

1 This paragraph draws heavily from Bassanetti, Iommi, Jona-Lasinio, Zollino (2004).
2 Put it differently, while aggregate value-added corresponds to the income generated by capital and labour, the value of final demand corresponds to the income generated by capital and labour plus the value of imported intermediate inputs (here we abstract from indirect taxes and subsidies).
3 The intuition is that while purchases and sales of domestically produced intermediate inputs are netted out at the aggregate level, there are no sales that offset imported intermediate inputs purchases.
(Gollop, 1983). Indeed, the existence of the aggregate production function, in the value-added model, requires that value-added separability must hold for the production function of gross output for each sector, that sectoral value-added functions must be constant returns to scale function and that they must be identical to the aggregate value-added function (Jorgenson, Gollop and Fraumeni, 1987).

The derivation of the growth accounting equation from the deliveries to final demand model, starts from an economy-wide production possibility frontier relating available primary inputs and technology to deliveries to final demand (see Gollop, 1983, and OECD, 2001):

\[ G(FD_t;KI_t,KO_t,L_t,M_t,A_{fd}^t) = 0 \]

where FD is an index of deliveries to final demand (in real terms) and M are real imported intermediate inputs.

The growth accounting equation for the deliveries to final demand model is:

\[ \frac{\dot{FD}_t}{FD_t} = \sum\frac{\dot{KI}_t}{KI_t} + \sum\frac{\dot{KO}_t}{KO_t} \]

where \( \frac{\dot{I}_t}{I_t} \) is the share of input \( i \) in the value of final demand and the rate of growth of final demand is defined as a Divisia index of the rates of growth of deliveries from individual industries.

Aggregate TFP growth measures the shift in the frontier holding primary inputs constant, where primary inputs include also imported intermediate inputs.

Gollop (1983) contrast the value-added model of economic activity and the deliveries to final demand model (both derived from a production possibility frontier framework) and shows that the two models give a rate of aggregate productivity growth that differ only for an amount proportional to inverse of the share of current price value-added to the value of final demand:

\[ \frac{A_{va}^t}{A_{va}^t} = \frac{\dot{FD}_t}{FD_t} \cdot \frac{A_{fd}^t}{A_{fd}^t} \]

1.1.2 Sources of growth at the industry level

Measuring the sources of growth at the industry level is a fundamental step to shed some light on the determinants of aggregate growth. For instance, industry level analysis is needed to answer the question of whether the major contributions to aggregate TFP growth came from ICT using industries, ICT producing industries or by industries less involved in ICT.

Starting from a description of industry technology in terms of a production function for industry gross output, it is possible to derive a growth accounting relation for each industry
(4) \[ Y_j^t = A_j^t F_j^t (K_{I_j}^t, K_{O_j}^t, L_j^t, X_j^t) \]

where \( Y_j \) is real gross output of industry \( j \), \( X_j \) are real intermediate inputs that industry \( j \) buys from other industries (both foreign and domestically produced)\(^1\).

The growth accounting equation for each industry is:

(5) \[ \frac{Y_j^t}{Y_j^{t-1}} = \text{sy}_{ki} \frac{K_{I_j}^t}{K_{I_j}^{t-1}} + \text{sy}_{ko} \frac{K_{O_j}^t}{K_{O_j}^{t-1}} \frac{L_j^t}{L_j^{t-1}} + \text{sy}_{l} \frac{X_j^t}{X_j^{t-1}} + \frac{A_j^t}{A_j^{t-1}} \]

where \( \text{sy}_i (i=ki, ko, l, x) \) is the share of input \( i \) in the value of gross output of industry \( j \).

### 1.1.3 Industry contribution to aggregate productivity growth

Domar (1961) first proposed a framework to link aggregate and industry level measures of TFP. The Domar’s aggregation formula expresses aggregate TFP as a weighted sum of industry level TFP’s, with weights equal to the ratios of the value of each industry gross output to the value of final demand:

(6) \[ \frac{A_i^t}{A_i^{t-1}} = \sum_j \frac{P_j^t Y_j^t}{P_j^{t-1} F_D} \frac{A_j^t}{A_j^{t-1}} \]

Hulten (1978) demonstrated that, under the assumptions of perfect competitions and constant returns to scale, the Domar’s aggregation formula actually measures the shift in the production possibility frontier over time (i.e. aggregate TFP derived from equation 2). Hulten (1978) derives the result in a closed economy framework: Gollop (1983) derives the same result for an open economy. Equation 6 shows that the shift in the production possibility frontier depends both on TFP growth in each industry and on shifts of economic activity among industries with different TFP growth rates\(^2\).

### 1.2 Measuring inputs

Following the approach first proposed by Jorgenson and Griliches (1967) and now become standard practice in growth accounting, all inputs are measured taking into account the heterogeneity in the marginal products of their individual types (i.e. the various types of capital goods and intermediate inputs and the various categories of the work force).

#### 1.2.1 Output and intermediate inputs

In a growth accounting framework, a first issue to deal with is the proper measure of output, since both value added and gross production can be considered. Indeed, as reported in section 2.1, the point

---

\(^1\) Note that industry level technology does not rely on the restrictive assumption of value-added separability.

\(^2\) Implicit in the Domar’s aggregation procedure there is the assumption that all industries pay the same price for the primary factors of production. If this assumption does not hold, then aggregate productivity depends also on the reallocations of primary inputs among sectors (Jorgenson, Gollop and Fraumeni, 1987). However in this work we do not account explicitly for interindustry shifts of resources.
Measuring productivity

is crucial at the level of sectors, where the gross production would be preferable to take fully into account the important changes in the contribution of intermediate inputs, as those recently coming, for example, from the rise in outsourcing. At the aggregate level, the issue is less relevant since in a closed economy as a whole contribution from intermediate inputs cancel out and both measures of output deliver pretty similar results. In an open context the same holds true if value added is corrected for imported intermediate inputs, as it is the case with final demand model.¹

Industry output is defined as industry’s gross output excluding intra-industry deliveries (this definition is usually referred to, e.g. by OECD, 2001 and by the Bureau of Labor Statistics; as sectoral output). Accordingly with the above output definition, industry intermediate inputs are measured net of all intra-industry deliveries.

At the level of the entire economy, sectoral output coincides with final demand. Aggregate gross output might be sold to domestic industries or to final demand: being exports to foreign industries part of final demand, aggregate gross output less output sold to domestic industries is equal to deliveries to final demand.

Industry output is evaluated at factor costs, while intermediate inputs are evaluated at purchaser’s price. These latter are measured as Tornqvist Index of intermediate inputs purchased by each industry distinguished according to the producing sector.

1.2.2 Capital input²

Measuring capital input³ requires a two-stage process. First, it is necessary to evaluate the flow of capital services generated by each type of asset during the accounting period, then to aggregate them to get a measure of the overall flow of services provided the stocks. The aggregation procedure requires an estimate of the price of the services provided by each asset (i.e. their user-cost). The flow of capital services generated by a capital good are not (usually) observable; therefore they have to be measured by a proxy. The standard practice assumes that the service flows are in proportion to the productive capital stock at constant price⁴.

For an asset whose service life is T years (i.e. an asset that remains in use in the productive process for T years), the productive capital stock is defined as a weighted sum of real past investment of the last T years, where the weights reflect the efficiency decay of the asset as it ages (i.e. the fact that older assets are less productive than newer ones because of wear and tear). The estimation of the productive capital stock must deal with the fact that the actual service life for assets put in place in a given year will not be the same for all the assets. In order to account for the heterogeneity in the service lives we assume that retirements follow a given distribution around the mean service life.

¹ See OECD (2001) for a formal derivation of these results.
² This paragraph draws heavily on Iommi, Jona-Lasinio (2002)
³ With respect to capital input, the same approach is taken both for ICT and non-ICT capital, so in this paragraph we refer generically to capital input. In our empirical analysis the methodology outlined below is applied separately to ICT and non-ICT asset types.
⁴ Hereafter we will refer to the productive capital stock evaluated at constant prices simply as the productive capital stock.
The user-cost is an estimation of the marginal cost that the producer incurs (implicit or explicit) to use one unit of capital good for one period (cost due to interest rate, economic depreciation, capital gains/losses and fiscal factors). Under the initial assumptions it also coincides with the price of using one unit of capital and the value of its marginal product.

Capital input is obtained as a Divisia index of the growth rates of the productive stock of the various assets:

\[ \frac{\dot{K}_t}{K_t} = \sum_{i=1}^{n} v^i t S^i_t / S^i_t \]

where is the productive stock of asset type \( i \), \( v^i = (u^i S^i_t) / \sum_{i \in n} u^i S^i_t \) is the cost-share of asset \( i \) in period \( t \), \( u^i \) its user cost, and \( n \) is the number of asset types\(^1\).

Aggregation is obtained by mean of a changing weights index and cost-share weights for each asset type, where the cost refers to the cost of using the asset during the accounting period (and not to the cost of buying the asset in that period).

A changing weight index is preferred to a fixed weight one because it is not affected by the substitution bias.

Cost-share weights allow accounting for the heterogeneity in the marginal product of each type of capital good (and so for changes in the composition of aggregate capital stock). In fact, under the standard growth accounting assumptions, differences in user-cost across assets reflect differences in their marginal products. So weighting the rates of change of the \( n \) asset-specific productive capital stocks with the relative cost shares is equivalent to assign a relatively larger weight to the rates of change of the assets that have higher marginal product and it allows to account for the substitution among different types of capital goods. Note that measures of capital input based on asset price weights (it is immaterial that they be the simple sum of the stocks of the various types of capital goods or changing weights indexes), fail to account for changes in the composition of capital stock.

Estimating productive capital stock requires that for each asset we choose a specific profile of the weights that reflect the efficiency decay of the asset as it ages (i.e. a specific age-efficiency function) and a specific retirement pattern (i.e. a retirement distribution and an average service life). For all assets we adopt the so-called hyperbolic pattern of efficiency decay, that is an age-efficiency profile that is concave towards the origin\(^2\) (i.e. efficiency falls at a rate that increases as the asset ages). The retirements distribution is modeled as a truncated normal distribution centered on the average service life.

The user-cost is measured as follows (see Esposito, Mastrantonio, 2004). The nominal rate of return is estimated from market interest rates (i.e., using productivity analysis jargon, it is an exogenous rate of return). More specifically \( r \) has been estimated as a weighted average of two interest rates taken as

---

\(^1\) Note that the cost refers to the cost of using the asset during the period and not to the cost of buying the asset in that period.

\(^2\) For a more detailed description of this issue and of the reasons for this choice see Iommi, Jona-Lasinio (2002)
a measure, respectively, of the cost of debt and of the opportunity cost implicit in internal sources of financing.

The other components of the user cost are asset-specific. The acquisition prices for new capital goods are calculated as the ratio of investment at current prices to investment at constant prices. The depreciation term at time t is obtained as the acquisition price of the asset at time t times the average depreciation rate during period t. The depreciation rate is obtained as the ratio of consumption of fixed capital at constant prices to net capital stock at constant prices. The capital gains-losses term is supposed to measure expected asset price inflation. Here we measure the change in the asset price as the first difference in investment deflator and expected inflation at time t is defined as a moving average of the changes in the asset price in the three years priors to t.

Both the productive capital stock and the user cost of capital have been estimated for nine types of capital goods that comprises non-residential gross fixed capital formation: six types of non-ICT capital goods (machinery and equipment; furniture; road transport equipment; air, see and rail transport equipment; non-residential buildings; other intangibles and services) and three ICT capital goods (hardware; communications equipment; software).

We have estimated productive capital stocks and their user cost of capital from 1980 to 2001.

However, since reliable statistics for inter-sector flows are presently available only since 1992 onwards, our analysis is focused on 1992-2003.

1.2.3 Labour input

Labour input is measured in terms of full time equivalent units. However, since data on hours worked have been just released, we will include them in our TFP measure in the very next future.

Labour quality differences are taken into account considering six types of professional qualifications and their related wages: registered employees by four status of employment (white and blue collars, clerks and apprentices), unregistered employees and self-employed. Basically, we assume that comparative income differences of each job position reflect all relevant skill diversities. However for a detailed description of labour input measure see Baldassarini, Di Veroli (2004).

1.2.4 Inputs shares

Labour cost is measured as the sum of compensation of employees from National Account and an estimation of the labour component of gross operating surplus. In order to estimate such component, we have followed the standard practice of assigning the same average compensation to self-employed as to employees.

Intermediate inputs cost is simply the sum of intermediate deliveries at current prices from other industries (both domestic and foreign).

Labour and intermediate input shares are the ratio of labour and intermediate input costs over the appropriate measure of output.
ICT and non-ICT capital shares are calculated as follows. First we have calculated the cost of capital services provided by each asset as the product of its productive capital stock times its user cost and we have summed up across ICT and non-ICT assets to get a first measure of the cost for ICT and non-ICT capital inputs. Then we have calculated a first measure of ICT and non-ICT shares as the ratio of their cost over output. Finally we have imposed that all factor shares sum to one by re-scaling ICT and non-ICT capital input shares.

2. Data description and implementation issues

In order to implement equations 2, and 5 we use the standard approach of approximating the continuous-time Divisia index by means of the Tornqvist index (i.e. continuous-time growth rates are approximated by first differences in the natural logarithm of the variable and continuous-time shares are approximated by the average of the share at time t and t-1). Being a changing weights index, the Tornqvist index is free of the substitution bias that affects fixed-weights indexes. Moreover it has the theoretically desirable property that it is an exact index for a translog structure of production (Diewert, 1976). Also the weights in equation 6 are measured as averages of two following years.

Our data refer to 29 sectors corresponding to the sub-sections of NACE Rev.1 classification for mining and manufacturing industries and to the sections for the other sectors, with the exclusion of Public administration and defence (section L) of Private households with employed persons (section P) and Letting of own property (group 70.2). Aggregate data are consistent with industry-level data.

3. Empirical results

Chart 1 shows the rates of growth of Italian per capita GDP and Total factor productivity over 1992-2003. As expected, TFP closely tracked the cyclical evolution of GDP over the whole period. At the beginning of the nineties, when the Italian economy entered the upturn that followed the 1992 currency crisis, TFP grew at remarkable rates, 2.7% in 1994 and 2% in 1995. But, with the exception of an increase of 1.4% in 2000, since 1996 TFP recorded fairly low rates of growth. During the last three years the poor performance of TFP can be attributed to the significant slowdown of TFP in the manufacturing sector, that recorded an average yearly rate of growth of -0.7%, and in some of the service sectors (for example, Trade and repairs).

Table 1 displays the average (yearly) contributions of factor inputs and TFP to output growth over 1992-2003. Across the whole period, real business output rose at an average rate of 2.08% per year. Output growth has been driven mainly by capital accumulation that contributed for 0.6 percentage points, of which 0.17 can be assigned to ICT capital. Intermediate inputs and TFP contributed for 0.5 percentage points each, while labour input accounted for modest 0.25 percentage points.

During the first half of the nineties, TFP was the main driving force of the Italian output growth, accounting for 1.4 percentage points. Labour input contributed for -0.5 percentage points to growth, while capital accumulation accounted for 0.42 percentage points (of which 0.07 pp are from ICT capital). In 1996-2000, instead output growth was mainly driven by a more intensive use of productive inputs associated with a rather small contribution of TFP. In the following years, primary factors of production continued to be the main drivers of output growth while TFP accounted for -0.8 percentage points.
Over the entire period input growth has been the source of nearly 60% of Italian growth, and TFP has accounted for approximately 40%. Nevertheless, looking at TFP and labour input in the sub-periods we notice that the contribution of TFP is higher the lowest the contribution of labour. This is particularly evident in the first half of the nineties when output growth was 1.7% and TFP is responsible for 1.4 percentage points while labour accounted for a negative share (-0.6%). Further, in 1996-2000, as the share of labour increased (0.7%) that of TFP decreased (0.7%). Therefore during the downturn (1992-96) Italian economic growth was driven mainly by TFP while through the upturn it was driven by factor inputs. As it will be evident in the following section, this result is supported also by the analysis at the industry level.

Across the whole period, the contribution of ICT capital good to output growth has been small if compared to those of the other factors, but it should be noticed that the contribution of ICT as a whole increased over the all period recording a small slowdown in 2000-2003.

In the second half of the decade, ICT capital services accounted for 0.25 percentage points to output growth that is a considerable result if compared with the same figures for 1992-96 (0.07 percentage points). Among ICT capital goods software is responsible for the largest share.

However, ICT capital, though its relative weight is small if compared to that of the other factors of production, acted as a stimulus for the Italian productive system. Actually, in the 1990s, even if non-ICT productive capital stock increased relatively less than ICT productive capital stock, the contribution to the growth of output of both types of capital grew at the same pace all over the period (see Bassanetti et al, 2004).

Another important feature of our results is that when the ICT contribution to output growth is low TFP is high and vice versa in the next period. Therefore the Italian data do not support the idea that the higher the ICT capital accumulation the higher TFP growth. Indeed huge ICT investment might require a time lag to have a positive impact on productivity growth. Further, ICT Italian production is low if compared to other European countries. Consequently, given that a fundamental contribution to economy-wide productivity growth comes from ICT producing industries, Italy could close the growth gap with other European countries trying to stimulate ICT production.


As clearly stated in Jorgenson and Stiroh (2000), ‘aggregate TFP gains reflect the evolution of the production structure at the plant or the firm level in response to technological changes, managerial choices, and economic shocks. These firm and industry – level changes then cumulate to determine aggregate TFP growth’. It is therefore extremely important to look at industry data to quantify the contribution of the productivity growth of each industry to aggregate TFP growth.

As it has been stressed in previous sections, productivity growth differs broadly among sectors. Indeed, positive productivity growth in one sector can be offset by negative productivity growth in another. The identification of the contribution of each sector to aggregate productivity growth is therefore a fundamental step to isolate the underlying sources of gains in productivity and to a better understanding of the forces driving economic growth. Further, this analysis allows to see if the

---

1 See Iommi et al (2002).
modest productivity growth of the Italian economy during 1992-2003 is a broad-based phenomenon that reflects losses in a majority of industries or if it depends on the reduced performance of a small but important group of industries.

In this section, we decompose the Italian economic growth across industries, we quantify the contribution of each industry to aggregate productivity growth and we examine the productivity performance of the individual industries that either produce ICT, use ICT or are relatively isolated from the ICT revolution.

As a first step, by means of equation (5), we measure the sources of economic growth for individual industries and we analyse the role of ICT capital accumulation as a driving force of industry growth. Table 2 reports estimates of the components of equation (5) for the period 1992-2003. For each industry we display the growth in output, the contribution of each input, productivity growth and the Domar weights calculated from equation (6). Excluding Agriculture and Fishing, industry productivity growth was highest in Electricity, gas and water and in Financial Intermediaries (FI) at 1.34% and 1.23% per year respectively. It is interesting to note that the rapid productivity growth experienced by FI was counterbalanced by a very low Domar weight of 0.09%, thus implying a contribution of 0.14 percentage points to aggregate TFP. On the other hand, the Trade sector recorded slower productivity growth, only of 0.01%, and even if it is the largest sector in terms of Domar weight (0.22), it contributed only for a small 0.002 percentage points to aggregate TFP. Industries where productivity growth is negative make negative contributions to economy-wide TFP growth. Mining industries accounted for the highest negative share lowering aggregate TFP growth by – 0.02.

The highest contribution of ICT capital is in Real estate and Financial intermediaries, where the new technologies accounted for 0.72 and 0.38 percentage points respectively of industrial output growth.

Charts 2-5, display the Domar aggregation from equation (6) for the whole period and for three subperiods. Summing up across 31 industries TFP annual growth rate is 0.72% for 1992-2003. Chart 2 shows how much the contribution to TFP growth differ across industries due to the variation in both industry productivity growth and Domar weights.

In conclusion, from our results it is evident that service sectors were important sources of TFP growth over the period. In particular, Transports and communications and F.I. played a fundamental role as a driving force of Italian productivity growth during the nineties. For what concern F.I., this result can be assigned to the fact that FI invested heavily in the new technologies, it recorded the highest contribution of ICT capital accumulation to its own output growth and finally it is the sector that contributed relatively more to aggregate TFP growth.

Further decomposing sectoral contributions to aggregate TFP growth into three sub-periods we can appreciate the dynamics of the composition of the above contributions. In 1992-96, service sectors were the driving forces of aggregate TFP growth. Financial intermediaries accounted for the largest share (3.1 percentage points), followed by Trade and repairs contributing for 1.05 percentage points (Chart 3). In the second half of the nineties, services remain the major source of aggregate TFP growth, but Transport and communication gained the pool position while Trade accounted for -0.1 percentage points to aggregate TFP (Chart 4). In 2000-2003, most sectors contributed negatively to aggregate TFP indicating the deep downturn of the Italian economy over 2002-2003 (Chart 5).
These outcomes prove the heterogeneity of sectoral contributions to TFP growth thus confirming how it is important to examine each industry individually to deeply understand productivity growth.

4. Concluding remarks and future research avenues

The paper investigates the growth performance of the Italian economy and its main determinants over the last decade, with special attention to the role of ICT. Although theoretically founded on standard growth accounting, the research has put forward a first implementation of the service flow approach in measuring capital input in Italy. To this purpose, a full data set has been preliminarily developed covering productive stock for a large variety of capital assets – included software, hardware and communication equipment – and the respective user costs; data range from 1992 to 2003, and refer to a detailed disaggregation of productive industries.

On the average of the last ten years the ICT accumulation accounted for a quite low share of output growth, barely below 15%; the contribution became larger in the second half of the nineties, growing until 1999 and decreasing thereafter. The sectors which mostly contributed to aggregate TFP proved to be services like Financial Intermediation and Transports; on the other side sectors like Energy and Constructions caused a huge drag. ICT played a major role in Real Estate and in Financial Intermediation sectors, was virtually neutral in Chemicals and Transports.

At this stage results are still experimental and wait for confirmation through methodological refinements, mostly aimed at: including a correction for the quality of labour input consistent with that of capital, controlling for quality of assets in deflating current prices values, filtering out the impact of cyclical fluctuations. These issues are on the top of our agenda for future research.
References


Chart 1 – Per capita GDP and TFP: Italy 1992-2003 (percentage variations)

Table 1 – Contributions to output growth

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth rate of output*</td>
<td>2.08</td>
<td>1.78</td>
<td>3.39</td>
<td>0.78</td>
</tr>
<tr>
<td>Contributions from*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICT Capital</td>
<td>0.17</td>
<td>0.07</td>
<td>0.25</td>
<td>0.20</td>
</tr>
<tr>
<td>Non-ICT Capital</td>
<td>0.49</td>
<td>0.35</td>
<td>0.58</td>
<td>0.55</td>
</tr>
<tr>
<td>Labor</td>
<td>0.25</td>
<td>-0.57</td>
<td>0.65</td>
<td>0.82</td>
</tr>
<tr>
<td>Intermediate Inputs</td>
<td>0.59</td>
<td>0.48</td>
<td>1.10</td>
<td>0.06</td>
</tr>
<tr>
<td>TFP</td>
<td>0.57</td>
<td>1.44</td>
<td>0.77</td>
<td>-0.84</td>
</tr>
<tr>
<td>Income sharesb</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICT Capital</td>
<td>3.97</td>
<td>3.47</td>
<td>4.06</td>
<td>4.46</td>
</tr>
<tr>
<td>Non-ICT Capital</td>
<td>22.36</td>
<td>22.69</td>
<td>22.91</td>
<td>21.63</td>
</tr>
<tr>
<td>Labor</td>
<td>54.10</td>
<td>56.20</td>
<td>53.04</td>
<td>52.09</td>
</tr>
<tr>
<td>Intermediate Inputs</td>
<td>19.57</td>
<td>17.64</td>
<td>20.00</td>
<td>21.82</td>
</tr>
<tr>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>Growth rates of*$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICT Capital</td>
<td>4.23</td>
<td>1.97</td>
<td>6.27</td>
<td>4.60</td>
</tr>
<tr>
<td>Non-ICT Capital</td>
<td>2.19</td>
<td>1.56</td>
<td>2.55</td>
<td>2.54</td>
</tr>
<tr>
<td>Labor</td>
<td>0.52</td>
<td>-0.97</td>
<td>1.23</td>
<td>1.58</td>
</tr>
<tr>
<td>Intermediate Inputs</td>
<td>3.08</td>
<td>2.62</td>
<td>5.70</td>
<td>0.29</td>
</tr>
</tbody>
</table>

* Average annual rate of growth

b Average of the period
### Table 2 – Sources of economic growth by industry

<table>
<thead>
<tr>
<th>Sectors</th>
<th>Growth rate of output</th>
<th>Contributions from</th>
<th>TFP</th>
<th>Domar weight</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>K-ICT</td>
<td>K-NON ICT</td>
<td>L</td>
</tr>
<tr>
<td>Agriculture</td>
<td>-0.43</td>
<td>0.01</td>
<td>0.42</td>
<td>-2.04</td>
</tr>
<tr>
<td>Fishing</td>
<td>0.00</td>
<td>0.01</td>
<td>-0.10</td>
<td>-2.53</td>
</tr>
<tr>
<td>Mining of energy producing</td>
<td>-0.37</td>
<td>0.07</td>
<td>3.58</td>
<td>-0.18</td>
</tr>
<tr>
<td>Other mining</td>
<td>-0.92</td>
<td>0.01</td>
<td>0.00</td>
<td>-0.62</td>
</tr>
<tr>
<td>Food, beverages and tobacco</td>
<td>1.32</td>
<td>0.05</td>
<td>0.23</td>
<td>0.00</td>
</tr>
<tr>
<td>Textiles</td>
<td>0.27</td>
<td>0.11</td>
<td>0.08</td>
<td>-0.74</td>
</tr>
<tr>
<td>Leather</td>
<td>-0.79</td>
<td>0.10</td>
<td>0.03</td>
<td>-0.55</td>
</tr>
<tr>
<td>Wood</td>
<td>2.71</td>
<td>0.03</td>
<td>0.02</td>
<td>-0.04</td>
</tr>
<tr>
<td>Pulp and paper</td>
<td>2.22</td>
<td>0.21</td>
<td>0.40</td>
<td>0.15</td>
</tr>
<tr>
<td>Coke and refined petroleum</td>
<td>-0.71</td>
<td>0.03</td>
<td>0.36</td>
<td>0.00</td>
</tr>
<tr>
<td>Chemicals</td>
<td>1.24</td>
<td>0.03</td>
<td>0.11</td>
<td>-0.02</td>
</tr>
<tr>
<td>Rubber and plastic</td>
<td>3.03</td>
<td>0.10</td>
<td>0.24</td>
<td>0.30</td>
</tr>
<tr>
<td>Non-metallic minerals</td>
<td>2.20</td>
<td>0.06</td>
<td>0.30</td>
<td>0.50</td>
</tr>
<tr>
<td>Basic metals</td>
<td>2.29</td>
<td>0.05</td>
<td>0.09</td>
<td>0.10</td>
</tr>
<tr>
<td>Machinery and equipment</td>
<td>1.95</td>
<td>0.08</td>
<td>0.13</td>
<td>0.14</td>
</tr>
<tr>
<td>Electrical equipment</td>
<td>1.33</td>
<td>0.12</td>
<td>0.22</td>
<td>0.09</td>
</tr>
<tr>
<td>Transport equipment</td>
<td>1.64</td>
<td>0.02</td>
<td>0.05</td>
<td>-0.42</td>
</tr>
<tr>
<td>Other manufacturing goods</td>
<td>2.39</td>
<td>0.03</td>
<td>0.09</td>
<td>-0.02</td>
</tr>
<tr>
<td>Electricity, gas and water</td>
<td>3.15</td>
<td>0.16</td>
<td>0.39</td>
<td>-0.58</td>
</tr>
<tr>
<td>Constructions</td>
<td>0.49</td>
<td>0.04</td>
<td>0.22</td>
<td>0.11</td>
</tr>
<tr>
<td>Trade</td>
<td>2.08</td>
<td>0.29</td>
<td>0.63</td>
<td>0.08</td>
</tr>
<tr>
<td>Hotels and restaurants</td>
<td>2.08</td>
<td>0.05</td>
<td>0.44</td>
<td>0.77</td>
</tr>
<tr>
<td>Transports</td>
<td>3.21</td>
<td>-0.02</td>
<td>0.61</td>
<td>0.04</td>
</tr>
<tr>
<td>Financial intermediation</td>
<td>3.77</td>
<td>0.38</td>
<td>-0.13</td>
<td>0.11</td>
</tr>
<tr>
<td>Real estate</td>
<td>5.14</td>
<td>0.72</td>
<td>0.52</td>
<td>1.69</td>
</tr>
<tr>
<td>Public administration</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Education</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Health system</td>
<td>2.25</td>
<td>-0.04</td>
<td>0.03</td>
<td>0.34</td>
</tr>
<tr>
<td>Other public services</td>
<td>2.44</td>
<td>0.25</td>
<td>0.59</td>
<td>0.94</td>
</tr>
<tr>
<td>Housing services</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>International organisations</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Chart 2 – Industry contributions to Aggregate TFP growth: 1992-2003

Chart 4 – Industry contributions to Aggregate TFP growth: 1996-2000

Chart 5 – Industry contributions to Aggregate TFP growth: 2000-2003
DEPRECIATION MEASUREMENT IN THE EU COUNTRIES\(^1\)

Bernd GÖRZIG

DIW Berlin,
Germany

E-mail: bgoerzig@diw.de

Executive summary

This paper compares the methodologies applied in estimating CFC in the EU 15 countries. The comparisons are based on the UNECE survey conducted in 2003. The results of the survey are discussed and supplemented with additional information on the estimates. In addition, a model to estimate the implicit service life assumptions in the EU 15 countries is presented and implicit service lives of assets applied by the countries are estimated.

Depreciation levels in the EU 15 countries vary considerably. Although there are a number of methodological differences between the countries’ estimates of CFC, it is difficult to decide to what extent these influence the estimated CFC values. Despite methodological differences with respect to the discard function, the depreciation schedule and the degree of disaggregation, the resulting implicit service lives for total CFC in the countries are surprisingly close together. It seems that a reconciliation of the estimation procedure might be necessary for only a few countries.

To improve the comparability of CFC across countries, it would be helpful to apply more standardised methods. The most important strategy for standardisation is to harmonise the asset and industry breakdown, as is being done for instance in the EU KLEMS (2006) project. This harmonisation will make it easier to separate structural from methodological influences.

The harmonisation of service life assumptions will remain an open question. It must be conceded that existing differences in service life assumptions across countries might have historical origins, and it should also be kept in mind that the service life of an asset is determined by economic factors. These might differ between countries. Moreover, there might also be differences in quality across countries for the same type of asset.

For industry-specific productivity analysis, it is important to make it clear to what extent model-based CFC is calculated using series of

— New investment, or
— GFCF, which includes purchases and sales of used assets, or

\(^1\) Research for this paper was funded by the EU’s Sixth Framework Programme under the EU KLEMS project.
Net additions to the capital stock of industries, which also include changes in industry affiliation and premature scrapping.

To assess adequately the influence of capital movements on the contribution of capital services to the output of industries, exclusive use of a PIM model does not seem to be an adequate solution. For industries, within intervals of about five years, model results should be combined with direct observation by a survey of capital stock and its age structure.

1. Overview

In common measures of TFP, depreciation is one of the most important variables for assessing the contribution of capital to overall productivity growth. An empirical comparison of the magnitude of depreciation or, as it is called in ESA’95, consumption of fixed capital (CFC) in the EU countries can help to understand the differing significance of CFC in the European countries. Figure 1 shows that the relationship between depreciation and net domestic product at factor costs varies considerably between the EU countries. It is low with an average of 11 per cent in Greece and more than twice as high at 24 per cent in Finland. In most countries, the ratio is quite stable over time. There are few exceptions, such as the case of Finland, where the ratio declined sharply in the 1990s.

Two disciplines can compete to explain these differences in the CFC ratios:

— Economists, and
— Statisticians.
Economists would strive to explain the differences in CFC ratios by the economic, demographic and geographic situation in the respective countries. However, the focus of this paper is on the question of the extent to which these differences might be attributed to different methodologies applied by the countries in measuring CFC.

CFC is an important variable not only in productivity analysis; it also has a direct impact on relevant political decisions. CFC is a component in calculating value added for the institutional sector of general government and for dwellings. Up to one third of total CFC can be attributed to these activities. Differences between countries in the methodologies used for measuring CFC therefore might have a direct influence on the level of GDP and in a number of important EU contexts, such as:

— Contributions to the Community budget,
— Assessment of the deficit criteria for public budgets.

If it is assumed that the deviations in CFC ratios result exclusively from methodological differences in calculating depreciation for non-market producers, it might be interesting to calculate GDP values under the assumption that the European average of the CFC ratio is applied for all countries. In this case, GDP in some of the countries (Greece and Ireland) could be up to six per cent higher than currently calculated. Conversely, GDP for Denmark and Portugal would be about three per cent lower (Görzig, 2006).

Alternatively, it could be assumed that deviations in the observed CFC ratios from the value of the EU average can be attributed entirely to measurement differences for market producers. In this case, the share of net operating surplus (including mixed income) in GDP would change considerably for some countries. The operating surplus would have to be reduced by two-digit rates in Ireland, Greece, Spain and Germany; it would be considerably higher in Denmark, Portugal and Finland. In most calculations of TFP, the overwhelming share of operating surplus (including mixed income) is taken as capital income. If it is assumed that deviations in CFC ratios are due to methodological differences in calculating depreciation for market producers, the weights of capital services in explaining total production would change considerably, and consequently so would the outcome of the TFP measures.

These examples make it clear that CFC measurement can influence output as well as capital input in the TFP measure. A qualified assessment of TFP measures as performed for instance in the EU KLEMS project makes it necessary to analyse the methodological differences between CFC calculations in the European countries.

2. Methodological differences in calculating CFC

The following section is based mainly on the results of an UNECE survey on this subject addressed to the National Statistical Offices of the UN members. The survey was conducted in 2003 and published in 2004 (UNECE, 2004). The responses by the National Statistical Offices to the questionnaire are very heterogeneous. It has been attempted to make them comparable in a systematic way. However, it is possible that some of the descriptions in this paper might not fully reflect the intentions of the respondents.
The following results of the UNECE survey will be discussed:

— Sources of depreciation and capital stock,
— Models applied for capital stock estimates,
— Depreciation schedules applied,
— Service life differences.

2.1. Sources

Two main sources for depreciation were reported:

— Direct observation,
— Estimates based on capital stock data.

Direct observation of depreciation can be based on a survey or on administrative data, such as tax balance sheets. Directly observed depreciation values are generally considered to be not in line with the valuation principles of the National Accounts (OECD, 2001). In general, depreciation values taken for instance from book-keeping information provided for the tax authorities or from a survey are valued at historical costs instead of current replacement costs. In addition, in many cases, observed values of depreciation do not reflect the requirement of the National Accounts that they be based on the foreseeable economic service life for the assets in question. Instead, extraordinary depreciation is often also included, resulting either from tax allowances or from catastrophic losses.

As far as it is known from the UNECE survey, out of the EU 15 countries, only Ireland has based its depreciation figures on tax reports. It also seems that a majority of the transition countries among the new EU Member States currently calculate depreciation figures based on direct observations. These are either survey-based or, in some cases, based on a kind of prescribed service life for governmental institutions.

According to ESA’95, depreciation should be based on capital stock data (ESA 6.04). Sources for capital stock figures can be

— Direct observations, and
— Cumulation methods.

Direct observations should according to ESA’95 have the first priority in calculating capital stock figures for a country. In practice, it seems that this method is rarely applied. Among most of the EU 15 countries, direct observation by surveys is reported for some assets by the Netherlands and France. The use of administrative records is reported by Denmark. The UK seems to have made some experimental surveys (West 1998), but this was not reported in the UNECE survey.

Administrative data frequently consist of physical data, generally used for long living assets such as dwellings and roads, which are valued at current replacement costs in order to calculate estimates for gross capital stock. To calculate net stock, additional assumptions are necessary. Major difficulties of this method, which also may be called the Quantity x Price (QxP) method, are the assessment of different qualities and the choice of adequate replacement prices. If conducted annually or at least
periodically, an advantage of the method is that movements in capital stock from one industry to another are recorded adequately.

What is commonly called the balance of fixed assets (BFA) approach is based on a survey in which enterprises report on capital stock at the beginning of the year, changes in capital stock due to net acquisition of new and used assets and reductions due to sales. In general, the capital stock reported is the net capital stock. If cumulated depreciation is reported, gross capital stock can be calculated from these figures as well. In many cases, units also report on depreciation and discards. Apart from the possibility that depreciation might be overestimated in the surveys, it also seems that reported discards are generally very low. It appears that assets which are to be discarded or scrapped have been sold by the reporting company and are reported as part of this item.

Table 1 – Methods applied by EU 15 Member States to calculate depreciation

<table>
<thead>
<tr>
<th>Country</th>
<th>Sources</th>
<th>Breakdown in Service Life Estimates</th>
<th>Model</th>
<th>Depreciation Method</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Assets</td>
<td>Industries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Austria</td>
<td>Net Stock</td>
<td>4 tan, 3 add</td>
<td></td>
<td>BEA</td>
<td>geo</td>
</tr>
<tr>
<td>Belgium</td>
<td>Gross Stock</td>
<td>4 tan, 3 add</td>
<td>&lt; 30</td>
<td>Discard function</td>
<td>lin</td>
</tr>
<tr>
<td>Denmark</td>
<td>Gross Stock / Admin. Reg.</td>
<td>&gt; 400</td>
<td>53</td>
<td>Discard function</td>
<td>lin</td>
</tr>
<tr>
<td>Finland</td>
<td>Gross Stock</td>
<td>5 tan, 1 add</td>
<td>31</td>
<td>Discard function</td>
<td>lin</td>
</tr>
<tr>
<td>France</td>
<td>Gross Stock / (Benchmark survey)</td>
<td>&lt; 30</td>
<td>Discard function</td>
<td>lin</td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>Gross Stock</td>
<td>&gt; 200</td>
<td>None</td>
<td>Discard function</td>
<td>lin</td>
</tr>
<tr>
<td>Greece</td>
<td>Gross Stock</td>
<td>4 tan</td>
<td>14</td>
<td>lin</td>
<td></td>
</tr>
<tr>
<td>Ireland</td>
<td>Tax Register</td>
<td>-</td>
<td>-</td>
<td>Not reported in ECE</td>
<td></td>
</tr>
<tr>
<td>Italy</td>
<td>Gross Stock</td>
<td>4</td>
<td>&lt; 30</td>
<td>Discard function</td>
<td>lin</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>Gross Stock</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Netherlands</td>
<td>Gross Stock / Benchmark surveys</td>
<td>5(11) tan 4 add</td>
<td>yes</td>
<td>Discard function</td>
<td>lin</td>
</tr>
<tr>
<td>Portugal</td>
<td>Gross Stock</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spain</td>
<td>Gross Stock</td>
<td>Tax Reg. / (Experts)</td>
<td>6 tan, 4 add</td>
<td>None</td>
<td>Discard function</td>
</tr>
<tr>
<td>Sweden</td>
<td>Net Stock</td>
<td>PIM</td>
<td>BEA</td>
<td>geo</td>
<td>Planned</td>
</tr>
<tr>
<td>UK</td>
<td>Gross Stock</td>
<td>Unknown</td>
<td>3 tan, 2 add</td>
<td>Some</td>
<td>Discard function</td>
</tr>
</tbody>
</table>

*Sources: UNECE, 200.*
In general, all data are reported at historical prices. In most transition countries a general revaluation of capital stock was made after the introduction of the market economy, such that the capital stock data of the surveys are currently valued at prices of the mid-1990s and thereafter. Although this approach allows revaluation at current replacement prices and the assessment of depreciation depends on the institutional structures of the country’s economy, such as the tax system, this kind of direct observation makes it possible also to assess movements in capital stock, an advantage which most of the model-based estimates lack.

If direct observation is not possible, ESA’95 (6.04) suggests a cumulation method, commonly known as PIM (Perpetual Inventory Method). It may be noted that the M in PIM stands for method, not for model. Companies apply this approach, while making occasionally a total inventory to check for unobserved losses. According to ESA’95, cumulation is done by adding and subtracting assets from a given capital stock. Given the definitions of stocks and flows of fixed assets in ESA’95, it can be argued that the survey-based approach as currently applied in some of the transition countries has all the features of a perpetual inventory method, while not being based on a formalised discard function.

Most countries which have reported using PIM apparently do apply a model, which is generally based on a formalised discard function. While the cumulation method, if properly performed, should yield the same results as direct observation, the main shortcoming when applying a discard function and similar approaches can be seen in the inadequate consideration of inter-industry movements of existing stocks. Not much is known about the extent and the way in which countries make separate calculations in the case of ownership changes (ESA 3.103). For some countries, there is evidence that the impact of capital movements between industries on the size of the capital stock of industries is not negligible, especially in the case of buildings.

In applying a formalised PIM model, service lives are assumed conceptually for homogenous types of asset. An industry-by-industry matrix of capital movements for every type of asset would be needed to calculate capital stock and capital services for industries. If service lives for a given type of asset are assumed to differ across industries, additional assumptions have to be made both for the assumed average service life and for the residual service life of the moved asset. A separate treatment and full coverage of acquisitions and sales of used assets is necessary to achieve a realistic assessment of the capital stock and capital services of an industry. It cannot be seen from the UNECE report to what extent the necessary data have been collected in the countries.

If the PIM model estimates are based on GFCF data, then in principle those parts of asset movements which are due to changes of ownership are included, at least on a net basis. However, for an adequate measurement of CFC, additional movements of capital such as premature scrapping of assets and changes in the industry affiliation (ESA 3.108) have to be included in the calculations. Premature scrapping of assets have been experienced in most of the transition countries, but also in certain industries in other countries, as a result of economic disturbances. Changes in industry affiliation arise, among other things, from mergers and splitting of enterprises. Many European countries have experienced such changes in the course of privatisation, mostly combined with the splitting of former state-owned companies into several smaller units with different main activities.

Insufficient coverage of capital movements is a severe obstacle for the analysis of productivity at industry level. If capital stocks and consequently capital services are based on formalised PIM models,
which for instance use only new investment as an input, then in the course of structural change of the economies, productivity and profitability of shrinking industries will be underestimated, and vice versa. While value added and employment will decrease in line with reduced activity in these industries, capital stock estimates as calculated by the PIM model will remain attributed to these industries, although in economic reality considerable amounts of capital stock have been moved to other industries. It is not known from the UNECE report to what extent countries are collecting data on changing industry affiliation and extraordinary scrapping. However, it can be expected that this part of capital changes primarily needs direct observation of the economy.

Coverage of capital movements might be less problematic at macro-economic level for bigger countries (Statistics Netherlands, 2005). However, for smaller countries, a considerable proportion of inter-industry movement of assets, especially equipment, is related to imports and exports. The comparison of big and small countries will therefore be hindered even at macroeconomic level.

Industry-based productivity analysis has to take into consideration the fact that both kinds of asset movement have an impact on capital stock for industries, so that CFC and consequently capital services for industries will be affected.

Survey-based information for some countries indicates that the acquisition of used assets might be of the order of one third of gross fixed capital formation. It is not clear from the UNECE survey how countries which apply a formalised PIM model treat sales and acquisitions of used assets and other volume changes. It seems that this item is mostly deemed to be negligible (OECD, 2001) or is primarily seen to be of importance for the institutional sectors (BEA, 1999; Schmalwasser, 2001).

2.2. Models

Three kinds of models are reported in the UNECE report. A PIM model based on a discard function is used by most of the EU 15 countries. The discard function chosen by most countries is bell-shaped, except in Spain, where a delayed linear function is used. Austria applies the BEA method; Sweden reported in 2003 that it intends to apply the BEA method in future.

While it can be shown that the specific shape of the discard function does not have so much impact on the outcome, it should be noted that in such cases, where countries are applying the discard function for highly aggregated assets, the function does not solely return accidental discards which can be insured against, as recommended by ESA’95 (6.04). The shape of the discard function in such cases also reflects the fact that in an aggregated discard function assets with different service lives are included.

2.3. Depreciation schedules

In line with the suggestions made in ESA’95 (6.04), most countries apply linear depreciation. Among the reporting EU 15 countries, Austria applies a geometric depreciation method and Sweden intended to do so.

ESA’95 suggests linear depreciation, but allows geometric depreciation if necessary. There are a number of reasons which suggest that geometric depreciation may be a more adequate depreciation
schedule. It is argued, for instance, on the basis of empirical and theoretical evidence that depreciation, e.g. the loss of value over time, is higher at the beginning than at the end of the service life of an asset. This may especially be true for assets which are exposed to keen competition and rapid changes in technology and/or taste.

On the other hand, for other types of assets, the contrary can be argued. Mainly for assets with a long service life, investors are frequently not motivated by an expected short-term return on capital, but by the expected increase in value of the asset. Investments in dwellings are a typical example where this kind of investment behaviour can be observed and explained theoretically (Görzig, 1998).

2.4. Service lives

Definition

A number of papers have proved that the level of depreciation depends heavily on the service life assumed for the asset. However, the notion of service life does not seem to be unique across the countries. In population statistics the average life expectancy is defined as the average age a member of a certain population can be expected to reach. It denotes the average number of years an individual will stay in the population. Formally, the average life expectancy can be calculated as the sum of all observed annual survival rates in a population. If direct observation of the age structure of a stock is available, the average life can be calculated from the observed values.

A model estimating the stock would use the observed average life as a parameter in calculating the survival rates of the members of the stock. This model should generate an age structure of the members of the stock which would return the observed value for the life expectancy. In most countries, models to calculate capital stock make use of this notion of service life.

Apart from this, a different notion of service life is in use for models which apply geometric depreciation. In the BEA calculations, the service life assumption is mainly a denominator, needed for calculating the depreciation rate (BEA, 1999; p. 32).

\[ \delta = \frac{R}{T} \]

The assumed declining balance rate \( R \) together with the service life assumption \( T \) determines the depreciation rate \( \delta \), which is in turn relevant for the size of the calculated net capital stock and the level of CFC. In this case, the service life assumption does not describe the average lifespan of an asset in the net capital stock. The service life assumption applied in a geometric depreciation formula defined in this way is not necessarily comparable with the notion of average service life in the sense of average life expectancy.

Sources

Most countries use several sources for their service life estimates. These are experts’ opinions for instance in the case of Spain, Italy, France and Germany. Information from the tax register is used by Spain, Germany, Belgium and others. Some countries use the results of other countries’ estimates, especially the US estimates. Frequently, use is made of administrative data (Germany, Denmark and
Other surveys of asset service lives (Cope, 1998) ask for data on more than two hundred different types of assets. The asset breakdown in the BEA estimates is about 150. For the German calculations, Destatis (Schmalwasser, 2001) uses more than 200 different types of assets. An idea of the magnitude of different service lives applied by firms might also be given by the fact that the German tables for tax service lives cover more than 2 000 different types of assets (BMF: AfA-Tabellen).

Another method of obtaining information on service lives is to ask for the age structure of the stock, preferably the gross stock (e.g. Czech Republic, Lithuania, Netherlands). In this case, values which are available in the companies’ accounting system can be aggregated and reported. Since values are normally reported at historic acquisition prices, additional assumptions and a revaluation at current replacement costs will be needed.

**Comparisons**

A varying degree of breakdown by asset and by industry makes it difficult to compare the assumed service lives across countries. In some cases, it might be possible to compare the assumptions for a specified asset for two selected countries. However, to arrive at an assessment of the impact of service life on the total deviation in CFC ratios, information of the aggregate impact of different service life assumptions is needed. A way to make this comparison would be to compare the assumptions at the level of higher aggregates, which have been used by all countries. For highly aggregated types of assets, information from the UNECE survey has been compared in Table 2. However, for most countries, only a range for the applied service life assumptions can be displayed. There is no information in the UNECE report on the weights individual service life assumptions have.

### Table 2 – Applied service life assumptions by selected EU 15 countries

<table>
<thead>
<tr>
<th>Country</th>
<th>Dwellings</th>
<th>Other Buildings</th>
<th>Transport equipment</th>
<th>Other equipment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>60</td>
<td>30 – 40</td>
<td>8 – 25</td>
<td>15 – 25</td>
</tr>
<tr>
<td>Denmark</td>
<td>75</td>
<td>50 (avg) – 59</td>
<td>14 (avg)</td>
<td>13 (avg)</td>
</tr>
<tr>
<td>Finland</td>
<td>50</td>
<td>20 – 70</td>
<td>6 – 25</td>
<td>5 – 27</td>
</tr>
<tr>
<td>France</td>
<td>25 – 60</td>
<td></td>
<td>7 – 15</td>
<td>9 – 21</td>
</tr>
<tr>
<td>Germany</td>
<td>74 (avg) 40 – 95</td>
<td>15 – 150</td>
<td>11 (avg) 8 – 25</td>
<td>13 (avg) 5 – 30</td>
</tr>
<tr>
<td>Greece</td>
<td>66 – 100</td>
<td></td>
<td>4 – 33</td>
<td>10 – 20</td>
</tr>
<tr>
<td>Italy</td>
<td>13 – 40</td>
<td></td>
<td>10 – 15</td>
<td>18 – 28</td>
</tr>
<tr>
<td>Netherlands</td>
<td>75</td>
<td>35 – 60</td>
<td>8 – 25</td>
<td>5 – 45</td>
</tr>
<tr>
<td>Spain</td>
<td>38</td>
<td>40 – 65</td>
<td>10</td>
<td>17</td>
</tr>
<tr>
<td>Sweden</td>
<td>75</td>
<td>25 – 80</td>
<td>2 – 35</td>
<td>10 – 40</td>
</tr>
<tr>
<td>UK</td>
<td>100</td>
<td>16 – 100</td>
<td>10 – 30</td>
<td>8 – 32</td>
</tr>
</tbody>
</table>

1 avg = weighted average over all assets.
2 UNECE (2004).
Weighted averages of the applied service life assumptions are given only in some cases. If the average service life of an asset is used to calculate linear depreciation values, then by definition the service life can be calculated by dividing gross stock by depreciation. In the case of geometric depreciation, this method is not applicable, since the service life definition in geometric depreciation schedules is not comparable with the one in the linear model.

A major improvement for obtaining better comparisons in productivity analysis across countries would be to have a more standardised approach in the asset and industry breakdown for all countries. This strategy of harmonisation has been chosen in the EU KLEMS project.

Model assessment of implicit service lives

Between levels and growth rates of investment and depreciation, there exists a reproducible non-linear relationship, depending on the assumed service life. One way to assess the implicit average service life applied in calculating total CFC would be to make use of an inverse PIM model, based on a standardised discard function. If time series of CFC and GFCF at constant and historical (current) prices for a given period are available, it is possible to calculate the respective shares of capital stock and depreciation which are the result of the published GFCF values. This calculation can be performed for different service lives. A gap will remain between calculated and published figures of CFC. To fill the gap for a given service life, the question can be asked how GFCF growth would have developed in these past years for which time series for investment are not made available.

The results of this approach are shown in Table 3. The OECD National Accounts provide time series on GFCF at current and constant (chain linked) prices for the EU 15 countries, covering the period from 1970 to 2004. In addition, CFC series at current replacement prices for the same period are available.

A simplified version of the capital stock model of the DIW (Kirner, 1970) has been applied to simulate the CFC figures for the countries in question. The DIW model estimates consistent values for gross and net capital stock at current (replacement) prices, acquisition prices and constant prices. Consistent values for discards and depreciation at current, acquisition and constant prices are also estimated (Görzig, 2006).

For every country the necessary time path for GFCF before 1970 can be calculated, given alternative service lives. It can be seen that, in general, the higher the service life assumption, the lower will be the resulting GFCF path before 1970. For some service lives, no realistic GFCF paths before 1970 could be found. This means that GFCF growth rates above 25 per cent would be needed to simulate the given CFC values. Growth rates of this dimension can also be a sign that the initial capital stock has been assumed to be very small. This may be the case in countries which experienced a rapid growth in GFCF in the observed period. In addition, it should not be forgotten that during the observation period CFC values in some years increased greatly due to inflation.

Although in most of the countries the period before 1970 was one of high capital accumulation, it should be noted that the calculated GFCF path is not comparable with possibly existing investment figures from other sources for this period. The calculated figures can only represent this changing part of the unknown GFCF values for the years before 1970 which is relevant for the CFC values in the
period thereafter. With an increasing backward distance from 1970, the weight of assets with a high service life increases, while assets with a low service life which have been part of GFCF in these years are not relevant for the period following 1970. Therefore, if the observed relationship between GFCF and CFC produces a very low service life the impact of pre-1970 GFCF values becomes very small.

Clearly, the calculated path of GFCF is also influenced by the divergence of the normalised model applied from the factual model in use for the respective countries. However, the general structure of the results will not be very different if major features of the model are changed, for instance by applying geometric instead of linear depreciation. The procedure described can be abbreviated by asking for the combination of service life assumption and GFCF growth in the past, which will give the best adaptation to the published time series of CFC. This non-linear problem can be solved by applying OLS to observed and simulated values in the framework of a Newton optimising procedure.

Table 3 – Assessment of implicit service life assumptions for published depreciation values – adaptation period for CFC 1970 – 2004

<table>
<thead>
<tr>
<th>Country</th>
<th>Growth rate of simulated GFCF at constant prices before 1970</th>
<th>Average service life</th>
<th>Average depreciation rate(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average service – years</td>
<td>Optimal adaptation of CFC series</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>Portugal</td>
<td>1</td>
<td>-1</td>
<td>-2</td>
</tr>
<tr>
<td>Denmark</td>
<td>7</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Belgium</td>
<td>5</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>UK</td>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Ireland</td>
<td>-</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>Netherlands</td>
<td>10</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Finland</td>
<td>7</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Spain</td>
<td>12</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>France</td>
<td>12</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Germany</td>
<td>12</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Sweden</td>
<td>-</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>Italy</td>
<td>-</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>Austria</td>
<td>7</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Greece</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Source: DIW Calculations.

\(^{a}\) No growth rates below 25% found.

\(^{a}\) At constant prices.

The results in Table 3 show a wide range of implicit service life assumptions: between 18 and 59 years. Some doubts may arise as to the extent to which CFC estimates of the countries have been made with a consistent model for the whole period since 1970. In addition, it should be noted that the data in the
OECD National Accounts statistics are qualified for most countries as estimated for the years before 1990. To what extent these estimates are based on a consistent CFC model for every country is not known. Additional country-specific information may be needed for better explanations. For Germany, for example, other sources suggest an implicit service life for the capital stock for the total economy of 32 years. However, in the data used here, the treatment of investment and depreciation for East Germany before unification may not have been solved consistently. Finally, as already mentioned, the service life assumptions applied in the countries may have changed over time. Therefore, additional simulations have been conducted for the periods 1980 to 2004 and 1990 to 2004.

The resulting pattern of service life differences between countries does not change much. There seems to be a stable group of countries with implicit service lives very close together, at around 30 years. These are the Netherlands, the UK, Ireland, Finland, Germany, France, Spain and Sweden, which represent quite a large share of the EU 15 economy. The highest service life was found for Greece (59 years), which was also an outlier with respect to the CFC ratios. For Finland and Ireland – opposite outliers with respect to the CFC ratio (Figure 1) – implicit service lives of similar dimension can be found (30 and 28 years), which are in the average range. Extremely low values seem to be applied in Luxembourg, Denmark, Portugal and Belgium. Comparatively high values were found for Italy and Austria.

Table 4 – Ranges of implicit service life for different periods

<table>
<thead>
<tr>
<th>Country</th>
<th>Average service life</th>
<th>Adaptation period for CFC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luxembourg</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>Denmark</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Portugal</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Belgium</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>Netherlands</td>
<td>29</td>
<td>27</td>
</tr>
<tr>
<td>UK</td>
<td>27</td>
<td>28</td>
</tr>
<tr>
<td>Ireland</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>Finland</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Germany</td>
<td>35</td>
<td>32</td>
</tr>
<tr>
<td>France</td>
<td>33</td>
<td>32</td>
</tr>
<tr>
<td>Spain</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>Sweden</td>
<td>36</td>
<td>34</td>
</tr>
<tr>
<td>Italy</td>
<td>38</td>
<td>37</td>
</tr>
<tr>
<td>Austria</td>
<td>38</td>
<td>38</td>
</tr>
<tr>
<td>Greece</td>
<td>59</td>
<td>59</td>
</tr>
</tbody>
</table>

Source: DIW Calculations.
Considering that the deviations of aggregate service lives are also influenced by compositional factors of GFCF, it can be concluded that, for most countries, service life estimates are fairly close together. However, even if the outlier Greece and the countries with extremely low service lives are ignored, the difference between a low service life of 25 years, e.g. for the Netherlands, and a high service life of 39 years, e.g. for Austria, is still quite high. The impact of this difference on capital stock estimates for these countries will be proportional to the relationship between service lives. This will exert a severe impact on productivity and profitability estimates.

For analytical purposes, it might be an interesting exercise to harmonise service life assumptions across the countries. This is being done for instance in the analytical module of EU KLEMS. This kind of simulation can help to identify the impact of different service life assumptions on TFP estimates. Furthermore, it must be conceded that differences in service life assumptions across countries might to some extent have historical origins and might not reflect factual differences. However, it should also be kept in mind that the service life of an asset is determined by economic factors. These might be different across countries. Moreover, there might also be differences in quality across countries for the same type of asset. Therefore, depending on the type of asset, service lives might differ. It could be imagined that internationally traded goods would have a more common service life across countries than some country-specific assets.

3. Conclusions

Depreciation levels in the EU 15 countries vary considerably. Although there are a number of methodological differences between the countries’ estimates of CFC, it is difficult to decide to what extent these influence the estimated CFC values. Despite methodological differences with respect to the discard function, the depreciation schedule and the degree of disaggregation, the resulting implicit service lives for total CFC in the countries are surprisingly close together. It seems that a reconciliation of the estimation procedure might be necessary for only a few countries.

To improve the comparability of CFC across countries, it would be helpful to apply more standardised methods. The most important strategy for standardisation is to harmonise the asset and industry breakdown, as is being done for instance in the EU KLEMS (2006) project. This harmonisation will make it easier to separate structural from methodological influences.

The harmonisation of service life assumptions will remain an open question. It must be conceded that existing differences in service life assumptions across countries might have historical origins, and it should also be kept in mind that the service life of an asset is determined by economic factors. These might differ between countries. Moreover, there might also be differences in quality across countries for the same type of asset.

For industry-specific productivity analysis, it is important to make it clear to what extent model-based CFC is calculated using series of

— New investment, or
— GFCF, which includes purchases and sales of used assets, or
— Net additions to capital stock, which also includes changes in industry affiliation and premature scrapping.
To assess adequately the influence of capital movements on the contribution of capital services to the output of industries, exclusive use of a PIM model does not seem to be an adequate solution. Within intervals of about five years, model results should be combined with direct observation by a survey of capital stock and its age structure for industries.

References


Bundesministerium der Finanzen, 2006, AfA-Tabelle für die allgemein verwendbaren Anlagegüter.


Kirner, Wolfgang, 1968, Zeitreihen für das Anlagevermögen der Wirtschaftsbereiche in der Bundesrepublik Deutschland, DIW – Beiträge zur Strukturforschung, Heft 5, Berlin.

OECD, 1993, Methods Used by OECD Countries to Measure Stocks of Fixed Capital, National Accounts: Sources and Methods, No 2.


UNECE, 2004, UNECE/Eurostat/OECD meeting on national accounts, Survey of national practices in estimating service lives of capital assets.

CRITICAL ASSESSMENT OF THE CONVENTIONAL APPROACH TO PRODUCTIVITY MEASUREMENT RESULTING FROM THE SNA

Bohdan WYZNIKIEWICZ
The Gdansk Institute for Market Economics
Poland

Abstract

The paper discusses selected conceptual and methodological issues of productivity measurement resulting from the SNA 93. Application of various measures of productivity level and dynamics (Gross Domestic Product and Gross Value Added) leads to unclear economic interpretation. The author discusses the application of two price systems for productivity measurement: firstly to measure productivity on the production (supply) side and secondly, to measure productivity on the expenditure (demand) side.

Lack of theory of productivity measurement leads to lack of consistency in statistical information. Interrelation between single and multifactor approaches to productivity measurement is considered from the conceptual point of view.

Finally, productivity is confronted with the related variables (labour intensity and capital intensity) in the input – output framework. A proposal to derive productivity measure from the input – output system is presented.
Executive summary

The paper discusses various aspects of productivity measurement as seen from the viewpoint of statistical practice resulting from the SNA 93 standards and an economic interpretation of statistical information. The central methodological issue is where to draw the line between, on the one hand, at statistical agencies delivering simple measures of productivity and, on the other, trying to be involved in applying econometric tools in a harmonized and internationally accepted manner. The latter course seems inevitable in the light of the progress in theoretical works on the total factor productivity concept and new phenomena in the global economy. Labour and capital do not adequately explain the pattern of productivity growth in contemporary economies.

As to general measures of productivity from the sectoral, national and international perspectives, it is concluded that the supply side is better measured by gross value added expressed in basic prices, while the demand side is better measured by gross domestic product expressed in market prices. In other words, two different measures of productivity are recommended.

Another recommendation is to measure both labour and capital productivity by applying indices with a constant structure of input or output factors. In our opinion, such measures provide a better explanation of productivity change than the traditionally applied indices with a changing input structure. Another measure proposed for productivity measurement and analysis is to apply inverted labour and capital intensity coefficients derived from the input-output scheme.

1. Introductory remarks and general considerations

The paper has four sections. The first section raises general issues of productivity measurement. The second section discusses a problem of SNA 93 measuring productivity with two ‘rival’ measures at national level: namely, gross domestic product and gross value added. The third section touches on the issues of structural change within the single productivity factor concept and the issue of statistical support for a multifactor productivity approach. The fourth and last section considers whether productivity in product groups could be derived from the input-output system in a similar way to that in which labour intensity and capital intensity are treated.

Productivity is a relative number, commonly defined as the ratio of a volume measure of output to a volume of input use (OECD, 2001). When it comes to the dynamics of the productivity, a general indicator defined in such a way creates several conceptual and methodological problems:

1. Neither the nominator nor the denominators of the productivity ratio are precisely determined or defined. In practice, for various types of analysis and data presentation, the choice of these variables of productivity is not made in a unified manner. Just to measure factor productivity alone, output is represented by gross output, gross domestic product and value added and input is represented by labour and capital (in various forms).

1 The author wishes to thank Professor Leszek Zienkowski, advisor to the President of National Bank of Poland, for his helpful hints and inspiring discussions while writing this paper.
2. Several factors (or groups of factors) influence both the nominator and denominator of the productivity ratio. Since a single factor approach does not explain the mechanisms and factors underlying the dynamics and change of productivity, a multi-factor approach for measurement is widely used as a way of obtaining a fuller explanation of the phenomenon.

3. Services and the non-market sector. Measuring the productivity of some service sectors seems to be problematic, while in the non-market sector output valuation is simplified and so is productivity measurement. Market services are not homogeneous. Most of them could be treated as goods, but there is a problem with trade and financial services where the analogy with goods is more tenuous.

4. Aggregation at national level and international comparisons due to the lack of agreed approaches. National practices differ in the methodology of productivity measurement.

The list of methodological problems resulting from conventional measurement of productivity is much longer and will not be dealt with in this paper.

One of the crucial conceptual issues of statistics in relation to productivity is the extent to which statistics should describe and explain productivity and its dynamics. Is it enough to provide a simple statistical measure (information – size and dynamics) or a set of measures (applicable for various objectives) of the phenomenon, or should statistics go further by breaking down (i.e. by calculating shares or contributions) the various factors (in accordance with economic theory or in an agreed manner) that determine the level of and changes in productivity?

The pressure for developments which make use of the latter approach is evident. There is an undisputable need to discuss what statisticians can and should do in this field. Productivity measurement, especially from the perspective of international comparisons, therefore presents a real challenge for contemporary statistics.

Economics cannot help much conceptually in this respect since there are various serious approaches and schools of thought. Work done in the field of economic science on the one hand demonstrates and deepens the complexity of the problem and, on the other hand, delivers various measurement techniques that are often based on an advanced econometrics approach. In the light of such developments we have to consider whether and where to draw the line between, on the one hand, statistics understood as the delivery of numbers by national and international statistical agencies and, on the other, the advanced statistical approaches developed by academic and research institutions.

Econometric techniques applied to seasonal adjustments of industrial production or gross domestic product time series have pointed the way forward and productivity measurement will probably become the next major topic where classical statistical tools will be enriched by an econometric approach. Carrying the debate forward in a theoretically justified and harmonized manner would be an important step towards that goal. The significant feature of seasonal adjustments performed by statistical authorities is that they are official, which is not the case of academic estimates based on the same data but applying a different methodology.

The complexity of productivity measurement, or more precisely of productivity dynamics measurement, is due to developments in contemporary real economies. A simple production function cannot be
II

Measuring productivity

Firstly, in advanced economies, services have long made up the bulk of social output, while productivity is conceptually and traditionally tied to industrial production and based on two production factors. It is presumed that productivity measurement in the services sector is more problematic than in manufacturing. Services, as a sector, are not homogeneous. There are those who believe that the output of non-market services, as conventionally measured in the SNA 93, should not contribute as an output for social productivity measurement. In other words, this part of GDP should be excluded from productivity measurement since there is no correct method to measure the dynamics of productivity. Such an approach would not solve the problem of the productivity of non-market services, since there will be an obvious inconsistency between total GDP and that part of GDP used to measure productivity. Achieving acceptance for other proposals (e.g. use of a set of various indicators or application of productivity dynamics resulting from market services sectors) would be difficult. The weakness of such proposals is that they either do not express productivity changes correctly or that they are not applicable in an international context.

As for market services all of them except trade and financial services should be treated similarly to goods when measuring productivity. Output in both trade and financial services, though measured in a conventional way, creates fewer doubts or controversies than in non-market services.

Secondly, globalisation and the technological progress in some sectors in recent years have boosted the importance of productivity factors other than traditional ones (i.e. labour and capital). A multi-factor approach to productivity measurement is being adopted in order to try and resolve this new situation.

2. Productivity in the SNA 93

Although the SNA 93 (UN, 1993) does not include a separate section on productivity measurement, the term ‘productivity’ is nevertheless mentioned many times in several chapters. The productivity concept is not restricted to labour as an input, and it is mentioned in the context of gross fixed capital formation and stocks of fixed assets (SNA93 section 15.96) and in the context of land (SNA93 section 10.51). There is only one statement that resembles a definition: ‘output per job would be an excessively crude measure of productivity and total hours worked is the preferred measure of labour inputs for the System’ (SNA93, section 17.11). Other labour measures are ‘jobs which are contracts ... to perform work in return for compensation’ and ‘full-time equivalent jobs’ (SNA93 section 15.102).

It should be pointed out that the SNA 93 (UN, 1993) makes several references to productivity analysis. However, there are no guidelines or hints as to which categories of outputs or inputs are preferred or should be applied for specific types of analysis. The OECD Manual (OECD, 2001) provides an important theoretical and practical contribution to productivity measurement. However, the guidelines presented are geared more to research ventures than to the work of national statistical agencies.

The ESA 95 does not contribute directly to the concept of productivity measurement. Like the SNA 93, the ESA 95 has no section on productivity; in fact, productivity is not even listed in the glossary. References to productivity are concentrated in the section on non-market activities in the ESA 95.
The SNA 93 actually allows freedom in the choice of matters discussed. Various measures of output and various measures of input could be used to measure productivity in the SNA 93. The choice of the measure depends on the focus and the purpose of the comparison. Gross output, gross domestic product and gross value added are normally used to represent output, while – on the input side – different variants of labour, including total population and capital, are applied. Productivity presented as per capita GDP says more about societies’ well-being than productivity presented in terms of the output / input ratio. The most common indicators are gross domestic product or value added, which are given preference over persons employed or hours worked.

In the context of international comparisons, the number of hours worked seems to be a better yardstick for measuring labour productivity than the number of persons involved in production. The arguments in favour of such a measure are obvious: national regulations differ as regards the statutory length of working time, and the number of hours worked by employees differs from one country to another. In some cases, statistics do not distinguish between full-time and part-time labour. There are, however, well-known difficulties with statistical observation and reporting of worked time. For practical reasons the number of employed persons must be taken into account as an input for productivity measurement; however, the number of hours should be singled out as the preferred variable.

A similar reasoning could be applied to measures of capital productivity. A certain proportion of fixed asset capital is permanently involved in the production process, while another part may only be involved for a few hours a day during a single shift. To be consistent and precise, therefore, statistics should take into account coefficients that adjust the value of capital to the number of shifts in which fixed assets are involved in production processes.

In terms of economic significance and interpretation of data, a category of output chosen as a measure of productivity seems to be more meaningful and important than an input category. Output used to describe productivity expresses the outcome of the measurement of a nation’s or an industry’s economic activity, while input merely expresses a unit to which output is attributed. Statistical data used for productivity measurement from the output side are normally derived from national accounts. Data used as input may originate from other sources and may not be consistent with national accounting rules, even though the SNA 93 recommends full concordance in this respect.

Gross output does not seem to be the variable of choice for measuring productivity. Researchers and statisticians criticise it for double counting of output (Zienkowski, 2001, van Ark, 2004). Moreover, regardless of which measure is applied, productivity must be expressed in real terms.

In practice, the choice of both output and input of productivity often depends on the availability of statistical data. Gross domestic product and gross value added are used to show labour productivity. Both measures are used interchangeably. The SNA 93 does not indicate which measure should be preferred.

Gross value added valued in basic prices is a measure of output that is recommended by the SNA 93, and consequently to be preferred for productivity measurement – at both sectoral level and at the national economy level. Such an approach should be fully accepted since it reflects the actual outcome of production. This measure has a clear economic meaning and interpretation. In other words, gross value added is the correct statistical productivity measure of the supply (production) side of GDP.
From the final demand perspective, the preferred and theoretically justified measure of productivity is GDP valued in market prices. Basic prices are not applicable in this case. Productivity expressed in terms of GDP per capita is not ‘pure productivity’ since it relates social output not to labour inputs, but to the number of members of society or simply social beneficiaries, only a certain proportion of whom were directly involved in the production process.

**It should be recommended to apply two different statistical measures of productivity at the whole economy level in two different price systems, instead of one universal measure.** The aim of analysis should be to decide which measure is applied. If statisticians agree that such a twin-track approach to productivity measurement is acceptable from a statistical correctness point of view, the problem then arises as to how to present this situation to public opinion.

The two measures deliver different numbers and different dynamics. The difference in levels between GDP and gross value added can be easily explained, since both variables describe different phenomena. However, the differences in dynamics are not easy to interpret, as dynamics for both variables demonstrate similar tendencies. Appendix 1 presents the difference in dynamics in the case of Poland. For higher dynamics (over 3 percent), GDP growth rate is higher than the GVA rate by 0.2 to 0.3 percent; for lower dynamics (around 1 percent), the differences in rates are not significant.

It should be concluded that statistical presentation of productivity in two price systems is desirable, since the two measures explain two different phenomena: first, productivity viewed as output per unit labour input and, secondly, productivity as a social benefit from economic activity.

### 3. Productivity and its factors

This section discusses the issue of decomposition of the factors influencing change in productivity over time.

Decomposition of the factors influencing productivity change can be considered on two levels of analysis. On the first level, the aggregate productivity index will be decomposed, showing the shift of the structure of both output and input. On the second level, the conventional multifactor model will be discussed.

#### 3.1. Aggregated index of the relative variable

Productivity indicators are general aggregated indices of relative variables. Output measure is divided by input measure. A general aggregated index of relative variables with \( a = x \ b \) and \( b = x / a \) can be presented as follows:

\[
X_{1/0} = \frac{X_1}{X_0} = \frac{\sum a_1 b_1}{\sum b_0} = \frac{\sum a_1}{\sum a_0} \frac{\sum x_1 b_1}{\sum x_0 b_0} = \frac{\sum a_1}{\sum a_0} \frac{\sum x_1}{\sum x_0} \frac{\sum b_1}{\sum b_0}
\]
This index can be decomposed using independently a system of four pairs of products of indices. In each pair of products of indices, one is of the Laspeyres type and second is of the Paasche type. In the first two pairs the first index has a constant structure of variable \( b \) (from period 0 or 1) and the second index presents the impact of structure change in variable \( b \) on the ratio \( x \). In the second two pairs the first index has a constant structure of variable \( a \) (from period 0 or 1) and the second index presents the impact of structure change in variable \( a \) on the ratio \( x \).

This kind of decomposition makes it possible to indicate how both changes in input structure and changes in output structure influence the change of the aggregated relative variable.

One pair of indices decomposing the general aggregated index of relative variables is the following:

\[
X_{1/0} = \frac{\sum_{x=0}^{1} x b_0}{\sum_{x=0}^{1} x b_0} \cdot \frac{\sum_{x=0}^{1} b_1}{\sum_{x=0}^{1} b_1} \cdot \frac{\sum_{x=0}^{1} b_0}{\sum_{x=0}^{1} b_0}
\]

or in symbol presentation

\[
X_{1/0} = b_{-\text{cont}(0)} X_{1/0} \cdot b_{\text{inc}(1)} X_{1/0}
\]

In the first index decomposing the general aggregated index, the structure of variable \( b \) is fixed and variable \( x \) changes, while in the second index variable \( x \) is fixed and is weighted by a changing structure of variable \( b \). Interpretation of these indices delivers more information on productivity change than the general aggregated index, since the impact of structural change (in output or input) on productivity is eliminated.

The typical general aggregated indices of relative variable \( X_{1/0} \) presenting productivity dynamics published by statistical agencies ignores the change in structure of both output and input. The Laspeyres-type index of constant input structure (from the base period) delivers more information. The same type of additional information on the character of productivity change delivers a twin Laspeyres-type index with constant output structure (from the base period). The formula of the latter index is derived in the same way as for the former index.

It is relatively easy to construct a statistical paradox using fictional numbers. The paradox denotes a situation in which the general aggregated index of the relative variable changes in the opposite direction from the change of the majority or of all individual indices of dynamics. Such a trivial statistical paradox, which may also occur in economic reality, could be easily explained by applying the decomposition of the general aggregated index described above. Interpretation of such indices shows the effects of change in structure of both variables and the direction of their change.

Here again, it might be difficult to explain to the general public the logic of the index with constant structure of input.
3.2. Multifactor Productivity Growth Approach

As defined in the OECD Manual (OECD 2001) multifactor productivity (MFP) relates a change in output to several types of inputs. MFP is [often] measured residually as that change in input that cannot be accounted for by the change in combined inputs. Multifactor productivity is a synonym of total factor productivity (TFP). Calculations on MFP are performed using econometric techniques.

The multifactor productivity concept has a wide range of applications in research and economic analyses and, so far, is not a subject of a direct concern for statistical offices. However, statistical offices should be aware of data needs. The desired solution is to elaborate guidelines for statistical data used for MFP analysis purposes. Such data should be as close as possible to the SNA standards.

Theoretical works on MFP are intensive, especially as regards modelling. The list of variables used as factors of productivity is relatively long. Factors are selected depending on data availability and on whether they are statistically significant for the productivity functions. Natural factors used in analyses are capital and labour. Other factors are treated as a residuum or are combined in the residua. Intermediate inputs – energy, materials, services known as KLEMS (OECD, 2001) – are also treated as factors explaining productivity. Other factors include ICT capital deepening, and non-ITC capital deepening – the growth of ICT and non-ICT capital per unit of labour (Donselaar et al, 2004). More advanced techniques decompose labour into factors such as education, health, experience (Becker and Kevin, 1992). There are also attempts to extend analysis to human and social capital, knowledge and management techniques as factors of productivity growth (Fukuyama, 1999).

At the current stage of discussion the lack of uniformity is an advantage rather than a shortcoming of the MFP analysis, since it makes analyses more complex. It seems plausible to remark that labour input, as the most important and classical productivity factor, should be separated from other factors of productivity.

4. Productivity in the input-output scheme

The SNA 93 suggests ten topics which should be analysed within the input-output framework. Productivity analysis is not included in the list, since productivity is not normally derived directly from the input-output scheme. The reason why productivity has not been included on the input-output analysis agenda seems to be trivial. Input-output tables have important constraints. They are compiled with a significant delay, they are not produced every year in each country and they have a fixed and sometimes heavily aggregated sector breakdown.

It is well known that the input-output system delivers input / output coefficients via the so-called Leontief inverse matrix. The many possible applications of the system include labour intensity coefficients and capital intensity coefficients, both direct and indirect, generated by a single unit of output.

Labour intensity (labour / output ratio) is the inverse ratio to labour productivity. The question arises as to why this simple relationship is not used to analyse productivity, at least on the sector level. Inverted labour intensity coefficients, both direct and indirect (full), may provide impetus for an analysis of productivity from a perspective which has not been explored so far. There would be a special interest
in the interpretation of inverted full (direct and indirect) labour productivity coefficients. The same approach should apply to capital intensity coefficients and capital intensity.

If productivity analysis were oriented more towards input-output relationships, this would create greater interest in the input-output tables and there would be pressure to use I-O techniques more intensively. If this is the case, input-output tables will have to be used more frequently.

References


**Chart 1 – GDP and GVA dynamics in Poland**

<table>
<thead>
<tr>
<th>Year</th>
<th>GDP</th>
<th>GVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>4.5%</td>
<td>4.3%</td>
</tr>
<tr>
<td>2000</td>
<td>4.2%</td>
<td>3.9%</td>
</tr>
<tr>
<td>2001</td>
<td>1.1%</td>
<td>1.2%</td>
</tr>
<tr>
<td>2002</td>
<td>1.4%</td>
<td>1.3%</td>
</tr>
<tr>
<td>2003</td>
<td>3.8%</td>
<td>3.6%</td>
</tr>
<tr>
<td>2004</td>
<td>5.3%</td>
<td>5.1%</td>
</tr>
<tr>
<td>2005</td>
<td>3.4%</td>
<td>3.2%</td>
</tr>
</tbody>
</table>

*Source: Central Statistical Office of Poland*
Impact of hedonic price measurement on productivity
THE USE OF HEDONIC REGRESSION TECHNIQUES FOR THE QUALITY ADJUSTMENT OF COMPUTER EQUIPMENT IN CONSUMER PRICES INDICES AND IN THE PRODUCER PRICE INDEX: THE UK EXPERIENCE

David FENWICK1
Consumer Prices and General Inflation Division
Office for National Statistics
1 Drummond Gate, London SW1V 2QQ
E-mail: david.fenwick@ons.gov.uk

Abstract

Compilers of price indices in National Statistical Offices strive to differentiate quality changes from price changes so that official price indices reflect only the underlying price change. The importance of this is most pronounced in hi tech goods where technology change can be rapid and where models are frequently upgraded with significantly better features so that the underlying price change can be very different to the actual price change displayed in the shops. Moreover different methods of quality adjustment and differences in the precise application of the same method can produce different results. It is important that proper consideration is given to these issues. The impact on measured inflation can be significant and also on real GDP growth through the use of the same price indices as deflators in the national accounts. The importance for ICT deflation and for the measurement of economic growth is especially important given the dramatic productivity gains. However, the lower deflators that are generated do not proportionately increase GDP growth rates because of other factors that come into play via external trade deflators.

This paper focuses on conceptual and methodological issues relating to hedonic regression techniques. It looks at the use of hedonic methods for the quality adjustment of computer equipment and in particular, personal computers. It compares and evaluates the various techniques available and the detailed methods of application.

---

1 The views expressed in this paper are those of the author and are not necessarily those of the UK Office for National Statistics.
1. Introduction

One of the most challenging issues facing the compilers of price indices is the separate identification and valuation of changes in the quality of goods and services. This needs to be done so that quality changes can be uniformly excluded from the index to ensure that the movement in the latter reflects only the underlying price change.

The issue is becoming of increasing importance\(^1\) and the challenge is most pronounced in high technology goods such as computers, where the turnover of goods is so great that it is very difficult to produce a stable and continuous index without a rigorous method of explicit quality adjustment. Models are upgraded, replaced by manufacturers or cease to be available in shops and are replaced with ones with different specifications. The turnover can also be rapid. In addition, these changes have been introduced against a background of falling prices. In these circumstances the precise method of quality adjustment is particularly important in ensuring that the true underlying price change is measured.

The paper covers both the theoretical considerations\(^2\) and empirical considerations underlying quality adjustment. It describes the method of quality adjustment using hedonic regression\(^3\) which has been used in the UK Producer Prices Index (PPI), Consumer Prices Index (CPI)\(^4\) and Retail Prices Index (RPI) to adjust for quality changes in computers from the start of 2003. In each case the use of hedonic techniques covers both desktop personal computers (PCs) and laptops\(^5\) but the paper focuses on PCs. Common hedonic functions are used for all three indices. Comparisons are made with other quality adjustment methods and the pros and cons are discussed.

The paper first gives the historical background on why hedonics was introduced into the price indices published by the UK Office for National Statistics. It then reviews the conceptual basis of hedonics

---

1 The reasons are threefold: the scale and pace of quality change is increasing, particularly from technological innovations; the lack of consistency between the methods of quality adjustment applied by different national statistics institutes plus the lack of coherence in the methods used within national statistics institutes across different indices; the various methods produce different results and are becoming increasingly accessible.

2 It is necessary at the outset to put ‘quality adjustment’ in context of the conceptual foundations of price indices. In practice ‘quality change’ can take many forms and is sometimes unobservable or not measurable, for example, if it is something, such as ‘ease of use’, which is not associated with a change in physical characteristic. A utility-based framework helps to explain, in the latter circumstances, consumer preferences for one model over another but the concept of utility is not necessarily an appropriate conceptual foundation for quality adjustment as this depends on the conceptual foundation of the index itself. Thus in the realm of consumer price indices a valuation of an observable characteristic based on ‘utility’ is appropriate for a cost-of-living index which measures the cost of maintaining the same level of utility or standard of living but is inappropriate for a fixed basket price index which measures the cost of buying the same basket of goods over a period of time. The differences are often expressed in terms of the difference between a cost-of-goods index (COGI), or non-COLI, and cost-of-living index (COLI). It is important to have a clear conception of the measurement goal. A Cost-of-Goods Index (COGI), or non-COLI, is an index which measures the cost of a fixed basket of goods over a period of time, the relevant basket described in term of physical characteristics and the cost. A Cost-of-Living Index (COLI) is an index which measures the cost of maintaining a constant level of utility over a period of time. A COLI is based on the assumption of a rational consumer with perfect knowledge of the retail market and who optimises their purchasing behaviour. It requires an estimation of the welfare function of individual.

3 The application of hedonic methods dates back to the first half of the last Century, when it was applied in the USA to calculate rental values taking into account the number of rooms and other amenities.

4 Also known in Europe as the Harmonised Index of Consumer Prices (HICP),

5 Hedonics was introduced for laptops from early 2005. The same approach as PC’s was used, based on ‘predicted’ prices.
in the context of different indices and then describes the choices relating to practical implementation and how they were made. It then goes on to describe the way hedonics has been applied in the UK for PCs\(^1\) and the impact on published indices. Finally there is a brief discussion on some of the outstanding issues, including the use of price indices as deflators.

2. **Historical background**

2.1. *Implicit Quality Adjustment: the default method*

Personal computers were first introduced into the basket of goods priced in the UK’s consumer and producer price indices in 1996 when the volume of sales first warranted it. It was also the first occasion when an explicit method of quality adjustment had been used.

Previously all quality adjustment was through the use of implicit methods, notably by imputation. Under the method of imputation a new base price is calculated for an item by assuming that its price change from the base month up until that month equals the average change in the elementary aggregate for that item. Thus if the price of a replacement item is £500 and the elementary aggregate index for that item (calculated excluding the product in question) in the relevant stratum is 108.34, the new base price is:

\[
\frac{500.00}{108.34 \times 100} = 461.51
\]

This procedure ensures that bringing in the new item has no effect on the elementary aggregate for that item in the month that it is introduced\(^2\).

Imputation is based on the assumption of similar price movements, i.e. the price of the new item will have moved in line with the price of similar items. This is equivalent to assuming that the pure price change from the replaced to the replacement item is the same as the price change for all similar items observed in the two time periods with any remaining portion of the price change being attributed to quality changes.

---

\(^1\) The paper has focused on the quality adjustment of hardware (PCs and laptops) using the physical attributes of the machines to produce a quality-adjusted price index. However, the increasing expenditure over the last twenty years on PCs and in the latter part of the period on laptops has also been matched by increasing expenditure on software. Despite this and the fact that the utility obtained from computers is as much a function of the software as of the hardware, very limited work had been undertaken by the international community of price index compilers on measuring software prices. For instance, in the UK there is not a separate index for software, rather it is taken into account indirectly with the inclusion of the operating system in the hedonic function for the PC or laptop and via the inclusion of anti-viral software in the CD-ROM item included in the ‘CD and Tapes’ section which with others are combined to produce the All Items Price Index. One of the exceptions is a quality adjusted software index using scanner data which is produced by Statistics Canada. It is constructed using monthly scanner data on prices and unit values for various software packages purchased by government and business. It is based on the matched models principle i.e. where price data files are produced for products that are present for any two successive periods. Hedonic quality adjustment is not used because Statistics Canada concluded that it was impractical because of the large number of characteristics associated with such software. The development of quality adjusted price indices for software certainly warrants further investigation.

\(^2\) In the case where an outlet closes, or refuses to allow further price collection, all items priced there are dropped. In that case, a new outlet is selected in the same location and new base prices are imputed for items priced in the outlet as shown above.
2.2. The introduction of PCs into the basket based on explicit quality adjustment using option (production) cost

This method of imputation continues to be applied to most of the items priced and in most circumstances the assumption of similar price movements can be considered reasonable. However, in the case of PCs with its associated rapid market development, the assumption is less realistic because of the greater scope for under-adjustment for quality. A reduction in price will be identified as a deterioration in quality and hence in essence an additional ‘hidden’ price increase. Thus the practice of implementing hidden price cuts when introducing new models leads to an upward bias. Market research indicated an underlying fall in computer prices had prevailed in recent years, as new cheaper models are introduced with better features, and the decision was taken to adopt explicit quality adjustment using option costing.

The explicit method of quality adjustment chosen was option cost for the CPI and RPI and production cost for the PPI:

— **Option Cost.** In option costing the retail cost associated with a change in specification is obtained from the cost of purchasing the change separately or as an added option. Fifty per cent of this is added to the price of the original model to give a price comparison that is independent of any changes in quality. Fifty per cent of the cost was applied for a number of reasons. In part this is because 100 per cent option cost could lead to an over adjustment for quality change, due to the fact that the cost of buying features separately is generally greater than buying them as a package. In addition it was considered prudent, when personal computers were introduced into the HICP in 1996, to take a proportion of the option cost given that it was the first time that this method had been used in a consumer price index.

— **Production cost.** This method adjusts the base month price to reflect the additional cost of the change in specification using production costs supplied by the manufacturer. The use of production costs ensures that allowance is made for the cost of specification changes even where these may not be apparent or of direct relevance to the consumer. For example, the cost of better circuitry in a computer to make full use of change in the size of memory would not appear in the retail specification but is relevant for the costs of the producer. When a manufacturer gives an assessment of the cost of adding the improvement, the full cost is used to adjust the base month price. In practice it could be difficult to obtain fully reliable production costs from the manufacturer, and where the manufacturer was unable to supply the production cost of the specification change a 50 per cent option cost methodology was employed as a fallback- the 50 per cent being based on a necessary allowance to account for economies of scale in production and variable margins.
Impact of hedonic price measurement on productivity

2.3. The move from option costing to hedonic quality adjustment

A number of drawbacks are associated with this method which became increasingly apparent as it was being applied:

— There can be a strong element of subjectivity which can impact on the end result. It can be observed that the application of option (production) cost necessarily involves an element of judgement. This includes the determination of the appropriate proportion of the option cost that should be taken, the identification of the relevant subset of options to price from the complete set available to producers or consumers and unravelling their individual basic cost.

— It is not totally reliable. If applied mechanistically, option cost can produce counter-intuitive results. This can be illustrated by way of example:
  
  • Assume that in month 1 the PC being priced contains a Celeron 1.4Ghz processor and that in month 2 this becomes a Celeron 1.7Ghz processor with a specification improvement in the processor, increasing in speed from 1.4Ghz to 1.7Ghz. Also assume that at the time when this PC was priced, the cost of buying the Celeron 1.4Ghz processor as a separate component was greater than the additional cost of the superior Celeron 1.7Ghz processor (which could be the case if the demand didn’t warrant general supply, for instance because of the need for a new motherboard to utilise the improved performance of a Celeron 1.7Ghz). In this instance, the method of option costing would have spuriously concluded...
from the reduction in option cost due to the lower Celeron 1.7Ghz cost that there had been some deterioration in quality.

— It was relatively resource intensive to implement in terms of market research and gathering together the necessary information as well as the iterations required each time there was a change in specification in one of the PCs being priced.

In addition the trend in quality adjusted computer prices in the UK as represented by the UK PC index using option cost was very different from the equivalent US index which used hedonics and fell at a much faster rate. Superficially this was at odds with expectations flowing from the fact that PCs are put together locally from components bought in a highly competitive international market and so the only cost differences between the UK and USA would be transportation, labour and taxes.

It was against this background and the increasing evidence that hedonics might be a better method of quality adjustment for PCs that we began to investigate the use of hedonics and undertook a pilot exercise. Initial results indicated that hedonics provided a more objective and technically more rigorous methodology.

2.4. The results of the pilot exercise.

Chart 1 shows for 2002 a comparison of the pilot hedonic PC index for the RPI with:

— 50% option cost.
— 100% option cost.
— the PC index for the USA as compiled by the BLS.

Chart 1 – RPI PC indices 2002
It can be seen that the results were very encouraging from an intuitive perspective. In particular:

— the hedonic based PC index was close both to the US PC index and to the 100% option cost lines (this mirrored the findings from earlier pre-pilot work).
— the continuing fall in PC prices shown by the hedonic index, which is in line with market expectations. Trade analysts report at the time a continuing price war taking place, with the merger of Hewlett Packard and Compaq only adding to the downward pressure on prices.

In addition, an analysis over an extended period indicated that the method delivered a stable, consistent and reliable index.

For PCs it was clear that hedonics offered an alternative and better method of quality adjustment. Hedonics was introduced in the CPI and PPI for adjusting for quality changes in PCs from February 2003 and in the RPI from February 2004. The technique has been used for digital cameras, also from February 2004, and has subsequently been used for laptops since February 2005.

The next section considers the choices that need to be made when applying hedonics. The interpretation of hedonics is very much tied up with the theory of price indices and the underlying conceptual issues and there are a number of options which are available relating to the detailed implementation of the technique.

3. The choices that need to be made in applying hedonics

3.1. The conceptual basis of hedonics

At its most simplistic hedonics is a statistical technique that exposes nothing more than a purely empirical relationship between the observable attributes of a good or service and the price paid\(^1\). This is insufficient in itself:

— Any application needs to be soundly based in economic theory and matters of economic interpretation need to be addressed;
— It is economic theory set in the context of the definition and purpose of an index that determines precisely how hedonics is applied;
— Different application methods can impact on the end result.

Thus the issue of the conceptual basis of hedonics is far from a trivial one. For instance, in the field of consumer prices the underlying conceptual basis of a cost-of-living index (COLI) differs from a Cost-of-Goods (GOGI) (non-cost-of-living index (non-COLI)) and all statistics agencies compute one or the other and sometimes both, or most likely a mixture of the two\(^2\). Not surprisingly, the conceptual basis of quality adjustment in the two types of consumer price indices differs:

---

\(^1\) Schultze Panel

\(^2\) A non-COLI reflects the extra money required in some period to purchase the same basket of goods and services that could be purchased by a given sum of money in some earlier period. A COLI reflects the minimum expenditure that a consumer faces from rising prices to achieve the same level of utility as in some earlier period. The UK RPI and CPI are not COLIs but rather variants of a non-COLI.
— For a fixed-basket (non-COLI) index the objective of quality adjustment is to allow ‘current retail value’ re-pricing back to the identical bundle of characteristics associated with the original item that was priced before it was replaced or its specifications changed. This is done from the empirical relationship derived from the hedonic function. The latter simply reflects the retail value in the market place of each individual component of the bundle of characteristics. No judgement needs to be made about the state of competition and the impact on pricing policy or the associated impact on consumer utility. The residual issues are essentially statistical and measurement ones that will be dealt with later in the paper. For example, the inclusion of an extra 64Mb of memory in a PC is a real change to the quality of a PC, which would add to the price. The sole concern is the quality of the good, and its effect on price – which under competitive market conditions is a function of production costs and the price the market will pay.

— For a COLI the aim is to measure the marginal utility to the customer. In a perfect market this would be reflected in the marginal cost to the consumer, as it could be assumed that what the customer pays reflects the utility value. But in an imperfect market the market valuation will generally be higher than the marginal utility value. So the application of hedonics without some (technically rigorous) adjustment for market imperfections will lead to an over-adjustment for changes in quality and a downward bias in the index. In these instances it may well be better to restrict the hedonics to a narrowly defined set of more homogeneous goods where the market is known to be competitive or use the implicit ‘class mean’ method to quality adjust. The latter assumes that the pure price change from the replaced to the replacement item is identical to that for items that continue to be priced and have been identified as being comparable replacements for the disappearing good. In an imperfect market this method at least has the advantage of utilising the more ‘perfect’ competition which should exist between highly ‘comparable’ goods.

For the PPI the objective is to measure the extent to which the differences in characteristics explain the difference in price between the old and new good. In this respect there is an exact parallel, in the family of consumer price indices, with a non-COLI. For PPI construction the coefficients represent estimates of resource cost as opposed to user value for a CPI. A number of empirical studies have indicated that the choice of method adopted for quality adjustment can have a significant impact on the results. But as with CPIs a number of complications manifest themselves; some of which are associated with the part of the production chain for which the PPI is measuring prices. For instance PPIs may be based on output prices at basic values (i.e. at the point they leave the place of production) or on input prices at purchasers values (i.e. when they enter the production process). There are restrictive assumptions underlying the different quality adjustment methods applied to each type of PPI and in addition there is a reconciliation problem with national accounts at constant prices. The issues are further elaborated in the Producer Price Index Manual: Theory and Practice¹.

In summary, the concept underlying hedonics is that the cost of making a product such as a PC is a function of the characteristics of that PC and that similarly the purchaser buys the bundle of characteristics at market price. Hedonics then relies on an initial identification of all possible characteristics for which a market valuation should be computed so that the significant characteristics can be identified and valued. But the decision to use hedonics and the precise way in which the hedonic function is applied will depend on the conceptual basis of the index. In some instances hedonics may not be the preferred method.

It should be noted that when hedonics is the preferred method for estimating quality adjustment for non-comparable replacements, practical choices need to be made on the detailed implementation—most importantly the choice of model, whether to use weighted or un-weighted prices data in its estimation and how to apply the model, so derived, to adjust the price index for quality change. The following sections address each of these issues in turn.

### 3.2. The choice of model

The functional form of the model is both a conceptual issue and a practical one. In practice, practitioners go for log linear formulation. A log-linear model is a multiplicative model in the price dimension and therefore may be considered the appropriate choice in a market context where attributes associated with an item or service may add value to one another. General observation suggests that the log-linear model is a good reflection of pricing in the retail market for PCs. This is because the cost of adding a new feature tends to be related to the underlying quality and price of a machine. For example, the addition of a DVD drive to an expensive PC costs more than for a cheaper PC, because a higher quality drive will be included in the more expensive PC. A non-log-linear model would fail to pick up these complex pricing relationships. It should also be noted that the multiplicative relationships underlying a log-linear model are more robust to general changes in price, and so have a longer life span—this represents a practical advantage when resources are scarce.

The ONS adopted a log-linear model.

### 3.3. Weighting

Whether to use weighted or un-weighted prices data in a hedonic regression and if the former whether to use volume or value weights is a contentious issue that involves both conceptual and practical considerations. It is also one on which an extensive amount of literature exist but on which there is as yet no strong consensus. Not surprisingly, the decision on weighting can have a significant impact on the result. This is pointed out by Jack Triplett, who in his handbook on hedonics favors the use of weighted data in some circumstances and un-weighted data in other circumstances. His argument is that for the dummy variable approach (see next section) a weighted hedonic function using sales values is appropriate since the objective is to measure a weighted average of a pure price change. For other methods of application of hedonic quality adjustment techniques he sees the arguments as more complex but concludes that a weighted index does not necessarily require a weighted function. He argues that in theory sales weights should be applied if there is justification in assuming that market shares reflect market value. If the latter does not apply because, for instance, variations in sales volumes are more indicative of market niches rather than competitiveness then un-weighted data is more appropriate, in large part because in these circumstances low sales models are useful in estimating the shape of the hedonic function. He also advises on the estimation of weighted and un-weighted regressions followed by an examination of the prices and characteristics of PCs with low market shares to decide which is the most appropriate. His argument then develops further before moving on to econometric issues relating to efficiency of estimation prior to concluding that the decision about weighting is not clear cut and that the more important issue is to deal with missing variables arising from a miss-specification of the model

---

of which the difference between a weighted and un-weighted function may be indicative. Where weighting is appropriate he favours volume weights for robust estimation and coping with missing variables.

In practice in many instances weights are not available so the decision is hypothetical. The hedonic functions computed by ONS use un-weighted data and are based on a sample of about three hundred computers.

3.4. Alternative implementation methods

There have been a number of references earlier in this paper to the different methods of applying hedonic quality adjustment in practice. There are two main modes of application: the direct (time dummy) method and the indirect method:

— The time dummy method pools all available data on a good over several time periods, fits a single regression model and uses the coefficients on the variables relating to time as a direct measure of the price index. The index is produced from the difference in the time dummy variables from period to period, i.e. the price ratio net of the quality component captured by the characteristics variables. It assumes that between any two periods the ‘underlying’ prices of all models [actual or potential] differ by the same ‘average’ percentage. Since regression coefficients involving the characteristics are held constant across periods, it also assumes that changes in marginal cost ratios or in consumer demand patterns are negligible over a short time period.

This time-dummy approach is favoured by many academics because it has some good statistical properties when employed on a single dataset, most particularly, it is argued that is maximises the use of the available data and it calculates regression coefficients from a much larger dataset. But it is not generally preferred by National Statistical Institutes (NSIs) because in practice indices constructed by this method tend to lack stability when employed over several datasets and over a period of time. As a consequence of using data spanning several time periods, characteristics which are virtually obsolete in the current time period can still have a coefficient attributed to them. This is a particular problem in an area such as PC’s where the technological developments occur frequently. More particularly, the use of this approach raises a potential revisions issue – when results are estimated for a new month, new estimates are available for previous months. For instance, the consumer value for a 10% increase in hard drive in a PC is unlikely to be the same now as it was a year ago. As noted earlier, the method also needs to be based on weighted regressions to prevent potential bias but that the necessary data for PCs is at a premium and may not be readily available on a regular basis to statistical offices.

For all these reasons this was not the approach adopted by ONS.

— Indirect method. This is the approach preferred by the majority of NSIs and is the one which has been adopted by ONS. The particular application of the indirect method that is being used for desktop PCs and laptops in the PPI and in consumer price indices is based on the use of predicted
prices. Predicted retail prices, derived from single reference period hedonic regressions, are compared with actual retail prices to derive an indirect estimate of the effect of quality change. The latter is then used to adjust, post-hoc, the observed price difference between the outgoing and the replacement item.

The following is an illustrative example, taken from real life, of how the predicted price approach to hedonic based quality adjustment can be applied in a situation where an individual model was priced in Period 1, but could not be found in Period 2. The replacement is close in quality, but has a single change in specification – an increase in processor speed.

<table>
<thead>
<tr>
<th></th>
<th>Period 1</th>
<th>Period 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor</td>
<td>Pentium D</td>
<td>Pentium D</td>
</tr>
<tr>
<td>Speed</td>
<td>3000</td>
<td>3400</td>
</tr>
<tr>
<td>Hard Disk Drive</td>
<td>320gb</td>
<td>320gb</td>
</tr>
<tr>
<td>Monitor Size</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>Actual Price</td>
<td>£1176.87</td>
<td>£1127.25</td>
</tr>
<tr>
<td>Predicted Price</td>
<td>£1451.91</td>
<td>£1484.31</td>
</tr>
</tbody>
</table>

Change due to quality changes = \[
\frac{\text{Predicted price new model}}{\text{Predicted price old model}}
\] = £1484.31 / £1451.91 = 1.022

New model base price = Old model base price x quality change
= £1176.87 x 1.022 = £1202.76

Adjusted PC Index = £1127.25 / £1202.76 x 100 = 93.7
(= 69.92 / 74.63)

Unadjusted PC Index = £1127.25 / £1176.87 x 100 = 95.78

The predicted price approach is preferred by a number of statistical offices, including the ONS, for two main reasons:

— for models where attributes are related, the omission of one can cause missing variable bias which can be problematic for individual coefficients, though not for the regression equation.

1 The alternative ‘direct’ method of applying hedonics is the individual coefficient approach. This uses the coefficients from single-reference period hedonic functions to adjust post-hoc the observed price of the replacement item to impute a new base price. It is claimed to have a common sense appeal but relies heavily on the reliability and stability of the individual coefficients estimated in the hedonic function. For instance multi-collinearity could be problematic. An over-arching requirement for this approach to work is to have efficient estimates of the individual hedonic coefficients. In contrast, the predicted price approach, as it uses the ratio of the overall prices predicted by the hedonic equation for the replaced and replacement items to adjust post-hoc the price of the replacement item, it tends to be more stable and the outcome less susceptible to the impact of multi-collinearity on the individual coefficients. Thus from a practitioner’s perspective the application tends to be more robust. Both approaches require an efficient estimator. Where the data is orthogonal the two estimators will converge.
Impact of hedonic price measurement on productivity

as a whole and it is the regression equation as a whole which is used under the predicted price approach.
— where a log price model is used (as is normally the case with PCs), bias is more problematic for the individual coefficients than for the model as a whole (even after a correction factor of half the standard error is applied).

4. Live application of hedonics

4.1. Operational and practical procedures

The ONS experience indicates that the advantages of the predicted price approach are realised in practice and in addition provide a systematic mechanism for routinely testing the validity of models. The results also confirm that the advantages are resilient to differences in market circumstances.

There are two main operational procedures relating to the practical estimation of the hedonic function:

— Running the regressions.
— Determining when the resulting model is out-of-date and needs to be re-run.

4.2. Running the regression

Un-weighted ordinary least squares linear regression estimates\(^1\) of the relationship between price and ‘characteristics’ are run using list prices from outlet websites for about 350 models. Before carrying out the regression the data is ‘cleaned’ and any features of a PC that are present in less than 3 models are grouped together under an ‘other’ category. This process is carried out to ensure that unusual features are not given a misleading coefficient on the basis of 1 or 2 quotes which might undermine the adjustment made for quality changes.

A form of stepwise regression is used where scope is given to the analyst to use market knowledge to aid in the selection of variables to be used in the regressions and get a locally optimal solution that also satisfies the expectations of the market place and which does not rely on the assumption that the explanatory variables are independent\(^2\). An iterative approach has been successfully adopted which follows closely Statistics Canada practice. It is repeated until a satisfactory result is achieved and consists of the following steps:

---

\(^1\) As already mentioned in the UK a regression against the log of price (which yields a multiplicative quality adjustment) is used rather than a straight linear regression (which yields an additive quality adjustment). This is because results of a multiplicative model can be applied to a longer time period and the model itself is a better reflection of pricing in the retail market in that retail pricing structures tend to be more multiplicative than additive.

\(^2\) Stepwise regression is an automatic procedure which works to produce an optimal least squares regression solution based on the assumption that the explanatory variables are independent. If this is not true and there is collinearity between the independent variable, as is frequently the case in hedonic regressions for price indices, the technique can produce either sub-optimal solutions, or ones in which the coefficients do not reflect the price relationships seen in the market place. It is for this reason a total reliance on regression diagnostic statistics (such as partial F-values and adjusted r-square’s) should be deferent to a more systematic, multistage system relying on human intervention at key stages.
1. Choose base values for dummy variables – these are left out of the regressions;
2. Run regression with all variables, and produce correlation matrix;
3. Examine correlation matrix, and associated statistics, to inform on collinearity between independent variables. Look at collinear pairs, and decide whether the variables can be combined, or one variable dropped;
4. Examine residuals for evidence as to whether to make continuous variables discrete dummies and partial residuals for evidence of the appropriate relationship between the dependent variable and individual continuous variables;
5. Re-run the regression with the amended variables;
6. Remove variables with low t-values (at this stage t<1);
7. Re-run the regression with the remaining variables;
8. Progressively add and remove variables, until a combination is reached that produces the best fit, with coefficients in line with market expectations (in particular positive coefficients for included variables);
10. Either remove the outlier observations, or add extra attributes, to remove their influence; Then re-run steps 1 to 9;
11. Look for evidence of missing variables. If this is present revisit data source for added information;
12. Group together dummy variables representing levels within categories that are not significantly different;
13. Run final regression.

4.3. Determining when the model is out-of-date

In addition systematic procedures have been put in place to ensure that the hedonic function is updated prior to it becoming out-of-date and, conversely, to avoid resources being used for unnecessarily updating the hedonic regression each month. We concluded early on in our preliminary research that it would be inadequate to place sole reliance on market information on emerging changes in specifications and features, to trigger an update in the hedonic model. This is illustrated by Chart 2 which shows how it was possible to miss a required model update, in this case in August, despite close monitoring of the market.
Impact of hedonic price measurement on productivity

Chart 2 – RPI PC Hedonic Indices 2002

The green line shows the index where the regression model was not updated until October, alongside one updated in August (the orange line). The green line is significantly lower than that produced using a new model from August. This is consistent with a situation where the earlier model is over-valuing the costs of quality changes, and so over quality adjusting. An analysis of predicted versus actual price, given in Chart 3, is revealing and shows the value of using such an analysis on a regular basis. It can be seen that there is a clear pattern in the difference between predicted price and actual price for the early part of the year with the pattern changing in August with a large and positive difference, suggesting that there had been a change in the market that needs to reflected in an updating of the hedonic model which was at that point over-valuing certain features of PCs. Updating the model in August corrected this bias.
The systematic procedures put in place to overcome this potential for bias from out-of-date models involve a monthly analysis of predicted prices against actual prices, in addition to market information.

Exploiting the systematic relationship which exists between the prices, predicted by the regression models, and the actual prices of PCs is key in ensuring that the regression is successfully quality adjusting for changes in the sample. From this example a clear pattern emerges, with the predicted price gradually increasing over time in relation to actual price, with the model updates in May and October resetting the relationship but that there can also be market shocks which require their own model updates.

Chart 5 looks at the relationship between predicted and actual prices together with 99% confidence intervals surrounding each model update¹.

---

¹ In the usual formulation, the standard error of model prediction = \( \frac{1}{n} \times \sqrt{\sum SE^2} = \text{STDI} \)

However, as the model produces predicted value of log price, the formulation becomes

\[
\text{Standard Error Price} = \text{Predicted} \times \sqrt{\left(\text{STDI}^2 - 1\right)} \\
\approx \text{Predicted} \times \text{STDI}
\]
Several points can be gleaned from this analysis. Most particularly, it emphasises the systematic nature of the change over time and the need to update the hedonics model at regular intervals, in this case three times a year in April, July and (possibly) December\(^1\). In addition, any decision on when to update a regression model must link an assessment of the degree of error in the predicted prices (as described above) with an analysis of the amount of quality adjustment taking place. In the case of the UK, this lead to the decision rule that the model will be updated if the predicted price analysis falls outside the 99\% confidence intervals and there is more than a 3 percentage point effect on the monthly index due to quality adjustment. The choice of three percentage points as a threshold is based on the fact that taking into account the expenditure on computers the PC index must be incorrect by more than 10 percentage points to have an impact on the published All Items Index.

5. The latest results and price trends for PCs in the CPI

In expenditure terms computers and their peripherals now account for 11.67 parts per thousand in the CPI basket of goods and services which is priced each month for computing the index. This represents a significant increase since PCs were first introduced in the index in 1996. It is also noteworthy that expenditure on computer games- where no attempt has been to quality adjust prices- accounts for over half this expenditure.

\(^{1}\) Note that the standard error calculation should in theory be adjusted to reflect the systematic nature of the relationship between predicted and actual price means. But for this analysis the approximation is considered satisfactory.
The last update to the PC model undertaken in May 2006 and is given in Annex A. The decision to update the model was undertaken because the difference between predicted and actual prices saw a significant shift in April. This is shown in the chart below. Market knowledge indicated that there were a number of new graphics cards on the market, and those that had been priced during the first few months of 2006 were being replaced and were no longer a regular feature on PC’s. This added to the conviction that an update to the hedonic model should be carried out.

An examination of the latest model indicates that the biggest features determining the price of PCs are brand, processor and certain types of graphics cards.

**Chart 5 – PC Confidence Intervals 2006**

The following chart shows the difference the new model makes to the subsequent indices compared with the previous model and illustrates well the potential for upward bias from the over-valuation of features which cease to be special and become standard after a period of time. It can also been seen that PCs continue to fall in price in real terms and this is currently due to improved features rather than a fall in the cash price.

If the UK experience is applicable to other countries then the steeper declines in computer price indices that countries can expect if hedonics is adopted will have a positive, albeit small overall impact on growth of GDP. The effects will carry through via Household Final Expenditure, while a lower PPI
computer index will be reflected in higher estimated output resulting from a depressed deflator. An impact can also be expected from hedonic trade price indices, which is likely to be broadly similar to the impact upon the domestic PPI. The net effect upon the balance of payments is likely to be relatively small, insofar as the effects upon exports and imports will work in the same direction.

Chart 6 – PC Hedonic Indices 2006

6. A final observation

There is now a reasonably extensive body of literature on the use of hedonics to quality adjust PC and laptop prices in CPIs and PPIs which is not matched for deflators.

Price indices and deflators are different entities within a wider family of statistics relating to prices. It is particularly pertinent to note against this background that the most recent and widely available guidance on deflators, the Eurostat Handbook on Price and Volume Measures in National Accounts\(^1\), and SNA93, pre-date the UN Manuals on Consumer Price Indices and Producer Price Indices\(^2\).

The significance of this fact is that the two manuals, which were developed in parallel, take advantage of the latest research into index number theory and practice and are essentially based on the same underlying economic and statistical theory. They represent a significant step forward in the theory and construction of price indices which is not fully reflected in the official literature on national

---


The two manuals provide a comprehensive and coherent overview of the latest thinking on the conceptual and theoretical issues associated with consumer and producer price indices and translate these into the available options for practical measurement.

The lack of progress in developing a methodology for deflators is illustrated by the limited attention given to constant price estimation in the SNA93 where only about 25 pages are given over to this subject. Similarly, the Eurostat handbook on Price and Volume Measures in National Accounts, whilst providing useful guidance on an industry by industry basis, goes no further in offering advice on a conceptual framework for deflators than recommending in general terms a Supply-Use approach. There is little generic guidance about when and where it is safe to use price indices to deflate a series they were not designed for.

The lack of guidance from the international statistical agencies is particularly odd given the fact that most economic modelling is expressed in constant price terms and international comparisons on a like for like basis are important. It may be assumed that a lot of practical work has been carried out by individual researchers but that the attention paid to systematisation and standardisation has been much less than that for price indices or the current price accounts.

The lack of guidance would not be a major concern if it wasn’t for the fact that as this paper has shown alternative approaches to constructing a price index can have a significant impact on the individual indices used for deflation and hence on GDP growth rates. Thus the recommendations of the System of National Accounts 1993 (SNA93) in respect of the measurement of volumes, insofar as they are prescriptive, have practical as well as theoretical implications for the use of CPIs as a family of deflators in the National Accounts. This extends beyond the provision of deflators for household expenditure. This is because the construction of price indices for deflators should allow for consistency and coherence across both production and expenditure approaches to GDP since the approaches are brought together to provide a single estimate of GDP volume. Thus the construction of price indices as deflators should allow for consistency and coherence across both approaches.

Producer Price Indices (PPIs) and Consumer Price Indices (CPIs) are used as deflators and need to be coherent and consistent in their construction with the system of national accounts.

1 SNA93, Chapter XVI Price and Volume measures.

2 For instance, in the UK national accounts the components of the Retail Prices Index (RPI) at a detailed level are used extensively in compiling constant price GDP, mainly to deflate much of household consumption expenditure, but also in the production approach to GDP, for instance to deflate those services which are mainly consumer services. In addition the all-items index is used where no suitable deflator is available, for instance in the deflation of value added by the UK Lottery.

3 SNA approach to deflation is:
   — GDP is primarily a measure of value added.
   — Changes in volume are not the same as changes in welfare
   — As input-output provides the most complete framework for constant price calculations, it follows that deflation of other approaches to GDP should be consistent with the approach to deflating value-added.
7. Conclusions

Our early investigations indicated that hedonics, if applied correctly, provided a much more robust method of quality adjustment than option costing but that a systematic approach to the updating of the hedonic function was necessary to avoid upward bias from the over valuation of the features making up computers which result from the use of out-of-date hedonic functions. This has proved to be the case. The UK has been applying this methodology to PCs since 2003 and to laptops since 2005. It continues to provide a resilient system of quality adjustment.

There is, however, a lack of guidance on a methodology for deflators which is illustrated by the limited attention given to constant price estimation in the SNA93.

This is particularly odd given the fact that most economic modelling is expressed in constant price terms and international comparisons on a like for like basis are important. The lack of guidance would not be a major concern if it wasn’t for the fact that as this paper has shown alternative approaches to constructing a price index can have a significant impact on the individual indices used for deflation and hence on GDP growth rates.

Key words:

inflation, quality adjustment, hedonics, concepts, implementation, dummy variable, predicted price, option cost, deflators.

**Analysis of Variance**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>30</td>
<td>54.7782</td>
<td>1.82596</td>
<td>206.42</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>320</td>
<td>2.83072</td>
<td>0.00885</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>350</td>
<td>57.60063</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Root MSE | R-Square | 0.9509 |
| Dependent Mean | 6.74523 | Adj R-Sq | 0.9463 |
| Coeff Var | 1.39436 |

| Variable | Label | DF | Parameter Estimate | Standard Error | Variance t Value | Pr > |t| | Inflation |
|----------|-------|----|--------------------|----------------|------------------|-------|----|-----------|
| Intercept | Intercept | 1 | 4.89134 | 0.09811 | 49.85 | <.0001 | 0 |
| Evesham | Evesham | 1 | 0.15830 | 0.01708 | 9.27 | <.0001 | 1.68393 |
| HP | HP | 1 | 0.14894 | 0.02936 | 5.07 | <.0001 | 2.09466 |
| PACKARD_BELL | PACKARD_BELL | 1 | 0.04713 | 0.02150 | 2.19 | 0.0241 | 1.30282 |
| Philips | Philips | 1 | 0.24969 | 0.03627 | 6.88 | <.0001 | 1.45164 |
| Sony | Sony | 1 | 0.48624 | 0.09715 | 5.00 | <.0001 | 1.06394 |
| Screen | Screen | 1 | 0.03478 | 0.00569 | 6.11 | <.0001 | 1.45164 |
| Pentium4 | Pentium4 | 1 | 0.10834 | 0.02304 | 4.70 | <.0001 | 3.53830 |
| PentiumD | PentiumD | 1 | 0.18439 | 0.02423 | 7.61 | <.0001 | 4.47752 |
| Athlon64 | Athlon64 | 1 | 0.13811 | 0.02733 | 5.05 | <.0001 | 2.24662 |
| Athlon64X2 | Athlon64X2 | 1 | 0.21688 | 0.02791 | 7.77 | <.0001 | 6.89813 |
| Speed | Speed | 1 | 0.00005517 | 0.00001558 | 3.54 | 0.0005 | 4.32645 |
| Ram | Ram | 1 | 0.00014316 | 0.00001484 | 9.65 | <.0001 | 2.57276 |
| SQRT HD | SQRT HD | 1 | 0.02550 | 0.00244 | 10.46 | <.0001 | 3.73446 |
| dvdrw2 | dvdrw2 | 1 | 0.12544 | 0.04674 | 2.68 | 0.0077 | 1.21719 |
| gf6200 | gf6200 | 1 | 0.15550 | 0.06007 | 2.59 | 0.0101 | 1.21332 |
| gf6600gt | gf6600gt | 1 | 0.30821 | 0.03610 | 8.54 | <.0001 | 1.29219 |
| gf7300le | gf7300le | 1 | 0.08825 | 0.04343 | 2.03 | 0.0430 | 1.86964 |
| gf7600gt | gf7600gt | 1 | 0.13966 | 0.02491 | 5.61 | <.0001 | 1.74780 |
| gf7800gtx | gf7800gtx | 1 | 0.35580 | 0.04107 | 8.66 | <.0001 | 1.30816 |
| gf7900gs | gf7900gs | 1 | 0.16083 | 0.04017 | 4.00 | <.0001 | 1.42607 |
| gf7900gt | gf7900gt | 1 | 0.21142 | 0.02450 | 8.63 | <.0001 | 2.24638 |
| gf7900gtx | gf7900gtx | 1 | 0.10080 | 0.04178 | 2.41 | 0.0164 | 2.28655 |
| ATI | ATI | 1 | 0.06971 | 0.01762 | 3.96 | <.0001 | 2.24265 |
| MB | MB | 1 | 0.00039416 | 0.0004924 | 8.00 | <.0001 | 2.13353 |
| XFiFatality | XFiFatality | 1 | 0.12249 | 0.02515 | 4.87 | <.0001 | 1.34826 |
| XFiXtremeMusic | XFiXtremeMusic | 1 | 0.14249 | 0.02381 | 6.04 | <.0001 | 1.51650 |
| XFiElite | XFiElite | 1 | 0.24903 | 0.05627 | 4.43 | <.0001 | 1.06474 |
| Speakers | Speakers | 1 | 0.04506 | 0.01255 | 3.59 | 0.0004 | 1.53118 |
| Onsite | Onsite | 1 | 0.05702 | 0.00523 | 10.89 | <.0001 | 1.44090 |
| RemoteContro | RemoteContro | 1 | 0.00888 | 0.01280 | 0.69 | 0.4882 | 1.43910 |

Brands – Evesham, HP, Packard Bell, Philips, Sony.
Processors – Pentium 4, Pentium D, Athlon 64, Athlon 64 X2.
Ram – Computer memory size.
SQRT HD – The square root of the Hard Disk Drive capacity
DVDRW2 – This indicates those PC’s with a DVD – Rewriter as a 2nd optical drive.
GF or ATI – All variables beginning GF or ATI are Graphics Cards.
MB – The Graphics card memory
XFi – All variables beginning XFi are soundcards.
Onsite – Type of warranty offered by outlets.

Rome31082006
ICT DEFLATION AND PRODUCTIVITY MEASUREMENT

Gavin WALLIS
Economic Analysis Division
Office for National Statistics (ONS)
1 Drummond Gate
London
SW1V 2QQ
E-mail: gavin.wallis@ons.gsi.gov.uk

Abstract

Constant price productivity growth is generally the object of interest for productivity analysis. The rapid value added growth in the ICT sector relative to the whole economy, the rapid growth in ICT capital formation, and the remarkable and rapid advances in the characteristics and capabilities of ICT products, mean that reliable productivity analysis depends strongly on good measures of deflation for ICT products. These ICT deflators need to reflect pure price change, so measured prices need to be adjusted for quality change. This paper looks at the impact and importance of ICT deflation that accounts for quality change on the measurement of productivity, with examples and analysis drawn from recent ONS work. The paper also looks at the development of input measures for multifactor productivity work, highlighting the importance of good quality deflators for ICT capital.

1 The views expressed in this paper are those of the author and not necessarily those of the Office for National Statistics.
1. Introduction

Measures of productivity growth are commonly used as core indicators of economic growth, prosperity and competitiveness. The simplest definition of productivity, and the one most commonly used, is the ratio of a volume measure of output to a volume measure of input. This simple definition does, however, conceal the various types of productivity measures that exist. A common distinction is between single-factor productivity measures, which relate output to a single measure of input, and multifactor productivity (MFP), which relate output to a set of inputs. In practice, the most common measure of productivity used is labour productivity, partly due to the difficulty of obtaining good measures of capital input. When capital inputs can be measured, long-term trends in capital-labour MFP often becomes the productivity measure of choice, as it is regarded as being a good measure of the long-term growth opportunities of an economy and also an indicator of inflationary pressures.

Whichever productivity measure we choose to focus on, it is clear that constant price productivity growth is the object of interest. We need both input and output to be measured in volume terms, meaning that good deflation becomes an essential part of accurate productivity measurement. This is particularly important when measuring productivity in high tech industries, which are generally intense users of Information and Communication Technology (ICT), where price change can be rapid.

This paper focuses on what are probably the two main challenges for statisticians and economists wishing to produce accurate and policy-relevant measures of productivity. These two challenges are the development of good price indices for the measurement of the volume of output and improved measurement of capital input, especially ICT capital. ICT deflation clearly plays a big part in both these challenges.

2. Growth in ICT

The rapid growth in ICT production and ICT use in major industrialised countries has been well documented. The contribution ICT makes to GDP and productivity growth needs to be considered in three different ways. The first is the direct effect on gross value added (GVA) of the ICT-producing industries. The second is the indirect effect of ICT investment on the GVA of ICT-consuming industries. The final consideration in terms of productivity analysis is the impact that ICT use has on the volume of capital input.

Figure 1 looks at the first effect and shows the growth of ICT sector GVA compared with the growth of UK economy GVA. The definition of the ICT sector used here is the one agreed by the OECD Committee for Information, Computer and Communications Policy (ICCP) in September 1998 and outlined in OECD (2002). This definition covers both goods, such as office machinery and communication equipment, and services, such as telecommunication and computer services.
Figure 1 – Gross value added (GVA) at current basic prices: ICT sector relative to the UK economy

Source: ONS, United Kingdom Input-Output Analyses, 2006 Edition

Figure 1 shows that, throughout the 1990s, ICT sector GVA grew rapidly compared with the growth in UK GVA. This illustrates the importance of ICT production and investment in driving UK economic and productivity growth. The problem with figure 1 is that it shows GVA at current basic prices. To assess the full contribution of the ICT sector to productivity growth, we need a constant price output measure. Figure 1 is based on UK input-output data, which are currently only available at current prices. As noted in OECD (2001a), the development of constant price input-output tables is an important step in producing more reliable productivity measures, especially at the industry level.

Given what we know about the rapid fall in the prices of ICT products over the period 1990 to 2004, figure 1 is understating the importance of the ICT sector in terms of economic and productivity growth. Some GVA volume measures are available for the UK to support this conclusion. They are not available at the disaggregated level required to identify the ICT sector, but are available for broad service industries. Figure 2 shows GVA volume measures for selected service industries and also for services as a whole. Real estate, renting and business activities includes two of the identified ICT-producing industries defined in OECD (2002); ‘renting of office machinery and equipment including computers’ and ‘computers and related activities’ (UK input-output group names are used here). Post and telecommunication includes another of the key ICT-producing sectors ‘telecommunications’.

It is clear from figure 2 that volume growth in the service industries which include ICT producers, such as post and telecommunication, has been very rapid compared with general services growth and also relative to other service industries.
It is important to be aware that the broad service industries shown in figure 2 encompass both ICT-producing industries and also ICT-consuming industries. This brings us to the second way in which ICT makes a contribution to GDP and productivity growth: the indirect effect of ICT investment on the GVA of ICT-consuming industries. Figure 2 shows that financial intermediation has also grown more quickly than services in general. Financial intermediation is one of the main ICT-consuming industries and is an industry that invested heavily in ICT during the 1990s. Recent work using firm level microdata (Bloom et al (2005), Clayton (2005), Sadun (2005)) has shown that industries and sectors that intensively use ICT have enjoyed more rapid productivity growth. This suggests that the indirect effect of ICT investment is contributing significantly to volume growth in some industries.

The discussion so far has focused on the growth of ICT in terms of its effect on output. In terms of long-run productivity growth and productivity measurement, it is important to also consider how the rapid growth in ICT has impacted on the volume of capital input. The starting point is understanding the importance of ICT investment relative to total investment in the economy. Measurement of the volume of capital input is discussed below.

**Figure 3** shows ICT gross capital formation relative to total gross capital formation. The rapid investment in ICT in the late 1990s is obvious. In 1992, ICT accounted for just over 13 per cent of total gross capital formation. By 2000, this had increased to just below 21 per cent. Since 2000 there has been an obvious turnaround with ICT accounting for just over 14 per cent of total gross capital formation in 2004. The common explanation for this pattern is that firms over invested in ICT in the run-up to the year 2000, due to fears over the well-publicised ‘millennium bug’. Such overinvestment would account for the weaker ICT investment growth after 2000.
Impact of hedonic price measurement on productivity

Another way to look at the importance of ICT is at its share of operating surplus (or profit) over time. **Figure 4** shows the composition of profit shares for the period 1950 to 2004 for the UK. These profit shares are taken from Wallis (2005), which reports capital services estimates for the UK.

**Figure 3** – Gross capital formation: ICT relative to total

![Graph showing ICT gross capital formation and total gross capital formation from 1992 to 2004.](image)

*Source: ONS, United Kingdom Input-Output Analyses, 2006 Edition*

**Figure 4** – Profit shares by asset type, UK, 1950-2004

![Graph showing percentage of profit shares by asset type from 1950s to 2000s.](image)

*Source: ONS, Wallis (2005)*
In figure 4, only computers are identified as a separate asset, with other ICT assets not separately identified from plant and machinery. However, figure 4 does show the rapid rise in the profit share of computers in recent years, from virtually zero in 1980-84 to around 10 per cent in the period 2000-04. This suggests that such ICT products have become an increasingly important part of capital input in the economy.

3. ICT deflation and quality change

The purpose of a deflator is to adjust nominal (current price) data into real (constant price) data. A traditional deflator (price index) is a matched model index using price quotes that track the price of the same good or service over time. The remarkable and rapid advances in the characteristics and capabilities of ICT present two problems in terms of price measurement. The first is how to deal with quality change and the second is how to account for the introduction of new goods and services. These two problems are clearly not mutually exclusive, as the distinction between a ‘new’ good and a new (and better quality) variety of an existing good is not clear.

For the purposes of deflation we are interested in the ‘pure price change’. Observed prices will incorporate both pure price change as well as quality change (observed price change = pure price change + quality change). For many goods and services, quality change is either very slow or non-existent and so observed price change coincides with pure price change. For many ICT goods, quality change is rapid, so to get a deflator that measures pure price change, we need to adjust observed price change for quality changes.

The importance of quality adjusting ICT deflators is most evident in computer hardware. Holdway (1999) reports that between 1993 and 1998, observed prices from desktop computers were fairly stable, but during this time CPU speed (MHz) increased by 1,263 per cent, system memory increased by 1,500 per cent, hard drive capacity increased by 3,700 per cent, and monitor size increased by 13 per cent. It is clear that, although observed prices fell very little, pure prices, once adjusted for rapid quality improvements, fell much more rapidly.

There are various ways of quality adjusting deflators, including matched models and hedonic regressions, but the underlying objective of all of these methods is to ensure that the estimated deflator reflects pure price change only. An accompanying paper to this, Fenwick (2006), describes the adoption and use of hedonic techniques for the quality adjustment of computer equipment in the UK consumer prices index (CPI) and producer prices index (PPI).

At this point it is worth mentioning double deflation. If we are interested in accurately measuring the impact of ICT on productivity and output, we need to ensure that both our output and input deflators are reflecting pure price change. Many ICT deflators fall quite rapidly and this increases the volume of real output for ICT-producing industries. For an industry that is a consumer of ICT, GVA and productivity would be inflated if the input price series for ICT were not quality adjusted. For this reason, both input and output price indices for ICT goods and services should be quality adjusted, a process known as double deflation. This is particularly important for productivity analysis as it ensures consistency between GDP and industry output.
4. Measuring the volume of output

The basic problem with using deflators that do not reflect pure price change, but also incorporate some quality improvement, is that volume measures of output will be underestimated. This will mean that traditional measures of productivity growth, such as labour productivity, will also be underestimated. Obviously this is particularly important for ICT-producing industries. For example, the quality adjusted producer price index (PPI) for computers and other data processing equipment has fallen from 385.8 in 1992 to 47.5 in 2004, where 2000 = 100.

Vaze (2001) presented ONS work on the impact of different treatments of ICT goods and services on measures of UK economic growth. It primarily considered the impact of using different price indices for ICT on UK growth. The analysis was conducted at a time when the ONS did not use hedonic methods to quality adjust price indices. Using US ICT deflators, the analysis highlighted the sensitivity of economic growth to the choice of ICT deflator and also to the method used to adjust ICT deflators for quality change. UK growth was estimated to be 0.1 per cent per annum higher over the period 1992 to 1998 when using the US ICT deflator.

Vaze (2001) also highlighted the importance of correctly identifying where the various ICT goods and services that are available appear in the National Accounts aggregates as mentioned previously. The output of the ICT sector falls partly in capital formation, which adds to GDP, and partly in intermediate consumption, which does not. The correct allocation between the two categories is necessary to avoid biases on the level and growth of GDP.

The UK adopted hedonic techniques to quality adjust computer equipment in 2003, as described in Ball and Allen (2003), and clearly this will have improved the measurement of the volume of output for the UK, especially for the ICT-producing industries. More recent work has looked at measures of software investment in the UK and accompanying software deflators (Chamberlin and Chesson (2006)).

Estimating software deflators, for deflating both input and output, that reflect pure price change presents a unique challenge. This is because software investment takes three different forms:

1. **Pre-packaged purchased software**: off-the-shelf purchased software
2. **Own-account software**: software developed and used in-house
3. **Custom software**: purchased software that can be customised in-house to better meet user requirements

A quality adjusted deflator for the first type of software can be produced using standard methods, such as hedonics. Own-account is much more difficult and generally has to be based on wage costs of labour working on own-account production. A deflator for customer software is generally a weighted combination of the own-account and pre-packaged software deflators.

The own-account and purchased software deflators from Chamberlin and Chesson (2006) are shown in figure 5. It should be noted that the purchased software deflator is based on the US software price index, rather than a direct estimate for the UK using hedonic techniques. This is due to a shortage of the necessary historical data. The US software price index is quality adjusted. For detail about the construction of these deflators, see Chamberlin and Chesson (2006).
Figure 5 – Purchased software and own-account software deflators

![Graph showing purchased and own-account software deflators from 1985 to 2003.](image)

*Source: ONS*

Figure 5 shows that the price of purchased software has fallen rapidly. This reflects the fact that the quality of purchased software has increased over time. The price of own-account software has shown a steady increase over the period shown. This illustrates the fact that it is based on the wage costs of labour working on own-account production. Own-account employees have enjoyed above average wage growth over the period. Figure 5 highlights both the importance and difficulty of measuring ICT deflation; the two different software deflators show very different patterns and these differences should be reflected within National Accounts aggregates for good productivity measurement.

The new estimates of software investment and software deflators in Chamberlin and Chesson (2006) are yet to be incorporated into the UK National Accounts, but they are expected to add around 1 per cent to GDP in current price terms. This will clearly have implications for the measurement of UK productivity growth and the adoption of a quality adjusted price index for software will clearly be a step forward.

5. Measuring the volume of capital input

Quality changes are generally positive for ICT goods and services, so failure to use deflators that reflect observed price change rather than pure price change will lead to the underestimation of capital input. This will in turn lead to overestimation of productivity measures that include capital as a factor of production.

MFP is probably the most common productivity measure that includes capital as a factor of production. MFP apportions growth in output to growth in the factor inputs, capital and labour, and growth in a residual which represents technical change. MFP analysis requires good measures of labour and capital input. The ONS has been developing these input measures as part of its wider development...
of productivity estimates (see Camus and Lau (2006)). Results for a quality-adjusted labour input measure can be found in Wallis et al (2005) and Goodridge (2006) and for capital services in Wallis (2005).

Capital services are the measure of capital input that is suitable for analysing and modelling productivity, including MFP analysis. This is because capital services are a direct measure of the flow of productive services from capital assets rather than a measure of the stock of those assets. In essence, capital services are a measure of the actual contribution of the capital stock of assets to the production process in a given year. This is in contrast to the wealth-based estimates of capital in the National Accounts, gross and net capital stock, which are essentially a measure of the value of the capital stock of assets.

The use of capital services for productivity analysis, rather than gross or net capital stock, is particularly important in light of the growing use of ICT. The remarkable and rapid advances in the characteristics and capabilities of ICT capital mean that prices have fallen rapidly and also that the ICT capital stock depreciates quickly. This means that the rental rates for ICT capital, which under perfect competition reflect the marginal product of capital, are very high relative to other assets. ICT capital provides a large stream of capital services, relative to other assets, over a short period of time.

Figure 6 shows capital services growth for computer hardware over the period 1987 to 2004. Capital services from computer hardware grew rapidly over the whole period, with an average growth rate of over 21 per cent. This compares with an average of just 3.6 per cent for whole economy capital services. Annual growth is lowest in 1991, but at nearly 8 per cent is still comfortably above growth in capital services from other assets for this period. See Wallis (2005) for further results.

Figure 6 – Annual growth in capital services, computer hardware and whole economy (all productive assets), 1987-2004

Source: ONS, Wallis (2005)
As shown in Wallis (2005), the time trend of capital services growth for computers also differs dramatically from other assets. The other asset types saw a fall in capital services growth in the early 1990s, associated with the recession in the UK. From figure 6 it is clear that there is not a fall in capital services growth in the early 1990s for computers and indeed growth in capital services actually shows a sustained increase in growth from 1991 to 2000. This is because the capital services estimates reflect both the increased quality of computer power as well as changes in the level of investment. For some of the years in the period 1991 to 2000, investment declined year-on-year but capital services still increased due to increased quality of computer power. Peak capital services growth occurs in 1998 with an annual growth rate of around 45 per cent.

When estimating capital services, the calculation of rental rates is key. The rental rate should reflect the marginal productivity of capital and, as such, the rental rate for computer hardware should reflect quality change, so long as the quality change has an impact on the marginal productivity of capital. With this in mind, it is informative to compare the rental rate for computers with the PPI for computers. We know that in terms of productivity measurement, if capital services are estimated correctly, then they provide the best measure of capital input and the rental rate reflect the price of that capital input. Figure 7 compares the rental rate for computer hardware with the PPI for computers (ONS code PQE).

Figure 7 – Comparison of rental rate for computer hardware and quality adjusted PPI for computers

Since 1997 the two series move very much in line. This is a good sign and suggests that the PPI for computers is an accurate measure of the price of capital input. Prior to 1997 the series diverge somewhat. Computers were first introduced into the PPI in 1996 and at the same time explicit methods of quality adjustment were introduced to replace the implicit methods used previously. With this in mind, it is not surprising that the PPI tracks the rental rate more closely after 1996. Calculation of capital services and rental rates is a complex modelling process and so we would not expect an exact fit of the PPI
and the rental rate. However, the comparison in figure 7 is informative as it suggests that better quality adjustment of deflators, which is a feature of the PPI for computers after 1996, and more so from 2003 onwards with the introduction of hedonics, does improve the accuracy of input deflators.

6. Conclusions

ICT products have become an increasingly important part of capital input in the economy and are making a major contribution to GDP and productivity growth. For accurate productivity measurement there is a need to ensure ICT deflators reflect pure price change. The remarkable and rapid advances in the characteristics and capabilities of ICT capital mean that pure prices have fallen rapidly. This means that measured prices need to be adjusted for quality change in order that both volume measures of output and volume measures of input are accurate.

Using deflators that do not reflect pure price change, but also incorporate some quality improvement, will lead to volume measures of output being underestimated. This will mean that traditional measures of productivity growth, such as labour productivity, will also be underestimated. On the other hand, capital input will be underestimated leading to overestimation of productivity measures that include capital as a factor of production.

The simple conclusion is that reliable productivity analysis depends strongly on good measures of deflation for ICT products. The impact of using poor quality deflators for ICT products depends, however, on which of the various measures of productivity is being used.
References


www.statistics.gov.uk/cci/article.asp?ID=1603


www.statistics.gov.uk/CCI/article.asp?ID=1401


www.statistics.gov.uk/cci/article.asp?ID=1235


www.statistics.gov.uk/cci/article.asp?ID=1464


www.oecd.org/dataoecd/16/14/1835738.pdf


www.statistics.gov.uk/CCI/article.asp?ID=1298
THE ROLE OF HEDONIC METHODS IN MEASURING REAL GDP IN THE UNITED STATES

Dave WASSHAUSEN and Brent R. MOULTON
Bureau of Economic Analysis
U.S. Department of Commerce
Washington DC, 20230 USA

Abstract

Accurate price indexes are crucial for preparing accurate estimates of real gross domestic product and corresponding productivity measures. The price index must capture price change for a ‘relevant’ market basket goods, while at the same time controlling for changes in characteristics and/or quality of these goods. Traditional price indexes (i.e. ‘matched model’) are well suited to capturing price change for goods that exhibit little or no quality change over time, however, for products whose characteristics and/or quality are changing rapidly (e.g. ICT goods), hedonic methods may be more suitable.

This paper provides a brief history of hedonic methods employed by U.S. statistical agencies and specifically examines the role of hedonic price indexes in the U.S. National Income and Product Accounts. It also attempts to dispel some popular misconceptions about hedonic methods.

1 The views expressed in this paper are those of the authors and do not represent an official position of the Bureau of Economic Analysis.
1. Introduction

The question is often asked, ‘What is the impact of hedonic price indexes on measures of real GDP growth?’ However, before one can attempt to answer this question, we must first answer the question, ‘What is the impact of hedonic methods on price indexes?’ It is often asserted that hedonic methods always result in a lower rate of price change than traditional ‘matched model’ methods and therefore overstate real GDP growth. However, there have been several studies in which the hedonic index has increased at a higher rate. Hedonic indexes share the same objective as traditional price indexes—to measure the rate of price change excluding (or holding constant) the effects of changes in quality. Both hedonic methods and traditional methods can be appropriate methods, depending on the circumstances. What is important is which method(s) yield the best price index for a given product.

Clearly, traditional matched model price indexes encounter problems when measuring price change for goods with rapidly changing quality and/or characteristics. One such problem is that the price changes observed for the matched models may not accurately reflect price changes for all models. That is, any new models that reflect the newest technology and therefore have different specifications will not be reflected in the matched model simply because they cannot be ‘matched’ to a model in the period prior to their introduction. For some types of products, particularly information and communication technologies (ICT) products, these models can represent a significant market share. A second problem with the implementation of a traditional matched model approach is that a model may be ‘matched’ to a model that is not in fact an exact ‘match,’ leading to a distorted price change. Hedonic methods address these challenges in a relatively systematic and objective way.

In this paper, we will first provide a brief history of hedonic price indexes, followed by a discussion on the hedonic price index basics. Next we’ll discuss the impact of hedonic price indexes on measuring real GDP, with a specific focus on computers and peripheral equipment. We also address some common misconceptions about hedonic price indexes and then provide some concluding remarks.

2. A Brief History of Hedonic Price Indexes

The origin of hedonic methods in U.S. official price statistics goes back to the famous article by the late Zvi Griliches (1961) that was published in the report of the Price Statistics Review (Stigler) Committee. Although Waugh (1928), Court (1939) and Stone (1954, 1956) preceded Griliches in developing and applying hedonic techniques,1 the work of Griliches was original in the following sense defined by Stigler (1955) when he wrote, ‘Scientific originality in its important role should be measured against the knowledge of a man’s contemporaries. If he opens their eyes to new ideas or to new perspective on old ideas, he is an original economist in the scientifically important sense.’ In this sense, the work of Griliches (1961) was surely original—as Lipsey (1990) observed, he took an unconventional method that was then on periphery of price statistics and demonstrated to the economics and statistics community that it could be used to address critical quality adjustment problems that previously had been considered intractable.

---

1 It was in fact Court who coined the term ‘hedonic pricing method’ while developing price measures for automobiles in the late 30’s. Court reasoned that using changes in average list prices would not account for the increased ‘welfare and happiness’ that society as a whole experienced with the rapid improvements in automobiles. His method took this ‘welfare and happiness’ of society into account by relating model prices to those characteristics that lead to consumer happiness (e.g. power, speed, interior room) and therefore became known as the ‘hedonic pricing method.’
Following Griliches, hedonic methods quickly grew to be a new branch of economic research, which is now far too vast to be easily surveyed. Articles that provide overviews of some of this literature include Griliches (1971, 1990), Triplett (1975, 1987, 2000), Berndt (1983, 1991) and Bartik and Smith (1987). We will leave it to others to comment on this literature, except to note that there have been a number of theoretical and empirical controversies. As Triplett (1990) observed, however, many of these controversies have counterparts in the traditional literature on economic index numbers—issues such as aggregation across individuals in constructing a social cost-of-living index, imperfect competition in constructing an output price index, and problems in separating demand shocks from supply shocks. Our view, which echoes Triplett’s, is that all quality adjustment methods are imperfect, but regardless of these imperfections, statistical agencies need to do a better job of quality adjustment, and so these controversies should not prevent agencies from using hedonic methods as one tool in quality adjustment. Many years passed before the economic literature had much impact on the U.S. official price statistics.

2.1. **Hedonic Price Index History for BEA**

In the mid-1980’s BEA collaborated with IBM to develop quality-adjusted price indexes for computers and peripheral equipment. The BEA-IBM price indexes for computer equipment and peripherals were introduced into the national income and product accounts in December 1985 (Cole et al. 1986; Cartwright 1986; Triplett 1986). The original indexes covered five types of computing equipment—computer processors, disk drives, printers, displays (terminals), and tape drives for the period 1972-84. Subsequently, a price index for personal computers was added, and a separate index was created for computer imports. The history and present status of the indexes are documented in U.S. Department of Commerce (2000).

It is interesting to note that the problem BEA addressed in its collaborative effort with IBM was more than just obtaining an improved method of quality adjustment. Prior to 1985, BEA simply had no acceptable price index for computers, so computers had been deflated by an index that was equal to 1 for all periods.

In the early 1990’s, BLS began publishing quality-adjusted producer price indexes (PPI’s) for computers and peripheral equipment. As these PPI’s became available, BEA used them to extrapolate the BEA computer price indexes. Eventually, BLS indexes were used for all of BEA’s quality-adjusted computer price indexes.

The BEA computer price indexes show notable declines for all periods; for 1959-2005, the average rate of price change for private fixed investment in computers and peripheral equipment is −16.9 percent per year. Although skeptics have occasionally questioned the rapid price declines, the BEA index has stood the test of time. Scholarly studies have generally found similar rates of price decline (for example, see Berndt, Dulberger, and Rappaport 2000; Aizcorbe, Corrado, and Doms 2000). Several other countries now regularly use the BEA computer price indexes to deflate the computer components of computer imports and capital formation in their own national accounts.

BEA’s next hedonic index was the price index for multifamily residential structures (de Leeuw, 1993). The issues for multifamily housing were the same as those for single family housing—severe heterogeneity in the characteristics of housing units leading to the use of an inadequate proxy as a
deflator. Considerable research was undertaken before an acceptable hedonic function was identified. The Census Bureau has generously assisted BEA in developing and maintaining this index.

After BEA introduced its quality-adjusted computer price index, a frequently mentioned criticism was that use of the index led to inappropriate measures of value added in the construction of computers because the prices of important inputs, such as semiconductors, were not similarly quality adjusted. In January 1996, BEA introduced new quality-adjusted price indexes for semiconductors, based on indexes for several types of memory chips and microprocessors that were developed by Grimm (1998). In the case of memory chips, Grimm used hedonic methods as a guide to constructing a matched model index. In the case of microprocessors, the hedonic function was used with the matched model to form a composite index, as described earlier.

In 1997, BEA introduced a hedonic price index for digital telephone switching equipment (Grimm 1996; Parker and Seskin 1997). The hedonic regression used data from the filings by regional telephone operating companies with the Federal Communications Commission and incorporated characteristics such as the location, type, and capacity of the switch (number of telephone lines).

In 1999, BEA expanded its concept of capital in the national accounts to recognize expenditures for software as fixed investment (Moulton, Parker and Seskin 1999). Because quality-adjusted price indexes for software did not then exist, BEA developed hedonic indexes for a portion of pre-packaged software—specifically, for spreadsheets and word processing programs (Parker and Grimm 2000). Because these types of software represent only a portion of pre-packaged software, the hedonic indexes were averaged with matched model indexes to form the pre-packaged software price index. Because data for estimating the hedonic function were available only for 1985-93, bias adjustments were applied to the matched model index for subsequent years.

In 2003, BEA introduced a hedonic price index for photocopying equipment (Wasshausen 2003). The hedonic regressions were biennial (to allow for changes in parameters over time) using data purchased from a trade source and incorporated characteristics such as copy speed, color capability, multi-functionality capability and recommended maximum copy volume.

In 2005, BEA hired a contractor to develop an improved price index for custom software. (The current price index is primarily a cost-based index that assumes roughly zero growth in multifactor productivity.) The contractor developed a number of price indexes for custom software using proprietary data on thousands of custom software projects. Custom software is a good example of a product where traditional matched model methods will simply not work. Custom software, by its nature, cannot be effectively matched from one observation to another. Accordingly, a hedonic price index based on biennial regressions where price was a function of project type and size, software type, software quality and fixed effects of clients was recommended (Goldfarb, Heller, White and Abel 2006). BEA has purchased updated proprietary data that was used in this study and is continuing to analyze whether or not this data may yield an improved custom software price index using hedonic methods.

In addition to developing several hedonic price indexes in-house, BEA uses the consumer, producer, import and export price indexes produced by BLS, the Census construction price indexes, and prices from several other agencies in deflating the national, regional, and industry accounts. In several
cases, BEA has also back-cast or made bias adjustments to indexes that have been shown to be biased, thereby maintaining a more consistent time series.

3. Hedonic Price Index: Basics

The term ‘hedonic methods’ refers to the use in economic measurement of a ‘hedonic function,’ \( h() \),

\[ (1) \quad p_i = h(c_i) \]

where \( p \) is the price of a variety (or model) \( i \) of a good and \( c_i \) is a vector of characteristics associated with the variety. The hedonic function is then used in one of several ways to adjust for differences in characteristics between varieties of the good in calculating its price index. The hedonic function is usually estimated by regression analysis.

There are basically two techniques for constructing a hedonic price index. The first technique, which we will refer to as the ‘regression price index,’ uses estimated coefficients on year dummy variables to estimate price change. For example, the hedonic function for photocopying equipment is as follows:

\[ P = \beta_0 + \beta_{MCS}x_{MCS} + \beta_{Color}x_{Color} + \beta_{MF}x_{MF} + \beta_{MV1}x_{MV1} + \beta_{MV2}x_{MV2} + \beta_t x_t + \mu \]

Where:

- ‘\( P \)’ is the natural log of price, ‘\( \mu \)’ is the error term and ‘\( \beta_0 \)’ is the constant.
- \( x_{MCS} \) is the natural log of copy speed.
- \( x_{Color}, x_{MF}, x_{MV1}, \) and \( x_{MV2} \) are dummy variables capturing the effect of color, multi-functionality and recommended copier volume.
- \( x_t \) is a dummy variable that captures the effect of time.

The estimated coefficient on \( x_t (\beta_t) \) represents the percent change of a constant-quality price in period ‘\( t \)’. That is, the change in price that is not explained by differences in speed, color, multi-functionality, or recommended copier volume (characteristics believed to account for ‘quality’) since they are controlled for within the equation.

The second technique for constructing a hedonic price index employs a mix of hedonic and matched model methods. These price indexes are sometimes referred to as ‘composite price indexes.’ Here, the matched model technique is used as much as possible. Only when a new model is introduced do we need to rely on hedonic methods. The BLS producer price index (PPI) for portable computers is a good example of this technique. Consider the following hypothetical example:

Assume the hedonic function for portable computers is a linear function as follows:

\[ p_i = h(ps_i, hds_i, \text{&} \ mem_i) \]

where ‘\( p \)’ is the price, ‘\( ps \)’ is processor speed, ‘\( hds \)’ is hard drive size and ‘\( mem \)’ is the amount of memory. In period ‘\( t-1 \)’ portable computer model 123, which has an 80 GB hard drive with 512 MB of memory and a processor speed of 2000 MHz, has a reported price of $1000. In
period ‘t’ model 123 has been replaced with model 123A and the only differences between the two is that model 123A has an increased processor speed of 2100 MHz and now costs $1200. Assume the estimated coefficient on \( p_s \) is 3.5 ($/MHz), therefore the increase in processor speed can be quantified in dollars as \((2100 	ext{ MHz} - 2000 	ext{ MHz}) \times 3.5$/MHz = $350. The quality-adjusted (QA) price change is calculated as \((P(t) - QA) / P(t-1)\) or \((1200 - 350) / 1000\), or -15 percent.

### Table 1 – Portable Computers Example

<table>
<thead>
<tr>
<th></th>
<th>Model 123, period ‘t-1’</th>
<th>Model 123A, period ‘t’</th>
<th>Estimated coefficient</th>
<th>Quality Adjusted Model 123A, period ‘t’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>$1,000</td>
<td>$1,200</td>
<td>---</td>
<td>$850</td>
</tr>
<tr>
<td>Processor speed</td>
<td>2000 MHz</td>
<td>2100 MHz</td>
<td>3.5</td>
<td>$350</td>
</tr>
<tr>
<td>Hard disk size</td>
<td>80 Gb</td>
<td>80 Gb</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Memory</td>
<td>512 MB</td>
<td>512 MB</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

BLS has other alternatives to using hedonic functions for dealing with changes in quality, however these methods rely on, perhaps, more subjective analysis. For example, one alternative used by BLS is to ask the respondent to provide a dollar estimate of the change in quality embedded in the new product.

There is an additional fundamental difference between these two hedonic techniques which lies in the weighting of the observations. In the ‘regression price index,’ the individual observations are essentially unweighted while the ‘composite price index’ reflects a weighted average of observations. Aizcorbe and Pho (2005) examined this issue of aggregating detailed observations for over 60 classes of consumer electronic and IT goods and found that differences in the weights do matter.

### 4. Impact of Hedonic Price Indexes on Measuring Real GDP

Hedonic price indexes are used to deflate a number of GDP final demand components, accounting for about 20 percent of nominal GDP in recent years. (See appendix A for a table presenting the components of GDP that are deflated with hedonic-type price indexes.) It should be noted that not all of these price indexes reflect hedonic methods for all years. In many cases, source data used to construct the hedonic function are not available for all years and a bias-adjustment is used to account for quality changes that may not be reflected in a traditional matched model price index. For example, the price index for prepackaged software reflects hedonic methods for years 1985-93; years 1994-present reflect a matched model price index with a bias-adjustment.

For many of the ‘hedonic components’ the difference between the matched model price and the hedonic price index is small and somewhat offsetting. For example, the introduction of hedonic price indexes by BLS slightly raised the rate of price increase for VCR’s and for rent but slightly lowered it for televisions.

The main component in which hedonic methods have had a significant impact is computers and
peripheral equipment. Focusing on private fixed investment in computers and peripheral equipment, the average annual price decline from 1990-2005 is 16.0 percent, although the rates of decline appear to be slowing down in more recent years (average annual decline for years 2001-2005 is 11.1 percent). Several previous papers and articles have examined the robustness and validity of hedonic computer price indexes, including a comprehensive literature review by Berndt and Rappaport (2000). Berndt and Rappaport’s review compared rates of decline for PC’s and Mainframes estimated by a dozen or so hedonic price index experts. Naturally there was some variation in reported declines over varying time periods, however, when detailed components are compared over similar periods the results are consistent, including BEA’s computer price index.

There is evidence that a ‘well constructed’ matched model index for rapidly changing high-tech goods could yield a price index that adequately controls for quality differences and that the price index is consistent with a quality-adjusted price index constructed using hedonic methods. Aizcorbe, Corrado, and Doms (2003) constructed price indexes for microprocessors using high frequency disaggregated data on models whose characteristics were constant over time and found that their matched model price indexes were remarkably close to those constructed using hedonic methods (table 2). Similar results were reported in Aizcorbe, Corrado, and Doms (2000) for personal computers. Silver and Heravi (2001, 2002) report similar findings using scanner data for washing machines and televisions. However, given that we often do not have the abundant data necessary to construct such a matched model price index, then the hedonic price index is the practical approach for measuring prices of rapidly changing goods or goods that by nature are heterogeneous (e.g. custom software or homes).

### Table 2 – Matched Model vs Hedonic Price Indexes

<table>
<thead>
<tr>
<th></th>
<th>Matched Model</th>
<th>Hedonic</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aizcorbe et al.:1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Desktop personal computers, 1993:I to 1998:IV</td>
<td>-30.3</td>
<td>-28.2</td>
<td>2.1</td>
</tr>
<tr>
<td>Microprocessors, 1993:I to 1999:IV</td>
<td>-18.0</td>
<td>-18.2</td>
<td>-0.2</td>
</tr>
</tbody>
</table>


Ideally, in order to measure the impact of hedonic price indexes on real GDP, we would have a second set of traditional matched model price indexes for all goods that are deflated using hedonic price indexes and we would re-compute real GDP using this second set of price indexes and then compare growth rates. Unfortunately this approach is not possible due to resource constraints. Borrowing from an analysis presented by Landefeld and Grimm (2000) we have prepared a shortcut one can take to estimate the effect that the hedonic quality-adjustment process for computers and peripheral has on measuring real GDP. Table 3 below shows that the average annual change in unit value for single-user computers over the years 2001-2005 is -4.9 percent. This estimate, which does not take any changes in quality into account, is derived from shipments data published in the ‘Current Industrial Reports’ by the U.S. Census Bureau (http://www.census.gov/cir/www/alpha.html). The NIPA private fixed investment quality-adjusted price index for single-user computers (a.k.a. PC’s) shows an average rate of decline 16.4 percent over the same years. The difference in the two declines
is 11.5 percent, which can be characterized as the affect of quality-adjusting the PC price index. Assuming this estimated ‘quality adjustment’ difference for PC’s is representative of all types of computers and peripheral equipment (or provides an upper bound estimate which is reasonable given that the price index for PC’s declines on average more rapidly than any of the other component) we can approximate the impact that hedonic quality adjustment has for all computers and peripheral equipment by multiplying the 11.5 percent ‘quality-adjustment’ difference by the average nominal share of final sales of computers to GDP (about 0.8 percent). The results of this exercise shows that the estimated impact of the hedonic quality adjustment is less than 0.1 percentage point of average annual real GDP growth over the period. This exercise is not intended to mitigate the role that hedonic methods have in constructing high quality price indexes; rather it is intended to illustrate that one should not attribute periods of robust real GDP growth, at least for the United States, to the use of hedonic methods to estimate quality-adjusted price indexes.

Table 3 – PC Prices: Average Annual Changes, 2001-05

<table>
<thead>
<tr>
<th></th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Census CIR average unit values</td>
<td>-4.9%</td>
</tr>
<tr>
<td>NIPA private fixed investment price index</td>
<td>-16.4%</td>
</tr>
<tr>
<td>Estimated quality adjustment</td>
<td>-11.5%</td>
</tr>
</tbody>
</table>

5. Misconceptions about Hedonic Price Indexes

The experience of the U.S. statistical agencies may help dispel several misconceptions about the application of hedonic methods. We list a few misconceptions below, in each case following with a brief discussion of the actual experience of applied hedonic methods.

5.1. Quality-adjusted Price Indexes Are Synonymous with Hedonic Indexes

Virtually all methods used to construct price indexes are designed to measure price change holding quality constant (an exception being unit-value indexes, which are not designed to hold quality constant). Hedonic methods hold quality constant by conditioning on characteristics in a regression analysis. There are other techniques, in addition to hedonic methods, for adjusting for changes in quality. For example, BLS may ask the respondent to provide a dollar estimate of the change in quality embedded in the new product and adjust the reported price for that new product accordingly so that it ‘matches’ the product reported in the previous period. A potential drawback to this technique is that an increase in price may be wholly attributed to the increase in quality. Alternatively, the reported value of the quality change may be based solely on the cost of incorporating that change, without reflecting the value perceived by the customer. Another quality-adjustment technique, recently adopted by BLS to construct the CPI for computers, is referred to as ‘attribute cost adjustment.’ In this technique, monetary values for the attributes that affect price (e.g. processor speed, amount of RAM, storage capacity, etc.) are obtained from original equipment manufacturers (OEM), websites and price compiler sites on the Internet and these values are used as the basis to determine appropriate

---

1 This calculation implicitly assumes that adjustments to percent changes in prices are exactly offset by adjustments to percent changes in quantities.
quality adjustment amounts.\textsuperscript{1} There is evidence that this technique yields results similar to those obtained using hedonic methods. A third technique is to impute a price change for the product that has experienced a change in quality. For this technique to be valid, the statistical agency must develop an imputation strategy that is consistent with the price-setting behavior of the seller. An example of such a strategy is the class-mean imputation method. The class-mean imputation strategy is often used for products where price change is closely associated with periodic introduction of new lines or models (e.g. motor vehicles) and direct quality adjustment is not feasible. With this method, price change is estimated from the price changes of other observations that are going through similar item replacement at the same time and were either quality adjusted directly or were judged directly comparable.\textsuperscript{2}

Traditional matched model methods hold quality constant by carefully specifying each variety in the sample and ensuring that exactly the same variety is resampled each period. There can be problems with matched model methods—the samples may be unrepresentative or out of date, the methods used to handle new or disappearing items may be sensitive to unusual price changes that sometimes occur when an item first appears or disappears, or the decisions made by the statistical agency analyst about how to treat the replacement of items in the sample may be faulty. These problems may lead to either upward or downward biases and hedonic methods are important because they can help address some of these problems. Several researchers—for example, Aizcorbe, Corrado and Doms (2003) & (2000)—have observed that matched model methods using good samples with up-to-date weights may give results similar to, or perhaps better than, hedonic methods.

5.2. **Hedonic Methods Are Opposed to Traditional (Matched Model) Methods**

In fact, most hedonic research at U.S. statistical agencies has led to the opposite conclusion—hedonic research has often led to improvements in sampling methods that have led to better samples, sample replacement strategies, or other improvements in the matched model indexes. The U.S. statistical agencies have found that hedonic analysis is a useful tool, whether used in the background as a guide to application of the matched model methods, or used directly in making quality adjustments for sample items that are being replaced. The articles by Liegey (1993) and Fixler et al. (1999) on the use of hedonics in the CPI for apparel include a number of examples of how hedonic methods can be used in both ways.

5.3. **Hedonic Methods Always Result in a Lower Rate of Price Change**

Many people think of the rapid decline of the BEA and BLS computer price indexes as representative of hedonic methods. The recently developed hedonic index for LAN equipment of Doms and Furman is another example of an item for which application of hedonic methods led to a much lower rate of price growth. On the other hand, several of the BLS hedonic CPI’s have resulted in higher rates of price change—as described above, the hedonic rent indexes were specifically designed to correct for a downward bias. Also, the apparel indexes and, perhaps more surprisingly, videocassette recorders are examples of items for which hedonic methods led to higher rates of price growth.

\textsuperscript{1} For more information on how BLS calculates the CPI for computers, please see http://www.bls.gov/cpi/cpifaccomp.htm.

\textsuperscript{2} For more information, please see http://www.bls.gov/opub/hom/pdf/homch17.pdf.
5.4. **Hedonic Methods ‘Solve’ the Quality Adjustment Problem**

Hedonic regressions are only as good as the data and modeling efforts that go into them. If an important new characteristic has appeared on the market, but is not included in the hedonic regression equation, there is no hope of using the hedonic function to adjust for the improvement in quality. Similarly, just as matched model methods may be biased if samples are out of date or unrepresentative, so also hedonic methods may be biased if estimated using unrepresentative samples. For both hedonic methods and more traditional methods, the statistical agencies must depend on knowledgeable staff who proactively keep track of new products and other market developments.

5.5. **Hedonic Methods Are Prohibitively Expensive**

Our experience has been that hedonic methods do require time and effort. Probably the most significant issue is collecting prices and detailed characteristics for a representative sample of models or varieties. We found that for a number of items—for example, televisions and rental housing—the data already collected by the statistical agency were perfectly adequate for hedonic analysis. For other items, special data collection or purchases of secondary data may be required; as Fixler et al. (1999) describe, the BLS CPI program is now employing both of these approaches specifically for use in the estimation of hedonic regressions. The actual regression analysis itself is often straightforward, and with training should be within the capacity of statistical agency staff in many countries.

6. **Conclusion**

Accurate price indexes are crucial for preparing accurate estimates of real GDP and corresponding productivity measures. The price index must be representative of the market and control for changes in characteristics and/or quality of these goods. Traditional matched model price indexes are well suited for capturing price change for goods that exhibit little or no quality change over time, however, for products whose characteristics and/or quality are changing rapidly (e.g. ICT goods) or are heterogeneous by nature (e.g. custom software), hedonic methods may be more suitable. Aizcorbe et al. (2000 and 2003) showed that matched model indexes constructed from highly disaggregated, high-frequency data can yield a price index that is remarkably close to a hedonic quality-adjusted price index constructed from the same dataset. However, such abundant datasets are not always readily available and therefore a hedonic price index may be preferable.

The role of hedonic methods in measuring real GDP has certainly been an important one and has grown significantly over the past 25 years. The incorporation of hedonic methods has sometimes led to higher rates of price decline (e.g. computers and semiconductors) than the traditional matched model price index, and has sometimes led to higher rates of price change (e.g. apparel, VCR’s and rent). While we are not able to precisely quantify the affect of incorporating improved prices that use hedonic methods, we did approximate the impact that quality-adjusting computer prices (arguably the most prominent hedonic price indexes with the greatest rates of decline) has on measuring real GDP. The results were minimal – on average less than 0.1 percentage point for years 2001-05.
References


## Impact of hedonic price measurement on productivity

### Appendix A – Components of GDP that are deflated with hedonic-type price indexes

<table>
<thead>
<tr>
<th></th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gross Domestic Product</strong></td>
<td>10,469.6</td>
<td>10,960.8</td>
<td>11,712.5</td>
</tr>
<tr>
<td><strong>Total hedonic components</strong></td>
<td>2,240.5</td>
<td>2,324.8</td>
<td>2,527.9</td>
</tr>
<tr>
<td><strong>Percent of GDP</strong></td>
<td>21.4%</td>
<td>21.2%</td>
<td>21.6%</td>
</tr>
<tr>
<td>Computers and peripheral equipment</td>
<td>198.4</td>
<td>201.5</td>
<td>221.4</td>
</tr>
<tr>
<td>PES</td>
<td>77.2</td>
<td>77.8</td>
<td>82.3</td>
</tr>
<tr>
<td>Exports</td>
<td>38.6</td>
<td>39.9</td>
<td>42.8</td>
</tr>
<tr>
<td>Imports(^1)</td>
<td>75.2</td>
<td>76.5</td>
<td>88.6</td>
</tr>
<tr>
<td>Government</td>
<td>7.5</td>
<td>7.2</td>
<td>7.7</td>
</tr>
<tr>
<td>Software</td>
<td>90.4</td>
<td>92.5</td>
<td>98.9</td>
</tr>
<tr>
<td>PES</td>
<td>80.4</td>
<td>82.3</td>
<td>88.6</td>
</tr>
<tr>
<td>Government</td>
<td>10.0</td>
<td>10.2</td>
<td>10.3</td>
</tr>
<tr>
<td>Structures</td>
<td>543.4</td>
<td>584.5</td>
<td>673.5</td>
</tr>
<tr>
<td>Private residential</td>
<td>297.2</td>
<td>344.0</td>
<td>416.3</td>
</tr>
<tr>
<td>Private nonresidential</td>
<td>156.0</td>
<td>148.7</td>
<td>161.1</td>
</tr>
<tr>
<td>State and local government</td>
<td>90.2</td>
<td>91.8</td>
<td>96.0</td>
</tr>
<tr>
<td>Telecommunications</td>
<td>19.2</td>
<td>15.0</td>
<td>14.6</td>
</tr>
<tr>
<td>PES</td>
<td>19.2</td>
<td>14.9</td>
<td>14.5</td>
</tr>
<tr>
<td>Photocopiers</td>
<td>2.4</td>
<td>3.0</td>
<td>2.8</td>
</tr>
<tr>
<td>PES</td>
<td>0.7</td>
<td>0.7</td>
<td>0.2</td>
</tr>
<tr>
<td>Exports</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Imports(^1)</td>
<td>1.1</td>
<td>1.8</td>
<td>2.0</td>
</tr>
<tr>
<td>Audio &amp; Video</td>
<td>52.1</td>
<td>53.1</td>
<td>56.8</td>
</tr>
<tr>
<td>PCE</td>
<td>242.9</td>
<td>247.9</td>
<td>262.4</td>
</tr>
<tr>
<td>Private Residential Equipment Investment</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>Apparel</td>
<td>242.9</td>
<td>247.9</td>
<td>262.4</td>
</tr>
<tr>
<td>PCE</td>
<td>242.9</td>
<td>247.9</td>
<td>262.4</td>
</tr>
<tr>
<td>Household appliances</td>
<td>32.0</td>
<td>32.8</td>
<td>35.3</td>
</tr>
<tr>
<td>PCE</td>
<td>27.0</td>
<td>27.7</td>
<td>29.9</td>
</tr>
<tr>
<td>Private Residential Equipment Investment</td>
<td>5.0</td>
<td>5.1</td>
<td>5.5</td>
</tr>
<tr>
<td>Rent</td>
<td>1,056.8</td>
<td>1,091.6</td>
<td>1,159.0</td>
</tr>
<tr>
<td>PCE</td>
<td>3.0</td>
<td>3.0</td>
<td>3.2</td>
</tr>
</tbody>
</table>

\(^1\) import are included as a positive since we are identifying the total dollar amount that is deflated with hedonic-type price indexes.

PES – Private fixed investment in equipment and software

PCE – Personal consumption expenditures
Measurement of productivity in the non-business sector
PRODUCTION MEASUREMENT IN MULTI-INPUT MULTl-OUTPUT CONTEXTS AND IN ABSENCE OF PRICES

Giuliana BATTISTI
Alfonsina IONA
Emmanuel THANASSOULIS¹
Centre for Performance Measurement and Management (CEPMMA)
Aston Business School
Aston University, Birmingham, B4 7ET, UK

Executive Summary

The issue of productivity attracts a great deal of interest both in academic and practitioner circles. Productivity is deemed to have a direct impact on the well-being of the citizens of a country and as such measuring it and understanding the factors that impact it is of paramount importance. This paper highlights some of the difficulties in using simple measures of productivity, especially in contexts where multiple inputs and multiple outputs prevail and prices are not available or even applicable as in many public sector services. The paper presents a linear programming based method for measuring productivity suitable for multi-input multi-output contexts in the absence of input-output prices which overcomes some of the problems encountered with the more traditional methods of measuring productivity.

Economists define productivity as the amount of output produced relative to the inputs used. Thus in a single-input single-output framework productivity is the ratio of output to input at a production unit. The problem arises when we deal with multi-input and/or multi-output cases where the foregoing notion of average product breaks down. Yet in many studies the most commonly used measure of productivity is ‘labour productivity’ measured as output per worker or per hour worked. Not only is it the easiest measure of productivity (capital and other inputs are much more difficult to measure) but also it is highly correlated with improvements in standards of living. However, labour productivity changes with the amount of other substitutes or complementary inputs (such as capital) and it is difficult to disentangle the effects of labour from those of other inputs on productivity. In short, labour productivity gives only partial information about a firm’s or a country’s overall productivity.

What is required is a measure of total factor productivity (TFP) which reflects all inputs and outputs used. To compute such a measure we need to summarise somehow multiple inputs used into an index measure and the same for the corresponding outputs. One way this can be done is by the use of input-output prices where those are available. However, two main issues arise: the first is whether all goods may be priced, while the second is whether prices are reliable indicators of the market value of goods. Many items, such as health and education, have no prices and/or quality is not clearly defined

¹ Corresponding author.
or reflected in prices. The absence of perfect competition for some goods or services presents a further problem as to whether prices reflect value-added or market power or lack thereof.

Recent research in the field of productivity measurement has advanced the **Malmquist Index** for measuring Total Factor Productivity change in a multi-input multi-output framework. These indices can be applied both in contexts such as the public sector where prices for inputs and/or outputs are not available and in contexts where such prices are available. Malmquist indices can be computed using **Data Envelopment Analysis (DEA)** which is a linear-programming based method for measuring efficiency and deriving other information such as performance targets, benchmark operating units which may prove examples of good operating practices, effects of scale size on productivity in multi-input multi-output contexts and so on.

We will use Figure 1 to illustrate the Malmquist Index. The figure shows for fictitious firms Water delivered (WDELA) and length of main (LEN) supported per £1000 of operating expenditure in periods t and t+1 respectively. Using the assumptions underlying DEA a *production space* is constructed from the observed data in period t and period t+1 respectively and the boundary of each space is shown. The production space contains all possible combinations of WDELA and LEN per £1000 in each period.

**Figure 1 – Measuring productivity change between period t and t+1.**

Consider now a company operating at F in period t and G in t+1. Its efficiency in period t is OF/OB reflecting its attained relative to its maximum possible output levels. The Malmquist index for the change in the productivity of this company between period t and t+1 is the **geometric mean of its efficiencies relative to the period t and t+1 boundaries respectively**. Relative to the period t boundary the efficiency of F is $\frac{OF}{OB}$ and of G $\frac{OG}{OE}$ and so one of the ratios to go in the geometric mean is $\left[ \frac{OG}{OE} \frac{OF}{OB} \right]$. In a similar manner taking F and G relative to the t+1 boundary we derived the second ratio to go in the geometric mean as $\left[ \frac{OG}{OI} \frac{OF}{OH} \right]$. Thus, for the company operating at F in t and G in t+1 we have

$$
\text{Malmquist Index} = \left[ \frac{OG}{OE} \frac{OF}{OB} \right] \times \left[ \frac{OG}{OI} \frac{OF}{OH} \right]^{0.5}.
$$
Note that the above index by algebraic manipulation can be decomposed so that

\[
\text{Malmquist Index} = \left( \frac{OG}{OI} + \frac{OF}{OB} \right) \times \left( \frac{OI}{OH} \frac{OH}{OB} \right)^{0.5} = \text{Efficiency change x Boundary shift.}
\]

The efficiency change term \( \left( \frac{OG}{OI} + \frac{OF}{OB} \right) \) if over 1 would suggest the unit has moved over time closer to the efficient boundary, or if under 1, it has fallen further behind the boundary. Similarly, the boundary shift term \( \left( \frac{OI}{OH} \frac{OH}{OB} \right)^{0.5} \) when over 1 (as in Figure 1) would suggest technology improvement in that the efficient boundary has shifted to a more productive location over time. Note that the boundary shift is the geometric mean of the shift at the output mixes in periods \( t \) and \( t+1 \).

The foregoing graphical illustration is operationalised by means of DEA. Software exist (e.g. www.deasoftware.co.uk) which yield directly the Malmquist Index of productivity change once the input output data are imported for the operating units. Moreover, the Malmquist index can be further decomposed to also reflect the impact, if any, of scale size changes between periods \( t \) and \( t+1 \).

In conclusion, traditional measures of productivity are problematic as they reflect partial productivity or rely on input-output prices. One way to overcome their problems is to use Malmquist indices which measure total factor productivity change over time in multi-input multi-output contexts without recourse to prices. Further, the indices can be decomposed to reflect the unit’s own change in efficiency, the movement of the efficient boundary and scale size effects.

**Abstract**

This paper addresses the issue of measuring productivity. It suggests that more research needs to be carried out to identify the drivers of the productivity growth, other than capital and labour, such as operational or human resources, management practices, organizational structures and ICT investments. Data of this kind are usually collected via ad hoc independent surveys at firm level as no official international source collects them. Therefore, accurate and harmonised longitudinal firm level data are needed to replace those from ad hoc small independent surveys. The paper also highlights some of the problems associated with the more traditional measures of productivity, notably due to the fact that they tend to be partial measures of productivity (e.g. labour productivity) or that they make recourse to input or output prices which may be inaccurate or not exist at all. The paper suggests the use of a Malmquist index to measure productivity changes over time. The advantage of such an index is that it can be computed without recourse to prices and it can be used in contexts where there are multiple inputs and multiple outputs. Thus, this index is seen as particularly suited for not-profit activities, including those provided by central and local government.

**Key words:**

Productivity, efficiency, Malmquist index
1. Introduction

The issue of productivity attracts a great deal of interest both in academic and practitioner circles. Productivity is deemed to have a direct impact on the well-being of the citizens of a country and for this reason measuring productivity and understanding the factors that impact it is of paramount importance. This paper highlights some of the pitfalls in using simple measures of productivity, especially in contexts where multiple inputs and multiple outputs prevail and prices are not available or even applicable as in many public sector services. The paper presents a linear programming based method for measuring productivity suitable for multi-input multi-output contexts in the absence of input-output prices. From this perspective, thus, the method overcomes some of the problems encountered by more traditional methods of measuring productivity.

A great deal of research has focused on productivity differentials and has attempted to understand the sources of productivity differences between different countries and across sectors (McKinsey 1998, O’Mahony 1998, O’Mahony and De Boer 2002, Broadberry and O’Mahony 2004, Griffith et al 2003, ESRC Seminar Series 2004, etc.). Is it the result of larger markets and organizations in some countries, and the economies of scale that thereby become available? Is it the result of regulatory regimes, e.g., in the ease with which companies can increase and reduce labour costs? Is it the result of different cultures regarding innovation and entrepreneurship? Is it a function of differential commitment to, and spend on, research and development? Or perhaps there exist different patterns of economic clusters of activity? Or alternatively, is the nature of the relationships between Universities and the private sector important? There are many competing explanations and a great deal of work has been undertaken at this relatively macro level of analysis (Porter and Ketels 2003).

Increasing attention is also now being paid to differences at the firm level, i.e., at a micro-level of analysis. Many studies underline that differences at firm level may be a function of how companies are managed. Indeed, according to several studies (Porter and Ketels 2003, Huselid 1995, Ichniowski et al. 1997, Leseure et al. 2004, Edwards et al. 2004) the use and the effectiveness of modern management practices\(^1\) may lead to a more efficient organisation or to multi-factor productivity change, as it happened in the US during the 1990s (Black and Lynch 2004).

Whilst on the one hand productivity differences exist across sectors and countries, on the other hand, the ongoing debate argues that these differences may be due to how productivity is measured, what exactly it is measured and how cleanly the various sectors can be delineated and compared. In fact, Statistical agencies and Academics use different measures of productivity and various methods to incorporate in productivity the effects of some important factors concerning firm organization and characteristics that are not systematically collected by any statistical source (i.e., management practices, organizational changes, skills and knowledge, etc.). We argue that the lack of harmonization of definitions and measurement methods, related to productivity, is generating results that are not necessarily comparable and, hence, an in depth analysis of measures and concepts is needed not only to understand the real level of a country’s productivity, but, most

---

\(^1\) See for example Battisti and Iona (2006) for a review of the literature on the impact upon firm performance of a range of practices, such as operational practices (Information and Communication Technology, Total Quality Management, Business Process Reengineering, Supply-Chain Partnering and Just-in-Time, etc.) and human resources management practices (such as Empowerment, Incentive pay-schemes, Team-based working and Learning culture including Skill development, etc.).
importantly, also to compare productivity across countries and derive important insights for the policy makers.

As a result, in this paper we outline some of the traditional ways of measuring productivity and we suggest a way to overcome some of their inherent drawbacks. The next section, numbered 2, reviews traditional ways of measuring productivity and highlights some of their drawbacks. Section 3 suggests ways of measuring productivity in a multi-input multi-output context when prices are not available, such as in the case of public services like education or health. Finally, Section 4 concludes.

2. Traditional Partial Productivity Measures and Associated Problems

Economists define productivity as the amount of output produced for inputs used. Such a concept is quite easy in a single-input single-output framework: it is the ratio of output to input at a production unit. The problem arises when we deal with a multi-input and/or multi-output case, where the foregoing notion of average product breaks down. Yet in many studies the most commonly used measure of productivity is ‘labour productivity’ measured as output per worker or per hour worked. Not only it is the easiest measure of productivity (capital and other inputs are much more difficult to be measured) but also it is the most highly correlated with improvements in standards of living. For economists, labour productivity is the key indicator of a country’s economic health and, over the long run, real income growth and hence standards of living must follow the labour productivity growth (Porter and Ketels 2003). Even Statistical Agencies use, as a measure of productivity, labour productivity according to the idea that long run growth depends on increases in employment rather than on intensity of work (HM Treasury 2000).

However, this measure of productivity presents a main drawback: since labour productivity changes with the amount of other substitutes or complementary inputs (such as capital), it is difficult to disentangle the effects of labour from those of other inputs on productivity. Labour productivity gives only partial information about a firm’s or a country’s overall productivity. Economists have attempted to overcome this problem by using a measure of productivity able to distinguish and take into account the contributions of all factors to the production. This measure of productivity is the Total Factor Productivity (thereafter TFP).

To compute the TFP of a production unit we need to summarise the multiple inputs used into an index measure and the same needs to be done for the corresponding outputs. One way to do this is by using the input-output prices, where those are available. However, even this measure of productivity presents a number of concerns as its accuracy will strongly depend on the quality of information available on prices of output and inputs and on the way in which outputs and inputs are measured.

In particular, economists and statisticians measure output by sales or value added per worker, or per hour worked, deflated by a price index. In other words, they sample a large number of items and see how prices change over time. Therefore, price is crucial in the measurement of productivity – and in comparing it across sectors – because it provides a measure of the value or information about the quality of goods sold. Nevertheless, that is not the entire story because two main questions still remain: the first is whether all goods may be priced, while the second is whether prices are reliable indicators of the market value of goods.
Regarding to the first question, measuring and comparing productivity may be problematic because in some sectors goods have no prices. That is, although sampling a large number of items and seeing how prices change over time may be relatively easy to do for say the manufacturing sector – where tangible goods provide more quantifiable information – problems can arise in defining the value of output in the services sector – where goods are more intangible – as in this sector many items, such as health and education, have no prices (Reynolds et al. 2005). It is also difficult to measure quality in this sector: defining the specification of the product as tightly as possible helps, but it may not be possible to match exactly across firms or countries.

Moreover, as pointed by Pilat (2005), what is crucial in computing the aggregate productivity of a country is the concern about the measurement of productivity of the retailing sector, not only because it is among the largest service industries, accounting for between 10% and 18% of value added in high-income OECD economies, but also because it is a sector where the available measures of productivity point to considerable variation across OECD economies. Like banking, its output in the national accounts is calculated as a margin, i.e. the difference between the value of products bought and products sold. But its treatment differs from other service industries, as there is no explicit consumption of retail services in the national accounts. Any such consumption is included in the value of the goods and services that are sold by the retail sector. This implies that any observed differences in the volume growth of retail services will thus affect the relative performance of industries, but not necessarily the aggregate productivity measures.

In European statistical practice (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. According to Pilat (2005), this practice ignores all changes in the quality of distribution services provided, by leading to a mis-measured volume of retail services. Often, because of this, in the business services sector, products are measured in inputs. Even in this case, comparing productivity across sectors may be problematic if the measurement of labour inputs is inadequate or different between sectors (Reynolds et al. 2005). For example, Reynolds et al. (2005), in explaining the heavy contribution of the retail sector to the low UK labour productivity, shows that there are significant measurement errors in the computation of the output and labour input in the retail sector. These measurement problems are due both to the fact that it is difficult to measure output and inputs accurately and to the fact that most of the extant studies are based on labour productivity rather than on TFP.

Regarding the second question (whether prices are reliable indicators of the market value of goods), recent studies in the retail sector (Griffith et al. 2005, Griffith and Harmgart 2005) argue that, whilst in competitive markets prices are supposed to accurately reflect the quality and cost of producing goods, in non competitive markets prices are usually overstated – they reflect not only the value of output or input but also the market power. If this happens in the output market, it will lead to an overstatement of productivity, and if prices are set non-competitively in the inputs markets they will lead to an understatement of productivity. Thus, comparisons of productivity across sectors may be biased.

In measuring productivity, other concerns arise when measuring productivity in the public sector. Some recent studies argue that assessing correctly the productivity performance in the public sector is crucial given that it has both a direct and indirect impact on a country’s aggregate productivity. The indirect impact arises from the fact that the public sector makes up a large share of the whole economy; whilst the indirect impact rests on the fact that the services produced by the public sector,
as education or health, have spill-over effects on the productivity of the private sector. However, measuring productivity in the provision of public services is quite difficult because public services are hardly ever sold in the open market, so it is not possible to observe market prices.

Also, often the measurement of public services’ price is susceptible to significant errors. For example, Berndt et al. (2000) show that there are important biases in the measurement of medical price changes and that these biases are large enough to have measurable consequences for the official overall price indices such as the GDP deflator and all-item Consumer Price Index. This, in turn, makes it difficult to value the output in this sector. Moreover, it is often difficult to define precisely the output in itself of many public services. For example, how to measure a healthy population, given that health may be affected by many factors such as dietary, demographic and socio-economic factors?

From a macroeconomic perspective, some studies have shown that the productivity measure may be biased if outputs or inputs are not accurately measured under waves of ICT investments. For example, Basu at al. (2003) show that when new technologies take place the need of organizational changes arises – organizational change that may be modelled as the accumulation of intangible complementary capital. This means that the firm is producing a stream of intangible output that constitutes gross investment in complementary capital. The problem is that some of this output is not typically measured in the national accounts, and contemporaneously investment in ICT may be associated with lower TFP as resources are diverted to re-organizational changes and learning. For example Violante (2003), that analysed the lack of productivity growth in the United Kingdom (and the high productivity growth in the United States), has found that during periods of strong investment in IT mostly output is unmeasured and so the true TFP growth may be underestimated. In contrast during periods where the economy has large stocks of IT and complementary capital inputs may be grossly under measured, and the true TFP growth is overestimated.

Concern has also been expressed by Battisti and Iona (2006) about the lack of harmonised official statistical sources for certain non capital inputs, such as organizational and management practices, ICT investments or management skills. Despite being key factors associated with the IT revolution and therefore the productivity growth of a country (see for example the milestone work of Milgrom and Roberts 1995 or Oulton 2002), they are not systematically collected by any statistical source. Consequently, most of the existing studies tend to rely on independent surveys carried out over small cross sections of firms and to use different definitions of the issue to be analysed. Battisti and Iona (2006) also argue that because of the wide range of definitions and measurements of productivity, and because of the different methodologies – parametric such as econometrics based and non-parametric such as linear programming methods – used to estimate the link between productivity and its main drivers, results arising from the empirical research are ambiguous and make it difficult to compare productivity across firms or countries over time and derive conclusions.

---

1 In the current literature productivity at the firm level is measured not only as labour productivity and TFP but also as Return on investment and sometimes it is proxied by measures of profitability, as for example ROCE (returns on capital employed), Tobin’s Q, net cash flow, GRATE (gross return on capital), quality, lead-time, and customer’s services. Similarly, management practices are measured with a wide range of practices. For example, JIT/TQM practices are measured with 17 different practices. Human Resources Management Practices are measured as incentive pay schemes, recruiting practices, work teams practices, flexible job assignment, skills and labour-management communications practices, or employee’s motivation practices. ICT or computerization is measured as the share of computer capital in the total capital stock, or as office, computing and accounting equipment for employee (see Battisti and Iona 2006).
Measurement of productivity in the non-business sector

Recent research in the field of productivity measurement has advanced methodologies for assessing Total Factor Productivity changes in a multi-input multi-output framework, using Malmquist indices. These indices can be applied both in contexts where prices for inputs and/or outputs are not available – such as the public sector – and in contexts where such prices are available. Therefore, the next section outlines how Malmquist indices can be used for measuring productivity change even when input output prices are not available.

3. The Malmquist Index for Measuring Productivity Change

Malmquist Indices are used to measure Productivity Change over time. As noted earlier productivity in the single input – single output case is measured by the average number of units of output per unit of input. In the multi – input multi – output case the notion of average product breaks down. However, since its introduction (Fare et al (1989) the Malmquist index has become one of the methods of choice for measuring productivity change over time in a multi-input multi-output context.

Malmquist indices can be computed using Data Envelopment Analysis (DEA). DEA is a linear-programming based method for measuring efficiency in multi-input multi-output contexts. A brief introduction to DEA can be found in the Appendix while a fuller introduction can be found in Thanassoulis (2001).

In what follows we use a graphical example to illustrate how a Malmquist Index of productivity can be computed and then decomposed.

Consider assessing productivity change over time of water companies on water distribution and assume for ease of illustration that water distribution is characterised by a single input – Operating Expenditure (OPEX) and two outputs – Water delivered (WDELA) and length of main (LEN). The two outputs reflect the quantity of water delivered and the geographical spread of the customer base served.

For simplicity we assume that water distribution is characterised by constant returns to scale (CRS). Therefore we can divide the output levels by OPEX to normalise OPEX to some unit such as £000 and so assess companies on efficiency by simply comparing them on their two outputs for a given standard level of input of £1000. Let the dots in Fig 3.1 represent the output levels of each company per £1000 of OPEX. Let us assume that we have data for two time periods, t and t+1 plotted within Figure 3.1. Using the DEA assumptions (and in particular that interpolation between companies yields other feasible in principle companies (see the Appendix) we create the production space enclosed by the LEN and WDELA axes and the piece-wise linear boundary ABCD and its extensions from A and D to the axes. We should draw a separate production space for each year t and t+1 but initially, to simplify matters, we assume that the two spaces turn up to have exactly the same efficient boundary as depicted in Figure 3.1.
Each company’s efficiency in each one of the years $t$ and $t+1$ can now be measured relative to the boundary. Consider now some company which operates at point $F$ in year $t$ and $G$ in year $t+1$. How has its productivity changed between period $t$ and $t+1$?

The company’s productivity growth is measured by the ratio of its year $t+1$ to its year $t$ efficiency.

Using DEA (see the Appendix) we conclude that the efficiency of company $F$ in year $t$ is given by the ratio of its observed to its maximum outputs and this ratio is $\frac{OF}{OB}$ in period $t$. Similarly the company’s efficiency in year $t+1$ is $\frac{OG}{OE}$. Thus its productivity growth is $\frac{OG}{OE} \div \frac{OF}{OB}$.

If this ratio is less than 1 the company’s productivity is lower in $t+1$ than in $t$ because the company is further from the efficient boundary in $t+1$ than it was in $t$ while the boundary itself has not moved. If the ratio is over 1, productivity has risen with time and if the ratio is 1 productivity has not changed.

Let us now take the more realistic and general case where not only companies move relative to the boundary over time but the boundary itself also moves. The situation that could arise is depicted in Figure 3.2.
To measure productivity change now we still use the ratio of the DEA efficiencies of the company when it is operating at F and at G. However, we now compute this ratio once relative to the t boundary and once relative to the t+1 boundary. The geometric mean of these two ratios is a measure of the productivity growth of the company in the same way the individual ratio was when the boundary had not moved.

Thus our measure of productivity growth for company F in Figure 3.2 is

\[
\left( \frac{\text{Ratio of DEA efficiencies of G to F relative to t boundary}}{\text{Ratio of DEA efficiencies of G to F relative to t+1 boundary}} \right)^{0.5} = \left( \frac{OG + OF}{OE + OB} \right) \times \left( \frac{OG + OF}{OI + OH} \right)^{0.5}
\]

This geometric mean is known as a Malmquist productivity index.

After algebraic manipulation the index can be decomposed as follows:

**Malmquist Index from F to G**, Fig. 3.2 = \[
\left( \frac{OG}{OI} \times \frac{OF}{OB} \right) \times \left( \frac{OI}{OE} \times \frac{OH}{OB} \right)^{0.5}
\]

Note that the first bracketed term \( \frac{OG}{OI} + \frac{OF}{OB} \) is the ratio of the period t+1 to the period t efficiency, reflecting the movement of the company relative to the moving boundary. The second bracketed term contains two ratios. The first \( \frac{OI}{OE} \) reflects the movement of the boundary between period t and t+1 at the mix of outputs of the firm in period t+1. The ratio \( \frac{OH}{OB} \) reflects the boundary movement at the output mix in period t. Thus the second bracketed term reflects the geometric mean of the movement of the boundary between period t and t+1.

We can generalise the foregoing approach by means of linear programming for multi-input multi-output situations so that the Malmquist index is as follows:

**MALMQUIST PRODUCTIVITY INDEX**

\[
\text{MALMQUIST PRODUCTIVITY INDEX} = \left( \frac{EF_{T+1}^{D_{t+1}}}{EF_{T}^{D_{t}}} \right) \times \left( \frac{EF_{T+1}^{D_{t+1}}}{EF_{T+1}^{D_{t+1}}} \right)^{1/2} (1)
\]

where, the notation \( EF_{T+1}^{D_{t+1}} \) stands for DEA efficiency of period t+1 data relative to the period t boundary of the production space and so on. The overall index value of 1 means constant productivity. A value over 1 means productivity growth and under 1 productivity regress.

This decomposition of the index in two parts is interpreted as follows:

\[
\frac{EF_{T+1}^{D_{t+1}}}{EF_{T}^{D_{t}}}: \text{This term, known as EFFICIENCY CATCH UP, reflects how much closer the company is to the period t+1 boundary compared to its position relative to the period t boundary.}
\]
This term known a **BOUNDARY SHIFT**, measures the distance between the period t and period t+1 boundaries as the geometric mean of the distances of the boundaries taken at the mix of output levels of the company in period t and period t+1 respectively. (This allows for the fact that the boundary may register varying shifts at different ‘mixes’ of driver levels. (The ‘mix’ is defined by the proportions drivers levels are to each other.))

The values of these two components when over 1 represent progress and under 1 regress. When they are 1 they represent no change.

Productivity change can be however not only due to efficiency change or boundary shift as reflected above, but also due to changes in scale economies. It is possible to decompose the above Malmquist productivity further to reflect changes in scale economies. Where constant returns to scale do not hold it is possible to use DEA to compute efficiencies **under variable returns** to scale (VRS). These efficiencies can never be lower than under constant returns to scale since the assumption of constant returns to scale (CRS) is more restrictive than that of variable returns to scale.

The ratio \([\text{CRS efficiency}] / [\text{VRS efficiency}]\) is known as the scale efficiency of a production unit.

Thus, let \(V_{EF}^{Dr_t}\) and \(C_{EF}^{Dr_t}\) be respectively the DEA efficiency of the data of period t relative to the variable and constant returns to scale boundary of period t. Then the Malmquist index of some unit \(j_0\) can be decomposed as follows:

\[
MI_{j_0} = \frac{V_{EF}^{Dr_{t+1}}}{V_{EF}^{Dr_t}} \times \frac{SC_{EF}^{Dr_{t+1}}}{SC_{EF}^{Dr_t}} \times \left[ \frac{C_{EF}^{Dr_{t+1}}}{C_{EF}^{Dr_t}} \times \frac{C_{EF}^{Dr_{t+1}}}{C_{EF}^{Dr_{t+1}}} \right]^{1/2}
\]

\(\uparrow\) Malmquist Index \(\uparrow\) ‘Pure technical efficiency catch up’ unit \(j_0\)
\(\uparrow\) ‘Scale efficiency catch up’ unit \(j_0\)
\(\uparrow\) ‘Boundary shift’ unit \(j_0\)

Where \(SC_{EF}^{Dr_t}\) is the scale efficiency of the unit defined as \(SC_{EF}^{Dr_t} = C_{EF}^{Dr_t} / V_{EF}^{Dr_t}\). Notation \(Dt\) and \(Tt\) is introduced above.

Given that DEA efficiencies can be computed in a multi-input multi-output context and without recourse to input or output prices this makes the Malmquist index as described in this section a very practical tool for measuring productivity change over time and identifying its components attributable to efficiency change, boundary shift and scale effects. Where input or output prices are available it is possible to compute a Malmquist index which will also reflect allocative efficiency changes. This index is beyond the scope of this paper but the interested reader may find details in Maniadakis and Thanassoulis (2004).
4. Conclusion

In this paper we have briefly reported on the debate about the sources of productivity differences across countries and across sectors. We have done so by focussing on the issue of productivity measurement that becomes crucial when productivity comparisons across sectors or countries are conducted. We believe that productivity comparisons are useful not only to inform us about whether a country is performing well but also to inform on the determinants of productivity growth (or slow down) and therefore how to target policy effectively.

We have underlined how differences in measured productivities do not necessarily reflect underlying differences in productivity both because of the several definitions and measurement methods of productivity and because of the several problems arising from the measurement of inputs and outputs. In fact, factors such as organizational structure and management practices at firm level or ICT investments are the factors currently recognised to be responsible for the productivity growth/slow down. However, despite their importance they are usually collected via ad hoc independent surveys and do not appear in harmonised standard official (international) statistics. This suggests that, on one hand, more effort should be put into the identification of the drivers of productivity growth (other than capital and labour) and, on the other hand, more accurate and harmonised longitudinal firm level international data is needed.

On the methodological side, it is increasingly argued how important it is to revise how inputs and outputs are combined and measured in productivity studies, especially for the non business sector. Whilst measuring productivity by means of the TFP allows the research to overcome all problems related to the traditional partial measure of productivity – namely Labour productivity – on the other hand, this measure presents the drawback of being strongly dependent on the quality of information available on prices. Therefore, given the important role of ‘price’ in measuring productivity, and the fact that not all goods may be priced, a Malmquist Index, commonly used in non-parametric performance measurement is proposed to measure productivity in a multi-input multi-output context when prices are not available such as is often the case in public services.
Appendix 1 – Brief Introduction to DEA

The seminal paper in DEA was by Charnes et al. (1978) who operationalised a definition of relative efficiency by Farrell (1957). Data Envelopment Analysis (DEA) measures the efficiency of a production unit by first estimating the maximum output levels feasible for the resource levels (inputs) that the unit is using or alternatively the minimum resource levels necessary to secure its output levels. If the former (i.e., output maximisation) orientation is adopted the ratio of the attained output levels to the maximum output levels estimated gives a measure of the efficiency of the unit. Likewise if the input orientation is adopted, the ratio of the minimum to the actual input levels gives a measure of the input efficiency of the unit. Where multiple outputs exist the ratio of corresponding maxima to actual outputs are computed and the maximum of these ratios is used as a measure of the output efficiency of the unit. Input efficiency is similarly measured by the maximum of the input ratios computed.

In order to compute the foregoing measures of efficiency what is needed is the estimation of the maximum output levels feasible for the input levels of a unit or alternatively the minimum input levels feasible for the output levels of the unit. This is done by building a feasible production space and then estimating the maxima or minima output or input levels respectively within that space. Though DEA is primarily suitable for multi-input multi-output contexts it is easier to illustrate graphically how it works by using a single-input single-output example.

Let us assume that the units denoted U1…U6 in Figure 1 use a single resource (e.g., expenditure) to secure a single output (e.g., number of students taught). To estimate either the minimum resource its output level would justify, or alternatively, the maximum output level its resource level would justify we assume that interpolation between observed units leads to other units feasible in principle even if not actually observed. For example all resource–output combinations on the linear segment U1U2 in Figure 1 represent interpolations of the units U1 and U2. It is assumed that the interpolation ‘units’ can be observed in principle. Using the interpolation assumption we construct a ‘production space’ as illustrated in Figure 1 which contains all ‘units’ feasible in principle even if not observed in reality. The boundary U1U2U3U6 of the production space corresponds to maximum output for any given level of resource or alternatively to minimum resource for given output level, and so it is referred to as the efficient boundary. Using the efficient boundary as a reference we can measure the efficiency of each unit. For example the minimum resource level justified by the output level of unit U5 is at point B on the efficient boundary.

Figure 1 – Measuring Efficiency by DEA
Thus the unit’s resource level can be contracted to the fraction $AB/AU_5$ and this gives a measure of the unit’s efficiency.

DEA is a generalisation of the approach sketched above with reference to Figure 1. The generalisation of the production space to multiple inputs and multiple outputs requires the use of linear programming. Apart from measures of efficiency, DEA makes it also possible to identify role model units such as those on the efficient boundary in Figure 1, whether a unit operates under increasing, constant or decreasing returns to scale and to address issues of most productive scale size and productivity change over time.

References

2. Basu S., Fernald J.G., Oulton N. and Srinivasan S. (2003), ‘The case of the missing productivity growth: or, does information technology explain why productivity accelerated in the US but not the UK?’, *NBER Macroeconomic Annual*.


PRODUCTION, PRODUCTIVITY AND COST PRICE OF PUBLIC SERVICES IN THE NETHERLANDS

Bob KUHRY

Social and Cultural Planning Office of the Netherlands

Executive Summary

The Social and Cultural Planning Office of the Netherlands has been involved in measuring the production of and demand for public services for more than 20 years. This line of research aims to generate information to facilitate the allocation of public funds by ministries and parliament. In the course of time, methods have been improved. Unfortunately, this is not the case with the quality of the available data. Moreover, the subject of production and productivity measurement is evasive since it is difficult if not impossible to evaluate product quality.

The second paragraph of this paper addresses measurement problems. The first and most vital step in evaluating the topic of this conference, productivity in the non-business sector, is the measurement of production itself. Next, production may be compared with the input volume to obtain an assessment of developments in productivity. Due to the absence of a market price mechanism in the public service sector, production can often only be evaluated in terms of physical quantities, such as the number of services rendered, or even the number of consumers of services. This may be related to the total exploitation expenditure or to the volume of inputs. Subsequently, an analysis in terms of partial or total factor productivity can be carried out, or alternatively, an analysis in terms of the relative cost price of products.

Measurement of production is very difficult in the case of (semi-)collective goods such as public administration in the narrow sense or defence. It is more feasible in the case of individual goods or collective goods in fields such as infrastructure or police and justice. In the case of final services, production can be measured either by value indicators or by physical indicators. In the first case, use is made of deflated costs or revenue figures. In the second case, counts of the number of performances or users can be used as product indicator. Even in the case of individual services, problems arise when assessing the heterogeneity, quality and effectiveness of the services involved. Especially a proper assessment of product quality may impose serious barriers on the interpretability of the outcomes. As a result, the use of a concept like labour productivity may become debatable.

The third paragraph discusses the outcomes of analyses on final services delivered to the public. Fields covered include health care, education, police and justice, social services, public transport and some segments of culture and recreation. Around 50 separate task fields are involved in this analysis. Of the public expenditure on final services, more than 90% is accounted for in this analysis.
In the period 1990-2003 the average yearly increase of the cost of these services, corrected for the price index of the gross domestic product, was 3.3% in the Netherlands, higher than the corresponding growth of the market sector (2.2%). This difference has especially accumulated in the recession years 2001-2003, when stagnation in the market sector occurred, but the public service volume was hardly affected. However, production in the public service sector increased only by 1.7% in contrast to the 2.7% production growth in the market sector. The difference is caused by an increase of the relative cost price of public services (1.6%) with respect to the average domestic product. At the same time, the relative cost price of market goods decreased with 0.5%. The reasons for the increasing relative cost price of public services is in part explained by the law of Baumol: a relative small increase of labour productivity, accompanied by a considerable increase of wages and of the price and volume of material inputs. One of the factors involved is the increasing average age of the labour force in public service.

The fourth paragraph discusses an application involving the measurement of the total production of the municipal government layer in the Netherlands. Municipalities have a wide variety of tasks including education, social services, cultural and recreational services, environmental services, housing, infrastructure, public transport etc. etc. Apart from final services, intermediate services and purely collective services play a part. In the analysis presented, about 60 task fields and more than 100 individual products are distinguished. An attempt has been made to incorporate all these products in a single framework of analysis. The method is essentially implemented using available key data on public services. The production of intermediate services is indirectly measured by reference to the production of the final services at which they are aimed. In other words, the costs of the intermediate services are consolidated into the costs of the corresponding final services. Most difficult is the inclusion of purely collective services. Two alternative solutions are presented: estimation of production by norm indicators (number of inhabitants, square miles etc.) or estimation by deflating the costs with a constructed price index. At the macro-level, municipal production decreased slightly over the period 1995-2003, mainly as a result of privatisation of services and changes in the tasks of municipalities. In the same period, an increase of the relative cost price of municipal services is observed with 1.3% per year. Over the whole period, labour productivity hardly increased whereas salaries, material and capital expenses showed a considerable increase. Results at the macro-level appeared to be quite robust to changes in the selection of product indicators.

Due to the problem of adequately capturing product quality in physical production measures, part of the increase of the relative cost price of public services might actually be related to an improvement of product quality. Nevertheless, the fact remains that the cost of these products per item or per consumer is increasing steadily and the increase of productivity is relatively slow, facts which have to be taken into account when planning the required volume and estimated costs of public services.

**Address:**
Dr. Bob Kuhry
Social and Cultural Planning Office of the Netherlands
PO Box 16164
2500 BD The Hague
Tel + 31 70 3407936
Fax + 31 70 3407044
E-mail b.kuhry@scp.nl
The paper is based on two recent reports:


1. **Introduction**

In the Netherlands, a number of planning offices has been established. These are governmental research institutions with a more or less independent status. The oldest one, the so-called Central Planning Office, has been founded shortly after the Second World War. The Social and Cultural Planning Office dates back to the seventies. Whereas the Central Planning Office has an economic orientation, the Social and Cultural Planning Office focuses at the welfare of citizens, using sociological as well as economic methods. The term ‘Planning Office’ is somewhat misleading. Even in the past, the goal has never been planning in the strict sense, but rather the analysis of social developments, policy evaluation and forecasting.

The Social and Cultural Planning Office of the Netherlands is involved in measuring the production and productivity of public services for more than 20 years.¹ This line of research aims to generate information to facilitate the allocation of public funds by ministries and parliament. Apart from an evaluation of production and productivity, a forecast of future demand for services plays a part in these analyses. A memorandum on the outcomes is presented each time a new government is formed. A more recent line of investigation, which is partly related to the first one mentioned, is the monitoring of the production of municipalities on the request of the Dutch Ministry of Interior Affairs. The emphasis on allocation problems means that an ‘helicopter view’ is required, since it should be possible to evaluate the merits of alternative modes of utilisation of scarce means. In this particular macro-level approach, emphasis is therefore not given to a detailed analysis of aspects of separate production processes, but to a broad characterization of overall relationships.

2. **Measuring public sector performance**

Any production process involves the resources deployed (=input), the process itself (=throughput), the product (=output) and effect (=final outcome).

The relationship between resources deployed and output delivered provides an insight into productivity, and the relationship between the output delivered and the objectives achieved reflects the effectiveness of the production process (see Figure 1).

¹ For our purpose, the public service sector is defined as an aggregate of public functions, which more or less correspond to industrial classes: public administration, defence, administration of social security, education and research, health care (but including the distribution of pharmaceutical products), social services, culture and recreation, housing and public transport. It should be noted that the definition of this sector is independent of the legal status of the producer, which may be public, private non-profit or commercial. Actually the share of commercial organisations in the Dutch public service sector was 20% and the share of private non-profit organisations 39% in 2002 (Kuhry and Pommer, 2006).
In the private sector, the production volume can be derived from the market value of the goods in question. Time series can be obtained by adjusting the result using a relevant price index to get value indicators.

Since services produced by the public sector are not generally traded in markets, their market value is usually unknown. In most cases, therefore, physical production indicators are used. This term refers to various types of indicator that can be used as a direct or indirect measure of production. They include:

1) performance indicators, which refer to the final product,
2) consumption indicators, which refer to the consumers of the services,
3) process indicators, which refer to the activities performed or interim products produced.

Performance indicators refer to the final product of service providers. As such, they are best suited for assessing the efficiency of services. Examples include the number of patients treated successfully, the number of pupils or students finishing their studies successfully in education and the number of concert performances.

Consumption indicators refer not so much to production in itself, as to the number of people consuming the products. They are therefore suitable for analysing and forecasting the demand for services. Examples include the number of hospital patients, the pupil numbers in schools and audience numbers at concerts. Evidently consumption indicators of use are only an indirect measure of the real production. However, due to severe regulations concerning the services in question, the number of pupils and the number of patients admitted are in fact a reasonable proxy for the production volume. If the number of hospital admissions or school pupils is chosen as the measure of production, it is implicitly assumed that the likelihood of recovery and examination pass rate are constant.

Process indicators can refer to tasks performed, and thus are a measure of the efforts of the staff concerned. Examples include the number of operations performed in hospitals, or the number of lessons taught in schools. In some cases, process indicators can be used as an indicator of production. In the case of home care, for example, production can be measured in terms of number of staff contact hours. This assumes that staff perform a constant amount of work per contact hour.
These indicators refer to the delivery of private goods and services to end users. It is often impossible to create an adequate performance indicator for purely public goods, which cannot be related to individual consumers. Exceptions are:

1) services such as police and justice, where the number of crimes solved, penalized offences, acts of assistance, convicted criminals and so on can be used as indicators of the production volume,
2) social security organisations, where the number of clients or the number of cases treated can be used as indicators of the production volume,
3) expenditure for infrastructure, which can be transformed into volume figures by using appropriate price deflators.

However, it is not easy, if not impossible to measure the performance delivered by services like public administration in the narrow sense and defence. However, some methods of analysis are available: the resources deployed (staff, money) for these purposes can for example be compared with total domestic production. This indicates what proportion of its production capacity a country sacrifices to maintain the services in question. Also, the resources deployed can be compared with certain norms, such as the size of the (relevant) population.

**Total productivity** refers to the relationship between the volume of production and the volume of resources deployed. **Labour productivity** concerns the relationship between production and the number of staff deployed. **Actual cost per unit product** is also important. This number is obtained by adjusting the total costs using a generic price index, and dividing the result by the production volume. The price index for household consumption, GDP or national spending can be used. This key figure shows the trend in the cost price of public services relative to the price development of GDP. In our analyses, we tend to compare the development of the production volume with the total expenditure involved, including material expenditure and outsourcing. In our municipal study (see paragraph 4) the central object under investigation is not the production and productivity of the municipal organs in the narrow sense, but the production and productivity of the service delivery financed by the municipal authorities. Obviously, such an approach does not easily fit within the value added framework of national accounting.

**Heterogeneity** often hampers measurement of production, when certain producers produce several products or types of product. In health care, for instance, we can look either at a patient’s required level of care or diagnosis category, in the police service we can look at types of crime, or in education at target groups.

In such cases it is paramount to find a single production measure to allow comparison with the resources used. This presents no problem in the private sector, because individual products can be weighted according to their value, the market price. This is not possible in most public services and individual production categories have to be weighted and aggregated on the basis of the resources involved. These are weighted using personnel input. One of the complications is that data on these weights are scarce, whereas some of the weights are rapidly changing. A prominent example is the weight of traffic offences, which is rapidly decreasing as a result of technical innovations. At the same time, the number of penalized offences is increasing dramatically.

---

1 Due to the differentiation in types of crimes and types of law cases, there is a considerable number of product indicators involved. These are weighted using personnel input. One of the complications is that data on these weights are scarce, whereas some of the weights are rapidly changing. A prominent example is the weight of traffic offences, which is rapidly decreasing as a result of technical innovations. At the same time, the number of penalized offences is increasing dramatically.
used (preferably total costs, or staff numbers in the absence of anything better) per unit product. In theory, therefore, it is perfectly possible to make adjustments for the heterogeneity of products. But this requires detailed production data to be available, and the analyst should be able to make sensible assumptions about the factors used for weighting.

One of the biggest problems associated with measuring the performance of public service providers lies in the fact that the quality of the products tends to be inadequately reflected in product indicators based on quantitative measurements.

**Quality** is a vague and complex concept that refers to the extent to which the characteristics of a product meet given requirements. It is useful to distinguish between:

1) objective and subjective measures of quality, and
2) system, process and product quality.

Objective measures of the quality of services might include the percentage of trains that run on time, the average call-out time for the fire service, the percentage of lessons cancelled in schools, the percentage of incorrect administrative decisions, the percentage of complaints upheld etc. However, many relevant aspects of quality cannot be measured objectively, such as medical staff’s manner towards patients and the correct following of procedures. Subjective quality assessments of products or the production process by questionnaires answered by users, supervisors (inspectors) or staff may provide extra information. In the Netherlands, quite some information of this type is available. Actually, my institute has launched a population survey on these topics this very month. There is also a relatively large number of international comparative studies of respondents’ opinions concerning public services, including the Eurobarometer, the European Values Study, the World Values Study, the European Social Survey and the World Competitiveness Yearbook.

Besides the quality of the product itself, the quality of the production process can also be relevant. One can indirectly test the quality of products by establishing whether the production process meets certain requirements (adequately trained staff, adherence to procedures, measures to assure quality testing). Secondly, certain aspects of the production process not directly related to the end product might be very important to the user. Clean toilets and good fire safety in schools, though not part of the actual education product (the acquisition of knowledge and skills), are nevertheless important conditions. System quality refers to the convergence of demand and supply. Imbalance of demand and supply can create waiting lists or inefficiency.

If quality is not adequately reflected in the measure of production, developments such as smaller class sizes in education and more staff in nursing homes can lead to a fall in observed productivity. Similarly, better educated and therefore more expensive teachers, or fewer residents per room in nursing homes can result in an increase in the cost price. On the other hand, the effects of such intended boosts to quality are often difficult to measure. Even when adequate measurement is possible (e.g. the average delay to train services or waiting lists for medical treatment), it often proves impossible to incorporate the outcomes satisfactorily in a workable product indicator. One therefore often has to make do with additional quality indicators in combination with quantitative production figures.
It is often much more difficult to relate production processes directly to \textit{effects (outcomes)} than to output. It is therefore useful to distinguish between objectives that can be measured objectively via the final product, and deeper, underlying social objectives. Direct objectives of education, for example, include achieving as many final examination passes as possible, the objectives of curative care might include successful completion of treatment, and those of the police service, solving crime. The better a product indicator reflects a direct goal of the production process, the more applicable it will generally be. In hospitals, for example, analyses often lose their focus because admissions, patient days and so on are often used as indicators rather than the number of successful treatments. The same applies to analyses of education, where pupil numbers are often used as product indicators instead of their school results. Examples of deeper, underlying aims include producing well-informed citizens and ensuring there is a well-educated supply of labour (education), helping people live a long and healthy life (health care) and fighting crime (police). This type of objective, which concerns the indirect effects of services, is also examined extensively in this report.

The term \textit{effect indicator} generally refers to key figures that describe the extent to which these underlying objectives are achieved. The degree to which this is the case, the \textit{effectiveness}, is often determined not only by the production process, but also by external factors. The more neutral term \textit{goal achievement indicator} is therefore often more appropriate in this context.

3. Public service sector: final services

\textit{Introduction}

This line of analysis has a long history. Around 60 different task fields are involved and more than 150 different products. Fields covered include health care, education, police and justice, social services, public transport and some segments of culture and recreation (see table 1). These services represent a major part of the public service sector. In terms of expenditure about 70\% of final public services is included in the analysis. Of the public expenditure on final services, more than 90\% is accounted for. Forecasting is an important aspect of these analyses and their use for policy advises, but not relevant in the present context.

The method uses data which are collected over the years as a matter of routine by Statistics Netherlands. Examples are the number of pupils by school type, the number of consults by physicians, the number of intakes by hospitals or the number of crimes solved by the police. In some cases, available data allow a more sophisticated analysis. For example, the tax office distinguishes more than 35 different types of products. In an analysis of educational production, pure counts of pupils can be replaced by measures involving information on study delay and study success. In health care, a more profound analysis is obtained if the diagnosis and the health status of patients are taken into account.

\textit{Key figures}

The costs of the services involved can be decomposed in production volume times cost per unit product. It is convenient to deflate the costs with the price index of the gross domestic product to obtain real costs. The growth rate of the real costs per unit product reflects the development of the \textit{relative cost price} of the service involved with respect to the average domestic product. Approximately, the following rule holds: the growth rate of real costs equals the growth rate of the production volume plus the growth rate of the relative cost price.
Figure 2 gives aggregate results for the public services involved. Real cost increases much faster than production volume. This discrepancy reflects the increase of the relative cost price of public services. Labour volume closely follows production volume, implying that there is hardly an increase in labour productivity.

**Figure 2 – Key figures for the public service sector**

![Graph showing production volume, real cost, and labour volume over years from 1990 to 2003.](image)

*Source*: Kuhry and Pommer 2006.
Table 1 summarizes the results of the analysis at an intermediate level of aggregation.

Table 1 – Public service sector: annual growth of real costs and production, 1990-2003

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>education and research (17)</td>
<td>0.8</td>
<td>2.1</td>
<td>0.2</td>
</tr>
<tr>
<td>- primary education (1)</td>
<td>0.8</td>
<td>2.8</td>
<td>-1.1</td>
</tr>
<tr>
<td>- special education (3)</td>
<td>1.5</td>
<td>1.7</td>
<td>-1.7</td>
</tr>
<tr>
<td>- secondary education (4)</td>
<td>0.0</td>
<td>2.6</td>
<td>0.7</td>
</tr>
<tr>
<td>- vocational/adult education (5)</td>
<td>-0.5</td>
<td>3.9</td>
<td>-0.1</td>
</tr>
<tr>
<td>- higher professional education (2)</td>
<td>2.5</td>
<td>0.4</td>
<td>1.1</td>
</tr>
<tr>
<td>- university education (1)</td>
<td>0.3</td>
<td>1.4</td>
<td>1.0</td>
</tr>
<tr>
<td>- scientific research (1)</td>
<td>1.6</td>
<td>0.8</td>
<td>1.7</td>
</tr>
<tr>
<td>health care (12)</td>
<td>2.3</td>
<td>1.2</td>
<td>0.4</td>
</tr>
<tr>
<td>- hospitals and specialists (2)</td>
<td>1.2</td>
<td>1.8</td>
<td>0.1</td>
</tr>
<tr>
<td>- extramural care (3)</td>
<td>1.3</td>
<td>1.1</td>
<td>-0.5</td>
</tr>
<tr>
<td>- mental health care (2)</td>
<td>2.0</td>
<td>2.6</td>
<td>0.1</td>
</tr>
<tr>
<td>- nursing homes (1)</td>
<td>1.5</td>
<td>1.9</td>
<td>-0.5</td>
</tr>
<tr>
<td>- care homes (1)</td>
<td>-1.0</td>
<td>2.3</td>
<td>-2.0</td>
</tr>
<tr>
<td>- home-care (1)</td>
<td>4.0</td>
<td>1.1</td>
<td>0.9</td>
</tr>
<tr>
<td>- care for the disabled (1)</td>
<td>2.3</td>
<td>2.9</td>
<td>-0.7</td>
</tr>
<tr>
<td>- issuing of medicine (1)</td>
<td>7.5</td>
<td>-2.1</td>
<td>4.6</td>
</tr>
<tr>
<td>police and justice (6)</td>
<td>2.7</td>
<td>2.2</td>
<td>-0.1</td>
</tr>
<tr>
<td>- police (1)</td>
<td>2.6</td>
<td>1.1</td>
<td>0.3</td>
</tr>
<tr>
<td>- fire service (1)</td>
<td>1.5</td>
<td>2.7</td>
<td>-0.1</td>
</tr>
<tr>
<td>- administration of justice (3)</td>
<td>2.2</td>
<td>6.0</td>
<td>-3.0</td>
</tr>
<tr>
<td>- prisons (1)</td>
<td>5.6</td>
<td>2.4</td>
<td>2.1</td>
</tr>
<tr>
<td>other (23)</td>
<td>1.2</td>
<td>1.5</td>
<td>0.2</td>
</tr>
<tr>
<td>- tax office (1)</td>
<td>3.5</td>
<td>-0.1</td>
<td>3.3</td>
</tr>
<tr>
<td>- administration of health care (3)</td>
<td>0.6</td>
<td>0.5</td>
<td>1.0</td>
</tr>
<tr>
<td>- administration of social security (5)</td>
<td>-1.3</td>
<td>3.1</td>
<td>-1.2</td>
</tr>
<tr>
<td>- asylum seekers (2)</td>
<td>15.7</td>
<td>1.7</td>
<td>0.4</td>
</tr>
<tr>
<td>- child care (1)</td>
<td>11.2</td>
<td>3.0</td>
<td>0.8</td>
</tr>
<tr>
<td>- social services (2)</td>
<td>1.6</td>
<td>-0.5</td>
<td>1.5</td>
</tr>
<tr>
<td>- culture (5)</td>
<td>-1.0</td>
<td>3.0</td>
<td>-1.9</td>
</tr>
<tr>
<td>- sport (2)</td>
<td>0.6</td>
<td>1.3</td>
<td>-3.7</td>
</tr>
<tr>
<td>- public transport (2)</td>
<td>1.2</td>
<td>0.3</td>
<td>2.5</td>
</tr>
<tr>
<td>total public service sector (58)</td>
<td>1.7</td>
<td>1.6</td>
<td>0.3</td>
</tr>
<tr>
<td>total market sector</td>
<td>2.7</td>
<td>-0.5</td>
<td>1.6</td>
</tr>
</tbody>
</table>

Source: Kuhry and Pommer, 2006
Source of underlying data: Statistics Netherlands, annual reports of organisations and ministries.
Production

The key numbers for primary and secondary education mainly reflect demographic trends in the period reviewed. Notwithstanding a considerable decrease of the relevant age group, enrolment in universities does not show a decline and enrolment in higher professional education even shows a considerable increase.

The increase of the volume of health care reflects a tremendous rise of the issue of medicine and a more moderate increase of home-care and of mental health care. The effects of ageing of the population are reflected in the volume growth of hospitals and nursing homes. Remarkable is the decrease of the number of clients in the care homes, which takes place despite ageing. This is due to a shift of the main function of these services from housing facility to health care institution.

In the field of police and justice, the rapid increase of the prison population contrasts with a more moderate increase in the number of crimes solved by the police. The former phenomenon is a result of longer sentences and less premature acquittals.

Among the rest group, a strong growth characterizes the administration and housing of asylum seekers and the capacity of child care. However, the number of asylum seekers shows a remarkable decrease in the period after 2000, due to a change in the admission requirements.

Figure 3 depicts the growth of the production volume for the main aggregates in the study. The relatively rapid growth of the market sector contrasts with the slow development of education (demographic causes) Health care is characterized by a moderate growth. The production of the police and justice sector slows a slow increase in the nineties, followed by a marked increase in the period 2000-2003. As we will see below, costs have increased considerably over the entire period, and the rapid increase in the ultimate period is partly due to an increased productivity of the organisations involved. Another striking phenomenon is the fact that the growth of the market sector stagnates in the recession years 2001-2003, whereas the public service sector show a continuous growth. In fact it is only in 2004 (not in the figure) that the public sector shows a delayed response to the stagnation in the market sector.
**Relative cost price and productivity**

The considerable increase of the real costs of the public service sector contrasts with a moderate growth of the production volume. This discrepancy is a result of the increasing relative cost price of public services. This aspect is illustrated in the second column of table 1 and in figure 4.
Whereas products from the market sector are characterized by a decrease of the relative cost price, public services are characterized by an average annual increase of the relative cost price of 1.6% with respect to the average domestic product. The figure for the health care sector (1.2%) is considerably lower than that for education and police/justice (both in the order of 2.1%). The history of the latter sector illustrates a poor behaviour in the nineties, followed by a considerable improvement in the years 2000-2003.

Table 2 gives a decomposition of the annual increase of the relative cost price in the public service sector and offers a comparison with the market sector. The reasons for the increasing relative cost price of public services is in part explained by the law of Baumol: a relative small increase of labour productivity, accompanied by a considerable increase of the price and volume of inputs. The main cause of the lagging labour productivity of many public services is the labour-intensive character of the services involved. For example, in education and health care labour cannot be simply substituted by machines or computers. However, a lack of incentives may also play a part. Another reason for the increasing relative cost price of public services is the increasing average age of the labour force in public service.

However, due to the problem of adequately capturing product quality in physical production measures, part of the increase of the relative cost price of public services might actually be related to an improvement of product quality. Nevertheless, the fact remains that the cost of these products per item or per consumer is increasing steadily and the increase of productivity is relatively slow, facts which have to be taken into account when planning the required volume and estimated costs of public services.
Table 2 – Decomposition of % average cost increase, 1990-2003

<table>
<thead>
<tr>
<th></th>
<th>public service sector</th>
<th>market sector</th>
</tr>
</thead>
<tbody>
<tr>
<td>increase real costs</td>
<td>3.3</td>
<td>2.2</td>
</tr>
<tr>
<td>- effect production volume</td>
<td>1.7</td>
<td>2.7</td>
</tr>
<tr>
<td>- effect relative cost price</td>
<td>1.6</td>
<td>-0.5</td>
</tr>
<tr>
<td>relative cost price</td>
<td>1.6</td>
<td>-0.5</td>
</tr>
<tr>
<td>- effect real contractual wage increase</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>- effect incidental wage increase</td>
<td>0.8</td>
<td>0.4</td>
</tr>
<tr>
<td>- effect labour productivity</td>
<td>-0.2</td>
<td>-1.3</td>
</tr>
<tr>
<td>- effect material and capital means</td>
<td>0.6</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Source: Kuhry and Pommer 2006.

Figures 5 and 6 show the corresponding development of labour productivity and total factor productivity for the main aggregates involved. The market sector shows a steady increase in labour productivity. In contrast, the labour productivity of public services hardly increases. Special cases are education and police/justice. In case of education, a rise of labour productivity in the early nineties is followed by a decline which is related to a policy decision to decrease class size in primary education. The latter may have resulted in a hidden gain of quality, but apparently not to an increase in effectiveness. In the case of police and justice, labour productivity declines in the nineties, but improves in the period 2000-2003. The pattern in figure 6, dealing with total productivity, strongly resembles that in figure 5 (labour productivity). However, for public services a decline in productivity, rather than a slight improvement is visible. This is caused by the increase of the volume of material and capital inputs per unit of product.
**Figure 5 – Labour productivity**

![Graph showing labour productivity](image)

*Source*: Kuhry and Pommer 2006.

**Figure 6 – Total productivity**

![Graph showing total productivity](image)

*Source*: Kuhry and Pommer 2006.
4. Municipal performance

Lately, the approach sketched in paragraph 3 has also been applied to a second field: the total production of the municipal government layer in the Netherlands. The increase of the so-called general municipal fund is at present indirectly determined by the increase of the expenditure of the central government. The Dutch Ministry of the Interior has launched a project to make the outputs corresponding to the available means more transparent. In the end, such an approach might give insight in the municipal needs.¹ The question raised can only be answered properly if a complete picture of municipal production can be drawn. However, municipalities have a wide variety of tasks including education, social services, cultural and recreational services, environmental services, housing, infrastructure, public transport etc. etc. Apart from final services, intermediate services and purely collective services play a part.

From 2001 onwards a number of reports have been published on this topic, of which the last one, Measurements for municipalities 2006, will be published by the end of next week. In this report, around 60 task fields and more than 100 individual products are distinguished (see appendix A for an inventory of the fields involved and a summary of the numerical results). An attempt has been made to incorporate all these products in a single framework of analysis. Although the approach leads to a number of recommendations to improve the available data, the method is essentially implemented using available key data on public services. The production of final services is partly measured with value indicators and partly with physical indicators (either indicators of performance or indicators of use). The production of intermediate services (‘overhead’) is indirectly measured by reference to the production of the final services at which they are aimed. In other words, the costs of the intermediate services are consolidated into the costs of the corresponding final services. Most difficult is the inclusion of purely collective services. It is not very useful to measure the production of these services by the number of civil servants involved or the number of reports written. Two alternative solutions are applied: estimation of production by norm indicators (number of inhabitants, square miles etc.) or estimation by deflating costs with a constructed price index.²

Rather crucial in the approach is a method of dealing with services which are only in part produced or financed by the government layer in question. In this case, the product can only in part be ascribed to municipalities. Since the analysis is carried out in relative terms, the growth rate of the production volume in comparison to the growth rates of the expenditure, problems only arise if the ‘share’ of municipalities is changing. To correct for such changes, the production is multiplied with this share, being the quotient of municipal expenditure to total costs. It can easily be shown that this implies that the ratio of municipal expenditure to municipal production equals the ratio of total costs and total production volume.³ This relationship only holds at the level of elementary products and not at the level of aggregates.

² In this approach, costs or revenues are deflated with a constructed price index, namely the price index of the gross domestic product corrected for the average annual increase of the relative cost price of public services (estimated at 1.3 %).
³ In other words, the relative cost price is a constant factor in the analysis, which in contrast to the production volume is not affected by mutations in municipal tasks.
Of the 58 task fields involved, production is measured in 17 cases by an indicator of use, 15 by a performance indicator, 13 by a value indicator and 12 by a norm indicator. The remaining case (the central apparatus) concerns overhead. The weakest link in the analysis, the norm indicators, represents about 12% of total expenditure. If norm indicators are replaced by an approach in terms of constructed value indicators, the outcomes of the analysis at the macro level are virtually unaffected. In fact, macro outcomes appeared to be quite robust to a considerable number of adaptations of the analytical scheme and product indicators. The aggregated results of the analysis are shown in figure 7.

Figure 7 – Real expenditure and production volume of municipalities, 1995-2004

The real expenditure of municipalities shows ups and downs. The decrease in the years 1995-1998 is a result of mutations in the tasks of municipalities, such as the privatisation of municipal housing corporations and a diminishing responsibility for care of the elderly. The increasing expenditure in the period 1998-2003 reflects the rise of municipal means in these years of economic prosperity. The decline in 2004 is a combined result of privatisation of public transport organisations and budget cuts as a delayed reaction on the recession in the previous years.

As was the case in the services analysed in paragraph 3, production follows expenditure at a distance. The difference between both lines represents the effect of an increase of the relative cost price of municipal services with respect to the average domestic product. The average annual increase is in the order of 1%, but was slow in the first and the last couple of years. Apparently, there is a relation with the availability of means: if the budget is restrained, the demand is met in a sober manner, large expenditures are postponed, the hiring of new personnel is restricted and salary levels are under pressure. As a result, municipal production actually decreased in the period 1995-2003.
When addressing the question whether the development of municipal services is sufficient to satisfy the need of citizens, one has to abstract from mutations of tasks. For the citizen, it is not of primary importance who is responsible for a certain task, but only if it is performed adequately. After a correction for task mutations, the pattern in figure 8 emerges.

**Figure 8 – Real expenditure and production volume of municipalities, 1995-2004 (corrected for task mutations)**

The figure shows that real expenditure is increasing continuously, with the exception of the years 2003 and 2004. The average annual increase is or the order of 1.5%. With the exception of the last year, there is also a steady but slow annual increase in the production volume (on average 0.7%), slightly higher than the population increase in this period. The increase of production volume is (on average, not in the last couple of years) much lower than that of the market sector and even considerably lower than that of the public service sector discussed in paragraph 3). Also the prospects for the developments in 2005 and 2006 (see dotted line in figures 7 and 8) are not very bright.

At the moment, there is only partial information on the decomposition of the expenditure in personnel, material and capital components. Also the available data on municipal personnel does not correspond to the demarcation in this paper since organisations which do not belong to the municipal apparatus, but are financed with municipal means, are also included in our analysis. The available data show an increase in the volume of municipal personnel, particularly in the years 2000-2002. This points in the direction of a slow or even negative development of labour productivity. Of course, as was pointed out in paragraph 3, this outcome may be partly due to an improvement of product quality which is not reflected in our product indicators.

*Source: Kuhry and Jonker 2006.*
## Appendix A

### Table 3 – Municipalities: analysis of net-expenditure and corresponding production, 1999-2004

<table>
<thead>
<tr>
<th>Task field (number of subcategories)</th>
<th>production (annual growth % 1999-2004)</th>
<th>real costs per unit product (annual growth % 1999-2004)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>general administration (4)</strong></td>
<td>-0.5</td>
<td>3.9</td>
</tr>
<tr>
<td>- city council (1)</td>
<td>-0.1</td>
<td>3.0</td>
</tr>
<tr>
<td>- central apparatus (1)</td>
<td>-0.1</td>
<td>4.2</td>
</tr>
<tr>
<td>- civilian affairs (2)</td>
<td>-1.9</td>
<td>4.0</td>
</tr>
<tr>
<td><strong>public safety (2)</strong></td>
<td>4.0</td>
<td>3.6</td>
</tr>
<tr>
<td>- fire service (1)</td>
<td>0.0</td>
<td>4.5</td>
</tr>
<tr>
<td>- other (1)</td>
<td>13.5</td>
<td>1.5</td>
</tr>
<tr>
<td><strong>roads and waterways (3)</strong></td>
<td>2.9</td>
<td>0.4</td>
</tr>
<tr>
<td><strong>education (exclusive accommodation, 10)</strong></td>
<td>-0.7</td>
<td>1.6</td>
</tr>
<tr>
<td>- primary education (2)</td>
<td>-0.7</td>
<td>2.5</td>
</tr>
<tr>
<td>- special education (2)</td>
<td>-8.9</td>
<td>2.7</td>
</tr>
<tr>
<td>- secondary education (2)</td>
<td>-2.4</td>
<td>1.2</td>
</tr>
<tr>
<td>- other education (4)</td>
<td>5.9</td>
<td>-1.0</td>
</tr>
<tr>
<td><strong>education: accommodation (4)</strong></td>
<td>0.2</td>
<td>3.2</td>
</tr>
<tr>
<td><strong>subsidised employment (2)</strong></td>
<td>0.5</td>
<td>-0.7</td>
</tr>
<tr>
<td><strong>social security (2)</strong></td>
<td>-1.6</td>
<td>-0.1</td>
</tr>
<tr>
<td>- administration (1)</td>
<td>-2.3</td>
<td>3.3</td>
</tr>
<tr>
<td>- transfers (1)</td>
<td>-1.5</td>
<td>-0.7</td>
</tr>
<tr>
<td><strong>social services (9)</strong></td>
<td>2.2</td>
<td>-0.4</td>
</tr>
<tr>
<td>- social services (5)</td>
<td>3.7</td>
<td>-3.2</td>
</tr>
<tr>
<td>- socio-cultural work (2)</td>
<td>0.7</td>
<td>3.9</td>
</tr>
<tr>
<td>- child care (1)</td>
<td>0.8</td>
<td>3.0</td>
</tr>
<tr>
<td>- caring homes (1)</td>
<td>-16.9</td>
<td>3.1</td>
</tr>
<tr>
<td><strong>provisions for handicapped (2)</strong></td>
<td>3.5</td>
<td>-2.3</td>
</tr>
<tr>
<td><strong>health care (4)</strong></td>
<td>7.0</td>
<td>3.8</td>
</tr>
<tr>
<td>- municipal health agency (3)</td>
<td>7.0</td>
<td>3.2</td>
</tr>
<tr>
<td>- nursing homes (1)</td>
<td>-19.3</td>
<td>0.0</td>
</tr>
<tr>
<td><strong>environment (3)</strong></td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>- administration of environment (1)</td>
<td>0.8</td>
<td>2.0</td>
</tr>
<tr>
<td>- sewers (1)</td>
<td>-0.4</td>
<td>1.6</td>
</tr>
<tr>
<td>- sanitation (1)</td>
<td>0.8</td>
<td>-0.5</td>
</tr>
<tr>
<td><strong>culture and recreation (9)</strong></td>
<td>1.7</td>
<td>0.9</td>
</tr>
<tr>
<td>- culture and sports (4)</td>
<td>0.6</td>
<td>2.2</td>
</tr>
<tr>
<td>- cultural heritage (2)</td>
<td>3.0</td>
<td>-0.1</td>
</tr>
<tr>
<td>- parks and sport fields (3)</td>
<td>2.9</td>
<td>-0.7</td>
</tr>
<tr>
<td><strong>housing, town and country planning (4)</strong></td>
<td>-5.9</td>
<td>2.0</td>
</tr>
<tr>
<td>- exploitation of housing (1)</td>
<td>-15.0</td>
<td>-1.8</td>
</tr>
<tr>
<td>- administration of housing (1)</td>
<td>0.2</td>
<td>3.7</td>
</tr>
<tr>
<td>- town and country planning (1)</td>
<td>0.8</td>
<td>6.3</td>
</tr>
<tr>
<td>- urban renewal (1)</td>
<td>-4.8</td>
<td>-0.6</td>
</tr>
<tr>
<td><strong>other services (2)</strong></td>
<td>-5.6</td>
<td>2.3</td>
</tr>
<tr>
<td>- public transport (1)</td>
<td>-8.9</td>
<td>3.5</td>
</tr>
<tr>
<td>- economic affairs (1)</td>
<td>0.6</td>
<td>0.2</td>
</tr>
<tr>
<td><strong>total municipalities (58)</strong></td>
<td>-0.1</td>
<td>0.9</td>
</tr>
</tbody>
</table>

*Source: Kuhry and Jonker 2006*

*Source of underlying data: Statistics Netherlands, annual reports of organisations and ministries.*
Accompanying the growth of the public sector in most European countries the last decades there has been an increased emphasis in recent years on improving the functioning of the public sector in order to promote accountability of use of resources and to promote efficiency and productivity. In order to succeed with such ambitions outputs and inputs have to be measured with sufficient coverage and quality. More often than not productivity in the public sector is not really measured based on independent measurement of outputs, but simply based on resource costs and assumed either to be constant or to grow with a certain percentage, e.g. set equal to the real wage increase.

For the parts of the public sector not participating in markets for their products it may be a difficult task in itself just to establish acceptable definitions and measurement of outputs. Current bookkeeping practices are often more designed to deal with inputs, i.e. the budget, than provide detailed enough information on the output side. Furthermore, for activities that produce multiple outputs with important elements of quality it can be a complex task to delimitate types of outputs and provide their measurement. Checking the existing production of output data at the micro level of parallel units within a public sector like income tax assessment and collection is the usual first step. One can usually find statistics covering number of tasks performed, like number of tax assessments from different tax payer groups recorded by local tax offices. Quality aspects are harder to quantify. One possibility would be to adjust for the complexity of tax assessment by criteria like the amount of tax involved or amount of information that has to be gathered.

Local tax assessment offices in Norway sort under the Directorate of Taxes that again sort under the Ministry of Finance. The Ministry made a direct request in 2003 to the Directorate of Taxes to work out one or more indicators of productivity in the sector. The first stage was to find indicators for outputs

1 The paper is based on a report (Førsund et al., 2005) of the project “Productivity studies in the Norwegian Tax Administration” done by the Frisch Centre for the Directorate of Taxes also published as Førsund et al. 2006.
and resources among the statistics produced by tax offices that are available to the Directorate. This process involved a number of people at the Directorate and resulted in a sharper focus on what kind of services that are actually produced and an awareness of weaknesses of the statistics and problems of comparability.

The Directorate of Taxes has been especially concerned about having the best possible data quality and reducing any ‘noise’ that may raise questions about the validity of the results. A recent theoretical development of the DEA method is to take explicitly into account the statistical properties of efficiency scores as estimators of unknown true scores by applying the technique of bootstrapping. This overcomes the inherent sampling bias of limited data sets under certain assumptions. Bootstrapping provides bias correction of the scores and confidence intervals, thus signaling the quality of the estimates of efficiency levels.

Data has been collected especially for the study for three years enabling us to also investigate productivity development using the Malmquist productivity index. The statistical technique of bootstrapping is applied to these index values for individual offices. The productivity change distribution ranges from a 20% decline to a 35% increase. Taken at face value the results indicate that units representing about half of the costs (in 2004) have had a productivity decline over the three years, while a half has had a productivity improvement. The confidence intervals tend to be wider for large units. Key information about uncertainty is provided by testing whether change is significantly positive or negative. About 20 % of the units had a significantly declining productivity, and 45 % significantly increasing productivity. The productivity development of individual offices, based on bias-corrected productivity measures, is classified into the four categories efficient cost increase, efficient cost savings, inefficient cost increase and inefficient cost savings. No unit belongs to the last group, and the second and third groups have the most observations, and about equal number in each group.

Abstract

The performance of local tax offices of Norway is studied over a three-year period applying Data Envelopment Efficiency analysis calculating Malmquist productivity indices. A bootstrap approach recently developed for DEA models is applied to establish confidence intervals for the individual indices enabling an identification of units having significant productivity decline or growth.

Key words: Tax office, Malmquist productivity index, DEA, bootstrap

JEL classification: C60, D24, L89
1. Introduction

Accompanying the growth of the public sector in most European countries the last decades there has been an increased emphasis in recent years on improving the functioning of the public sector in order to promote accountability of use of resources and to promote efficiency and productivity. In order to succeed with such ambitions outputs and inputs have to be measured with sufficient coverage and quality. More often than not productivity in the public sector is not really measured based on independent measurement of outputs and inputs, but simply based on resource costs and assumed either to be constant or to grow with a certain percentage, e.g. set equal to the real wage increase.

For the parts of the public sector not participating in markets for their products it may be a difficult task in itself just to establish acceptable definitions and measurement of outputs. Current bookkeeping practices are often more designed to deal with inputs, i.e. the budget, than provide detailed enough information on the output side. Furthermore, for activities that produce multiple outputs with important elements of quality it can be a complex task to delimitate types of outputs and provide their measurement. Checking the existing production of output data at the micro level of parallel units within a public sector is often the first step. One can usually find statistics covering number of tasks performed, like for local tax offices number of tax assessments from different tax payer groups. Quality aspects are harder to quantify. One possibility for tax offices would be to adjust for the complexity of tax assessment by criteria like the amount of tax involved or amount of information that has to be gathered.

Local tax assessment offices in Norway sort under the Directorate of Taxes that again sort under the Ministry of Finance. The Ministry made a direct request in 2003 to the Directorate of Taxes to work out one or more indicators of productivity in the sector. This paper is based on the response to this request (Førsund et al., 2005) and a further elaboration in Førsund et al. (2006). While the latter paper deals with both efficiency and productivity, the present paper focuses on productivity and extends this analysis. The first stage was to find indicators for outputs and resources among the statistics produced by tax offices that are available to the Directorate. This process involved a number of people at the Directorate and resulted in a sharper focus on what kind of services that are actually produced and an awareness of weaknesses of the statistics and problems of comparability. The main purpose of the study from the Directorate’s point of view was to get estimates of efficiency and productivity levels as a basis for finding explanations for differences and to formulate actions that can be taken to reduce inefficiencies and improve productivity.

Productivity is normally perceived as the ratio of outputs to inputs, and in the presence of multiple inputs and outputs these must be weighed together to single numbers for outputs and inputs respectively in order to measure total factor productivity (TFP). There are two main strategies to follow regarding creation of weights (Førsund, 1997). The index approach utilises exogenous information; the standard weights are the prices of outputs and inputs. Well known such productivity indices are the Törnqvist index and the Fisher index. The technology approach is based on information about the production technology, and weights are explicitly or implicitly deduced from the technology. It should be noticed that when the possibility of inefficient operations is assumed only the technology approach has been used. Our approach is based on the technology approach of estimating the transformation of inputs into outputs. Due to lack of knowledge of functional forms and consequently the need of letting the
data speak maximally concerning the nature of the transformation, a non-parametric method termed Data Envelopment Analysis (DEA) is employed.¹

The Directorate of Taxes has been especially concerned about having the best possible data quality and reducing any ‘noise’ that may raise questions about the validity of the results. A recent report from a British Working Party of Performance Monitoring in the Public Services has, as one of the recommendations that reported performance measures should always include measures of uncertainty (Bird et al., 2005). Since the intention of calculating efficiency scores for individual tax offices is to use them to find explanations for efficiency and productivity differences, and to help identify possible measures leading to productivity improvements, it is important that the calculations are based on best available methods. A recent theoretical development of the DEA method is to take explicitly into account the statistical properties of efficiency scores as estimators of unknown true scores by applying the technique of bootstrapping (Simar and Wilson, 1998, 2000). This overcomes the inherent sampling bias of limited data sets under certain assumptions. Bootstrapping provides bias correction of the scores and confidence intervals, thus signaling the quality of the estimates of productivity levels.

Data has been collected especially for the study for three years enabling us to investigate productivity development using the Malmquist productivity index. The statistical technique of bootstrapping is applied to these index values for individual offices. The productivity change distribution ranges from a 20% decline to a 35% increase. Taken at face value the results indicate that units representing about half of the costs have had a productivity decline over the three years, while a half has had a productivity improvement. The confidence intervals tend to be wider for large units. Key information about uncertainty is provided by testing whether change is significantly positive or negative. About 20% of the units had a significantly declining productivity, and 45% significantly increasing productivity. The productivity development of individual offices, based on bias-corrected productivity measures, is classified into the four categories efficient cost increase, efficient cost savings, inefficient cost savings and inefficient cost increase. No unit belongs to the third group, and the second and fourth groups have most units.

The paper is organised in the following way: Section 2 presents the methods used for estimating the productivity scores including bootstrapping. In Section 3 the data set is presented and the specification of the output and input variables that could be established discussed. The empirical results for productivity developments are discussed in Section 4. Section 5 concludes.

2. Methodology

The point of departure for defining the production technology that will be the basis for measuring productivity is to formulate a production possibility set accommodating also inefficient operations. Let $x$ be a vector of inputs and $y$ be a vector of outputs, then the production possibility set at time $t$ is defined as:

$$S^t = \{(x, y) \mid x \text{ can produce } y \text{ at time } t\}$$  \hspace{1cm} (1)

¹ The non-parametric method was originated in Farrell (1957) and further developed into the tool in use today in Charnes et al. (1978)
General requirements on the technology set are that it is convex and exhibits monotonicity. In the presence of inefficient operations the relevant technology reference is the efficient border of the technology set. This border set will be termed the *frontier production function*.

Building on the idea in Malmquist (1953) of proportional variation of variables when measuring change, Caves et al. (1982) introduced the bilateral Malmquist productivity index based on the ratio of Farrell (1957) efficiency measures for the two units (e.g. the same unit measured for two different time periods). Efficiency is measured against the same frontier technology. A brief overview of the Farrell efficiency measures will therefore be given before defining the Malmquist index.

**The Farrell efficiency measures**

The *Data Envelopment Approach* (DEA) is especially suitable as a benchmarking tool in a setting of multiple inputs and outputs, where price information on outputs are not available, and there is no firm knowledge about the shape of the transformation function relating outputs to inputs. Following Farrell (1957) the production structure is based on a convex production possibility set \( \hat{S}_t \), as an estimator of the true production possibility set \( S_t \) in (1), defined empirically by enveloping the observations as tightly as possible by a piecewise linear convex outer boundary (see Banker et al. (1984) for the properties of the empirically defined set).

Figure 1 illustrates two different shapes of frontiers and the corresponding production possibility sets. The ray through the origin is a constant return to scale (CRS) frontier and the piecewise concave graph is a variable returns to scale (VRS) frontier. The frontiers as boundaries of production possibility sets correspond to the neoclassical notion of an efficient transformation function between inputs and outputs. Efficiency measures for observations are in general based on the distance between an observation and the boundary. Following Farrell (1957) there are two

**Figure 1 – Definitions of efficiency measures**

![Diagram of efficiency measures](image-url)
Measurement of productivity in the non-business sector

basic radial directions to go from an observation to the frontier: keeping outputs fixed and moving to the frontier by a proportional reductions in inputs, or keeping inputs fixed and moving to the frontier by a proportional expansion of outputs. With reference to Figure 1 the efficiency measures are defined for an inefficient observation P by first finding a reference point on the frontier. The reference point is found by using linear programming to place a convex piecewise frontier as close as possible to the observations (Banker et al., 1984). Then by the nature of the estimation method the reference points on the frontier are linear combinations of estimated best practice observations. Using the notation in Figure 1 the estimators of input- and output oriented efficiency measures are defined as follows with reference to the VRS frontier:

\[ \hat{E}_1 = \frac{P}{P_{1\text{ref}}} \quad \hat{E}_2 = \frac{P}{P_{2\text{ref}}} \]  

(2)

where P is observed productivity \((y/x)\) and \(P_{1\text{ref}}, P_{2\text{ref}}\) are the productivities at the estimated frontier. We see that, following Farrell, the measures are restricted to be between zero and one. The general nature of Eq. (2) in the case of multiple outputs and inputs follows from restricting inputs, respectively outputs to move proportionately from the observation to the reference points.

An important reference for calibrating productivity change is the concept of technically optimal scale (Frisch, 1965). This is the scale where the returns to scale is one, and is illustrated in Figure 1 as the tangent point \(P_{\text{tops}}\) of the CRS line and the VRS frontier. A measure of technical productivity for unit \(i\) is then defined as the following ratios between productivities

\[ \hat{E}_3 = \frac{P}{P_{\text{tops}}} = \frac{P}{P_{1\text{crs}}} = \frac{P}{P_{2\text{crs}}} \]  

(3)

Productivity measurement

The Malmquist productivity index (Caves et al., 1982) is developed for discrete time and defined by using the Farrell efficiency scores for two different periods \((u \text{ and } v)\) for a unit. The DEA estimator of the Malmquist productivity index is:

\[ \hat{M}^s_{di}(x_u, y_u, x_v, y_v) = \frac{\hat{E}_d^s(x_u, y_u)}{\hat{E}_d^s(x_v, y_v)} \hat{S}^s, \quad d = 1, 2, \quad i = 1, \ldots, J, \quad u, v = 1, \ldots, T, \quad u \neq v \]  

(4)

Here the index for the frontier technology is \(s\), the index for the orientation is \(d\), the index for the unit is \(i\), the index for the two periods is \(u\) and \(v\), the number of time periods is \(T\) and the efficiency score estimators are conditional upon the estimator for the production possibility set. The productivity interpretation of the Malmquist index follows from noting that the definition of the efficiency measures involved implies that observed productivity is compared with maximal productivity at the border of the technology set, keeping either inputs or outputs fixed. The Malmquist index captures

---

1. The linear programmes for calculating the efficiency score estimates can be found in Forsund et al. (2006).
2. In general the technically optimal scale point may not be unique, i.e. the CRS line may coincide with a segment on the frontier, but the scale elasticity will be one along such a segment (Forsund and Hjalmarsson, 2004a).
3. The definition of the true Malmquist index is the same as in (4) with true efficiency scores conditional on the true set \(S^*\) instead of the DEA estimates.
the relative change in efficiency for two periods, and since the technology frontier is the same this relative measure has the interpretation of productivity change.

In order to calculate the efficiency scores the data set must be extended to include time periods. This is straightforward: the observations used to support the frontier indexed $s$ must be specified (for example observations for a specific year), and then the unit $i$ from two periods is used as the observation in two separate efficiency calculations, one for each period $u$ and $v$. It is then possible that the efficiency scores become greater than one.\footnote{If a year is chosen as the basis for the reference technology the VRS specification will not yield feasible solutions for observations having smaller inputs than the smallest observed for the reference set.}

Productivity as measured by the Malmquist index, based on the production possibility set (1) above, may be influenced by changes in the scale of the operation, but two units that have the same ratio of outputs to inputs should be viewed as equally productive, regardless of the scale of production. Doubling all inputs and outputs keeping input and output mixes constant does not change productivity, even though the technology $P^S$ has variable returns to scale. The relevant reference set is therefore one that is homogenous of degree 1 in the input-output vector, and the set that fits closest to the technology is its homogenous envelopment defined by

$$S^e = \{(x, y) | (\gamma x, \gamma y) \in S^S \}$$

(5)

where $\gamma$ is a free positive scalar. Only if the underlying technology exhibits constant returns to scale will the sets defined in (1) and (5) be equal. This technology is illustrated in Figure 1 by the ray from the origin labeled CRS.

Focussing on productivity measures over time we will be interested in a measure of TFP that is not influenced by underlying scale changes of the units. Such a measure is obtained by only using the technology information about changes in the technically optimal scale (TOPS) (Frisch, 1965) over time. Such a subset of the technology set is defined by (Førsund and Hjalmarsson, 2004b):

$$TOPS^e = \{ \{(x, y) | \varepsilon(x, y) = 1, (x, y) \in S^e \} \}$$

(6)

where $\varepsilon(x, y)$ is the scale elasticity function. From classical production theory we know that the productivity is maximal at optimal scale where returns to scale, $\varepsilon$, is one, thus this is a natural reference for productivity changes over time. The TOPS set is also the intersection of the efficient boundary of the technology $S^e$ (the frontier production function) and the boundary of its envelopment $S^*$. In Figure 1 the TOPS set is just the point labeled $P_{tops}$.

Just as the productivity should be unchanged if the input-output vector is proportionately scaled, a measure of productivity should double if outputs are doubled and inputs are kept constant, and vice versa. A productivity measure should therefore be homogenous of degree 1 in outputs and of degree -1 in inputs. A measure of productivity change over time, based on all the inputs and outputs for the unit in question for two periods, should similarly be homogeneous of degree 1 in outputs from the
last period and inputs from the first period and homogeneous of degree (-1) in outputs from the first period and in inputs from the second period; i.e. if outputs in the second period or inputs in the first period doubles the TFP measure should also double, and if outputs in the first period or inputs in the second period doubles then the TFP measure should be reduced to one half. Using the subset TOPS is one way of obtaining the required homogeneity properties of a Malmquist productivity index.

An illustration in the two-period case is provided in Figure 2. Observations of the same unit are indicated by P1 and P2. The two corresponding VRS frontiers are drawn showing an outward shift indicating technological progress. A further question is whether to use the envelopment of the technology of a single year or several years as the reference for the productivity index. One consideration is whether the resulting productivity measure is circular. We will be interested in pointing to years with strong or weak productivity growth, so we need circularity in order to interpret the results in such a way (Gini, 1931). We will use as a reference technology an average technology by using the envelopment of all technology frontiers as a fixed reference frontier, i.e. \( S^* = U, S' \), thereby fulfilling the circularity condition while at the same time utilising technology information from all time periods. In Tulkens and van den Eeckaut (1995) this type of frontier was termed the intertemporal frontier. As is common in indices, performance is calculated using information that may not have been available in the first period, but this is consistent with retrospective evaluation.

Using a linear homogeneous envelopment implies that the orientation of the efficiency index does not matter. The Malmquist index then simplifies to:

\[
\hat{M}^*_i(x_{iu}, y_{iu}, x_{iv}, y_{iv}) = \frac{\hat{E}^*_i(x_{iu}, y_{iu})}{\hat{E}^*_i(x_{iv}, y_{iv})} \frac{S^*}{S}, \quad i = 1, \ldots, J, \ u, v = 1, \ldots, T, \ u \neq v
\]
Measurement of productivity in the non-business sector

where superscript $s$ here indicates that all data is used as the technology reference set. The Malmquist productivity estimator is conditional on the estimator for the linear homogeneous envelopment set in (5).

**Bootstrapping**

It is well known since Farrell (1957) that a piecewise linear envelopment of data as tight as possible ‘from above’, obeying some basic properties of production possibility sets, results in a frontier estimator that is pessimistically biased. We have a limited number of observations or realisations of an unknown technology and the frontier rests on outlier observations. Since the DEA method is based on enveloping the observations as tightly as possible there may be potential realizations of the unknown technology that are not appearing as actual observations. The efficiency scores are correspondingly optimistically biased. The sampling bias for a given observation can be expected to be higher the lower the number of other observations in the sample. Banker (1993) proved in the one input – one output case that as the number of observations goes towards infinity, the distance between the DEA estimate and the true efficiency score goes towards zero, i.e. the DEA estimator is consistent. In Simar and Wilson (2000) generalisations to multiple outputs and inputs are reviewed. The DEA frontier estimate is based on the best-observed practice, but this is a biased estimate of the best possible practice in any real-world (finite sample) situation. We know, however, that the bias is non-negative, in the sense that the DEA estimated efficiency is higher than or equal to the true efficiency. Following Simar and Wilson (1998) the Data Generating Process (DGP) assumes in general that the sample observations $(x_i, y_i)$ are realisations of independent identically distributed variables on the production possibility set with the probability density function $f(x,y)$. In our setting with e.g. a radial output-oriented efficiency variable $E_{2s}$ distributed on $(0,1]$ we assume that the observations are generated by randomly drawn efficiencies from the true efficiency distribution, with exogenously given output levels and input mixes. There is a strictly positive probability of drawing observations close to all parts of the true production frontier, and the DEA assumptions (no measurement error, convexity, free disposability) hold. In the following a homogenous efficiency distribution is assumed, i.e. the efficiency distribution is independent of output scale and input mix, but this can be relaxed with a more complicated DEA bootstrap methodology (Simar and Wilson, 2000).

Bootstrapping is a way of testing the reliability of the dataset, and works by creating pseudo replicate data sets using resampling (Efron, 1979). Bias correction in DEA using bootstrapping, following Simar and Wilson (1999), is based on an assumption by analogy on the distributions of the estimators, implying that the difference between the Malmquist estimator based on pseudo data and the DEA-based estimator is distributed like the distribution between the DEA estimator and the true Malmquist index, assuming estimators to be consistent:

$$\tilde{M}^s(u,v) = M^s(u,v) \left( \tilde{S}^s \right) \left( \tilde{S}^s \right), \quad u,v = 1,...,T, u \neq v$$

Here $M^s$ is the true unknown efficiency, $\tilde{M}^s$ is the original DEA estimate, $\tilde{S}^s$ and $\tilde{S}^s$ are the theoretical production possibility set and its DEA estimate, respectively. The bootstrapped estimate is obtained in the following way: the empirical distribution of the efficiency scores from the original DEA run for calculating efficiency scores making up the Malmquist index (7) is used to estimate a smoothed distribution by a kernel density estimate (KDE), using reflection to
to have a consistent estimator of the efficiency score distribution at the boundary of the distribution where the efficiency score is 1. The pseudo observations are then created by projecting all inefficient observations to the DEA frontier (output orientation here) and drawing randomly an efficiency score, $\theta_i$ for each unit from the KDE distribution, $y_i^{ps} = (y_i / \hat{E}_i)\theta_i$. A new DEA frontier is then estimated on these pseudo observations $(\hat{x}_i, y_i^{ps})$, each generated by mimicking the original DGP, as if the original DEA estimated frontier were the true frontier. The new frontier must lie on the inside of the original DEA frontier. We make 2000 such draws and establish 2000 new DEA frontiers, resulting in 2000 pseudo sample efficiency estimates for each observation. The Malquist index is calculated for each run according to (7).

Building upon (8) the bias of the Malmquist productivity estimator can be estimated. However, it is pointed out in Simar and Wilson (2000) that the bias correction may create additional noise in the sense that the mean square error of the bias-corrected score may be greater than the mean square error of the uncorrected estimator. This was the case for the tax office data at hand. We will therefore use the original DEA estimates for the Malmquist index.

This problem motivated Simar and Wilson (1999) to suggest a direct way to calculate the confidence intervals. The confidence interval limits may be defined by

$$\Pr(-\hat{b}_{ai} \leq \hat{M}_i^*(u, v) - \hat{M}_i^*(u, v) \leq -\hat{a}_{ai} \mid \hat{S}_i) = 1 - \alpha, \ i = 1, \ldots, J, \ u, v = 1, \ldots, T, \ u \neq v \quad (9)$$

The estimates for the limits are found from the distribution of $(\hat{M}_b(u, v) - \hat{M}_i^*(u, v))$ for $b = 1, \ldots, B$ by sorting in increasing order and finding the values matching the chosen degree of confidence. The estimated $(1 - \alpha)$ confidence interval for the true efficiency score $M_i(u, v)$ is then

$$\hat{M}_i^*(u, v) + \hat{a}_{ai} \leq M_i^*(u, v) \leq \hat{M}_i^*(u, v) + \hat{b}_{ai}, \ i = 1, \ldots, J, \ u, v = 1, \ldots, T, \ u \neq v \quad (10)$$

3. Data

The Directorate of Taxes is responsible for assessing taxes and for collecting them. The local tax offices in Norway use about 60% of all labour of the Directorate, and are responsible for tax assessment for all types of income tax. In that connection the tax offices are also responsible for keeping track of changing addresses of persons and companies. A motive for collecting primary statistics at the level of a local tax office is then that an updated address register of people and firms is necessary for the quality of tax assessment. Such statistics are also collected to help other public sectors. A motive for collecting data on outputs is to be able to keep track of the work load of a tax office by the central decision-making unit. This is necessary in order to obtain a realistic picture of the local activities and control the allocation of resources to offices.
The present exercise is restricted to use pre-existing data. In view of the observation made in the introduction of difficulties with measuring inputs and outputs in the public sector not operating through markets, it is pertinent to ask if the available data is good enough for the purpose of measuring efficiency. The Directorate has answered cautiously affirmative since statistics of the main activities in the form of many detailed indicators are kept for internal use. As mentioned in the introduction the Directorate of Taxes has had an extensive discussion about the most relevant measures for outputs and inputs. Furthermore, the data set has been controlled in several ways, e.g. finding extreme values by inspecting the distribution of variables, abnormal changes from year to year, etc., and this internal process of data control in several ways, e.g. finding extreme values by inspecting the distributions of has ensured an acceptable quality of the data.

Although the data are not collected primarily to serve the purpose of efficiency and productivity studies of offices the existing output data is not based on input costs, but constitute independent quantity measurements and thus may be used for such studies. One could wish more information on quality both of outputs and inputs, but since efficiency and productivity studies of the nature reported here and in Førsund et al. (2006) are new to the Directorate, the effort of gathering relevant variables for the study based on existing data was seen as enough for a first step.\footnote{In Bakli et al. (2001) the importance of developing data specifically for efficiency and productivity studies if such efforts are to be taken seriously is emphasized.}

The list of the variables chosen for the study together with some key information about the variables is given in Table 1. Only one input is specified; the total use of resources measured in money. Six outputs are specified representing the main activity areas. Main activities are to

Table 1 – The data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Year</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Range</th>
<th>Sum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>X: The cost of deployment of resources including manpower, offices and current expenses. The cost has been adjusted for compensation in the budget for special circumstances, like rent and travel costs.</td>
<td>2002</td>
<td>2 804 888</td>
<td>171 593 294</td>
<td>168 788 406</td>
<td>1 226 341 592</td>
<td>12 513 690</td>
<td>18 291 279</td>
</tr>
<tr>
<td></td>
<td>2003</td>
<td>2 998 829</td>
<td>177 198 456</td>
<td>174 199 627</td>
<td>1 250 478 743</td>
<td>12 759 987</td>
<td>18 985 940</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td>2 884 602</td>
<td>172 265 680</td>
<td>169 381 078</td>
<td>1 247 754 940</td>
<td>12 732 193</td>
<td>18 437 839</td>
</tr>
<tr>
<td>Y1: Number of people relocated during the year registered by home address and number of immigrations and emigrations.</td>
<td>2002</td>
<td>633</td>
<td>97 028</td>
<td>96 365</td>
<td>602 963</td>
<td>6 153</td>
<td>10 686</td>
</tr>
<tr>
<td></td>
<td>2003</td>
<td>633</td>
<td>101 186</td>
<td>100 523</td>
<td>611 812</td>
<td>6 243</td>
<td>11 127</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td>804</td>
<td>110 497</td>
<td>109 693</td>
<td>643 080</td>
<td>6 562</td>
<td>11 973</td>
</tr>
<tr>
<td>Y2: Number of false registrations detected by control activities.</td>
<td>2002</td>
<td>0</td>
<td>799</td>
<td>799</td>
<td>3 783</td>
<td>39</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>2003</td>
<td>0</td>
<td>1 526</td>
<td>1 526</td>
<td>4 701</td>
<td>48</td>
<td>156</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td>0</td>
<td>3 299</td>
<td>3 299</td>
<td>6 925</td>
<td>70</td>
<td>337</td>
</tr>
</tbody>
</table>
process tax returns from individuals and returns from the two types of businesses that are specified; self-employed and limited companies. In addition one variable covering treatment of complaints and two variables covering activities checking the information about addresses are included.

Concerning the one input cost variable the dominating expense is labour (about 80%, as is also the case for Belgian tax offices reported in Moesen and Persoons, 2002). The evaluation in the Directorate of Taxes was that it would require too much effort to make a more detailed breakdown (e.g. labour, office space, equipment, materials) comparable between units. One requirement for applying the DEA method fruitfully is that the units are using resources to produce a common set of outputs under the same external conditions. In order to take into consideration that some input activities are non-discretionary depending on location (driven by e.g. population, population density, size of geographical area) and as such not contributing to the measured outputs, the input data has been corrected by deducting items like office rent, special administrative expenses and travel expenses. This procedure is supported by the fact that the objective of the Directorate of Taxes is to find explanations for efficiency and productivity differences that in the short run can be used to improve the performances of offices. In the longer run the importance of external conditions should be studied more closely.

We have chosen to pool the data for the 98 units for the three years for which we have observations since it seems reasonable to assume that the technology is stationary over the three-year period. The cost input variable is adjusted using the consumer price index.

4. Productivity development

Due to the short time span we have data for and lack of information about development of frontier technology for tax offices we have assumed that the technology is the same for all years. This means that when we measure the productivity development for an office it is the change in efficiency relative to the optimal scale that will constitute the productivity change. In the definition of the Malmquist index Eq. (7) the technology index $s$ refers to the pooled sample, and the years $u$ and $v$ for a unit may be bilateral combinations of the years 2002, 2003 and 2004. We have assumed that the true values of the Malmquist index are independent over time and have followed the bootstrap procedure outlined in Section 2 (Simar and Wilson, 1999). Due to increased mean-square errors applying bias-correction the original DEA estimates are used, but confidence intervals are established following the procedure outlined in Section 2.
The productivity development for the units between 2002-2003, 2003-2004 and 2002-2004 are set out in Figure 3. All the 98 tax offices are shown, represented by histograms with the width proportional to the input costs (for 2004). The histograms are sorted for ascending values of the productivity index. In addition the lower and upper 95% confidence interval limits are shown. The horizontal line at the value of 1 delimitates units with productivity decline and increase.

In Panel (a) the development from 2002 to 2003 shows that units representing about half of the costs have had a productivity decline, starting at an index value of 0.78, meaning that the productivity has declined with 22% for this tax office, while the other half have had a productivity increase up to 38%. The group of offices with the highest decline and the highest increase are both rather small. Large and medium sized offices show mostly productivity decline with some exceptions. The confidence intervals show that the large and medium-sized units have the widest intervals. Both the groups with the highest productivity decline and the strongest increase have the narrowest intervals, implying that their productivity developments are rather accurately estimated. A strategic question is whether decline and increase are significant. This can be tested simply by inspecting whether the value of 1 is contained in the confidence interval. It is especially for the large units with wide confidence intervals that a hypothesis that the productivity level has remained constant cannot be rejected.

The development shown in Panel (b) for 2003–2004 reveals that more units have productivity decline distributed over about the same interval as the previous period. The units with decline now represent 58% of the input costs (in 2004). The productivity growth is markedly weaker.
Measurement of productivity in the non-business sector

Panel (b) – 2003-2004

Panel (c) – 2002-2004
with just a few units achieving higher growth than 12%. But even a more pronounced difference is revealed by inspecting the confidence intervals. They have widened, especially for large units, thus implying that for units representing about 2/3 of the input costs a hypothesis of constant productivity cannot be rejected.

The development over the whole period is set out in Panel (c). Since there are only three years of observation one should be careful interpreting trends. The productivity change distribution ranges from a 20% decline to a 35% increase. Taken at face value the results indicate that units representing about half of the costs (in 2004) have had a productivity decline over the three years, while a half has had a productivity improvement. This can be visualized by comparing the ‘triangles’ formed by the areas between the end of the histograms for productivity changes and the line of 1, i.e. the triangle on the left below the line for productivity decrease and to the right above the line for increase. The average (unweighted arithmetic) is a growth of 4%. Among units with productivity improvement the small ones dominate. Some average sized units have had slight improvements while others have experiences decline. The confidence intervals shown as step curves in Panel (c) show a pattern of relatively small intervals for small sized units while the intervals for the larger units tend to be much wider (with a few exceptions). The implication is that we can trust the results for the small units, but that we must be careful when using productivity figures for the large units. More reduced data densities in the neighbourhood of large units make the determination of the productivity score more uncertain. Some of the medium and large units with productivity decline have upper limits of the confidence interval above 1, so a hypothesis that these units have had zero productivity growth cannot be rejected. A closer inspection reveals that 25 units have had a significant productivity decline, and 45 units significant increase, while for 28 units a hypothesis of constant productivity cannot be rejected, representing 18%, 36% and 54% of input costs respectively. Both groups with significant decline or growth are thus dominated by units smaller than the average, while in the group of units for which a hypothesis of constant productivity cannot be rejected we have that the units are markedly larger than the average including the largest unit in the sample. But the second largest unit is an exception and is the only large unit with significant growth, both in the period 2003 – 2004 and for the whole period 2002 – 2004.

By sorting the units according to ascending value of the higher confidence interval limit, respectively the lower confidence interval limit the structure of the group with offices with declining, respectively increasing productivity can be illustrated as in Figure 4. Panel (a) shows
Figure 4 – Identifying units with significant productivity decline respectively increase 2002-2004

Panel (a). – Units with significant productivity decline 2002-2004

Panel (b). – Units with significant productivity growth 2002-2004
the group with significantly declining productivity as the units located below the horizontal line of 1. We see that the group is rather homogeneous as to the small size. In Panel (b) the group with significant productivity increase is found above the horizontal line of 1. We see that this group also mainly consists of small units, but that one large one, in fact the second largest, has had productivity growth. Visual inspection of Panels (a) and (b) shows that the group with increase is markedly larger than the group with decline (the size scale is the same).

Comparing the change in the resources used and the productivity score provides a further characterization of the nature of productivity growth (see Førsund and Kalhagen, 1999). In Figure 5 a scatter diagram of productivity change from 2002 to 2004 is shown together with the relative change in input costs. To the left of the origin costs have decreased from 2002 to 2004 while to the right costs have increase. The total range is from -20% to +23%. Together with the horizontal line of 1 delimitating the units with productivity decrease and increase respectively the lines form four quadrants numbered I to IV. In quadrant I units have had both productivity growth and cost increase. Such units may be said to have experienced efficient cost increase. The unit (indicated by an arrow) with the highest cost increase has had a productivity growth of 5% and has expanded the costs with 14%. The unit with the highest productivity increase with 13% has had a cost increase of 3%. The units in quadrant II have also had productivity growth, but experienced cost reductions. This may be termed efficient cost savings. The unit with the highest productivity change has had an increase of 35% (maximal of all units) and reduced the costs with the maximal cost savings of 20%. Another unit with the second highest productivity growth of 32% has had a cost decrease of 9%. In quadrant III productivity decrease is combined with cost decrease.
This is *inefficient cost savings*. There are no units in this quadrant. Units in quadrant IV have the worst of both worlds with decreasing productivity and increasing costs. This is *inefficient cost increase*. The unit with the highest productivity decline, 14%, has had the maximal cost increase of 23%. The unit with the second-highest productivity decline of 13% has had a cost increase of 9%.

The size structure of the units located in the different quadrants can be revealed by making a histogram representation using as the sorting criterion ascending values of the relative cost increase. In Figure 6 the group with efficient cost savings is readily identified as the units located above the horizontal line of 1 starting from the left. Most units are rather small with also some medium-sized units. The group of units with efficient cost increase consists of units located above the horizontal line of 1. There is one large unit and two medium-sized units in this group. Units below the line all belong to the group with inefficient cost increase. We notice that most of the large (including the largest) and medium-sized units belong to this group.

### 5. Conclusions

Productivity measurement in the public sector may be based on a top-down approach or a bottom-up approach. The advantage of a bottom-up approach followed here is that existing primary-data collection at the micro level of parallel service production units can be utilized, not only to measure the aggregate productivity performance of interest for external use like reporting the figure of 4% productivity increase on the average over the three-year period, but also to reveal the productivity performance of individual units, thus providing necessary information for providing explanations for differences in productivity performances across micro units.
The main objective of performance measurement of units is to characterise productivity performance in such a way that ways of improving performance can be found. This is of especial importance for a public service production sector not selling the services in a market and facing accountability and stakeholder interest in performance. The present study has shown that is of crucial importance to use methods that enables us to make a statistical assessment of the uncertainty of productivity estimates that are the ‘engine’ of performance measurement over time. The results show that large units may easily appear with a better productivity performance than they should have if uncertainty is explicitly introduced. Establishing confidence intervals for productivity performance makes it possible to test hypotheses about declining or increasing productivity in a rigorous way.

The productivity results reveal changes even over short periods. Part of the changes must be attributed to internal budgeting procedures naturally lagging real changes in tasks that are mainly exogenous. As more time periods accumulate productivity analyses should provide more valid information on inherent qualities of tax offices expressed by the labels efficient cost increase, efficient cost savings, inefficient cost savings and inefficient cost increase used in this study. Taken at face value units representing half the costs have had productivity decline and half have had productivity increase in the range of -20% to +35%, and resulting in an overall productivity increase of about 4%. The range of change may seem somewhat surprising for such a short period. For any policy actions it should be noted that the confidence intervals for the large units are wide, while they are narrow for small units. It is also of interest to note that both small and large tax offices are found in both the two groups of offices with significant decline and increase of productivity respectively. Therefore causes of productivity differences cannot be attributed to size in general.

The type of performance evaluation performed in this study reveals inefficiency and productivity structures, but does not provide ready explanations of causes for the revealed differences. This is left for further research. A good start will be to study the units appearing as the units with the best productivity performance in Figures 4 – 6, and check e.g. their pattern of use of resources and composition of outputs compared with the average in order to generate hypotheses about factors explaining productivity differences.

References


MINUTES OF THE WORKSHOP IN LONDON ON ‘IMPROVEMENT OF NON MARKET OUTPUT IN EDUCATION AND HEALTH
London, 3-5 October 2006

Alain Gallais
OECD

1. This workshop was co-organized by OECD, the British ONS and the Ministry of Finance of Norway. The attendants were numbered between 160 and 180. Some health experts left after the first day and a half, as it was planned to be consistent with the ‘OECD health accounts working group’ in Paris, October 5th-6th. Other experts (of national accounts, purchasing power parities, education, public services efficiency fields) remained the whole two and a half days. All documents, including PowerPoint presentations are available on: http://www.oecd.org/document/34/0,2340,en_2649_33715_36450978_1_1_1_1,00.html.
A full report of this workshop, including oral interventions, will soon be established, under the coordination of Peter C. SMITH, expert on health economics from the University of York and chairman of the UKCeMGA advisory board. The UKCeMGA proposed a further cooperation with OECD after the workshop, to be precised. Next similar workshop is forecast in Paris, 7th-8th of June, 2007.

2. The relationship between output and outcome was not studied specifically during one session, but evoked in quite all, as this problematic is identical to this of quality adjustment. It was during session 6 (‘Education – international databases and comparisons’) that this topic was more precisely treated, because of the detailed formula for Purchasing Power Parities relying on PISA results, and the explanation of what exactly was to pick from outcome to determine output. A debate proved that there was not yet a consensus between policy makers, education experts and national accountants on the taking into account of ‘inequality’ of knowledge and skills within one country according to PISA, for instance. Outcome is multi-dimensional, output is one-dimensional (because it is resumed in monetary terms): the link is necessarily complex.

3. Similarly, each contribution to each session used words as ‘activities’, ‘output’, ‘outcome’, ‘productivity’, ‘efficiency’ according to its own terminology, which was slightly different. An attempt of definitions was given by the contribution of Wouter Van DOOREN in session 1, but it was recognized that further improvement needs a clarification of this terminology, common to all fields as far as possible.

4. The countries’ experiences in implementing the output methods in education and health were quite useful to learn that 1) almost all European countries have applied or will apply soon output methods in education and health, but 2) few apply quality adjustments.
5. From a PPP point of view, or for solutions on quality adjustment, health was less precisely treated than education. The occurring of a new task force ‘health-specific PPP’ led by OECD has been exposed, and under condition of being funded by European DG SANCO would be in charge of developments in this topic. The background document that I redacted with Alwyn PRITCHARD of UK provided a solution to the objections of health experts with the current output method applied in national accounts.

6. The treatment of PISA for cross-countries comparisons suggested to education experts a similar treatment of national sources on examination data, and an upcoming working group in Den Hag (Netherlands) in December could be in charge of such a topic.

7. During the conclusion session, François LEQUILLER exposed the next steps (in particular the next workshop in Paris, 7th-8th of June, 2007) and the table of contents of what should be the ‘OECD manual on best practices in output measurement for education and health’:

- Synthesis of national experiences and results
- Definitions, principles and terminology (input / output / outcome)
- Then for education and health respectively:
  a. Definitions, principles, terminology
  b. Review of (best) stratification and quantity indicators
  c. Review of best quality indicators
  d. Temporal formulas
  e. Spatial framework and formulas
  f. Verification of consistency between market and non-market methods

An interim manual will be written for the end of March 2007, then debated in the Paris workshop and the final manual will be presented in the next OECD working party on national accounts, in October 2007.
Annex:
Definitive agenda of the Joint OECD/ONS/Government of Norway workshop ‘Measurement of non-market output in education and health’

*Tuesday, 3 October (9h30 – 17h30)*

**Morning**

**Session 1 – Measuring non-market output, efficiency and productivity.** Chair: F. LEQUILLER

| 1.1 | The national accounts perspective, time and spatial comparisons (PPP) : the projects of Eurostat and OECD (P. KONIJN, Eurostat/C1, A. GALLAIS, OECD/STD/NAFS) |
| 1.2 | Stakeholders : the key to success in measuring the productivity of government (J. GRICE, ONS/UKCeMGA) |
| 1.3 | Meeting the need for more timely, detailed and better quality data on government expenditure (V. ALDIN, ONS/UKCeMGA) |
| 1.4 | Non-market services : the view from the national accounts (D. CAPLAN, ONS) |
| 1.5 | StatRes, a Norwegian project for measuring and displaying public sector performance (M. Strømgren, FAD) |
| 1.6 | The use of indicators on output in management and policy applications (W. Van DOOREN, university of Leuven) |
| 1.7 | 12h30 : Welcome by K. DUNNELL, National Statistician |

**Afternoon**

**Session 2 – Health – recent experience in measuring output growth.** Chair: T. ATKINSON

| 2.1 | The Dutch experience on measuring health output and labour productivity (F. KLEIMA and A. CHESSA, CBS) |
| 2.2 | Measuring health output and productivity in the UK : an essential element of public accountability (C. E. LITTLE and P. LEE, ONS/UKCeMGA) |
| 2.3 | The Portuguese experience : estimation of the output of hospitals according to different methodologies (I. QUINTEL, INE) |
| 2.4 | Measurement of Health Output – experiences from the Norwegian National Accounts (A.-L. BRATHAUG, SSB) |
| 2.5 | Review of European and OECD countries experiences, from questionnaires on health services launched in June 2006 (F. MALHERBE, Eurostat/C1 and A. GALLAIS, OECD/STD/NAFS), followed by a general discussion on health output methods recent experience |
|  | Background document : The Czech experimental calculations in health (Eva JEDLICKOVA, CZSO) |

**Session 3 – Health – quality indicators and value of output.** Chair: A.-L. BRATHAUG

Discussant : M. WEALE

| 3.1 | OECD project on Health Care Quality Indicators (E. KELLEY, OECD/ELS/HD) |
| 3.2 | Health quality indicators and value of output (A. SIMKINS, DH) |
| 3.3 | Measuring the Output of Health in the United States (M. CHRISTIAN, BEA) |
**Wednesday, 4 October (9h30 – 17h30)**

**Morning**

Session 3 – (continuation if needed)

Session 4 – **Health** – cross-country comparisons: towards health specific PPPs
Chair: P. SCHERER

| 4.1 | International Comparisons of Prices and Volumes in Health Care among OECD Countries (M. HUBER, European Center for social welfare policy and research) |
|     | Background document: Proposal of an output method for PPP on (non market) health services (A. GALLAIS, OECD/STD/NAFS and A. PRITCHARD, ONS/UKCeMGA) |

**Concluding Health Session.** Chair: P. SCHERER

| 4.2 | Implications for OECD work on Health Accounts (E. OROSZ) |
| 4.3 | 'tour de table' of health and PPP experts |

**Afternoon**

Session 5 – **Education** – experience in measuring output growth. Chair: P. KONIJN

| 5.1 | Measuring education output and productivity in the UK (M. LUCIANO, ONS, UKCeMGA) |
| 5.2 | The point of view of the British Ministry of Education (D. GARNISS, Dfes) |
| 5.3 | Measuring the Education Function of Government in the United States (M. CHRISTIAN, BEA) |
| 5.4 | Methodology for Measuring Education Output Using a Human Capital Approach in the U. S. (B. FRAumeni et al., university of Southern Maine) |
| 5.5 | Volume measures of education in the Norwegian National Accounts (K.-W. DAM and K. SØRENSEN, SSB) |
| 5.6 | Assessing Performance within Higher Education in Norway (O.-J. SKODVIN et al., Norwegian Ministry of Education and Research) |
| 5.7 | The Czech experimental calculations in education (Eva JEDLICKOVA, CZSO) |
| 5.8 | Review of European and OECD countries experiences, from questionnaires on education services launched in June 2006 (F. Malherbe, Eurostat/C1 and A. GALLAIS, OECD/STD/NAFS) |


**Thursday, 5 October (9h30 – 13h30)**

**Morning**

### Session 6 – Education – international databases and comparisons. Chair: J. GRICE

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1</td>
<td>PISA database for pupils of 15 years of age, and similar data for other ages (including projects). Output and outcome. (M. DAVIDSON, OECD/EDU)</td>
</tr>
<tr>
<td>6.2</td>
<td>Inter-country comparisons of volumes and prices: non-market education services (A. GALLAIS, OECD/STD/NAFS and A. PRITCHARD, ONS, UKCeMGA)</td>
</tr>
<tr>
<td>6.3</td>
<td>Project of updating the UOE collection (L. MEJER, Eurostat/F4)</td>
</tr>
</tbody>
</table>

### Session 7 – other non-market services. Chair: F. LEQUILLER

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1</td>
<td>Can the output of collective services be measured? The case of Public Order and Safety (P. CULLINANE, ONS/UKCeMGA)</td>
</tr>
<tr>
<td>7.2</td>
<td>A new approach to measuring the provision of adult social care (J. EBDON, ONS/UKCeMGA)</td>
</tr>
<tr>
<td>7.3</td>
<td>Productivity in the central government sector of Finland – the case of the Ministry of Transport and Communications 2000-2004 (J. HEIKKINEN and M. LEHMUSKOSKI, Statistics Finland)</td>
</tr>
</tbody>
</table>

#### Background document: Volume calculations of individual government production in the National Accounts of Sweden (B. MAGNUSSON, Statistics Sweden)

### Session 8 – Round table for conclusions. Chair: P. SCHREYER

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1</td>
<td>Round table for conclusions (J. GRICE, J.-A. PERSSON, F. LEQUILLER, P. SCHERER, M. DAVIDSON, P. KONIJN)</td>
</tr>
</tbody>
</table>
Human capital, knowledge and their impact on measuring productivity
KNOWLEDGE PRODUCTIVITY

DESIGNING AND TESTING A METHOD
FOR MEASURING KNOWLEDGE PRODUCTIVITY IN ORDER
TO GIVE DIRECTION TO KNOWLEDGE MANAGEMENT
INITIATIVES

Christiaan STAM
INHOLLAND University
Centre for Research in Intellectual Capital
PO-box 261, 1110 AG Diemen
The Netherlands
de Baak – Management Centre VNO-NCW
Professional Development Centre
PO box 88, 3970 AB Driebergen
The Netherlands
+31 (0)6 51 989280
stam@intellectualcapital.nl

Abstract

The nature of production has changed. As a consequence, organizations have lost sight of the sources of productivity and productivity developments. Main objective of this research is to design and test a method for measuring knowledge productivity in order to give direction to knowledge management initiatives. This research project is based on the paradigm of the Design Sciences. This paper presents the methodological and theoretical framework, the initial design of the method and the results of the first two case-studies.
Executive Summary

1. Introduction

The changing nature of economy – from labour and capital intensive to knowledge intensive – puts management for a problem. Although we are still speaking about ‘products’ and ‘productivity’, ‘an ever increasing share of GDP resides in economic commodities that have little or no physical manifestations’ (Youngman, 2003: p.7). Output has become intangible, which makes it hard to quantify. Moreover, production processes have become diffuse, because the traditional clear distinction between input, transformation and output have faded, which makes it very difficult to relate changes in output to changes in the use of means of production. Today’s production process is circular rather than linear, because the result of the process also serves as input to the new processes. Any product is also raw material. Finally, the result of today’s production process is characterized by a high degree of variability of outcome, which makes comparison and objective measurement impossible (Stam, 2001). Consequence of this change is that managers have lost sight of the sources of productivity and productivity growth, which leads to distorted resource allocation and poor (external) communication about organizational performance.

Starting point of this research is that productivity, in the sense of efficient and effective use of resources, is the main determinant of profitability and competitiveness. Therefore it is of vital importance that organizations have a clear sight on the drivers of productivity and productivity developments in order to explain and improve organizational performance.

2. Research objective and questions

Main objective of this research is to develop a practical method to help organizations to measure knowledge productivity and to assess the quality of the preconditions for knowledge productivity in order to give direction to knowledge management initiatives. Underlying assumption of this research objective is that measurement leads to better understanding and better resource allocation, which eventually leads to better organizational performance. In this respect the act of measuring can be seen as an intervention that contributes to improving organizational performance (Kaplan & Norton, 2004). The main aim of measuring knowledge productivity is not to look backward, but to improve future performance. Another assumption is that knowledge management initiatives directly contribute to increasing the preconditions for knowledge productivity and indirectly to improving organizational performance. Based on this research objective, the main research question was defined as: how to measure knowledge productivity in order to give direction to knowledge management initiatives?

3. Research methodology

Major element of this research is to design a method and test the effectiveness of the method in practice, therefore this research follows the paradigm of the Design Sciences developed by Van Aken (1994; Van Aken, 1996, 2004a, 2004b). The ultimate objective of the Design Sciences is to develop valid and reliable knowledge to be used in designing solutions to problems in the field in question (Van Aken, 2004b: p.225).
In the design sciences the typical research product is the heuristic prescription, technological rule or solution concept. Although the research will be driven by and will take place around local problems, the applicability of the solution concepts will be non-local. This means that the solutions can be used to solve similar problems (class of problems) in similar contexts (class of contexts). Solution concepts are typically studied within its intended context of application, in order to be as sure as possible of its effectiveness, also under the influence of less well-known factors. Therefore, the typical research design is the multiple-case. ‘Through multiple case-studies one can accumulate supporting evidence which can continue until ‘theoretical saturation’ has been obtained’ (Van Aken, 2004b: p.235).

This research follows the developing multiple case study in which the solution concepts are developed and tested in collaboration between the researcher and the people in the context of application. The developing multiple-case study operates as a learning system, based on the reflective cycle: each case is analyzed, lessons are drawn and improvements are made before the method is tested again. This process is repeated until sufficient supporting evidence has been obtained.

4. Design of the method

The design objective within this research is to develop a practical method to measure knowledge productivity in order to give direction to knowledge management initiatives.

Application domain

Based on the distinction made in the Danish Guidelines (STI, 2003) and by Andriessen (Andriessen, 2004), this method primarily aims at solving internal management problems related to measuring and managing intangible resources (class of problems). This method will be tested in Dutch medium-sized (50-250 employees) knowledge intensive professional service firms (class of contexts).

Requirements

The next step within the research process was to formulate a set of requirements for the new method. Based on Van Aken (1996) and Weggeman (1995), Andriessen (2004) we made a distinction between limiting conditions, functional requirements, operational requirements and design limitations.

Method

The concept of Knowledge Productivity has been subject of different interpretations. Van Lakerveld (2005) signals three different approaches: the epistemological, the economic (organizational science), and those who stress the importance of learning processes. The main distinction between the latter two seems to be the answer to the question: ‘is knowledge productivity a process or an outcome?’ Depending on the choice that is made between these two, Stam and Evers (2004) signal two dominant complementary approaches, which they call economic and process approach. Whereas the economic approach dominantly seems to aim at identifying and measuring knowledge productivity, the process approach dominantly aims at identifying the necessary preconditions for enhancing knowledge productivity. Because the aim of this research is both to measure and improve knowledge productivity, this method combines both the economic and process approach. In order to measure knowledge productivity, this research uses Zegveld’s (2000) quantitative framework. In order to assess the
quality of the preconditions for knowledge productivity, this method uses Kessels’ (1996) concept of the Corporate Curriculum. Both approaches are combined in a knowledge productivity framework.

Figure 1 – Knowledge Productivity Framework

This Knowledge Productivity Framework served as a theoretical starting point for the development of the first version of the method.

5. Testing the method

The method will be tested by the researcher in 4-6 iterations (alpha testing) and by third parties in 1-2 iterations (beta testing). At this point two iterations have been completed. The first test was in an IT consultancy firm with 110 employees. Based on this first test the method was improved and tested again in another consultancy firm with 43 employees (focus on SME-related topics).

6. Conclusions

Based on the first two iterations we have drawn conclusions about the results and effects of the method, about the requirements and the design of the method and about the application domain. These conclusions serve as a starting point for further improvement and testing of the method in the next iterations.

7. Acknowledgement

This PhD-research project is a joint initiative of the Center of Intellectual Capital Research of INHOLLAND University and de Baak – Management Center of the Dutch Federation of Industries (VNO-NCW).
References


Introduction

The sources of productivity have always been the main subject of economic debate because it is the main determinant of profitability and competitiveness. In order to improve productivity we should be able to identify the sources of productivity. The past decades our production process has changed. Traditional factors of production, like natural resources, labour and capital have lost significance. At the same time the importance of intangible inputs, like information and knowledge, rose. Knowledge has become the main ingredient in products and services (Drucker, 1993; Castells, 1996; Nonaka & Takeuchi, 1997). Consequence of this transformation is that managers have lost sight of the sources of productivity and productivity growth, which leads to distorted resource allocation and poor (external) communication about organizational performance (Eustace, 2000; Blair & Wallman, 2001; Eustace, 2003; CEC, 2006). Important underlying cause of these problems is the fact that we do not have the concepts to reveal and communicate about this new type of productivity. Moreover, management does not have the methods and tools to reveal the effectiveness of knowledge-based production processes (Sveiby & Lloyd, 1988; Edvinsson & Malone, 1997; Stewart, 1997; Sveiby, 1997) and subsequently they do not know how to improve knowledge productivity. The lack of available information about the effective use of knowledge hinders management to design effective policies aiming at improving organizational performance. It is of vital importance that management and organizations have a clear sight on the drivers of productivity and productivity developments in order to explain and improve organizational performance. Therefore we need a new theory that puts knowledge in the centre of the wealth creating process (Drucker, 1993). Aim of this research is to contribute to this challenge.

Research objective and question

Main objective of this research is to develop a practical method to assess the quality of the sources of knowledge productivity in order to give direction to knowledge management initiatives. Underlying assumption is that knowledge management initiatives aim at improving knowledge productivity (Nonaka & Takeuchi, 1997; Weggeman, 1997). This means that knowledge productivity is the result of knowledge management and one of the drivers of organizational performance. The latter implies that knowledge management directly contributes to improving knowledge productivity and indirectly to improving organizational performance.

Figure 1 – Relationship between knowledge management, knowledge productivity and organizational performance
Another assumption of this research is that measurement leads to better understanding, better communication and better resource allocation, which eventually leads to better organizational performance. This assumption is in line with the concept of intellectual capital measurement (Edvinsson & Malone, 1997; Stewart, 1997; Sveiby, 1997) and the concept of performance management (Kaplan & Norton, 1999, 2004). In this respect the process of measuring is more than only assigning scaled numbers to items (Swanborn, 1981). The process of measurement should also be seen as an intervention that contributes to improving organizational performance. Through measurement, organizational performance gets a body, which makes communication and interpretation possible (Mouritsen et al., 2002). In this sense measurement supports the organizational learning cycle.

As the research objective is to acquire knowledge about how to measure knowledge productivity in order to give direction to knowledge management initiatives, the main research question is how to develop a practical method to measure knowledge productivity in order to give direction to knowledge management initiatives?

**Research methodology**

Major element of this research is to design a method and test the effectiveness of the method in practice, therefore this research follows the paradigm of the Design Sciences developed by Van Aken (1994; 1996; 2004b; 2004a). The core mission of the Design Sciences is to develop general knowledge which can be used by professionals in the field in question to design solutions to their specific problems. Driving force of the paradigm of the design sciences is the utilization problem (Susman & Evered, 1978; Van Aken, 2004a) or rigor-relevance dilemma (Argyris & Schon, 1991; Andriessen, 2004b). ‘Management theory is either scientifically proven, but then too reductionistic and hence too broad or too trivial to be of much practical relevance, or relevant to practice, but then lacking sufficient rigorous justification’ (Van Aken, 2004b: p. 221). Therefore, Van Aken (2004b) proposes to make a distinction between Description-driven and Prescription-driven research programs. Whereas the former aims at explaining problems, the latter aims at generating knowledge to be used in designing solutions to solve problems.

The first important characteristic of the Design Sciences is that they are motivated by solving problems. The second distinguishing characteristic of the Design Sciences is the normative or prescriptive nature of the outcome of a research program. Whereas the typical outcome of descriptive research are algorithmic prescriptions (‘if you want to achieve Y in situation Z, then perform action X’). In the design sciences prescriptions are of a heuristic nature. This means that they should be used as a solution concept. A solution concept is a general prescription, which has to be translated (by the professional in the field) to a specific problem at hand. (‘if you want to achieve Y in situation Z, then something like X will help’). These heuristic prescriptions pay respect to the belief that problems are always context related. Although problems and situations might be very similar, this is not a guarantee that a solution that worked in situation A also works in situation B. Therefore, the third main characteristic of the Design Sciences is that the research is justified by pragmatic validity. Whereas the descriptive research leads to propositions which are accepted as true on the basis of the evidence provided, the indeterminate nature of heuristic technological rules makes it impossible to prove its effects conclusively. However, testing of the technological rule in its intended context can lead to sufficient supporting evidence or theoretical saturation (Van Aken, 2004b).
In the design sciences the typical research product is the heuristic *prescription, technological rule* or *solution concept*. Although the research will be driven by and take place around local problems, the applicability of the solution concepts will be non-local. This means that the solutions can be used to solve similar problems in similar contexts. Solution concepts are typically studied within its intended context of application, in order to be as sure as possible of its effectiveness, also under the influence of less well-known factors. Therefore, the typical research design is the multiple-case. Every case serves a specific purpose within the overall scope of inquiry and therefore follows a *replication* logic (Yin, 2003). ‘Through multiple case-studies one can accumulate supporting evidence which can continue until ‘theoretical saturation’ has been obtained’ (Van Aken, 2004b: p.235). Moreover, the multiple-case study operates as a learning system, based on the reflective cycle: each case is analyzed, lessons are drawn and improvements are made before the method is tested again. This process is repeated until sufficient supporting evidence has been obtained.

**Figure 2 – The reflective cycle, Van Aken (1994)**

The reflexive cycle (Van Aken, 1994) should be seen as a combination of the regulative-cycle (problem-solving cycle) of the professional aiming at solving a unique and specific problem and the research cycle of the scientist aiming at the development of knowledge which can be used in a class of comparable problems.

**Literature review**

The past decades a *resource-based view* of the firm emerged. Authors like Penrose (1959), Hamel and Prahalad (1990; 1994), and Stalk et al (1992) contributed to this new strategic paradigm. The resource-based view of the firm is based on the assumption that firm-specific competencies have
Human capital, knowledge and their impact on measuring productivity

become the most important source of sustainable competitive advantage. The traditional competitive environment was stable and transparent. It was characterized by clear defined markets, customers and competitors. Today’s competitive environment however, is characterized by dynamic changing markets and fast changing customer demands. More and more competition has become the ability to anticipate on these changes and thus asks for a more dynamic strategic approach as an alternative to the traditional industry-based view. ‘In the short run, a company’s competitiveness derives from the price/performance attributes of current products (...). In the long run, competitiveness derives from an ability to build, at lower cost and more speedily than competitors, the core competencies that spawn unanticipated products. The real sources of advantage are to be found in management’s ability to consolidate corporate wide technologies and production skills into competencies that empower individual businesses to adapt quickly to changing opportunities’ (Prahalad & Hamel, 1990: p.81). So core competencies are the main resources and can be seen as the collective learning in the organization. The ability to learn is a necessary precondition in order to be able to continuously adapt to changing customer demands.

Authors like Nonaka and Takeuchi (1991; 1997), Leonard-Barton (1995), Sveiby (1988; 1997; 2001), Davenport and Prusak (1998), and many more, claim that knowledge is the most important resource. This knowledge management literature can be seen as a further specification or extension (Bontis, 2002) of the resource-based view into a ‘knowledge-based theory of the firm’ (Grant, 1996). ‘In an economy where the only certainty is uncertainty, the sure source of lasting competitive advantage is knowledge’ (Nonaka & Takeuchi, 1997). However, this does not mean that the knowledge-based view is a synonym for the resource-based view. Most important and fundamental difference is that the resource-based view only implicitly refers to knowledge, whereas the knowledge-based view gives extensive elaborations on the nature and definition of knowledge and the way it should be managed.

Like knowledge management, the concept of knowledge productivity can also be seen as a further specification of the resource-based view. Difference however, is the belief that the competitive advantage of organizations does not come from knowledge itself, but from knowledge productivity, or the extent to which knowledge has been put into use. It was in The Production and Distribution of Knowledge in the United States that Machlup discovered the importance of knowledge as a product. In his recalculation of the national product of the United States, he discovered that total knowledge production in 1958 already accounted for almost 29 per cent of adjusted GNP (Machlup, 1972: original publication in 1962). Moreover, the ‘knowledge-industry’ was not only the largest industry, but also grew faster than the traditional industries. These conclusions led to the observation that there should be some relationship between knowledge, value creation and economic growth.

It was Drucker who translated the macro-economic discovery of Machlup to organizational level, because ‘We know that productivities are created and destroyed, improved or damaged, in what we call the ‘micro-economy’: the individual enterprise, plant, shop, or office’ (Drucker, 1981). Moreover, he realized that the real productive power of organizations is determined by the ability of the so called knowledge workers to make knowledge productive. In The post-capitalist society (Drucker, 1993) he stressed the importance of the development of a new economic theory that puts knowledge in the centre of the wealth creating process. According to Drucker, the competitive advantage of businesses increasingly depends on the ability of organizations to make the knowledge worker more productive. Therefore, ‘knowledge-worker productivity is the biggest of the 21st-century management challenges’ (Drucker, 1999: p.92). In order to face this challenge we should develop a theory that
discloses the relationship between the productivity of knowledge workers and the environment. ‘Not to see the forest for the trees is a serious failing. But it is an equally serious failing not to see the trees for the forest. One can only plant and cut down individual trees. Yet the forest is the ‘ecology’, the environment without which individual trees would never grow. To make knowledge productive, we will have to learn to see both forest and tree. We will have to learn to connect’ (Drucker, 1993: p.180). To get better insight in the sources of productivity we will have to see the interaction between people and their environment.

**Preconditions for knowledge productivity**

Closely related to and in line with Drucker, it was Joseph Kessels (1996; Kessels, 2001b) who introduced the concept knowledge productivity. ‘Knowledge productivity concerns the way in which individuals, teams and units across an organization achieve knowledge-based improvements and innovations’ (Harrison & Kessels, 2004: p.145). Whereas Drucker interpreted knowledge worker productivity as a management challenge, Kessels puts the individual in the centre of his theory. Main underlying assumption of this concept is that ‘the character of labour is changing: routine work is more and more taken over by machines and computers. The work that remains requires independent decision-making and creative thinking; the physical activities of employees are being replaced by mental and social activities. (...) As this change of the character of labour takes place, it is inevitable that the workplace turns into a learning environment. (...) The conditions for good work become similar to the conditions for good learning’ (Kessels & Van der Werff, 2002: p.20). So knowledge productivity requires a good learning environment.

In order to help organizations improve their knowledge productivity, Kessels introduced the Corporate Curriculum: ‘the plan for learning to increase knowledge productivity, leading to constant improvement and radical innovation, and ultimately to economic advantage’ (Kessels, 1996; Kessels & Van der Werff, 2002). The Corporate Curriculum should not be seen as a formal educational or training curriculum. ‘Rather, it involves transforming the daily workplace into an environment where learning and working can be effectively integrated. It facilitates the creation of a rich and diverse landscape that encourages and supports employees in the learning they need to do in order to continuously adapt and to innovate’ (Harrison & Kessels, 2004: p.155). The Corporate Curriculum consists of all the intended and not intended conditions that affect the learning processes among workers in organizations (Van Lakerveld et al., 2000) and identifies seven critical learning functions (Kessels, 1996; Keursten et al., 2004b):

1. Acquiring Subject Matter Expertise and professional knowledge directly related to the organization’s business and core competencies
2. Learning to identify and Solve problems by using the acquired subject matter expertise.
3. Cultivating Reflective Skills and meta-cognitions that contribute to finding, acquiring and applying new knowledge.
4. Securing Communication Skills that provide access to the knowledge network of others and that enrich the learning climate within the workplace.
5. Acquiring skills for Self regulation of Motivation and affection related to working and learning.
6. Promoting Peace and Stability, in order to enable specialization and incremental improvement.
7. **Causing Creative Turmoil in order to stimulate innovation.**

According to Van Lakerveld et al. (2000), distinction can be made to those learning functions that directly refer to the learning processes (1 to 5) and those that refer to the conditions of learning (6 and 7). Within the 5 functions that refer to the learning processes we can make another distinction between those that dominantly refer to the knowledge processes (1-3), and those that dominantly refer to the knowledge workers (4 and 5). The result is that we can make a distinction between three different kinds of learning functions: those related to the individual (competences and motivation), those related to the knowledge processes (subject matter expertise, solve problems, reflection), and those related to the organizational environment or conditions (calm and stability, creative turmoil).

Figure 3 – Three layers of the Corporate Curriculum

This model tries to pay respect to the human-centred definition of knowledge of Kessels. ‘It gives centre stage to the person of the learner, active within a complex of relationships, engagements and commitments with others’ (Alred & Garvey, 2000). Therefore the inner circle represents the learning functions that are dominantly related to the individual. The outer circle represents the learning functions that are dominantly related to the organizational environment or conditions for learning. The circle in between can be seen as a combination of the inner and the outer circle and represents the learning functions which are dominantly related to the organizational learning cycle or knowledge processes (knowledge spiral (Nonaka & Takeuchi, 1997), knowledge value chain (Weggeman, 1997), etc.) as defined by the knowledge management literature. Moreover, these knowledge processes are both related to the people and the organization. They are both human and structural capital.
The result of knowledge productivity

As we have seen, Kessels defines knowledge productivity both as a process and as a result. ‘Knowledge productivity has been described as the key ability of an organization, a team, or employee, to signal relevant information and to develop new competencies. These new competencies are to be applied to the incremental improvement and radical innovation of work processes, products and services’ (Kessels & Van der Werff, 2002: p. 20). Whereas he, and many others, extensively elaborate on the first part of his definition, only very little has been said about the second part which refers to the result of knowledge productivity. Moreover, hardly any research has been done on the relationship between the preconditions of knowledge productivity (Corporate Curriculum) and the results in terms of incremental improvements and radical innovations.

Starting point of the concept of knowledge productivity is the precondition that organizations in the knowledge economy, to be successful, should continuously improve their processes, products and services, and radically renew from time to time (Drucker, 1993; Nonaka & Takeuchi, 1997). Based on Walz and Bertels (1995), Kessels (2001b) makes a distinction between gradual improvements and radical innovation. ‘Gradual improvement (involving adaptive learning) elaborates on what is already present and leads to additional refinement and specialization. Radical innovation (involving investigative and reflexive learning) involves breaking with the past and creating new opportunities by deviating from tradition’ (Harrison & Kessels, 2004: p.157). These different types of learning are acknowledged as the result of knowledge productivity (Keursten et al., 2004b; Van Lakerveld, 2005). Similarly distinction is made between single-loop versus double-loop learning (Argyris & Schon, 1978), ‘higher-level’ and ‘lower-level’ learning (Hedberg, 1981; Fiol & Lyles, 1985); ‘learning I’ and learning II’ (Bateson, 1972); ‘generative’ and ‘adaptive’ learning (Senge, 1992); ‘tactic’ and ‘strategic’ learning (Doddson, 1993), ‘reflexive’ and ‘adaptive’ learning (Guile and Young, 1999 in: Harrison & Kessels, 2004). All of these distinctions refer to incremental improvements to existing practice on the one hand, and radical rethinking of basic goals, norms, and paradigms on the other hand.

The distinction between gradual and radical innovations can also be related to the exploitation/exploitation dilemma (March, 1991). This dilemma represents the two strategic options a company has: exploitation of old certainties or exploration of new possibilities. These dual innovation strategies are acknowledged by many authors as well. Von Krogh et.al. (1994) distinguish between an organization’s need to survive (maintain its position in its current environment) and its need to advance (forge ahead in an emerging new environment) (Harrison & Kessels, 2004). Abell (1999) summarizes these innovation strategies as ‘competing today while preparing for tomorrow’. Based on the punctuated equilibrium (Eldredge & Gould, 1972) and its application on organizational development (Tushman & Romanelli, 1990; Tushman & O’Reilly III, 1996), Zegveld (2000) makes a similar distinction between incremental and radical change. ‘The essential difference between incremental and radical change is that incremental change is about aligning and can be related to the process of production and value creation, while radical change is about the process of forming a company’s perspective and the process of forming resources’ (Zegveld, 2000: pp 26-27). According to Zegveld incremental innovation is about ‘doing things better ’ and radical innovation is about ‘doing better things’. In order to detect incremental and radical innovation, Zegveld (2000) developed a Quantitative Framework for measuring exploitation and exploration.
Knowledge productivity framework

The concept of Knowledge Productivity has been subject to different interpretations. Van Lakerveld (2005) signals three different approaches: the epistemological, the economic (organizational science), and those who stress the importance of learning processes. The main distinction between the latter two seems to be the answer to the question: is knowledge productivity a process or an outcome? Depending on the choice that is made between these two, Stam and Evers (2004) signal two dominant complementary approaches, which they call economic and process approach. Whereas the economic approach dominantly seems to aim at identifying and measuring knowledge productivity, the process approach dominantly aims at identifying the necessary preconditions for enhancing knowledge productivity. Because the aim of this research is both to measure and improve knowledge productivity, this research combines both the economic and process approach. In order to measure knowledge productivity Zegveld’s (2000) Quantitative Framework is used. In order to assess the quality of the preconditions for knowledge productivity Kessels’ (1996) concept of the Corporate Curriculum is used.

In his research about the Corporate Curriculum, Van Lakerveld (2005) finds evidence for the positive relationship between the learning functions of the Corporate Curriculum on the one hand and quality improvements and innovative potential on the other hand. However, he does not make the step from identifying improvements and innovation to organizational performance. Therefore he concludes: ‘the question remains whether organizations with a rich Corporate Curriculum perform better’ (Van Lakerveld, 2005: p.189, translation CS). Future research should elaborate on this topic. Similar conclusions are drawn by Keursten et.al. (2004b). Based on a large reconstruction research in sixteen case studies they conclude that creative turmoil is the driver of the innovation process. Moreover, the reconstruction also revealed a positive relationship between the quality of the Corporate Curriculum and successful innovation processes. However, the case studies did not provide any evidence about the deliberate use of one or more of the elements of this concept. Therefore, they conclude, it would be interesting to find out if a deliberate use of this concept leads to a better performance.

Learning is not a process of gradual change, but intermittent change. Relative long periods of relative stability are alternated with short periods of fundamental change. As Zegveld is mainly concerned with the result, he argues that periods of relative stability can be recognized by incremental innovation, whereas periods of fundamental change can be recognized by radical innovation. As Kessels is mainly concerned with the conditions that lead to innovation, his reasoning is the other way round. ‘Gradual improvements benefits from conditions of relative stability and the time to reflect on what is needed in order to improve current operations and processes. Radical innovation is more likely to flow from the creative turmoil (…)’ (Harrison & Kessels, 2004: p.157). Based on the conviction that good conditions for learning eventually result in incremental and radical innovation, Kessels developed a set of learning functions as preconditions for knowledge productivity (Corporate Curriculum). Based on the same distinction between incremental and radical change, Zegveld developed a Quantitative Framework for measuring exploitation and exploration. Within this research, these two approaches are combined in a Knowledge Productivity Framework (Figure 4). This theoretical framework was used as a starting point for the design of a practical method to measure knowledge productivity and to assess the quality of the preconditions of knowledge productivity in order to give direction to knowledge management initiatives.
The design objective within this research is to develop a practical method to measure knowledge productivity in order to give direction to knowledge management initiatives. The process of designing and testing the method consists of four phases (Andriessen, 2004a): defining the application domain; creating a list of requirements; designing the method; evaluating the design. This section follows these steps in designing a first draft of the method, which is called the Knowledge Productivity Enhancer.

Defining the Application Domain

The first step is to define the context of application, which consists of a class of problems and a class of contexts. Relevant distinctions between different types of problems are given in the Danish Guidelines (STI, 2003b), Andriessen (2004a), and RICARDIS-report (CEC, 2006). All sources make a distinction between internal management problems and external reporting problems. Andriessen even adds a third class of problems related to statutory and transactional issues. As the main aim of the Knowledge Productivity Enhancer is to give direction to knowledge management initiatives, the method seems to fit into the class of problems aiming at solving internal management problems. This method is designed for and will be tested in Dutch medium-sized knowledge intensive professional service firms (class of contexts).

Creating a List of Requirements

The second step of the design cycle is to create a list of requirements for the new method. Van Aken (1996; 2004c), Weggeman (1995) and Andriessen (2004a) make a distinction between four types of requirements. Within this research, three types of requirements should be taken into account. Functional requirements are the performance requirements related to the problems that should be
solved by applying the method. Operational requirements are the requirements from the perspective of the user. Design limitations are the boundaries set by the preferred solutions. Limiting conditions are the technical requirements the environment places on the method when applied. Whereas this method is not subject to limiting conditions, an overview is given of the first three types of requirements.

Functional requirements define the results the method needs to produce. These requirements follow primarily from the problem definition of this research. Therefore, the main requirement of the method is that it should give direction to knowledge management initiatives. In order to do so, the KP-Enhancer should:

1. Create awareness about the importance of knowledge productivity for improving organizational performance.
2. Reveal and assess the quality of the preconditions for knowledge productivity.
3. Measuring the result of knowledge productivity in terms of incremental and radical innovations.
4. Generate possibilities for improvement.
5. Develop a strategy for knowledge management.
6. Improve communication about knowledge productivity.

Operational requirements define the ease of use of the method from the perspective of the users of the method. First of all, the method should be practical and user-friendly. The method should ‘speak for itself’, should be easy to understand and easy to apply. Second, as organizations are reluctant to invest in new concepts or methods, the Knowledge Productivity Enhancer should require a limited amount of time. Third, as the aim of the method is to improve the management of knowledge resources, operational requirement is also that the method makes knowledge productivity tangible in the sense that it can be managed. Fourth, in order to assess progress, the method should be replicable, without help from outside, for the (knowledge management) professional.

Finally, the design limitations are the boundaries of the method set by the preferred solutions. The designer of the method often sets these boundaries. This method, following the outcome of the initial literature review, has three design limitations. First limitation is that the assessment of the preconditions of knowledge productivity will be based on the concept of the Corporate Curriculum (Kessels, 1996). Second limitation is that the measurement of knowledge productivity in terms of incremental and radical innovations is based on Zegveld’s Quantitative Framework (Zegveld, 2000). Third design limitation is that the process of giving direction to initiatives for improvement will be based on the Danish Intellectual Capital Statement Model (STI, 2003b).

**Designing the Method**

The Knowledge Productivity Enhancer combines the concept of the Corporate Curriculum, Zegveld’s Quantitative Framework and the Danish Guidelines for Intellectual Capital Statements and consists of three phases. Aim of the first phase is to identify problems and set the objectives for applying the method. Main aim of the second phase is to gather data for analysis of the current situation. The third
phase gives direction to knowledge management initiatives. An additional fourth phase, in which progress is monitored, is added to stress that applying the method is only the start of a longer process. As it takes time to monitor progress, this phase is not included in this research project.

**Figure 6 – Design of the method within the context of the multiple case study**

Within this research project, applying the method should be seen as a part of the learning cycle of the developing multiple case study. Each case-study is preceded by a *call for cases* and each case study is concluded with an assessment of the effectiveness of the method. These two steps make the connection between the application of the method in practice (practice stream) and the theoretical reflection (knowledge stream) which is characteristic for Design Based Research (Andriessen, 2006). Aim of the practice stream is to solve specific problems. Aim of the knowledge stream is to develop knowledge that can be transferred to comparable situations.

**Call for Cases**

Aim of this step is to find a suitable context of application. In order to qualify as a case study within this research, organizations should fit into the class of contexts for which the method is designed and the problem at hand should fit into the class of problems for which the method is designed. If both criteria are met, then the organization qualifies to serve as a case. The result of this step is a strong indication that the organization qualifies as a case study and is willing to apply the method.

**Intake (phase 1)**

If an organization seems to qualify as a case study, the next step is to verify this assumption. Yin (2003) refers to this step as ‘screening case study nominations’. The intake is based on a semi-structured interview with the sponsor and/or contact person. Main questions to be answered are:

1. Does the organization fit into the class of contexts for which the method is designed?
2. Does the problem at hand fit into the class of problems for which the method is designed?
3. Is this the right moment to apply the method and do we get the necessary support?

If the organization qualifies as a case study, the intake is used to make a formal planning for applying the method. In this sense, the call-for-cases and the intake can be compared with the acquisition of a consultancy assignment. Moreover, these steps provide the opportunity to establish a good working relationship with the ‘customer’. The intake takes about one hour and the interviewees do not have to prepare for the interview. The report of the interview is verified by the informants and serves as a reference point for evaluating the effectiveness of the method (see below).

**Analysis of Current Situation (phase 2)**

Aim of this phase of the method is to gather data. The output of this phase serves as input for the next phase in which we will analyse the data in order to give direction to knowledge management initiatives. The analysis of the current situation is based on Zegveld’s Quantitative Framework and Kessels’ Corporate Curriculum. The former measures the result of knowledge productivity in terms of incremental and radical innovations. The latter reveals the preconditions for enhancing knowledge productivity.

**Measuring incremental and radical innovations (2a)**

Due to practical reasons, within this research Zegveld’s analysis has been simplified. Main simplification has been the reduction of the number of years from twelve to five. According to Zegveld (interview September 2005), a series of five years would be the minimum to recognize a pattern.

**Incremental innovation (exploitation)**

As we have seen above, incremental innovation relates to single-loop learning and is about incremental improvements to existing practice. Within Zegveld’s quantitative framework, incremental innovation is detected by measuring the (in)consistency of the application of a generic strategy. Reasoning behind this is that different generic strategies result in different performance outcomes, which means that a shift from one to another strategy will be reflected in the financial data. Moreover, a change in the focus from one to another generic strategy should be explained by the (implicit or explicit) desire to improve existing practice. Based on Porter (1980) and Karnani (1984), Zegveld makes a distinction between an efficiency strategy, added-value strategy and a volume strategy. An efficiency strategy implies the aim of continuously reducing costs and thus increasing value per unit turnover and therefore can be detected if most of the created value can be related to a decrease of company specific costs per unit turnover. Successful deployment of an added value strategy implies a trade-off between the premium price the customer is willing to pay and the increased cost development as the result of extensive research, product design and intensive customer support to gain additional margins. Therefore, an added value strategy can be detected if most of the created value can be related to an increase of added value per unit turnover. Finally, a volume strategy can be detected when most of the created value can be related to a higher turnover without changing the added value per unit turnover or company specific costs per unit turnover (Zegveld, 2000). Companies are either volume or performance driven. If they are performance driven, they can either have a focus on added-value or on efficiency.
Aim of the Quantitative Framework is to detect stability or instability on a longitudinal basis. In order to do so ‘financial data has been chosen since it can be related to specific developments in the value chain and since financial data is widely available’ (Zegveld, 2000: p. 50). These financial data is used to detect a change in the selection of one of the three generic strategies (volume, efficiency and added value). The two excluding aspects of exploitation are defined as:

1. Stability in exploitation
   Stability implies that a substantial part of the development of the total operational value of a company can be related to a single generic strategy.

2. Change in exploitation: incremental innovation
   Incremental innovation implies that no one single generic strategy realises a substantial part of the development of the total operational value of a company.

According to Zegveld (2000) stability and change are two excluding aspects of exploitation, companies can either be categorized as stable (related to a single generic strategy) or as incremental innovation companies. Only if the figures detect a change in the focus from one to another strategy, then we speak about incremental innovation. ‘Incremental innovation can be the result of a change of deployment from one generic strategy towards a different generic strategy or can be the result of an absence of a generic strategy within the period of analysis. A discontinuity in the deployment of a generic strategy can be the result of explicit or implicit choices by management or by core stakeholders and will affect the process of production and the process of value creation’ (Zegveld, 2000: p.49). Based on Luehrman (1997), Zegveld (2000) proposes to measure the development of exploitation by measuring Operational Cash Flow (OCF) and the contribution of the three generic strategies to the build up of OCF (see Appendix I). If the figures reveal a shift in the contribution to OCF from one to another generic strategy, the company qualifies as an incremental innovation company.

Radical innovation (exploration)

Radical innovation relates to double loop learning and is about the process of forming resources, which leads to a radical new perspective to the company. Like exploitation, radical innovation is a neutral concept and can lead either to creative destruction or creative accumulation. Moreover, like exploitation, radical innovation may be the result of explicit or implicit developments.

In his research, Zegveld (2000) investigates the relationship between the company’s perspective and the development of resources. Based on several studies he concludes that the build-up of resources is stable when, from the company’s perspective, no fundamental changes occur within the firm’s basic orientation towards its resources (customers, employees, partners and shareholders). However, a different perspective on resources leads to a situation where a different and new positional advantage and different competencies are developed, which subsequently leads to a different build up of resources. This new perspective or radical innovation can be detected by measuring the stability of the build-up of resources.

According to neo-classical theory growth is driven by exogenous changes in the different factors of production. However, using only two factors of production (labour and capital) could not fully explain economic growth. Therefore, more and more economists focused on innovation and the development
of knowledge as the (endogenous) source of continuous productivity increase in order to explain the productivity gap (Solow, 1957; Romer, 1990). Nowadays, the assumption that changes in productivity that can not be explained through changes in labour or capital are caused by knowledge, seems to be generally accepted. ‘Starting with the neo-classical model by Solow, different authors have developed different models for calculating the surplus or residual value and hence the Total Factor Productivity (TFP). (...) The residual value is defined as the creation of additional output above the level of the different inputs or resources’ (Zegveld, 2000: p.65). In economic theory a change in the build up of residual value is a determinant of change of perspective. Therefore, based on Solow, Zegveld proposes to use the concept of Total Factor Productivity (TFP) to calculate the residual value caused by knowledge.

Figure 6 – Translating TFP into TRP (Zegveld, 2000)

<table>
<thead>
<tr>
<th>Model</th>
<th>Total Factor Productivity (TFP) (Solow)</th>
<th>Total Resource Productivity (TRP) (Zegveld)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output</td>
<td>Private non-farm GNP</td>
<td>Added value of companies defined as: Turnover minus all outsourcing intermediate goods and services. Output is defined as: Employment costs + Depreciation + Net profit.</td>
</tr>
<tr>
<td>Capital</td>
<td>Employed capital</td>
<td>Depreciation</td>
</tr>
<tr>
<td>Unity</td>
<td>Manhour</td>
<td>Employees (fte)</td>
</tr>
<tr>
<td>Periodicity</td>
<td>Yearly</td>
<td>Yearly</td>
</tr>
<tr>
<td>Correction</td>
<td>Inflation</td>
<td>None</td>
</tr>
</tbody>
</table>

The contribution of Zegveld has been that he applied the concept of TFP to companies instead of countries by broadening the concept of TFP to Total Resource Productivity (TRP). ‘By adapting the TFP model to companies, the residual value is defined as knowledge or intellectual capital’ (Zegveld, 2000: p.53). TRP measures the accumulation of knowledge and the build-up of the residual within the company and may provide insight into how well a company allocates and exploits its resources. Therefore, the build-up of the residual is a determinant of change of the perspective. Reasoning behind this is that discontinuity in the residual build-up at company level should be interpreted as a shift in the deployment of knowledge or intellectual capital. A radical shift in the development of the residual build-up implies a more radical impact due to the deployment of new intellectual capital and can therefore be defined as radical innovation. Radical innovation is defined as a new combination of resources which leads to a new perspective for the organization.

The aim of TRP is to detect stability or change in the perspective of the organization. The two excluding aspects are defined as:

1. Stability in exploration
   Continuity of the perspective of a company and hence continuity of the different stakeholders in relation to the company. Results in a longitudinal continuous build-up of total resource productivity.

2. Change in exploration
   Change of the perspective of a company and hence a discontinuity in the importance of the different core stakeholders in relation to the company.
To detect radical innovation, a change in build-up of knowledge should be observed, whereas ‘a discontinuity of the company’s perspective results in a longitudinal discontinuous build-up of the total resource productivity’ (2000: p.59). According to Zegveld, a sudden positive shift in the development of company specific TRP implies radical innovation or a radical change in the perspective of the company. ‘This change of perspective is related to the development and successful initial deployment of residual value or intellectual capital related to this new perspective’ (Zegveld, 2000: pp.103-104). This reasoning follows the resource-based view of the company which argues that innovation is about new combinations of resources. Within Zegveld’s model, ‘residual value is defined as ‘intellectual capital’ (…) or company specific knowledge which is developed by combining and recombining resources’ (Zegveld, 2000: pp.70-71). Based on the calculation of residual change (see Appendix I) we can determine whether a firm qualifies as a radical innovation firm or not. Radical innovation can be recognized by a significant change in the residual build-up.

Assessment of the quality of the preconditions for knowledge productivity (step 2b)

Parallel to measuring incremental and radical innovations data is gathered about the quality of the preconditions for knowledge productivity. Aim of this step is to reveal the sources of knowledge productivity and get better insight in the current situation. Assumption is that the quality of the preconditions determines the extent to which incremental and radical innovations will be achieved. Therefore, the quality of the preconditions is an indicator of future performance and not directly related to Zegveld’s quantitative analysis. The assessment of the quality of the preconditions for knowledge productivity is based on Kessels’ Corporate Curriculum (Kessels, 1996). As we have seen, the Corporate Curriculum consists of seven learning functions.

Learning function 1: Subject matter expertise

The first learning function has been defined as ‘Acquiring subject matter expertise and professional knowledge directly related to the organization’s business and core competencies’ (Kessels, 1996; Keursten et al., 2004a). In a sense, the first learning function covers the main part of the concept of knowledge management (Nonaka & Takeuchi, 1997; Weggeman, 1997; Davenport & Prusak, 1998; Stam, 2004). Subject matter expertise stresses the importance of ‘strategic grounding’ (Stam, 2004) as it is about knowledge which is directly related to the main work processes and work-related objectives (Keursten, 2001; Keursten et al., 2004b; Van Lakerveld, 2005). Moreover, subject matter expertise is both about tacit and explicit knowledge (Polanyi, 1974; Kessels, 2002), and it is about the way knowledge is developed, shared and codified throughout the organization (Kessels & Keursten, 2001; Keursten, 2001). Subject matter expertise is about the strategic grounding and processing of knowledge and therefore asks for knowledge-based strategies and facilitation of knowledge processes.

Learning function 2: Solving problems

The second learning function has been defined as ‘learning to identify and deal with new problems using the acquired subject matter expertise’(Kessels, 1996; Keursten et al., 2004a). From a knowledge management perspective, solving problems refers to the process of applying (Weggeman, 1997; Davenport & Prusak, 1998), combining (Spek & Spijkervet, 1994; Nonaka & Takeuchi, 1997) or exploiting (Sprenger et al., 1995) knowledge. Within these processes, which are at the ‘end’ of the knowledge value chain, knowledge is put into use, or in other words: made productive. All other
knowledge processes support this second learning function. Second distinguishing characteristic of this learning function is that it stresses the gap between existing subject-matter expertise (as a result of the first learning function) and the knowledge that is needed in order to find solutions for new challenges. Solving problems is the competency with which this gap can be closed. However, the gap will never be closed entirely. New situations always ask for new interpretations of existing knowledge, therefore the need for the ability to solve problems will remain. Moreover, solving problems is a personal capacity and cannot be separated from its context (Sveiby, 1997). If the context is complex and dynamic, the professional’s work should be based on a body of knowledge which has to be interpreted and applied depending on the specific circumstances (Weggeman, 1997). Solving problems is about the ability to renew and stretch expertise and therefore asks for creativity and room for experimenting with new ways of working.

**Learning function 3: Reflective skills and meta-cognitions**

The third learning function has been defined as ‘cultivating reflective skills and metacognitions to find ways to locate, acquire and apply new knowledge’ (Kessels, 1996; Keursten et al., 2004a). The main message of this learning function is that we should not only learn how to develop, share and apply knowledge (first two learning functions), but also reflect on the effectiveness of these processes (Kessels & Keursten, 2001). Meta-learning reflects an organisation’s attempts to learn about (and improve) its ability to learn (Argyris & Schon, 1978). Main questions related to this learning function are: Why are we good in solving problem A, and why is it that we do not know how to handle problem B? What can we learn from our experiences and can we do it better? Reflective skills are necessary in order to learn from past processes (Van Lakerveld, 2005). This learning function enables organizations, teams and individuals to manage their own learning processes. ‘How can we improve our ability to develop, share and utilise knowledge in the workplace, and help others to do so’ (Harrison & Kessels, 2004: p.156). From a knowledge management perspective, this learning function refers to the process of evaluation (Weggeman, 1997; Stam, 2004). Moreover, this process makes the connection to the concept of the learning organization (Senge, 1992). Reflection stresses the idea that the output of the process also serves as input for a new (production) cycle (Nonaka & Takeuchi, 1997; Zack, 1998). Reflective skills are of vital importance for the development of meta-cognitions. Important preconditions for the development of reflective skills are open communication, constructive feedback and creating time and space to look backward (Kessels & Keursten, 2001).

**Learning function 4: Communication skills**

The fourth learning function of the Corporate Curriculum has been described as ‘acquiring communicative and social skills that help people access the knowledge network of others, participate in communities of practice and make learning at the workplace more productive’ (Kessels, 1996; Keursten et al., 2004a). Communication skills stresses that knowledge is processed through people. More and more research is being done to identify the critical skills of the knowledge worker (Sprenger et al., 1995; Tissen et al., 1998; A. Abell & Ward, 2000). Important skills are the ability to communicate and collaborate, as it is through communication and collaboration that knowledge is developed and shared. Another aspect of this learning function is the extent to which the environment supports knowledge sharing. From a knowledge management perspective, this aspect refers to the preconditions for knowledge management in terms of structure and culture, as these aspects have an important impact on the knowledge processes and the knowledge friendliness of the company.
Human capital, knowledge and their impact on measuring productivity

(Weggeman, 1997; Stam, 2004). Communication skills is about the ability to communicate and collaborate and the knowledge friendliness of the organization in terms of structure and culture.

Learning function 5: Self regulation of motivation

This fifth learning function has been defined as ‘acquiring skills to regulate motivation, affinities, emotions and affections concerning working and learning’ (Kessels, 1996; Keursten et al., 2004a). This learning function, also at the heart of the Corporate Curriculum, is the most implicit learning function (Keursten et al., 2004a) and refers to the importance for knowledge workers to identify personal themes and ways to develop these. It is about skills that give meaning to learning and enhance commitment (Kessels, 1996), because ‘in a knowledge economy it is useless when a manager says: Be smarter, or show more creativity! Being smart and creative depend heavily on personal interest’ (Kessels & Van der Werff, 2002). People are only smart if they want to be (Harrison & Kessels, 2004: p.156). Personal interest is closely related to the process of inspiration, passion or motivation and sense-making (Leenheers, 2004). In their reconstruction study, Keursten et.al. conclude that ‘personal motivation and affinity with a particular topic was the driving force behind innovations and improvements’ (Keursten et al., 2004b: p.167). Moreover, a positive correlation has been found between attention to intrinsic motivation and the performance of individuals in the learning process (Vansteenkiste e.a. in: Van Lakerveld, p.16). Self regulation of motivation puts the locus-of-control with the individual, because it implies that the extent to which organisational objectives are achieved, heavily depends on personal entrepreneurship. ‘A personal entrepreneur works from an intrinsic passion and primarily strives for personal interest. He has the ability to organize his work in such a way that it suits his personal preferences. He sees himself as a firm, although he is an employee’ (Rondeel & Wagenaar, 2002: p.123, translation CS). Although motivation can not be ‘managed’, it can be supported by providing space for personal entrepreneurship.

Learning function 6: Peace and stability

The sixth learning function has been described as ‘promoting peace and stability to enable exploration, coherence, synergy, and integration’ (Kessels, 1996; Keursten et al., 2004a). This learning function refers to the need for incremental improvements through further specialization (Ansoff & Sullivan, 1993; Harrison & Kessels, 2004). Peace and stability gives employees the opportunity to explore existing knowledge and search for possibilities to apply this knowledge into their daily practice. Moreover, peace and stability also refers to the need for time for reflection, learning and knowledge sharing. Time and peace provide the opportunity to reflect on the efficiency and effectiveness of processes, products and services. Peace and stability provides a context in which people can experiment, without direct consequences. Peace and stability provides the certainty and the time which is necessary for specialization and improvement (Van Lakerveld, 2005).

From a knowledge management point of view, this learning function refers to the organizational need for a certain degree of redundancy in creating knowledge. Redundancy means that the knowledge level within the organization exceeds the minimum level of knowledge needed to perform the necessary tasks. (Nonaka & Takeuchi, 1997). ‘Lack of redundancy and time to reflect exploit existing (intellectual) resources, and consume these without generating new knowledge. Lack of peace and stability results in impoverishment of intellectual assets’ (Kessels, 2001a; Kessels & Van der Werff, 2002). However, the drawback of this learning function is that ‘too much peace and stability might
bring about overly one-sided specialization and an excessive internal focus’ (Kessels, 2001b; Keursten et al., 2004a). In this sense, Sveiby (1997) argued that stability should be seen as a counter balance of growth and renewal.

Peace and stability is an important precondition for knowledge productivity in general and incremental innovation in particular. Important elements of this learning function are specialization, time to reflect and redundancy.

*Learning function 7: Creative turmoil*

The seventh learning function has been described as ‘causing creative turmoil, which leads to radical innovation’ (Kessels, 1996; Keursten et al., 2004a). Creative turmoil refers to the need for creativity as a driver of innovation and improvement (Shapero, 1985). The cause of the turmoil is often ‘an existential threat: a matter of winning or losing, surviving or going under, being in or out of the game’ (Harrison, 2004, p.156). Although Van Lakerveld (2005) found a positive relationship between work-pressure and learning, not all pressure is creative turmoil. Creative turmoil is mainly recognized by pressure which is caused by ‘the importance that is attached to the outcome of the process or because people themselves feel a strong urge to solve a particular problem’ (Keursten et al. p.168/18).

Many authors have referred to the necessity of a certain degree of ‘strategic ambiguity’ (Nonaka & Takeuchi, 1997), ‘strategic imbalance’ (Itami, 1991), ‘strategic distance’ (Senge, 1992), ‘strategic confusion’ (Stacey, 1995) or ‘strategic disorder’ (Levy, 1994) as a requirement of creativity. According to Senge (1992) ‘distance’ between vision and reality is the source of creative tension as distance makes it necessary to take action in order to come closer to the objective. Similar reasoning can be found in Itami (1991) and Nonaka and Takeuchi (1997). A certain degree of chaos, disorder or even failure may prevent complacency, and could stimulate organizations to stretch beyond their strategic comfort zone. Moreover, creative chaos can stimulate individuals to fundamentally change their ways of thinking and create new knowledge.

Keursten et al. (2004b) argue that external pressure is important to really make a difference in daily work. However, not all unrest is creative turmoil and too much creative turmoil may yield a thousand new ideas but leaves little opportunity to elaborate on them, and will therefore limit innovation. Creative turmoil without the time to reflect will lead to ‘destructive chaos’ (Schon, 1983). This implies that the sixth and the seventh learning function should be in balance. Creative turmoil is seen as a precondition for creating radical innovations. Main prerequisite for this learning function is strategic ambiguity.

*Survey*

The main elements of the seven learning functions of the Corporate Curriculum (see Appendix II) have been operationalized into 10 statements about the current situation within a company. All employees within the organization are invited to give their perception about these statements. Aim of the survey is to make analysis of the current situation possible and generate input for the next phase, in which the findings are translated into a Knowledge Productivity Statement.
Create a knowledge productivity statement (phase 3)

Main aim within the third phase of this method is to generate a *Knowledge Productivity Statement* (KPS). As the aim of this phase is to give direction to knowledge management initiatives, the process of generating a KPS is based on the process of generating an Intellectual Capital Statement, as developed by the Danish Ministry of Science Technology and Innovation (STI, 2003b).

**Intellectual capital statement model**

An Intellectual Capital Statement consists of four elements, which are described in the *New Guideline* (STI, 2002, 2003a, 2003b). The first element is the *knowledge narrative*. ‘A narrative is a plot about a certain phenomenon. It shows the sequence of a set of events, it dramatizes the linkages between these events, and it points out not only the ‘good’ things that characterize the phenomenon but also the crucial ‘bad’ elements that have to be avoided to make the point of the narrative succeed’ (Mouritsen et al., 2002: p.14). The second element are the *management challenges*, which are the challenges that have to be overcome in order to implement the knowledge narrative. The third element are the *initiatives*, which are the actions that can be taken to do something about the management challenges. Finally, the fourth element are the *indicators*, which monitor the progress of initiatives. They make initiatives visible by making them measurable.

**Figure 7 – The intellectual Capital Statement Model**

![Knowledge Narrative](image)

**Source:** STI (2003), Intellectual Capital Statements – The New Guideline

Though the Intellectual Capital Statement Model is designed to create intellectual capital statements, it seems to fit the concept of knowledge productivity and the aim of this research for several reasons. First, this model is designed to translate strategy into knowledge-based action. Intellectual capital statements translate the knowledge narrative into activities that the firm has to put in place to enhance the performance of its knowledge resources (Mouritsen et al., 2001). This aspect supports the aim of this research to give direction to knowledge management initiatives. Second, intellectual capital statements give knowledge an object, which makes it possible to monitor and manage knowledge.
resources. ‘To achieve this, the intellectual capital statement’s knowledge narrative is related to a monitoring system, which identifies the knowledge management activities’ (Mouritsen et al., 2002: p.20). Intellectual capital statements help to make the object of intangibles clear by creating a language for thinking, talking and doing something about the drivers of companies’ future earnings (Roos et al., 1997; Mouritsen et al., 2002). ‘Counting and numbering are means by which knowledge may be drawn forth as an object that has features, attributes and aspects. It is by counting the development of these aspects that knowledge management activities get a form – and a practice. Only when attached to numbers is it possible to identify and communicate, in a reasonable form, what knowledge is all about’ (Mouritsen et al., 2002: p.19). This aspect supports the objective of solving internal and external communication problems, because when translated to indicators the complexity of knowledge is reduced, and made manageable and communicable.

Third, the lack of existing models for monitoring knowledge resources asks for a model that includes the logic of reading. Reading an intellectual capital statement is different from reading a financial statement, because the intellectual capital statement does not have the institutions that make certain readings conventional, as in the case of the financial statement. ‘The logic of reading the indicators can therefore not be ‘outside’ the document but has to be made part of it’(Mouritsen et al., 2001).

Fourth, managing knowledge resources is a process. The step-by-step approach of the model guides the manager through the process of formulating a statement. Participating in this process is at least equally important as the outcome. The aim of intellectual capital statements is to visualize the path towards realizing the knowledge narrative. In other words, they create the infrastructure required to make the knowledge narrative possible (Mouritsen et al., 2002). However, the main benefits of the process do not come from the statement itself, but from the act of preparing the statement (STI, 2003b; Andriessen, 2004a). This aspect supports the assumption that the process of measuring is more than assigning scaled numbers to items. There is more to an intellectual capital statement than the indicators (Mouritsen et al., 2001). However, aim of the indicators is to monitor the implementation of interventions related to the knowledge narrative and take actions accordingly.

Generating an intellectual capital statement means filling in the different elements of the model. Whereas it seems as if the process of creating such a model is very linear, it has been designed as an iterative process. ‘Preparing intellectual capital statements is a creative activity where a meaningful whole is gradually developed. The important thing is not to get locked in one place’ (STI, 2003b: p.19). In order to avoid the work stagnating to a halt, the Danish Guidelines suggest working with the entire model at the same time. ‘The work is completed when you have a coherent explanation. You should be continuously critical of all the elements, and they should be reassessed until everything is coherent’ (STI, 2003b: p.22). As it is often difficult to immediately determine the knowledge narrative, the guidelines suggest starting with the existing activities and underlying objectives, before determining the narrative. However, more important than the sequence of activities is to be open for surprises and new insights. Another important aspect of working with this model is coherence between the different elements. The Danish Guidelines suggest using words like ‘therefore’ and ‘because’ to illustrate coherence. Moreover, at each stage of the process it is important to ask the question ‘why?’, because the answer to this question illustrates the relationship with the other elements.
Knowledge productivity statement

The process of generating a Knowledge Productivity Statement (KPS) follows the process of the Intellectual Capital Statement Model. However, some minor modifications have been made. The main difference is that the KPS does not start from the concept of intellectual capital and its different types of intangibles, but from the concept of the Corporate Curriculum and its seven learning functions. Starting point of the process is the output of the analysis of the current situation (phase 2). Generating a KPS takes place in several workshops with a selection (max. 10) of the respondents to the survey. Developing a KP-statement is a process that consists of four steps:

1. Make an inventory of existing initiatives and underlying challenges. As the Corporate Curriculum serves as starting point, the inventory focuses on initiatives and challenges related to the seven learning functions. In a sense, these elements could be translated into ‘learning challenges’ and ‘learning initiatives’.

2. Develop a knowledge narrative. This narrative or knowledge strategy expresses the company’s ambition to increase the value a user receives from a company’s goods or services and translates the so called use value into knowledge resources. The latter are needed to give direction to the challenges.

3. Reformulate challenges, select initiatives. Based on the ambition (knowledge narrative) and the quality of the preconditions for knowledge productivity (outcome survey), we now can reformulate challenges and initiatives. Which initiatives deserve priority? Which initiatives should be launched? Which can be eliminated?

4. Define indicators. Finally, after the narrative, challenges and initiatives have been completed, indicators are defined to monitor the progress of the initiatives.

Once fully completed, the analysis will be presented in a Knowledge Productivity Statement, analogue to the Intellectual Capital Statement Model above. This statement serves as the main output of the Knowledge Productivity Enhancer.

Based on the experiences with applying the Danish Guidelines, some additional operational requirements should be taken into account when generating a KP-statement. First, a KP-statement is a communication tool and therefore should be easy to read. The complete statement should fit on one page and should be understandable for all employees. Second, the number of challenges and initiatives should be limited to those that are perceived to be of greatest importance to realizing the organizational objectives. The total amount of initiatives in the final statement should be realistic to implement. Therefore, the number of challenges should be limited to 4 or 5. Finally, the indicators should be SMART: specific, measurable, action-oriented, realistic and a measure of time.

Evaluating the Design

Within this research, the design of the method is an iterative process. After each case study the effectiveness of the method will be evaluated. The lessons learned will be used to improve the design, before applying it in the next case study. The evaluation will be based on the experiences of the
researcher, a semi-structured interview with the customer (case study) and a statistical analysis of the outcome of the survey.

The interview is used to develop understanding and give meaning to the outcomes of the case study. Semi-structured interviews are variously described as informal, unstructured, narrative or non-directive and are in essence a conversation between two or three people – the interviewer and the informant(s) (Mador, 2003). Semi structured interviews are a particularly useful method of collecting data where the nature of the research is exploratory and the focus is upon gaining understanding from the perspective of the actors involved.

Main elements within the evaluation are the application domain (class of problems and class of contexts), effectiveness (results) of the method, functional and operational requirements of the method, and the design of the method.

**Testing the KP-Enhancer**

The aim of this research is to design and test a method. Therefore, after the design, the next step was to test the method in various iterations using the developing multiple-case study. This final section of the paper elaborates on the findings within the first two tests.

**First Iteration: Testing the Initial Design**

The KP-enhancer was first tested by IT Solutions BV, a Dutch ICT service provider specialized in Oracle database systems and Java-technology. IT Solutions BV has about 100 employees divided over four business units: Education, Internet Solutions, Consultancy and DBA Solutions. The method was applied between October 2005 and January 2006. As this was the first time the method was applied, the aim within this case study was to test the initial design of the method. The main problem within this company appeared to be an internal control issue. By applying this method, IT Solutions BV expected to get a better understanding of the sources of knowledge productivity and how to improve knowledge productivity. The main effects of applying the KP-enhancer within this case-study were increased awareness among employees about the importance of knowledge productivity, and increased involvement of employees in translating strategy into action. The most appreciated part of the method was the phase in which we generated a KP-Statement. Although the method generated some progress in the thinking about preconditions for knowledge productivity and indicators for measuring knowledge productivity, this progress was perceived as insufficient.

The main difficulties within this first iteration were caused by applying Zegveld’s Quantitative Framework. As a result of the reduction of the analysis to a series of five years, the outcome was seriously disturbed by one deviating accounting figure (2002). This disturbance made the conclusions about the stability or instability questionable, which resulted in rejection of the outcome by the participants of the workshops. Increase of the number of figures would probably have reduced the disturbing effect of 2002. However, due to practical problems we were not able to extend the number of measures. Therefore we decided to ignore the outcome of this analysis and focus on the analysis based on Kessels’ Corporate Curriculum. Another problem was caused by Zegveld’s conceptualization of incremental innovation. As most of the participants interpreted incremental innovation as everyday (minor) improvements of existing practice they had difficulties with connecting this concept to a
shift in generic strategies. Moreover, they did not accept the fact that the analysis did not detect incremental innovations (i.e. minor improvements). Another finding in this case study was that the method lacked coherence. Although the different parts (i.e. survey, quantitative analysis and KP-statement) were suggested to form a coherent method, the relationship between the three was not clear to the participants. On the one hand this lack of coherence was caused by the difficulties with applying the Quantitative Framework. On the other hand, the method seemed to lack connecting elements.

Based on the experiences in the first iteration, the design of the method was changed on three points. First, we improved the reliability of the questionnaire by increasing alpha-scores per learning function through deleting several items in the survey. Second, in order to bridge the gap between Kessels’ learning functions and Zegveld’s Quantitative Framework, we developed a set of additional statements about incremental and radical innovations. Third, in order to increase coherence between the second and the third phase, we developed a board-game to generate collective findings. This paper focuses on the latter improvement.

In order to strengthen coherence it seemed as if the method should generate more ‘tangible’ output through collective analysis of the outcome. Therefore, to discuss the outcome and to collect shared findings that could generate direction to a KP-statement, a board-game was developed. Main questions to be answered in this game are: ‘what result(s) are you aiming at (incremental innovations and/or radical innovations)?’ and ‘which learning function(s) need to be improved in order to realize this result?’ The game is played after the presentation of the outcome of the survey. The game is played in several rounds. The first round each participant is asked to put a ‘result’ card on either incremental or radical innovations. Everyone is asked to elaborate on his/her choice. At the end of this round everyone is asked to reconsider his/her choice. The second round, each participant is asked to put a ‘priority’ card on the learning function which needs to be improved first. Again everyone is asked to elaborate on his/her choice and finally everyone is asked to reconsider their initial choice. The third and fourth rounds are repetitions of the second round. The result of the game is a visualization of the importance of the different results that the company is aiming at (incremental and/or radical innovations) and the kind of improvements which are needed to achieve this result, given the outcome of the analysis of the learning functions. These collective findings serve as a starting point for generating a KP-statement.

**Second Iteration: Testing Coherence**

Midfield Consultants is the second organization where we applied the KP-Enhancer. Midfield has 43 employees (36 consultants and 7 support staff), divided over three locations. The focus of Midfield is on Small and Medium-sized Enterprises (SME’s). Specific SME related topics are company succession & transfer, franchise & commercial co-operation, expropriation & real estate consultancy and project & subsidy management. The method was applied between February 2006 and May 2006. As the improvements after the first iteration were mainly meant to improve coherence, the main aim of the second iteration was to test coherence between the different elements of the method. Midfield’s main motive to apply the KP-Enhancer was that the management expected the method to raise awareness about the importance of two internal projects (age-conscious personnel policy and securing knowledge and networks), give direction to these internal projects and involve people in implementing the activities related to these projects.
Again, like in the first iteration, the process of generating a KP-statement was perceived to be the most valuable element of the method. It was perceived valuable because Midfield finally managed to get their strategic ambitions on paper. Moreover, the fact that it fitted on one page and that it was action-oriented was also highly appreciated. After completion of the method, the outcome was presented at a special meeting to all employees. As a result many people committed themselves to one or more projects. In this sense the outcome appeared to be very useful and in line with the expectations. To a lesser extent, the method had also been applied to improve corporate reputation and the ability to attract new employees. Asked after the effects related to these issues, it appeared that the KP-statement had successfully been used in interviews with potential employees in order to introduce the company and the challenges they were facing. In this sense it seems as if the KP-statement also contributed to improve external communication.

Main finding of the first iteration was that the method lacked coherence. The efforts to integrate the various elements within the method after the first iteration, seems to have worked out very well within this case-study. The process of generating a KP-statement was perceived as a very ‘natural next step’ after the analysis of the current situation. The collective findings – generated during the KP-board game in the first workshop – seem to have bridged the gap between analysis and action. The game played an important role in generating these findings and shift gear from passive to active participation. However, the KP board game also generated friction and resistance in the process, which almost resulted in the termination of the method. The introduction of the method and the presentation of the outcome of the analysis during the first workshop confirmed the initial expectation with some of the participants that the method was too ‘scientific’. According to some of the participants, the time that was invested in reflection and in generating collective findings lacked practical relevance. These hesitations about the effectiveness of the method eventually led to an outburst of anger by one of the participants. It seems as if the game at the end of the first workshop served as the trigger. According to this participant, the workshop was ‘childish, repeating things we had already done before and a loss of time’. At first this outburst resulted in an awkward moment which seemed to endanger the continuity of the process. However, the discussion that followed also resulted into increased awareness of the problem they were trying to solve (lack of ability to reflection). In this sense the outburst of anger had been very productive and contributed positively to the final outcome.

Although the analysis of the current situation should have been based on both the survey and the quantitative analysis, Zegveld’s Quantitative Framework was not used in the process of analysing the current situation within this case study. The main reason for this was the fear that presenting these ‘academic’ calculations (after the crisis) would further endanger the continuity of the process. Though the assumption was that insight in these measurements was necessary to analyse the current situation, the method seemed to work very well without these measurements. It seems as if the quantitative analysis is not a necessary element within the method. Moreover, the outcome of the survey seems to generate sufficient input for making a KP-statement. Therefore, combined with the experiences in the first iteration, we should consider leaving out this element in the next iteration.

The KP-statement was appreciated for its communicative power. The fact that it made the implicit strategy explicit, that it fitted on one page and that it gave an overview of actions to be taken, made it very useful within this context. ‘It does not present anything new, but it brings existing elements together in a convenient arrangement’. However, in order to stress the relationship with organizational ambitions, an additional column was added including the mission statement, strategic objectives and the collective findings of the analysis of the current situation.
After initial hesitations about the effectiveness of the method and a crisis early in the process, the method seems to have been very effective in the sense that it lead to the main intended results *(raising awareness, give direction, and involve people)*. The survey was received very positively and the KP-statement, as tangible result of the process, was highly appreciated. Important finding of this iteration is also that the method seems to work very well without the Quantitative Framework.

**Further Research**

The objective of this research is to acquire knowledge about how to measure knowledge productivity in order to give direction to knowledge management initiatives. The main focus of this research until now has been on the theoretical design of a method and testing the initial design in two iterations. Within the contexts of the first two case-studies, this method seems to have generated satisfying results. Therefore, the main focus within the next iterations will shift from the development of the method to the implementation of the method. Main objective within these iterations will be to acquire knowledge about implementing the KP-Enhancer. Consequently the main question to be answered will be how to implement the KP-Enhancer? In order to answer this question, special attention will be paid to the next issues:

1. The method seems to contribute to collective sense making. What happens if people within an organization go through a process like this? (Weick, 1995) How to contribute to the construction of meaning?

2. In line with earlier research (Van Lakerveld, 2005), the first two iterations confirm that Reflection is a problem within many organizations. This method forces organizations to practice reflection. How to reflect, when reflection is the problem?

3. As the KP-Enhancer is an intervention method, then what are the specific requirements with respect to the interventionist?

From a methodological point of view, the third question also raises the issues of objectivity and bias. As this research is based on participative observation, the central methodological problem (or dilemma) is balancing adequate subjectivity with adequate objectivity (Bruyn, 1966). How to maintain enough distance to be able to locate the contextual experiences in a wider theoretical and social context? Related to the issue of objectivity, and perhaps helpful in answering this methodological question is the issue of plausible rival explanations (Campbell & Stanley, 1963). What are alternative conclusions that can be taken from the same data and which of the alternative explanations seems to be the most plausible?

**Acknowledgement**

This PhD-research project is a joint initiative of the Center of Intellectual Capital Research of INHOLLAND University and de Baak – Management Center of the Dutch Federation of Industries (VNO-NCW).
Appendix I – Detecting incremental and radical innovation

Detecting incremental innovation

**Calculating generic strategies**
- Volume (V) = turnover
- Efficiency (E) = turnover/employment costs
- Added Value (A) = added value/turnover

(Added value is defined as employment costs, plus depreciation, plus tax, plus interest and other financial costs, and net profit.)

**Calculating Operational Cash Flow (OCF)**
- Performance = A-(1÷E)
- OCF = V*(A-1/E)
- Change in OCF = OCF^t1 – OCF^t0

**Calculating Cash Impact (ci)**
- ciV = (V^t1*(A^t1-1/E^t1))-(V^t0*(A^t0-1/E^t0))
- ciA = (V^t0*((A^t1)-1/E^t1))-(V^t0*((A^t0)-1/E^t0))
- ciE = (V^t0*(A^t0-1/(E^t1)))-(V^t0*(A^t0-1/(E^t0)))

Detecting radical innovation

**Data needed**
- Output = added value (O)
- Capital = depreciation (C)
- Number of employees (e)

**Calculate labour and capital productivity**
- Output/employee (Oe)
- Capital/employee (Ce)

**Calculate residual change**
- Change of output/employee (dOe = Oe^t1 – Oe^t0)
- Change of capital/employee (dCe = Ce^t1 – Ce^t0)
- Capital/Output (C/O)
- Residual change (dR=1+(dOe-C/O*dCe))

Source: (Zegveld, 2000).
Appendix II: Operationalization of the Corporate Curriculum

Learning functions:

1. **Subject matter expertise**
   - Knowledge grounded in strategy
   - Effective knowledge processes (develop, share/codify)

2. **Problem solving**
   - Effective knowledge processes (apply)
   - Ability to renew and stretch (creativity)

3. **Ability to reflect**
   - Effective knowledge processes (evaluate)
   - Reflective skills

4. **Communication skills**
   - Competences of the knowledge worker
   - Culture of knowledge sharing
   - Structure for knowledge sharing

5. **Self regulation of motivation**
   - Space
   - Personal entrepreneurship
   - Management support

6. **Peace and stability**
   - Room for specialization
   - Time for reflection and sharing
   - Organizational redundancy

7. **Creative turmoil**
   - Strategic ambiguity
Human capital, knowledge and their impact on measuring productivity

II

Literature


Human capital, knowledge and their impact on measuring productivity


Wagenaar (Eds.), *Kennis maken, leren in gezelschap*. Schiedam: Scriptum.


**Key words:**

MEASURING AND DRIVING PRODUCTIVITY – A KEY TO SUCCESS FOR INNOVATIVE COMPANIES

Monika KIRCHER-KOHL
CFO, Infineon Technologies Austria AG

Abstract

The measurement of productivity is one enabler in the management of successful companies:

1. Measuring: The first step is to identify and then measure the ‘right’ indicator.
2. Comparing: By comparing the data with internal and external benchmarks, additional potential can be identified.
3. Steering/Controlling: In order to actively manage productivity, the right levers and key success factors need to be identified as well as initiatives derived.

In this contribution, we will give an overview of the productivity indicators currently used by our company and discuss not only their advantages, but also their limitations. While the productivity measurement of production processes is well established and meaningful indicators are in place, the measurement of innovation and R&D productivity is more complex. The indicators currently used for these processes include purely financial indicators (cost/engineer) as well as technical indicators such as cycle time and execution quality, but a single indicator combining those two factors and thus measuring innovation efficiency is still missing.

In a knowledge based society, human capital is becoming the most important production factor, especially from a European point of view. Statistics and traditional financial analysis such as income and profit & loss statements are retrospective tools. For innovative companies such as Infineon, it is increasingly important to look forward and to find methods to measure and assess potentials and future opportunities. To address this, Infineon Austria will be publishing its first intellectual capital statement this year.

Regarding the Lisbon goals, the challenge is to ‘measure to steer/control’: We must find the appropriate indicators (Example education: Is it sufficient to only measure the average length of education? How can we measure what students are learning?) in order to pull the right levers and to enable the successful implementation of initiatives to make the European Union the most competitive and dynamic knowledge-driven economy by 2010.

There will always be industry and company specific indicators for measuring productivity, but companies will increasingly employ official statistical data (e.g. regarding (continuous) education and training) as input factors for their internal analysis.
Summing Up
1. Productivity is quite an important measure, not only for national purposes but also for international comparisons. But comparisons show that there are often obvious differences, and there still seems to be some doubt as to whether these reflect actual developments or whether they are caused by different methodological approaches.

2. The seminar has covered various aspects concerning productivity measurement in a number of countries and supra- or international institutions.

The first session was concerned with the more general approach to productivity measurement. A model for measuring overall productivity and its implications was presented. It was shown that productivity can traditionally be split into different parts, such as labour and capital productivity, but also that other possible input factors can be placed into relation to output. For capital productivity, one important issue is to calculate real figures by using depreciation of the capital stock. Another presentation developed the fact that depreciation – closely connected with service life – differs greatly across European countries. The last presentation referred to productivity measurement in National Accounts. It became clear that approaches for the purposes of SNA and ESA are still very different. The pros and cons of the various components were discussed, with special emphasis on the real implications.

In general the first session showed that:

- there are substantial differences in calculating the various shares of productivity components;
- a distinction should be made between demand and supply side approach. Especially the two regularly used output measures should be treated differently. There was a proposal to measure gross domestic product (GDP) in market prices and gross value added in basic prices;
- capital goods should be treated according to the number of shifts in the production processes;
- depreciation methods and service lives should be harmonised as far as possible in the European Union (EU).
3. The impact of hedonic price measurement on productivity was another major focus of the seminar. When output figures include products with rapid quality changes over short periods, the problem is that price differences do not just reflect the real changes. In such cases, price changes are too low compared with real changes without respect to quality improvements. There are various ways of calculating real price changes more accurately. One quite well developed method is ‘hedonic price measurement’, a sophisticated approach to eliminating quality changes. It is often used to calculate price developments of computers. In a few countries like the USA or Germany, this approach is used for some important products in terms of their share of overall value and the rapidity of quality changes. These approaches need quite a high level of expertise on product changes over time and are quite costly. To sum up, the following proposals were made:

– to implement hedonics methods for a limited number of products at least EU-wide;

– to ensure comparability of hedonic price measurement;

– to create a central data base for updating SNA deflating indicators.

4. In most countries the non-business sector is a significant element in National Accounts. The health and education sectors, for instance, contribute considerably to the output/outcome of an economy. In National Accounts the focus is on output, but the public health and education sectors are concerned with outcome. The main problems are that there are no market-related output data available or even calculable, coupled with a lack of market prices for input factors, especially labour. It was stressed that:

– a clear distinction has to be made between output and outcome;

– there is an urgent need for approaches and methods for calculating output and input volumes;

– a contribution should be made to the ongoing discussion in inter- and supranational institutions/organisations about improved coverage of such areas in official statistics.

5. Knowledge has become an important input factor in terms of economics. It is crucial for developed countries as well as for the European Community. Growth, employment and competitiveness depend enormously on the active use of knowledge and information in production processes. Knowledge can feature in a number of ways. Of special interest is, on the one hand, the amount of investment in or expenditure on research and development or innovation. On the other hand, we have what is known as ‘human capital’, which is concerned with the level of knowledge of employees and managerial staff, and the up-to-dateness of their skills and operational experience. The seminar showed how improvements in skills can be achieved in learning processes at the micro-level and how they affect productivity. Another presentation showed how learning processes can be adjusted in official statistics. The third presentation dealt with real problems of internationally active firms. These global players are eager to keep up their level of innovation and to develop their staff’s potential. The major improvements in this area might, arising from the ensuing discussion, be:
– to introduce a satellite system for knowledge in the National Accounts;

– to cover statistically the efforts of firms and individuals to improve staff’s level of knowledge;

– to develop appropriate deflators for observing and/or calculating real developments;

– to develop approaches to the depreciation of knowledge.

At the end of the seminar, discussion focused on whether the question mark in the seminar title ‘Are we measuring productivity correctly?’ could be eliminated. The answer was no. In general the conclusions were that

– volume measures are needed not only for products but also for services, and especially for the non-market sector;

– partial productivity indicators are important. But with regard to total factor productivity, further attention and investigation has to be directed to their interdependencies, how they interact and how they should be interpreted with respect to users’ needs;

– the quality of productivity calculations (e.g. the components) is very important. Hedonic methods should be applied to a limited number of products in all EU countries. Operational definitions are necessary to improve coverage of the public sector and of specific services;

– knowledge in terms of R&D and human capital is important for productivity measurement.

Even though it is quite clear that not all problems can be solved immediately, the measurement of productivity can only be improved if ongoing work is extended and continued as quickly as possible. It is particularly important that the methods and all data used for calculating productivity should be harmonised and made comparable, at least for the Member States of the European Union.
Reaction from Eurostat
I would like to thank CEIES and all participants for what I think was a most useful and very stimulating seminar. We addressed all key issues related to productivity. They concern concepts and data, but also strategic orientations for National Statistical Institutes, possibilities of sharing experience and the need for good communication.

For all of us, there is no doubt about the importance of productivity for economic and monetary policy. That is why we statisticians must make every effort to measure it correctly. Our meeting has brought together experts in the field as well as experienced users and data providers. The Chair of the Sub-Committee has presented an excellent summary of the sessions and the main conclusions.

In the following, I would like to address the most important points.

Productivity measures relate the volume of outputs to the volume of inputs. Hence, the correct measurement of volumes is at the heart of the productivity concept. In the case of market activities, this leads immediately to the discussion on the most appropriate deflation methods. In the case of non-market production, the issue is the direct measurement of output volumes.

As regards deflation, participants made the point that the national accounts manuals SNA93 and ESA95 are not satisfactory. I think that this criticism is justified. While major progress was made with the Eurostat Handbook on price and volume measures in national accounts, more guidance will be necessary. In particular, when revising the SNA and the ESA in the coming years attention should be paid to the need for precise orientations.

The development of direct volume measures for the non-market sector has been at the centre of national accountants’ improvement efforts over the last years. The focus is on education and health. In the European Union, comprehensive work is ongoing. It is based on several legal acts which require Member States to implement better measures. A lot has already been achieved, but the challenge remains big. In particular, how best to take quality changes into account is top of the agenda. Eurostat intends to organise a workshop next year which should pave the way for further progress in the areas of education and health.

The quality issue, however, concerns the measurement of output volumes in the case of market production, too. In this context, the use of hedonic methods has been strongly promoted. There are already a number of Member States that apply hedonics for selected goods. I think our discussions made clear that hedonics are an important tool for quality adjustment, but the method does not provide a miracle solution to the quality problem. It needs to be used intelligently, can be resource intensive
and requires in-depth experience. Sharing of experience between NSIs is indispensable in order to learn from best practice and to ensure a harmonised approach across countries. A co-operation project on hedonics has been launched in the area of HICP. National accountants will certainly benefit from that.

The importance of a correct measurement of the volume of inputs, in particular labour input, for productivity is often underestimated. Eurostat’s experience is that uncertainties related to hours worked (the preferred measure for labour input) have a major impact on the reliability of productivity figures. In the last two year, substantial progress has been recorded due to a joint Eurostat/OECD initiative together with Member States. What still needs to be tackled is to capture the changing composition (quality) of the labour input.

The role of research and development and of human capital was underlined in one of our sessions. As regards R&D, I would expect significant improvements to come with the new SNA and ESA. The intention is to treat the output of R&D as capital formation. This, together with the introduction of the concept of capital services, will help to grasp the influence of research and development on productivity. By contrast, the concept of human capital will not be part of the innovations in the revised national accounts manuals. There are still too many problems to be solved. It could even be argued that a sufficiently reliable measurement cannot be expected in principle.

A special issue related to the measurement of capital input for industries/sectors was highlighted in our meeting. Estimates of capital stock and consumption of fixed capital are normally derived via the perpetual inventory method (PIM). The PIM does not automatically take account of acquisitions and disposals of existing fixed assets. Given that today we increasingly observe such movements between industries/sectors (due to privatisation, for example) to what extent existing PIM estimates by industry/sector must be adjusted to correctly reflect ownership of the assets needs to be examined.

In recent years, the use of multi-factor productivity measures (in contrast to single-factor productivity such as labour productivity) has become more widespread. Multi-factor productivity (MFP) is much more model-based than single factor productivity. It is a concept that is often preferred by researchers, only a few statistical institutes publish estimates at present. One reason is certainly that the approach to measuring capital services (the appropriate rate of return, in particular) is not yet sufficiently harmonised. The next SNA/ESA explicitly addresses capital services. This will lead to a convergence of measurement practices. Eurostat does not produce estimates of multi-factor productivity, while DG ECFIN has been doing so for quite some time. The first priority for Eurostat is to improve the official estimates of single-factor productivity. As soon as an international consensus on MFP is in sight, we will reconsider the publication of official Eurostat estimates.

Before I come to an end I want draw your attention to the Commission’s EU KLEMS project. We have heard it being mentioned several times during the seminar. The EU KLEMS project on growth and productivity accounts brings together NSIs and the research Community with the aim to set up a comprehensive productivity data base for industries. The first release of the data base is scheduled for March 2007. Eurostat intends to disseminate results on its website.

My impression is that during this seminar we have discussed a wide range of important issues on the measurement of productivity. We have been presented the state of the art, we have learned about
achievements, but we also identified quite a number of points that need to be improved in order to be able to give a full-hearted yes-answer to the title question of the seminar. There’s work ahead for statisticians. Let’s tackle it.
List of Participants
## LIST OF PARTICIPANTS

<table>
<thead>
<tr>
<th>Name</th>
<th>City, Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHDE Pasi</td>
<td>Helsinki, Finland</td>
</tr>
<tr>
<td>AHNERT Henning</td>
<td>Frankfurt, Germany</td>
</tr>
<tr>
<td>ARDI Carolina</td>
<td>Rome, Italy</td>
</tr>
<tr>
<td>BALDASSARINI Antonella</td>
<td>Rome, Italy</td>
</tr>
<tr>
<td>BATTISTI Giuliana</td>
<td>Birmingham, UK</td>
</tr>
<tr>
<td>BERNARDINI PAPALIA Rossella</td>
<td>Bologna, Italy</td>
</tr>
<tr>
<td>BLAIR Sheena</td>
<td>Luxembourg</td>
</tr>
<tr>
<td>BREGAR Lea</td>
<td>Ljubljana, Slovenia</td>
</tr>
<tr>
<td>CAMILLERI Reno</td>
<td>Valletta, Malta</td>
</tr>
<tr>
<td>CHAILOS Vassilis</td>
<td>Nicosia, Cyprus</td>
</tr>
<tr>
<td>CICCONE Julien</td>
<td>Luxembourg</td>
</tr>
<tr>
<td>COSTANTINO Cesare</td>
<td>Rome, Italy</td>
</tr>
<tr>
<td>DANUSĖVIČS Henriks</td>
<td>Riga, Latvia</td>
</tr>
<tr>
<td>DE HAAN Jan</td>
<td>Voorburg, Netherlands</td>
</tr>
<tr>
<td>DEMIR Ömer</td>
<td>Ankara, Turkey</td>
</tr>
<tr>
<td>DI MARIA Charles-Henri</td>
<td>Luxembourg</td>
</tr>
<tr>
<td>DI VEROLI Nadia</td>
<td>Rome, Italy</td>
</tr>
<tr>
<td>FENWICK David</td>
<td>London, UK</td>
</tr>
<tr>
<td>FERDMAN Clara</td>
<td>Örebro, Sweden</td>
</tr>
<tr>
<td>FØRSUND Finn</td>
<td>Oslo, Norway</td>
</tr>
<tr>
<td>FREIRE Idílio</td>
<td>Lisbon, Portugal</td>
</tr>
<tr>
<td>GALLAIS Alain</td>
<td>Paris, France</td>
</tr>
<tr>
<td>GÖRZIG Bernd</td>
<td>Berlin, Germany</td>
</tr>
<tr>
<td>GRAF PUCKLER Botho</td>
<td>Cologne, Germany</td>
</tr>
<tr>
<td>HAGSTEN Eva</td>
<td>Stockholm, Sweden</td>
</tr>
<tr>
<td>HIRVONEN Timo</td>
<td>Helsinki, Finland</td>
</tr>
<tr>
<td>IOMMI Massimiliano</td>
<td>Rome, Italy</td>
</tr>
<tr>
<td>IONA Alfonsina</td>
<td>Birmingham, UK</td>
</tr>
<tr>
<td>Name</td>
<td>Affiliation</td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>JONA-LASINIO Cecilia</td>
<td>Rome, Italy</td>
</tr>
<tr>
<td>KAFAS Antonis</td>
<td>Nicosia, Cyprus</td>
</tr>
<tr>
<td>KALNINA Elita</td>
<td>Riga, Latvia</td>
</tr>
<tr>
<td>KAZEMIER Brigt</td>
<td>Voorburg, The Netherlands</td>
</tr>
<tr>
<td>KIRCHER-KOHL Monika</td>
<td>Villach, Austria</td>
</tr>
<tr>
<td>KODRIVČ Borut</td>
<td>Koper, Slovenia</td>
</tr>
<tr>
<td>KOEV Eugen</td>
<td>Helsinki, Finland</td>
</tr>
<tr>
<td>KOTULICS Tamas</td>
<td>Budapest, Hungary</td>
</tr>
<tr>
<td>KRASNIQI Mimoza</td>
<td>Pristina, Kosovo</td>
</tr>
<tr>
<td>KUHRY Bob</td>
<td>The Hague, The Netherlands</td>
</tr>
<tr>
<td>KULS Reinhard</td>
<td>Frankfurt, Germany</td>
</tr>
<tr>
<td>LAMEL Joachim</td>
<td>Klosterneuburg, Austria</td>
</tr>
<tr>
<td>LOUKAIDOU Marinala</td>
<td>Nicosia, Cyprus</td>
</tr>
<tr>
<td>MACLEAN Ian</td>
<td>Esher, England</td>
</tr>
<tr>
<td>MALHERBE Francis</td>
<td>Luxembourg</td>
</tr>
<tr>
<td>MANTEGAZZA Suzanna</td>
<td>Rome, Italy</td>
</tr>
<tr>
<td>NICOLAOU Maria</td>
<td>Nicosia, Cyprus</td>
</tr>
<tr>
<td>NIMANI Artan</td>
<td>Pristina, Kosovo</td>
</tr>
<tr>
<td>ONDRUS Vitezslav</td>
<td>Prague, Czech Republic</td>
</tr>
<tr>
<td>ORIGER Marc</td>
<td>Luxembourg</td>
</tr>
<tr>
<td>OWCZARZAK Radoslaw</td>
<td>Dublin, Ireland</td>
</tr>
<tr>
<td>PAST Harald</td>
<td>Vienna, Austria</td>
</tr>
<tr>
<td>PICOZZI Luisa</td>
<td>Rome, Italy</td>
</tr>
<tr>
<td>QELA Jeton</td>
<td>Pristina, Kosovo</td>
</tr>
<tr>
<td>RAMM-SCHMIDT Ralf Eiler</td>
<td>Helsinki, Finland</td>
</tr>
<tr>
<td>RECKTENWALD Joachim</td>
<td>Luxembourg</td>
</tr>
<tr>
<td>RUANE Frances</td>
<td>Dublin, Ireland</td>
</tr>
<tr>
<td>RUKSENAITE Jurga</td>
<td>Vilnius, Lithuania</td>
</tr>
<tr>
<td>SARAZIN Ines</td>
<td>Ljubljana, Slovenia</td>
</tr>
<tr>
<td>SCHEIDHAUER Marie-Paule</td>
<td>Luxembourg</td>
</tr>
<tr>
<td>Name</td>
<td>City</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>SCHNORR-BÄCKER Susanne</td>
<td>Wiesbaden</td>
</tr>
<tr>
<td>STAM Christiaan</td>
<td>Diemen</td>
</tr>
<tr>
<td>SZUKICS Gábor</td>
<td>Budapest</td>
</tr>
<tr>
<td>THANASSOULIS Emmanuel</td>
<td>Birmingham</td>
</tr>
<tr>
<td>TOPRAK Ömer</td>
<td>Ankara</td>
</tr>
<tr>
<td>TRONTI Leonello</td>
<td>Rome</td>
</tr>
<tr>
<td>TUDINI Angelica</td>
<td>Rome</td>
</tr>
<tr>
<td>VAN DER WIEL Henry</td>
<td>The Hague</td>
</tr>
<tr>
<td>VAN WOUWE Martine</td>
<td>Antwerp</td>
</tr>
<tr>
<td>VANAGA Alla</td>
<td>Riga</td>
</tr>
<tr>
<td>VERGARA Francisco</td>
<td>Paris</td>
</tr>
<tr>
<td>WÄCHTER Gerhard</td>
<td>Luxembourg</td>
</tr>
<tr>
<td>WALLIS Gavin</td>
<td>London</td>
</tr>
<tr>
<td>WAGNER Karin</td>
<td>Vienna</td>
</tr>
<tr>
<td>WASSHAUSEN David</td>
<td>Washington</td>
</tr>
<tr>
<td>WYZNIKIEWICZ Bohdan</td>
<td>Warsaw</td>
</tr>
</tbody>
</table>