Quality aspects in price indices and international comparisons: Applications of the hedonic method

Proefschrift

ter verkrijging van het doctoraat in de
Economische Wetenschappen
aan de Rijksuniversiteit Groningen

Peter Hein van Mulligen

RIJKSUNIVERSITEIT GRONINGEN

QUALITY ASPECTS IN PRICE INDICES AND INTERNATIONAL COMPARISONS: APPLICATIONS OF THE HEDONIC METHOD

Proefschrift

ter verkrijging van het doctoraat in de
Economische Wetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. F. Zwarts,
in het openbaar te verdedigen op
donderdag 25 september 2003
om 14.15 uur

door

Peter Hein van Mulligen
geboren op 3 april 1974
te Vlagtwedde

Promotor	Prof.dr. H.H. van Ark
Co-promotor	Dr. M.P. Timmer
Beoordelingscommissie	Prof.dr. E.J. Bartelsman
	Prof.dr. M. Silver
	Prof.dr. E. Sterken

Publisher
Statistics Netherlands
Prinses Beatrixlaan 428
2273 XZ Voorburg

Printed by
Statistics Netherlands – Facility Services

Cover design
WAT ontwerpers, Utrecht

Information
E-mail: infoservice@cbs.nl

Where to order
E-mail: verkoop@cbs.nl

Internet
http://www.cbs.nl

The author is fully responsible for the text of this report; the text does not necessarily correspond with the official point of view of Statistics Netherlands.

Key figure: X11
ISBN: 903572698 7
Product code: 6008203002

Acknowledgements

The roots of this dissertation lie in the work I carried out for my M.A. thesis in 1997. My purpose at that time was to construct quality-adjusted purchasing power parities for the construction sector for several European countries, among which the Netherlands. The use of the so-called 'hedonic method' was an explicit goal of the research, as it seemed the only justifiable method to use in such a heterogeneous sector of the economy. Unfortunately, adequate data on the construction sector were not available, reducing the empirical nature of the study.

Not deterred by the data problems I encountered, I took up a more ambitious project in the form of a Ph.D.-research under the same supervisor, Bart van Ark. As my M.A. study, this research was embedded in the ICOP group at the Faculty of Economics of the University of Groningen, where the comparison of productivity between nations is a key topic. The ICOP research, like any kind of research that involves comparing prices of different products, was sometimes criticised because it did not adequately take into account quality differences. Given its place in ICOP, the initial approach of the research would be the industry-of-origin approach rather than the expenditure approach.

Rare are those theses where the final output perfectly corresponds with the initial research proposal, and this one is no exception. The industry-of-origin approach proved to be too narrow to explore the quality issue, and consumer prices could not be neglected. Therefore, the focus shifted to the quality issue in general, and also price index numbers were covered. This is were the second location where I carried out my research comes in: Statistics Netherlands, where I worked on my research from September 2001 to June 2002. Statistics Netherlands co-financed my research, and made possible the publication of this thesis.

Both at the University of Groningen and Statistics Netherlands, many people provided me with helpful comments, advice and assistance. First of all, I want to thank Bart van Ark and Marcel Timmer, my supervisors in Groningen. Bart's broad knowledge on everything related with economic measurement, his focused comments and his pool of ideas were essential for the writing of this thesis. Marcel entered the project at a later stage, but his role in the completion of this thesis was not less vital. At Statistics Netherlands, I owe much thanks to Jan de Haan, from whose expertise on price index numbers in general and hedonics in particular I benefited enormously.

In addition to these people, many others contributed in important ways to this study. In particular, I would like to thank Heymerik van der Grient, Robert Inklaar, Stefan Linz, May Hua Oei, Eddy Opperdoes, Peter Ruchti, Jack Triplett, Kees Zeelenberg and various discussants and participants in the conferences and seminars where I presented parts of my work. I am grateful to Eric Bartelsman, Mick Silver and Elmer Sterken for finding the time to com-

ment on the draft of this thesis. Moreover, I would like to thank Statistics Netherlands for providing me with the opportunity to finish my thesis there.
Finally, I wish to express my enormous gratitude to Helen Visser, without whom this thesis would have been strictly impossible. Helen provided me with her support in many different ways, and even put up with me having to spend several days from home each week for three and a half years. Therefore, I wish to dedicate this thesis to her.

Peter Hein van Mulligen
Rotterdam, July 2003

Contents

List of tables in the text

List of figures in the text

1. Introduction

The revival of the debate on price index measurement

Since the early days of the measurement of economic activity, national accountants have been struggling with the problem of observing and integrating continuous changes in the real world in their measurement framework. One of the areas of continuous methodological debate concerns changes in product samples, improvements in the quality of new products and the decomposition of value change of these products into a price and a quantity component.

In 1996, this debate received a new stimulus from the publication of the findings in an advisory report to the Senate Finance Committee of the 'Boskin-committee' with respect to the Consumer Price Index (CPI) in the United States (Boskin *et al*., 1996). The main conclusion of this report was that the U.S. CPI overestimated true inflation by 1.1 percentage points per annum on average, with a plausible range of between 0.8 – 1.6 percentage points.[1] Upward biases due to the late pickup of new products in the sample at a time when price declines from the early introduction onwards had already occurred, and the failure to account for real price declines due to quality improvements accounted for the lion's share of the overestimation of the inflation rates, namely for 0.6 percentage points.

The Boskin report has led to a major debate in the United States about the use of the CPI as a measure of the cost-of-living index (COLI), which among other things is used for indexing social security payments. Although the estimates of the Boskin-commission have been criticised on methodological grounds, most agreed that the measurement of the CPI needed to be improved.[2] It is also clear that the problem does not only concern the consumer price index, and it is not restricted to the United States only. For example, a study on Germany concluded that the inflation rate in that country had an upward bias of about 0.75 percentage points per annum, of which about 0.5 percentage points was due to unmeasured quality improvements (Hoffmann, 1998). Although he provides no numerical estimates, Oulton (1998) states that the error in the U.K. retail price index is of the same general order. Researchers in other countries also concluded that their CPIs were probably overestimated, although precise estimates are hard to generate.[3]

In fact the debate brought back to the foreground a long-standing issue on how to account for quality differences in comparisons of prices, income and output. The quality issue is relevant whenever prices are compared of products that are not identical. This is true for both price indices, which compare prices over time, and international price comparisons such as purchasing power parities, although the bulk of research has concentrated on the former. In this thesis I will focus on the following key issues: first, I will study the impact of differences in quality on intertemporal and international price comparisons. Second, in this thesis I will compare two methods of price measurement,

namely the conventional matched model method and the hedonic method. These methods will be assessed both theoretically and empirically on their suitability to deal with the 'quality issue'.
Below follows a brief introduction on the topics of the quality issue and the hedonic method, after which the structure of the thesis is set out.

The quality issue

Why is the measurement of price indices so important? The most obvious and important reason is that the CPI in most countries counts as *the* measure of inflation. Inflation estimates are among the most important economic statistics, and governments and economic interest groups (such as unions and employers' federations) use it for wage bargaining, and to adjust all kinds of subsidies, welfare schemes and so on. Another important function of price indices concerns their use as deflators for nominal income and output. [4)] They are applied to derive growth in real income, and to deflate output series, which in turn are used to calculate changes in real output and productivity. The latter application is the focus of this thesis. If price indices are overestimated or underestimated, real growth figures have a bias of the same magnitude, but of opposite sign. Gordon (1990) provides an extensive list of issues to which price measurement is related: the accuracy of historical economic growth rates; the measurement of growth in total factor productivity; the measurement of capital/output and investment/output ratios; the need for relative prices for studies of the demand for durable goods; assessment of the terms of trade of countries; measurement of the user cost of capital equipment; the measurement of the payoff to research and development; and accurate measures of the relation of transaction to list prices, to properly measure fluctuations in real demand and output.
Mismeasurement of prices may not only lead to biases in comparison over time but also across space. For example, the use of different methodologies across countries can easily lead to misinterpretations of the comparative growth performance. In a much-quoted article, Wyckoff (1995) analysed data on price indices for the computer manufacturing industry for twelve OECD countries and compared these indices over the period 1968 to 1992. There appeared to be a huge spread between the indices of the countries, amounting to a range of more than 150 percentage points for the 1980s only. A similar study by Eurostat (1999a) for the European Union for the 1990s found a smaller but still substantial spread. At an aggregate level, some of these differences may of course be due to differences in the composition of output, but at the most detailed level of narrowly defined industries or products, output should be relatively homogeneous.

Wyckoff distinguished several possible explanations for the spread in price indices he found, and concluded that there was only one plausible explanation for the magnitude of the problem. This concerns the difference in methodologies used to calculate the computer price index. It appeared that the countries with the fastest decline in prices of computer equipment (notably the United States) used a methodology that was different from the standard methods that are employed by the countries which showed a much slower price decrease or even an increase in computer prices. The standard method, which is usually referred to as the 'matched model' method, only matches prices of identical products that are observed in both periods of the price index. Quality changes and the occurrence of entering or exiting products are ignored. The alternative method, referred to as the 'hedonic' method, makes use of information on the characteristics of products for which it estimates implicit prices. Since the 1980s this method has been applied for the U.S. computer industry and other high-tech goods in the United States. In particular for computers the price declines based on the hedonic method are much faster than for the traditional matched model method.

The extensive use of hedonic price measurement in the United States against standard matched model methods in most European countries, has led to a controversy on the interpretation of the comparative growth performance of the German economy vis-à-vis the United States. In a monthly report of the Deutsche Bundesbank (2000), the German central bank, a claim was made that a large part of the difference in GDP growth between the United States and Germany can be explained by the fact that the statistical agencies in the U.S. make use of the hedonic method to adjust for rapid quality changes in information and communication technology (ICT) equipment, whereas such an adjustment is not made by German agencies. Another report of the Bundesbank claims that in case German national accountants – in a similar way as in the U.S. – would have applied hedonic deflators for IT goods, real GDP growth in Germany would have been only 0.2 percentage points lower than in the U.S. for the period 1996 to 1999 (Deutsche Bundesbank, 2001).

But it is unlikely that the overall gap in German and U.S. GDP growth is for a large part caused by different methodologies to deflate ICT output, given the relatively small size of the ICT producing industry relative to total GDP. Landefeld and Grimm (2000) estimate that the increasing quality change of personal computers only adds, at most, about 0.25 percentage point to annual GDP growth in Germany for the period 1995–99. Another country that uses the hedonic method to deflate ICT prices is France. Lequiller (2001) puts an even smaller estimate on the 'growth effect' of quality change: only 0.04 percentage point of annual GDP growth for the period 1995–98 can be explained by a downward adjustment of the deflators for computers and software.[5] Quality-adjusted price indices therefore have some impact on overall GDP growth, but they cannot explain a large part of growth differentials between countries. Although the impact on a total economy level is modest, at the detailed level of the ICT industry itself, such methodological differences may make a substantial difference (van Ark *et al.*, 2002).

The quality problem is not only important for international comparisons of the change in prices, but also for comparisons of price levels across countries. Currency conversion factors that adjust for differences in relative price levels across countries, known as purchasing power parities (PPPs) or unit value ratios (UVRs), are used to compare relative output, per capita income and productivity levels in different countries. The procedure to calculate conversion factors is essentially similar to that for price indices, i.e. prices of identical or similar goods are compared, and the aggregate of these ratios provides the currency conversion factor to obtain international comparisons of real output and productivity. As for price changes, this procedure critically assumes that the products of which the prices are compared are identical. If this is not the case, a quality bias is introduced, just as is the case with price indices. This bias will cause the real output or expenditure of one country relative to another country to be overestimated, and therefore lead to overstating its relative per capita income or productivity level. The issue has been of major concern in the International Comparisons of Prices (ICP) programme, which is presently carried out by Eurostat, OECD and the United Nations to provide PPP measures for GDP and subcategories of GDP at expenditure level (such as consumption, investment and government expenditure). [6)] The quality problem has also been a key methodological issue in international comparisons of productivity levels, such as those in the International Comparisons of Output and Productivity (ICOP) programme at the University of Groningen. [7)] In both ICP and ICOP most price comparisons are essentially based on matching 'comparable' products, with in some cases ad-hoc adjustments when quality issues clearly stood out. [8)]
A method that is sometimes used in an international context is to measure quality by looking at relative prices (or unit values) of internationally traded products, using import and export statistics. [9)] In this method, a distinction is made between 'up-market' products which have an export unit value which is substantially above some average level, and 'down-market' products, with export unit values substantially below that level. Countries that produce (or export) more 'up-market' products are assumed to produce more products with higher quality. This method is not useful in the construction of unit value ratios, since the share of up-market and down-market products does not tell us anything about the unit values themselves. Moreover, the method is very dependent on the level of aggregation that is used, as explained in van Mulligen (1999).

The hedonic method

The focus of this thesis is on procedures to make quality adjustments in price comparisons, both across periods of time and across countries. As already indicated above, broadly speaking, there are two ways to calculate relative price levels: the *matched model* method and the *hedonic* method.

The first is the traditional method of price comparison. Only prices of similar products are compared, although small differences in product specifications are sometimes assumed to be negligible. More serious problems occur when (new) products appear for which there is no counterpart in the other period or country. Such products simply cannot be matched and either have to be left out of the sample or require certain assumptions concerning the differences in quality with other products.

The hedonic method makes use of regression analysis to construct quality-adjusted price indices. The basic idea is that a product consists of several characteristics, and that these characteristics are what buyers are mainly interested in (Lancaster, 1971). [10] For example, in the case of a computer, the buyer might consider the performance of its processor, the size of its hard disk or the capacity of the graphical card. When the quality of a good changes, the product characteristics change in quantity (and vice versa), and the implicit price change of that characteristic as well. A hedonic function estimates the effect of the change in each of the characteristics on the price of the good. Any remaining price change is considered the 'pure' price change, unaffected by differences in quality). [11]

Following the pioneering work of Waugh (1928) and Court (1939), Griliches (1961) brought the hedonic method under the attention of economists and statisticians. Subsequent work at the Bureau of Economic Analysis (BEA) (Cole *et al.*, 1986) started using the method in the national accounts deflator for computers (Cartwright, 1986). Nowadays, the use of the hedonic method is much more widespread, and by some estimates (Landefeld and Grimm, 2000; Moulton, 2001), components where the hedonic method is (partly) applied now account for 18% of GDP.

The main advantage of the hedonic method is that it makes an explicit, objective quality adjustment in the price index. The biggest problems associated with the hedonic method are the lack of suitable data in some product areas but even more so in services, the fact that there has not yet evolved a 'standard methodological practice', difficulties associated with the choice of relevant characteristics, and the inherent resistance of many statisticians to measuring 'unobservable' price changes.

Although the hedonic method has now been used for more than a decade in several U.S. price statistics, most European statistical agencies have only recently introduced it at best. Some countries have implemented the U.S. deflator for computer equipment, adjusted for exchange rate fluctuations, but apart from France (mentioned above) and very recently Germany, none has constructed their own hedonic index.[12] Indeed the debate on the advantages and disadvantages has been ongoing, even though many statistical agencies in the European Union as well as Eurostat, the statistical agency of the European Union, are now seriously considering the introduction of hedonic price measurement in their statistical system.

Structure of the thesis

The structure of this thesis is as follows. Chapter 2 provides a discussion of how to conceptualise the 'quality problem' in the case of price measurement. A description is given of matched model methods and of how statistical agencies, which extensively apply this method, deal with changes in quality or the presence of goods that cannot be matched. It also provides an illustration of the quality problem in the dimensions of price index construction and the calculation of relative international price levels.
The hedonic method will be discussed in detail in Chapter 3. First, the theory behind the hedonic method will be reviewed, followed by a description of how hedonic functions can be used to calculate quality-adjusted price indices and purchasing power parities. Although the research on hedonic price indices is far more extensive than the literature on hedonic purchasing power parities, some pioneering studies on both topics are commented upon.
Chapters 4 and 5 are of a more empirical nature. Two commodities, on which a lot of hedonic work has been carried out, will be studied in more detail, namely computers and automobiles. Computers are the topic of Chapter 4. The main body of the empirical work is the construction of quality-adjusted price indices for several types of computer equipment in the Netherlands. These results will be compared with matched model indices, to assess the differences. Recently, it has been indicated that the hedonic method is not really needed for computers (Aizcorbe *et al*., 2000); a claim which is also reviewed in Chapter 4 with the use of detailed scanner data for computers.
In Chapter 5, the focus is on price measures for automobiles from the production side in six major car-manufacturing countries: Japan, the United States, Germany, France, the United Kingdom and Italy. With hedonic regressions, quality adjusted industry PPPs for cars across the countries will be computed for this industry.
Chapter 6 provides a summary and comments on the need, usefulness and feasibility of adjusting price comparisons for quality, both through time and across countries. Recommendations with respect to the use of the hedonic method in price index numbers and international price comparisons will be made.

Notes

[1] This estimate takes into account improvements which had just been carried out by the Bureau of Labor Statistics (BLS). Prior to these improvements, Boskin *et al*. estimate an average upward bias of 1.3 percentage points per annum for the period 1978–96.

[2] Critiques of the commission's findings can be found in, for example, Moulton and Moses (1997), Abraham (1997), and Bureau of Labor Statistics (1997). In fact the Boskin report spawned a massive amount of discussion and research. Interesting studies in this respect are Boskin *et al*. (1998), Schultze and Mackie (2001) and Lebow and Rudd (2001).

[3] See Folkertsma (1998) for the implications on the Dutch CPI, and Brachinger *et al.* (1999) for the Swiss case.

[4] It should be noted that for deflation purposes, the producer price index rather than the consumer price index is used. Hoewever, the quality issue for both is similar.

[5] Note that real GDP growth in the U.S. was substantially bigger than in France during this period, which puts the different estimates into perspective. Also, the ICT sector in the United States has a much bigger output share than the French equivalent.

[6] See, for example, Kravis, Heston and Summers (1982) for the most detailed account of quality problems in the ICP programme.

[7] See van Ark (1993, pp. 34–38) and van Ark and Timmer (2001) for a discussion.

[8] See, for example, van Ark and Gersbach (1994) for a discussion of quality adjustments in the manufacturing productivity study of the McKinsey Global Institute (1993).

[9] Illustrations are provided in Freudenberg and Ünal (1994) and Minne and Noordman (1999).

[10] Hence price measurement is mainly based on the perception of product characteristics by the user. See Chapter 3 for a more detailed discussion.

[11] This would ideally be the case. In practice, many unobservable, price-determining characteristics will remain present in the price index. This will be discussed more extensively in Chapter 3.

[12] An exception is the hedonic index for newly constructed buildings in the Netherlands, which was adopted by Statistics Netherlands in 1994.

2. The quality problem in price comparisons: history, concepts and applications

2.1 Introduction

When measuring economic growth, the analyst is primarily interested in changes in 'real' income, output or expenditure. Hence the nominal (money) value of output needs to be adjusted for the change in prices. [1] The nominal value of output consist of two components, namely prices and volumes or quantities. Since this thesis is dealing with price index numbers, I will use an index notation for the change in nominal value, where V stands for the index of nominal value, P is the general price index and Q is that of volume:

$$V = P*Q \tag{2.1}$$

To assess the real growth rate of output, we are only interested in the change in volume, which excludes inflationary effects:

$$Q = V/P \tag{2.2}$$

Therefore, we need measures of the price changes in output or expenditure, called deflators or price indices. These measures of price changes often contain a bias, by not adequately adjusting for changing quality of the output. If quality increases remain unobserved, price measurements are biased upwards, and volume changes are proportionally biased downwards, and vice versa when quality deteriorates. Using equation (2.2), we express the price component in a quality component (A) and a quality-adjusted price change component (P'):

$$V = (P'*A)*Q \tag{2.3}$$

The main problem now – which is the main focus of this thesis – is to measure A, and move it from the price to the volume component. In doing so, we consider a quality increase as an increase in volume, rather than an increase in price:

$$V = P'*(Q*A) = P'*Q' \tag{2.4}$$

where Q′ is a quality-adjusted measure of the volume component. Returning to equation (2.2), we then get:

$$Q' = Q^*A = V/P' \tag{2.5}$$

Hence to achieve a measure of real output growth, which is adjusted for changes in quality, we ultimately need quality-adjusted measurements of price change. This chapter discusses different methods of price measurement, focusing on how quality adjustments are carried out with these methods.

Price indices are generally based on a sample of products in two or more periods. The conventional method for calculating price indices is called the 'matched model' method (Triplett, 1986). According to this method, prices for products with the same product specifications that can be observed in two periods are used in the index only. In this way a price difference between both periods is a pure price change, which cannot be attributed to a change in the quality of the product. As has been reiterated by the Boskin commission, when models improve in terms of quality or when new models emerge, matched model indices will be biased. The Boskin commission was not the first to address the quality issue, however. Already as long ago as in 1915, Wesley Mitchell identified the quality problem in the U.S. wholesale price index (Banzhaf, 2001).

One problem of the matched model method is that the price changes for matched models do not capture the price change for all models. Typically, only 'established' models that have been around for some time are included in this method (obsolete as they may be), while new and better models are excluded. These new models are likely to cause a price change, which is not reflected in the price change of continued products. This problem is called the 'outside the sample' problem (Triplett, forthcoming), a terminology that will be adopted in this thesis.

A second problem is that models that are not identical are nevertheless matched, for example because of forced replacements. What generally happens when an item drops from the index, is that the statistician looks for an item that replaces the disappeared one. When the original item and its replacement are not identical, a bias occurs at the moment of replacement. This problem is analogously called the 'inside the sample' problem. The difficulty of the 'inside the sample' and 'outside the sample' problems is that they substitute for one another: the more one tries to avoid the 'outsider' problem, the more likely it will be that the 'insider' problem becomes more important.

The purpose of this chapter is to provide an overview of the standard methods of price comparison, and the associated problems when changes in quality arise. First, a short overview of price index measurement and the associated quality problem will be presented in Section 2.2, together with some of the different types of indices that are used. Section 2.3 gives a description of the matched model method, the most widely used method to construct price indices. Several implicit and explicit ways to deal with changing quality exist for this method, which will be briefly touched upon. An alternative method to correct for differences in quality is the hedonic method, which makes use of regression techniques. This method will be shortly introduced in Section 2.4, and will be discussed in more detail in Chapter 3. Griliches' (1961) use of the hedonic method also gave renewed rise to the so-called user value vs. resource cost debate, which will also be dealt with in Section 2.4. Section 2.5 discusses the importance of the quality problem in international price comparisons. Section 2.6 then provides some illustrations of the quality issue both in an intertemporal and a cross-country context. Section 2.7 concludes.

2.2 *Price index numbers and quality: theory and history*

Cost-of-goods vs. cost-of-living indices

The measurement of price changes has been a topic of interest for a long time. The earliest known example dates from 1707, when William Fleetwood, the Bishop of Ely, compared the value of money for an Oxford student for that year and for the year 1460, using a fixed basket of goods (Diewert, 1993, p. 34).
The general price index problem is how to derive an overall price change of a multitude of different goods and services. Using arithmetic means, the price index P^{01} for the overall price change between periods 0 and 1 can be expressed as follows:

$$P^{01} = \sum_{i=1}^{N} w_i \frac{p_i^1}{p_i^0} \tag{2.6}$$

where p_i^0 and p_i^1 are the prices of good i (i = 1,...,N) at times 0 and 1, respectively. The weight w_i reflects the relative importance of good i in the index. How different prices are weighted in the index determines the index number properties, which is a major topic in the price index literature. The standard procedure is the use of quantities sold of the products that are in the index. Quantities are the natural counterparts of prices, and they ensure that products that are consumed more frequently obtain a larger share in the index.[2)]
Formal price index formulae, which employed quantity weights, were first developed in the 19th century, of which the most famous and most widely used are the Laspeyres and Paasche formulas.
The Laspeyres index calculates the average price change between periods 0 and 1 of a fixed basket of N goods with prices p_i, using the corresponding quantities q_i from the base period 0:

$$P_L^{01} = \sum_{i=1}^{N} w_i \frac{p_i^1}{p_i^0} = \frac{\sum_{i=1}^{N} p_i^1 q_i^0}{\sum_{i=1}^{N} p_i^0 q_i^0} \tag{2.7}$$

with $\quad w_i = \frac{p_i^0 q_i^0}{\sum_{i=1}^{N} p_i^0 q_i^0}$

The Paasche index uses the same fixed basket, but instead applies quantities from the current period 1 as weights:

$$P_P^{01} = \sum_{i=1}^{N} w_i \frac{p_i^1}{p_i^0} = \frac{\sum_{i=1}^{N} p_i^1 q_i^1}{\sum_{i=1}^{N} p_i^0 q_i^1} \tag{2.8}$$

with $\quad w_i = \frac{p_i^0 q_i^1}{\sum_{i=1}^{N} p_i^0 q_i^1}$

or alternatively, if one prefers to use value shares from period 1:

$$P_P^{01} = \left[\sum_{i=1}^{N} \frac{w_i}{p_i^1 / p_i^0} \right]^{-1} \tag{2.8'}$$

with $\quad w_i = \frac{p_i^1 q_i^1}{\sum_{i=1}^{N} p_i^1 q_i^1}$

Since both the Laspeyres and Paasche indices use weights from a given period, 0 or 1, they are fixed weights indices and therefore do not allow for substitution effects. Generally, buyers shift from more expensive items to relatively cheap ones. A Laspeyres index then overestimates price change, as it gives too much weight to the more expensive items, which were purchased more in period 0. Likewise, a Paasche index will generally underestimate actual price change, since it gives too much weight to the cheaper items, which are purchased more in period 1. This phenomenon is sometimes called the 'Gerschenkron effect' after Alexander Gerschenkron (1962) who studied it in an international comparative framework.

In the early 20th century, Fisher (1922) proposed the geometric average of both Laspeyres and Paasche indices as the 'ideal index', to provide the best approximation of the actual price change. This index is now known as the Fisher index. Together with the Törnqvist index,[3)] these indices are the most widely used in empirical research. For example, a variant of the Laspeyres index (called the Young index) is the formula mostly applied in the computation of consumer price indices by statistical agencies (de Haan and Hoven, 2001).[4)] Because the indices presented here focus exclusively on the prices or costs of goods (or services), they are sometimes referred to as 'cost-of-goods' indices.

In addition to these indices, there are many more price indices available. [5] To judge which formulae are more appropriate to measure price change, the so-called 'test approach' is often used. Diewert (1987) lists ten different tests to judge the suitability of index numbers. Whereas no single formula satisfies all tests, some formulas satisfy more tests than others. Since different formulas satisfy different subsets of tests, and no general agreement exists on which subset of tests is the 'best', the test approach does not lead to a single price index that is superior to others. [6]
An alternative approach to judge the suitability of tests is the economic approach (Diewert, 1987, 1993). This approach uses the assumption on optimising behaviour by consumers and producers and is based on the economic theory regarding consumers and producers. This has led to 'cost-of-living' indices P_C, which keep the level of utility of a representative consumer constant:

$$P_C^{01} = \frac{E(p^1, U^0)}{E(p^0, U^0)} \tag{2.9}$$

where $E(p,U)$ is the expenditure function of a representative consumer with utility level U, facing price level p. Without going into details on the economic approach to price indices, this approach leads to price index numbers that are *exact* and *superlative*. In the definition of Diewert (1976), a price index is exact if it equals the ratio of costs to obtain a given level of utility for a representative consumer in different periods. [7] A price index is superlative if it is exact for a flexible unit expenditure function $e(p)$, which is the minimum expenditure of obtaining one unit of utility. As it turns out, only three price indices are superlative: the Fisher, the Törnqvist and the Walsh indices. [8] As with the test approach, the economic approach does not lead to a single best index formula. However, empirical research shows that the difference between the superlative indices is usually very small in a time series context. This is not necessarily the case with cross-sectional price comparisons. In addition, the test approach is also more inconclusive in the case of spatial comparisons than for time series (Diewert, 1999).

Fixed base indices vs. chaining

The review above only considered price index formulae for two adjacent periods. We now consider the case where there are more than two consecutive periods. To calculate price indices over longer periods of time, the two methods that are mostly used are fixed-base indices and chain indices.
With the fixed base principle, one particular period is chosen as the base period, and all other periods are linked to this base period, call this period 0. The base period determines which goods are selected for the index, and only price changes of these items are tracked over time. If we now want to measure the price change from period 1 to period 2, we divide the index P^{02} by the index P^{01}. In practice, statistical agencies change their base or benchmark period every few years, using a different basket of goods for which prices are collected in each new base year.

With the chain principle, all price indices for adjacent periods are calculated, and no period is singled out as the base. If we therefore want to measure the price change from period 0 to period 2, we multiply, or 'chain', the indices P^{01} and P^{12}.
The chain principle offers the advantage of not treating a single period asymmetrically, like the base period principle does. It better allows for substitution effects, so that the Gerschenkron effect present in a fixed-base index will disappear. Moreover, if chaining is combined with frequent resampling, the index allows for better treatment of new and disappearing items, thus substantially reducing the outside the sample bias, although it cannot be eliminated altogether. Because of these advantages, the chaining principle is the preferred principle in the System of National Accounts 1993. In price index handbooks, the chaining principle is endorsed as well, albeit more implicitly (Eurostat, 2001). [9)]

The quality issue in historical perspective

In the history of price index research, the problems caused by changing quality were recognized at an early stage. While reviewing the U.S. wholesale price index in 1915, Wesley Mitchell stressed that the problem of heterogeneous products extended beyond the question of changing quality of a particular item (Banzhaf, 2001). He also recognized the problems caused by differences in prices of the same article in different locations and across different sellers. To overcome these problems, the price collector should select a representative sample of all price quotations, and must have sufficient technical knowledge regarding the products, to ensure that prices of the same variety are collected (Banzhaf, pp. 349–350). [10)] However, in 1915 Mitchell did not suggest explicit procedures to measure quality to adjust price indices for quality changes.
During his second review of U.S. official price indices in 1944, Mitchell again addressed the problem of rapidly changing quality, as did von Hofsten (1952) who focused on the case of Sweden. Both authors noted that because of the diversion of resources to war efforts, many products deteriorated in quality, which was not take into account in the official price indices. [11)] Hofsten suggested to use characteristics of products to measure their quality, and to make explicit adjustments to price indices on the basis of these characteristics.
The quality problem was more explicitly addressed by the Stigler Committee (Stigler *et al*., 1961), who made several recommendations on how to tackle quality changes. The recommendation that has received most attention in retrospect, was to perform regression analysis to calculate implicit prices of characteristics that determine the price of the product. This method became known as the *hedonic method*. A staff paper by Zvi Griliches (1961) which used the hedonic method was included with the Stigler report, although this method was first employed by Waugh (1928) and Court (1939). The purpose of the hedonic method, namely explicitly adjusting for quality changes by using characteristics of products, was consistent with the overall suggestion of the

Stigler Committee, namely that the consumer price index (CPI) should be an utility indifference-based cost-of-living index, where utility is generated by goods and their characteristics. [12] The Bureau of Labor Statistics (BLS) rejected Mitchell's recommendation, and only recently explicitly endorsed the cost-of-living framework (Banzhaf, p. 362; U.S Bureau of Labor Statistics, 2000).
The debate whether the CPI should be based on the cost of goods or the cost of living still has not been resolved. A recent contribution to the debate is the Schultze report (Schultze and Mackie, 2001), which argues that the CPI should be based on a cost-of-living framework, albeit a conditional one. External factors that take place outside the universe of private goods (e.g. the environment, crime rates, public goods, and so on) should be held constant for the calculation of the CPI.

2.3 *The matched model method* [13]

The most commonly used method to obtain measurements of price differences, either between periods of time or between countries, is the *matched model* method. Until recently, nearly all price indices constructed by statistical agencies were made using this methodology, and it was implicit in the indices presented in the previous section. From the perspective of statistical agencies making price index numbers, the basic procedure is the following. A sample of sellers and products is selected, each with an initial period-price for each product. For later periods, the price for the same product sold by the same seller is collected; both the physical characteristics and the point of selling are therefore held constant, to reduce heterogeneity in the index. Prices for both periods are matched for each product/seller combination, and the final price index is computed from all these price ratios. [14]
Most price index compiling agencies have adopted the Laspeyres-formula approach for the CPI, like the one in equation (2.7), although the Paasche formula is occasionally used as well. As mentioned in the previous section, the Laspeyres index is a static formula: because it uses a fixed composition of goods and services, it does not allow for substitution between different products. Extending beyond the bilateral case, the Laspeyres index (2.7) for time t can also be written as:

$$P_L^{0t} = \sum_{i=1}^{N} w_i^0 \left(\frac{p_i^t}{p_i^0} \right) = \frac{\sum_{i=1}^{N} p_i^t q_i^0}{\sum_{i=1}^{N} p_i^0 q_i^0} \tag{2.10}$$

where $w_i^0 = \dfrac{p_i^0 q_i^0}{\sum_{i=1}^{N} p_i^0 q_i^0}$

is the value share of product i in the total consumption basket in the base period.

As mentioned in the previous section, the base period for the price index not only determines the value shares by which the price ratios are weighted, but also which goods are selected. As it is impossible to include each and every product sold in every store in the CPI, generally samples are drawn to represent total consumption. Random sampling is usually involved, but at the level of individual commodity prices, sometimes a cut-off sample is taken, which implies that only the most frequently sold models are selected. This ensures a good representation of a large part of total consumption, something not necessarily achieved with random sampling (although random sampling may yield unbiased estimators, unlike cut-off sampling). [15)]

The 'prices' that are included in the index are actually averages of prices of products with the same specification. [16)] These are usually not physically identical products, but rather individual commodities with a certain specification. Depending on the commodity, this specification allows for a certain amount of heterogeneity. For the PPI, Gordon (1990) provides the examples of wire rods and colour television sets. In these examples, the specification of the former is technically very detailed, leaving little room for variation, whereas the description of the latter leaves much room for interpretation. According to Gordon, "for relatively homogenous commodities, it is possible to make the specification so detailed that it is quite unlikely that quality change slips through unnoticed" (p.83). The more heterogeneous a commodity and the more associated technical innovation, the harder it becomes to provide a specification that covers each feature. Because of this procedure, (2.10) needs to be rewritten into a version that is closer to actual practice:

$$P_L^{0t} = \sum_{j=1}^{M} w_j^0 \left(\frac{\overline{p}_j^t}{\overline{p}_j^0} \right) \tag{2.11}$$

where $\overline{p}_j^0$ and $\overline{p}_j^t$ are the average prices of a specified commodity j (j=1,...,M) at times 0 and t, and w_j^0 the associated share in consumption in the base period.

These specified commodities are the lowest level of aggregation in the price index, and are sometimes referred to as 'basic headings'. Individual price collection takes place on the level of selected items within the basic heading. For each basic heading, the price $\overline{p}_j$ is derived by using prices of the different items within the basic heading across the outlets that are in the CPI sample.

An example, taken from de Haan and Hoven (2001), may clarify this. The Dutch CPI consists of six levels of aggregation, the lowest being the basic headings, called articles. One category at the fifth level is audio equipment, which contains eight basic headings or articles. One of these is a CD player of a specific brand. On the basis of market shares, one particular model of this brand is selected for the index. For this basic heading, the specification is rather detailed, something which is not the case for every basic heading. [17)] Prices for this particular CD-player are collected across different outlets in different towns, and the unweighted arithmetic average is used as the price for this basic heading. The price index of such a basic heading is therefore equal to the ratio of the average prices in two periods. The index weight w_j^0 that is attached to each basic heading is based on budget interviews, which are held across 2,000 households.[18] Market shares are used only to determine which particular good or goods to observe for a basic heading. The index weights are revised every five years, and are usually implemented three to five years after each base year. The 1995 weights were only implemented in the CPI in January 1998, and the 2000 weights in January 2003. This long lag is for the largest part due to the time it takes to implement the results from the household budget interviews in the weighting system. Starting from January 2004, however, the weighting system in the Dutch CPI will be revised annually instead of every five years.

Since prices are only matched for commodities sold by the same retailers, a lot of 'hidden' price determining factors are held constant. Prices are not only determined by quality aspects of the commodities themselves (depending on the amount of detail in the specification), but also other factors, which are largely determined by the point of selling, like service provided and location of the outlet. This also means that if changes appear in the types of retailing services, additional biases can be introduced in a matched model index. One could argue that the purchase of a commodity consists of the physical commodity itself and of the services included in the transaction. Therefore a match is only possible when both parts of the transaction are unchanged through time. The same book by the same author of the same publisher of the same print run is simply not the same commodity if it is purchased online at Amazon, in the local supermarket or in the academic bookstore downtown.[19] The focus of this thesis, however, is on biases introduced by quality differences of the goods themselves.

As discussed in the previous section, the quality bias comes in two ways: the 'inside the sample' problem and the 'outside the sample' problem. The first problem occurs when there are quality changes in the models that are already in the index. The second problem is relevant when new products appear, which are not part of the sample and therefore not (yet) in the index. Following price index methodology, it is often difficult to distinguish between the two problems. For example, if a new computer model is introduced, is this just an existing model with somewhat different specifications, or is it considered to be a new product? Strictly speaking, from the matching point of view it should be considered as a new product, since only identical products that can be matched are part of the index. However, as the 'specification pricing' method still leaves some room for variation of characteristics, the dividing line between 'inside the sample' items and 'outside the sample' items remains blurred.

Quality adjustment procedures in the matched model methodology

When researchers construct a price index over longer periods of time, separate indices of shorter periods are usually linked. This method is known as splicing, and usually no particular attention is paid to which items are used for the separate indices. Price statisticians, however, pay considerable attention to the items that are in the index. If the price of an item cannot be observed any-

more, price statisticians look for a replacement item to fill the gap, as described in Section 2.1. Because the new item and the item it replaced are generally not identical, the price statistician faces a quality problem.
This is illustrated with three consecutive periods. At times t-1 and t, the price sample contains the prices of J items i, where i = 1,..., m-1, m. At time t+1 item m has been replaced by item n. So in all three periods, there are J observations, of which J-1 are matched for the entire period. The quality adjustment problem is to find a way to match these samples, while maintaining the matched model methodology. Statistical agencies use various methods to deal with this issue, and the choice of a particular method appears rather subjective. Some possible methods, that are employed by statistical agencies, are discussed here: the overlapping link method, the imputed price change method, the direct comparison method, the link-to-show-no-price-change method and explicit quality adjustment methods. All terminology used below is taken from Triplett (forthcoming).

The overlapping link method

For the overlapping link method, it is assumed that item n is already available at time t. We now compute two indices: one for period t-1 to t, which contains observation m; and one for t to t+1, which contains the replacement item n. These two indices are linked to obtain the price change between periods t-1 and t+1, although they do not consist of the same sample. For simplicity, this is illustrated with an unweighted geometric average price index. The price change from t-1 to t+1 is then equal to:

$$\begin{aligned}
P^{t-1,t+1} &= \left(\prod_{i=1}^{m}\frac{p_i^t}{p_i^{t-1}}\right)^{\frac{1}{J}} * \left(\prod_{i=1}^{n}\frac{p_i^{t+1}}{p_i^t}\right)^{\frac{1}{J}} \\
&= \left[\left(\prod_{i=1}^{m-1}\frac{p_i^t}{p_i^{t-1}}\right) * \frac{p_m^t}{p_m^{t-1}}\right]^{\frac{1}{J}} * \left[\left(\prod_{i=1}^{m-1}\frac{p_i^{t+1}}{p_i^t}\right) * \frac{p_n^{t+1}}{p_n^t}\right]^{\frac{1}{J}} \\
&= \left[\left(\prod_{i=1}^{m-1}\frac{p_i^{t+1}}{p_i^{t-1}}\right) * \frac{p_n^{t+1}}{p_m^{t-1}} * \frac{p_m^t}{p_n^t}\right]^{\frac{1}{J}} \qquad (2.12)
\end{aligned}$$

Prices of m and n actually cannot be matched directly (as appears in (2.12)). The first two terms between brackets in the final part of (2.12) together comprise an index based on the average prices of the two unmatched price sets at periods t-1 and t+1. The third element, the ratio of the prices of m and n at time t, p_m^t/p_n^t, serves as an implicit quality adjustment for the biased price ratio p_n^{t+1}/p_m^{t-1} . Therefore this implicit quality adjustment can be seen as a measure of the amount by which the quality adjusted price index (2.12) increases at a different rate between periods t-1 and t+1 compared to a simple ratio of average prices of the unmatched price sets (Triplett, forthcoming).

This quality adjustment may be larger or smaller than the correct quality adjustment, depending on the reason why item m was replaced by n. For example, if item n has a better price/performance ratio than m, then the quality adjustment is smaller than the correct adjustment. The major disadvantage of this method, however, is the lack of data. Usually computers m and n are not available in a same period, so that not all price sets that are needed for this method are available. This is the main reason why the overlapping method is only rarely used by statistical agencies. Instead a variant is used, which will be discussed next. The methods discussed in the remainder of this section will focus on the two period case, as the assumption for these methods is that there is no overlapping period.

The imputed price change, implicit quality adjustment (IP-IQ) method

A method that is frequently used when new items are introduced in the index, is what Triplett (forthcoming) calls the 'imputed price change, or implicit quality adjustment method'. At the time when an item disappears from the sample and is replaced by a new one, the ratio of their prices is deleted from the index, and replaced by the price ratio of all unchanged items that are in the index.[20] Thus, the price index of a new item is assumed to be the same as all existing items. The price index from period t to t+1 then becomes:

$$P^{t,t+1} = \left(\prod_{i=1}^{m-1} \frac{p_i^{t+1}}{p_i^t} \right)^{\frac{1}{J-1}}$$

In essence this method is the same as the overlapping link method except that the obsolete and new items themselves are dropped from the index in period t, because an overlapping link cannot be made. Therefore this method makes an implicit quality adjustment. A bias is introduced when this implicit quality adjustment is not equal to the actual quality change. When the implicit adjustment is larger than the actual quality change, the price index is biased downwards.

Suppose that general prices are rising and quality is increasing, in which case one would expect an upward bias. But since the ratio p_n^{t+1}/p_m^t is larger than the price change of all unchanged items (because n also has a better quality than m), and this larger price change is deleted from the index and interpreted as a change in quality, not as price change, this introduces a *downward* bias in the index: i.e., the IP-IQ method overadjusts for quality change.

This downward bias can be illustrated with the following example: suppose prices of a certain commodity are constant over time, and only the introduction of a new item introduces a new (higher) price. Since the price index of the other items is unity, the index of the new good becomes unity as well. In the next period, the new item has become an existing one, with unchanged price. In this example, the price index will always be unity, although price changes do occur. This may be especially relevant in oligopolistic markets, where consumers have incomplete information, and producers can exercise their market

power. In this way, producers can set their prices of new products above the amount that would be warranted by the increase in quality.
What matters with the IP-IQ method is that it introduces a bias into the price index when the implicit quality adjustment is not equal to the *actual* quality change, not necessarily when the actual quality change is not equal to zero. When the implicit adjustment is larger than the actual quality change, this bias is downwards, when it is smaller, the bias is upwards.

The direct comparison method

When the prices of observations m and n are not available at the same time, their missing ratio has to be estimated. Rather than imputing the price change using matched items, one can assume that there is no quality difference between m and its replacement n, in which the quality adjustment is unity. However, if quality is improving, the direct comparison introduces an upward bias in the price index. It is against this method that the main body of criticism of the matched model is aimed. The direct comparison is the clearest example of how the inside the sample bias comes about: prices are matched of items that are not identical. The price index from period t to t+1 then becomes:

$$P^{t,t+1} = \left[\left(\prod_{i=1}^{m-1} \frac{p_i^{t+1}}{p_i^t} \right) * \frac{p_n^{t+1}}{p_m^t} \right]^{\frac{1}{J}}$$

Although the direct comparison method is used very frequently by statistical agencies, it appears that it is only applied when quality differences are minimal, so that the bias introduced by this method is also relatively small (Triplett, forthcoming).

The link-to-show-no-price-change method

Whereas the direct comparison method assumes that the price difference consist only of inflation, and assumes zero quality change, the link-to-show-no-price-change method makes the opposite assumption. It assumes that the price difference between observation m and its replacement n is entirely due to quality change, and inflation equals zero. Therefore the implicit quality change is equal to p_n^{t+1}/p_m^t , and the adjusted price ratio is set to 1. When prices are rising, this method biases the index downward, whereas it introduces an upward bias when prices are falling, regardless of the actual quality change. Like the direct comparison method, this method implies a strong inside the sample bias. The price index from period t to t+1 then becomes:

$$P^{t,t+1} = \left[\left(\prod_{i=1}^{m-1} \frac{p_i^{t+1}}{p_i^t} \right) * 1 \right]^{\frac{1}{J}}$$

The use of this method is restricted, and even banned in the Eurostat regulations concerning the computation of the Harmonised Index of Consumer Prices (HICP) (Eurostat, 2001). Statistical agencies do not employ it when the price of the replacement item is lower, but its quality is judged higher.

Explicit quality adjustments: option pricing and judgmental quality adjustments

Unlike the previously discussed implicit methods, the option pricing method makes an explicit quality adjustment in the price index. It is applied when an item disappears from the index, and a replacement is needed. Furthermore, the new item differs from the old one only in the sense that it contains an extra feature, or 'option', the old one didn't have. The price of the optional feature when purchased separately is estimated or enquired from the manufacturer, and is subtracted from the new price. As optional features generally are more expensive when purchased separately then when part of a commodity, this method tends to introduce a downward bias. To correct for this, for example Statistics Netherlands generally uses half of the option price to subtract from the price of the model. However, this is an arbitrary amount, the bias of which is uncertain but likely not zero. This method is applied regularly by statistical agencies, but only when quality change is deemed small and the option price can be estimated (Gordon, 1990, and Chapter 4 of this thesis).
As implied by the name, when a judgmental quality adjustment is applied, the price statistician makes an estimate of the monetary equivalent of the better quality of the replacement item. Sometimes this is based on estimated production costs,[21)] other times more subjective or even arbitrary adjustments are made.[22)] Regardless of the actual adjustment, it depends on the subjective judgment by the price statistician, and is therefore not to be recommended.

Summary

For all variants of the matched model method that are used to adjust for quality differences, the major problem stems from the fact that items which cannot be matched appear regularly in the index. This problem is especially relevant when the index is based on a fixed base periods that changes only every couple of years, which is standard practice in official price statistics. Silver and Heravi (2002b, pp. 2–3) criticise the matched model methodology with a fixed base period on the grounds of its 'static sampling universe': it restricts the index to models which exist in both the base and the current period. It does keep the quality of matched products constant, but when goods disappear from the index and new goods come on the market and no correction is made for this in the index, the index becomes ever less representative for the actual price change. It is a static sample of a 'dynamic universe'.

The methods discussed above offer some correction for this problem, but all of them lead to inside or outside the sample problems. A possible solution is switching to a chained index principle, which eliminates the representativeness problem somewhat, depending on the frequency by which items are resampled. But even with a chained principle, prices of items cannot be matched at the moment they enter or exit the market.

Therefore, it is worthwhile to consider alternative methods to the matched model approach in order to assess for quality changes in price indices. To this date, the hedonic method has been the most popular alternative to the matched model method.

2.4 *The hedonic method: a solution?*

Except for the option pricing and judgmental methods, all methods to adjust for quality differences in a price index discussed above make only implicit quality adjustments. A method that does this more explicitly is the *hedonic* method. This method was pioneered by Waugh (1928) and Court (1939) and brought to the foreground of economic analysis by Griliches in 1961 as a staff paper to the report of the Stigler Commission.

The basic idea of the hedonic method is that price changes of heterogeneous goods or services can be analysed by disaggregating them into more elementary units that better measure the nature of the good, namely its characteristics. Lancaster (1971) defines characteristics of goods as "objective properties of things that are relevant to people". Rather than goods, characteristics are seen as homogeneous economic variables, which together form heterogeneous goods. Lancaster draws the analogy of shopping carts. Consumers can only buy carts with a fixed set of products in them, but they value a cart by the products that are in it. Hence the term 'hedonic', which was coined by Court (p. 107n):

> Webster's New International says "Utilitarianism, seeking the good in the greatest happiness of the community as a whole, is the chief hedonistic doctrine." Thus, Hedonic price comparisons are those which recognize the potential contribution of any commodity, a motor car in this instance, to the welfare and happiness of its purchasers and the community.

Therefore, characteristics are what consumers are interested in. These are crucial for the hedonic method. For example, when buying a computer, consumers are not necessarily interested in the computer per se, but more in its computational speed and storage capacity. This notion is the cornerstone of the hedonic index theory; the hedonic hypothesis states that "heterogeneous goods are aggregations of characteristics, and economic behaviour relates to the characteristics" (Triplett, 1987).

The hedonic method makes use of regression analysis to estimate the effect of individual characteristics, the determinants of quality, on a product's price:

$$p_i = h(z_i) + \varepsilon_i \tag{2.13}$$

where $h()$ is a function of the quality characteristics z_i of product i, and e_i a random error term.
This section will introduce the hedonic method in a general way with the use of Griliches' 1961 article on car prices. This paper led to a wide debate on how to handle quality issues in general, the so-called 'user value, resource cost' debate (Triplett, 1983), which will be dealt with in the second part of this section. The hedonic method and its applications are tackled in more detail in Chapter 3.

Griliches' 1961 study [23)]

In this article, Griliches built a small theoretical framework taking account of quality aspects in the construction of price indices and to use this framework for regressions on the prices of automobiles and their characteristics in the United States from 1937 to 1960. In short, Griliches' aim was "to find out whether this method [of adjusting price indices for quality change] is feasible and operational and whether the results are promising and different enough [from the unadjusted price indices] to warrant the additional investment in resources that are needed to acquire the necessary data". (Griliches, 1961, p. 137)
To make the method feasible for his purpose, Griliches used multiple regression techniques on data for various models of cars with different specifications, sold at different prices, to derive the implicit prices of all the model characteristics he distinguishes. These characteristics included a car's weight, its horsepower, its length and several dummy variables concerning the type of engine, transmission and optional features like power steering etc.. The basic idea is then to interpret the coefficients as implicit prices of each characteristic to adjust for the quality change in the overall price change of cars between the studied periods. Alternatively speaking, in this way, Griliches states, one can calculate what the price of the good would have been had there been no changes in quality (Griliches, p. 138).
As with other explicit quality adjustment methods discussed above, these regression equations can be used to estimate the price of a new bundle of characteristics (or qualities) that was not available in the previous period, provided that this new bundle does not contain any new characteristics or qualities, but only differs from other models in the quantity of each characteristic. Especially in sectors where the quality of goods changes rather rapidly or where there are many different varieties of a good under the same basic heading, like in the case of cars, the failure to pick up new quality characteristics can be a major drawback. But Griliches argues that even if one type of a product does contain some (completely) new qualities, "the equation can be used to estimate the change in price...and half a loaf may be better than none" (Griliches, 1961, p. 140).
A way to resolve this may be to set the quantity of these 'new' qualities at zero at the time it was not yet available or to use dummy variables to identify the presence or absence of this 'new' quality characteristic.
Having estimated the coefficients, Griliches then defines an index of quality change for each variety, given by the ratio of the predicted price of the variety based on the combination of qualities in the latter period and of the predicted

price based on the combination of qualities in the former period. For all varieties, these ratios are aggregated into one 'quality change index'. With this index, he calculates the 'true price index', given by the ratio of the observed price index and the quality change index.
In the next chapter the selection of explanatory variables in the model is discussed in more detail. In this respect it is noted here that Griliches' study suffers from the fact that one of the main characteristics of cars which he uses to fit quality adjusted price indices, i.e. weight, is only a proxy for the really relevant characteristics. The weight of a car is by itself not very important to both producers and consumers, but there may (and probably will) be underlying more meaningful characteristics that are reflected in the weight of a car (including some which are also part of the regression analysis, like length and horsepower). Weight therefore bears no one-to-one relationship with all these other characteristics. Hence the coefficient for weight will be a biased estimate of the actual coefficients of the proxied characteristics. Griliches acknowledges this, which is why he introduces some dummy variables for these non-quantifiable characteristics to correct for the fact that large cars may be more expensive than the average car, for other reasons than weight and length alone. Another problem with carrying out a hedonic regression analysis for cars is the multicollinearity between the characteristics, like weight and length: the longer a car, the heavier it becomes, ceteris paribus. To remedy this problem, one must be very careful in deciding which variables to use in the regression.
Despite such criticisms (and some more will be reviewed below), Griliches' work set the stage for a lot of subsequent work on the use of hedonics techniques to account for quality changes in price index construction. The results from many of these studies were sufficiently different from the official CPI figures to signal the importance of the quality problem. Griliches' main conclusion is that "over the whole period since 1937 the CPI may be overestimating the rise in automobile prices by at least a third" (Griliches, 1961, p. 138); if the quality adjustment is based on the beginning period weights to make it better comparable to the CPI, then "about three fourths of the rise in automobile prices in the CPI since 1937 could be attributed to quality improvements" (ibid.). In fact, the price of an average car, adjusted for the improvement in quality, actually declined in the period 1954–1960.

The user-value vs. resource cost debate

Despite the fact that the results found by Griliches stressed the urgency of a resolution to the issue, not everyone agreed on the superiority of the hedonic method. For example, Milton Gilbert (1961) argues that changes in the quality of products are not very relevant in constructing price index numbers. Although Gilbert acknowledged the importance of the notion of quality change for economic welfare and production theory, he did not see a large role for this concept in the construction of price index numbers, since, as he put it, "there are strict limits to the kind of changes in quality that can be brought within the scope of index number measurement" (Gilbert, 1961, p. 287). According to Gilbert, researchers who apologise for not including quality changes in their analyses do not see that 'any conceptually broader goal is no longer quantitative' (ibid.). The underlying concern here is that if quality improvements are brought into the sphere of quantitative measurement, this would eventually make it "impossible to construct measures of output and price changes that are useful to the study of economic growth" (ibid.).

One of Gilbert's main points is that the concept of economic growth and thus economic welfare can only be studied by analysing whether there are really more goods now than there were before, and not if these goods are 'better'. He states that the change in quality should be measured by the changes in costs from units of constant quality. When there are quality improvements, these must be assessed by the additional costs involved in producing those items. These amounts are then counted not as a price increase but as an output increase. If the price of the good has changed with a different amount than the cost of the quality change, then this difference reflects a true price change.

This distinction by Gilbert between 'cost' and 'price' is a not an actual distinction, as what Gilbert proposed is exactly what Griliches is doing in his study to analyse the causes of the increases in the prices of cars. What Griliches calls the implicit price of a characteristic or quality, Gilbert is calling an 'additional cost' or an 'output increase'. According to Griliches the consumer faces a trade-off between the price of a product and its quality characteristics, and the hedonic function allows one to quantify the price-characteristics combinations at which the consumer is indifferent. In contrast Gilbert argued that the consumer will always be better off with a car at lower prices, because he obtains the same output for a smaller part of his budget. The difference is that Griliches relates to the characteristics of commodities as generating utility, whereas according to Gilbert, only final commodities as a whole (not their characteristics) can generate utility to the user. According to Gilbert, the cost-free changes in quality must not be treated as a quality change because these cannot be identified with a change in output and hence are unmeasurable; the fact that the new model may be of more value to the consumers is of no importance to Gilbert. This difference in perspectives is the cornerstone of the 'resource cost versus user value' debate.

In short, Gilbert's four major objections to incorporating quality change in the construction of price index numbers are: quality cannot be measured; only the goods or services themselves, not their characteristics, can and should be measured; changes that are measured should be measured by costs and not by value to the purchaser, expressed in prices; and welfare measures should be based on goods and services only.

In a paper published in 1964, Griliches criticised Gilbert's view. According to Griliches, Gilbert's first point is not very helpful, since quality changes are in fact being measured. Probably they are not measured well enough or maybe even wrongly, but the fact that it seems to improve empirical results makes Gilbert's criticism besides the point (Griliches, 1964).

With respect to the second point, Griliches argued that there is more to the data than meets the eye of the statistician. Griliches takes as an example the case of artificial fertiliser. A farmer who buys a sack of fertiliser is not per se interested in the number of pounds contained in the sack, but in the plant nutrients contents of it. It is only a small step to discern several plant nutrients contained in the fertiliser and to attach an implicit price to them. The fact that there probably have to be made some imputations is no reason not to carry out such procedures, as in the construction of national accounts and official price index numbers imputations are also often used.

Griliches also did not agree with Gilbert that the value of the characteristics of a good or service to the purchaser is of no importance in measuring changes in prices and output, for which the fertiliser example set out above is equally relevant. As for the last point, welfare is not only determined by production, but by utility as well. Clearly, society is better off having safer cars, faster computers and better fertilisers, even when the total number of cars, computers and fertilisers produced and consumed does not change.

Denison (1989) devotes an entire chapter to the deflator for computers calculated by the BEA. Following Cole *et al*. (1986), BEA uses hedonic methods to calculate a deflator for computer equipment in the National Income and Product Accounts (NIPA) deflator from 1986 onwards (Cartwright, 1986). Denison objects to the BEA deflator for computers, as according to him, deflation measures should not be based on the ability of capital goods to contribute to production, but on the costs to produce these capital goods. By its very nature, the BEA hedonic deflator for computers measures capital on the basis of its ability to contribute to production, and is not based on cost changes. According to Denison, such measures wrongly attribute advances in knowledge to an increase in capital input. In the case of the hedonic deflators for computers, the rising quality should be counted as an advance in knowledge (and thus as an increase in multi-factor productivity), and not as a price decline in the nominal capital stock, and thus as a growth of real capital.

Denison's objection is aggravated by the fact that the NIPA deflators for other (non-ICT) capital are still largely based on cost measures, and are therefore incompatible with the one for computer equipment. According to Denison, this is inconvenient and illogical.

The preceding discussion makes clear that, like Gilbert, Denison is an advocate of the resource cost position without an adjustment for the productivity impact of inputs. However, even when user value is concerned, like with capital input measures and consumption, hedonic price deflators are unsuitable as well, according to Denison. He bases this opinion on the grounds that "… the [hedonic] method for computers is a deviation from that used for other consumer products as well as for other capital goods. It prevents a general description of the national product estimates." (Denison, p. 37).

But Denison's position on this point is questionable. It is not clear why a better method of price measurement should be dismissed simply because it is different from standard practice, without evaluating the differences. On the other hand, if different deflators are inconsistent, productivity gains may be attributed to the wrong industries, as Triplett's (1996) example of semiconductors shows. This possible inconsistency is a major reason why statistical agencies, elsewhere than in the U.S., have been slow to adopt the hedonic method for price indices that would benefit most from it.

In addition, Denison disregards the fundamental proposition of the hedonic method, namely that heterogeneous products, like computers, are aggregates of characteristics which constitute actual output (Lancaster, 1971). The hedonic method accounts for increases in characteristics, whereas traditional price measurements only take final products into account. If one is willing to accept the proposition that characteristics equate output,[24)] than one should adopt the hedonic method to construct price measurements for goods with rapidly changing compositions of characteristics.

Already before the critical notes of Denison, Triplett (1983) tried to resolve this issue, known as the 'user-value' (supported by Griliches) versus 'resource-cost' (supported by Gilbert) debate, as follows. He notes that there are two different uses of price data, namely as an input price measure (i.e. from the users' perspective) and as an output price measure (i.e. from the producers' perspective). Input measures relate to goods and services when they are bought or used (like the CPI). Output measures, like the producer price index (PPI), focus on production and costs. Since inputs and outputs imply different theoretical price index treatments, Triplett argues that advocates of both views are correct. Researchers who are mainly interested in output measures are right to use the resource-cost approach, while those who are primarily concerned with input uses are correct to use the user-value approach.

However, in the long run, user value is important to producers as well, even if the user value of a particular characteristic has no associated resource cost. Suppose two firms manufacture the exact same product with the same characteristics bar one, and this characteristic has no associated resource cost. Output measures for both firms, e.g. productivity, will be equal, as their resource costs are the same. However, when consumers do value the particular characteristic by which the products of both firm differ, then they will prefer the variety that includes it, and if nothing changes, the firm that produces the good without it will go bankrupt. This implies that the value of output is not only determined by costs, but should be related to user value as well from an economic perspective. This point was stressed by Gordon (1990), who in his hedonic price studies for several products selected characteristics on the criteria of both user value and resource cost.

From an output deflation point of view, there is another reason why output and input price measures are not independent. A good example is the case of the use of semiconductors as an intermediate input in the computer manufacturing industry, illustrated by Triplett (1996). Regardless of how the output price index for the computer manufacturing industry is calculated, a quality adjustment is needed in the price index of semiconductors, which are one of the most important inputs in this industry.

Indeed the recent debate on quality issues in price measurement has strongly focused on the information and communication technology (ICT) industry. Wyckoff (1995) was one of the first to point out that the introduction of hedonic price measurement for the computer industry in the NIPA of the

United States led to a much faster decline of deflators for the computer industry than in other OECD countries. This sharply declining price deflator would imply that real output in this industry has risen proportionally, with equal implications for productivity growth. As Triplett points out, this implication is false, however. Most of the quality increase in computers can be attributed to the strongly increased quality of semiconductors. If no adjustment is made for this quality change in semiconductor inputs, the fast productivity growth of the semiconductor producing industry is wrongly assigned to the computer manufacturing industry. Triplett indicates that productivity growth in computer manufacturing is substantially smaller when accounting for the increasing quality of semiconductors. Output deflators, which underestimate the price decline in semiconductor inputs, grossly overstate real output growth in computer manufacturing. For this reason, the U.S. Bureau of Economic Analysis (BEA) recently started using quality-adjusted deflators for semiconductors to deflate computer output (Moulton, 2001).

I have dwelled at length on Griliches' article and the ensuing debate for two reasons. The first reason is that, although Griliches was not the first author to make use of a regression analysis to make quality adjustments, his analysis brought the hedonic method under the attention of a wider audience including economic analysts and statisticians, which finally resulted in the adoption of the hedonic method in several official price indices in the United States, notably that for IT-equipment.[25] Secondly, in his attempts to revive the quality debate, Griliches encountered a lot of criticism for the way he tackled rapid quality changes. Some authors, like Gilbert, disapproved of the notion of quality adjustment entirely. The ensuing debate has had a lot of influence on how economists think about the quality issue, so in that respect Griliches' article was a pioneering study as well. His methods may seem simplistic compared to some of the more recent sophisticated hedonic analyses, discussed in Chapter 3, but it provided the spark that got the hedonic engine running. Although they have been intertwined since the discussion between Gilbert and Griliches, the use of hedonics is a separate issue from the user value - resource cost debate. The focus of the latter was whether or not to explicitly adjust for quality changes that are valued by users, and whether or not commodities themselves or their characteristics are the relevant units of measurement. The hedonic method is just one way of explicitly measuring quality changes, by looking at a product's characteristics.

2.5 *Quality adjustments in international price comparisons*

The above concentrated mainly on the effect of quality changes on price indices. However, quality differences are relevant whenever prices are compared, be it between periods of time or across space. Quality differences across countries affect comparisons of the relative levels of output when measured on the basis of currency conversion factors like purchasing power parities (PPPs) or unit value ratios (UVRs), or when price indices are compared internationally.

Kravis (1986, p. 21) already mentioned the problems associated with international comparisons of heterogeneous goods. Hence differences in quality of a certain product need to be reflected in the specified 'basket' of goods and commodities on which the currency conversion factor is based.
Exchange rates are generally not suitable for comparing relative output levels (van Ark, 1993). First, they do not indicate the comparative value of currencies in the production or consumption of all goods and services. Second, exchange rates in principle only refer to price relatives for tradable goods and services. Finally, exchange rates can be subject to substantial short-term fluctuations and capital movements.
Whereas PPPs are based on comparisons of expenditure prices of specified items, industry-of-origin studies use ratios of unit values (UVRs) [26] which are based on matches of product groupings. This section provides a short description of both methods, and the way in which the quality problem is relevant in either approach. The second part of this section provides illustrations of the effect of quality differences across time and space for two goods: buildings and personal computers.

The expenditure approach: purchasing power parities (PPPs)

A purchasing power parity can be defined as the number of currency units required in one country to buy goods equivalent to what can be bought with one unit of the currency of another country (Kravis *et al*., 1982). PPPs are therefore based on relative expenditures on different commodities, and are suitable to compare GDP levels of countries or levels of expenditure per capita for subcategories, such as food, clothing, housing, etc..
International comparisons of prices based on the expenditure approach have been regularly carried out since WW II. Important examples of the earliest studies are those at the OEEC, the predecessor of the OECD, during the 1950s. [27] The purpose of these studies was to compare real expenditure by category in the United States and Europe. For some goods in these studies, purchasing power parities were derived by calculating the ratios of average values divided by a ratio of quantities. This approach was more comparable to that used in industry-of-origin comparisons, such as Paige and Bombach (1959), than to the recent PPP programmes. For items for which no reliable quantity information was available, PPPs were directly derived by comparing prices of specified items: 'specification prices'.
In later expenditure studies, the 'specification prices' approach was almost exclusively used, in particular since the beginning of the International Comparison Project (ICP), a joint project of the World Bank, the United Nations and the University of Pennsylvania. ICP was carried out in three phases, for 1967, 1970 and 1975 (Kravis *et al*., 1975, 1978, 1982). In the 1980s, the OECD used PPP estimates from Eurostat and did a survey of several other OECD members to derive PPPs for all OECD countries (Hill, 1984; Ward, 1985). Since then, expenditure PPPs for OECD countries are constructed on a five-year basis (and since 1990 on a three-year basis) for a large number of expenditure cate-

gories, the so-called basic headings.[28] At the basic heading level expenditure values are available to obtain real expenditure levels.

Below the basic heading level, specification prices are averaged to provide PPPs, and these basic heading PPPs are then weighted at expenditure shares to obtain aggregate PPPs. An example from the ICP Handbook (United Nations, 1992) of a basic heading is 'fresh fruit' (basic heading 1.1.1.06.1). Within this basic heading prices are collected of several specified products, e.g. oranges, tangerines, lemons, bananas, apples, cherries, grapes, and so on. The selection of specified products is based on three considerations. First, the sample of specifications is based on a final product classification, since there is generally a near complete overlap in this classification between countries. Second, a minimum number of specifications is determined for each different category. As homogeneous categories are less complex, the number of specifications needed to capture the price structure of basic headings in these cases is lower than for more heterogeneous categories. However, in ICP, "the homogeneity or heterogeneity of basic headings should be judged not so much in terms of number of items but in terms of their relation to individual price ratios. If the dispersion of price ratios is small, only a few specifications may be necessary even when it appears that the basic heading is quite heterogeneous in terms of number of distinguishable items. Where price ratios have substantial dispersion, a larger number of specifications is needed" (United Nations, p. 32).

Finally, the specific items are selected on two principles. The first principle, the so-called 'characteristicity' principle is that goods with the largest expenditures are adopted: items selected for a basic heading should be characteristic for each country (United Nations, 1992). This is equivalent to the 'outside-the-sample' problem discussed above. The second principle is that for each country identical items are selected, which is equivalent to the 'inside-the-sample' problem. Apart from circumstantial factors, there are four conditions mentioned in the ICP Handbook (p.31) for identity: first, the unit price of the item should relate to the same size (e.g. the price of one litre of milk); second, the physical and functional properties should be the same; third, the types of outlet where the matched items are sold should be the same; and fourth, delivery conditions should be the same as well.

The characteristicity and identity principles create to some extent a trade-off which, according to the ICP Handbook, is 'perhaps the most important issue in price selection' (United Nations, 1992, p. 30). Identical products can easily be matched, but they may still have the problem of lack of characteristicity. Identical products that have a large expenditure share within a basic heading in one country may have smaller shares in other countries, and therefore be less characteristic. For a relatively homogeneous group of countries, like the European Union, this may be less relevant, but for comparisons across the entire world, it will be more problematic.[29] To resolve this conflict, two solutions are possible (United Nations, pp. 51–53).

First, if one wants to stick to the characteristicity criterion, one can compare prices of items that are not identical, but are equivalent in use. Such products are sometimes referred to as common products (Gilbert and Kravis, 1954).

These products are usually called by the same product name in different countries, but have somewhat different characteristics. An example is vegetable oil, which in some countries has an olive base, in others groundnuts, corn, sunflower kernels, etcetera. Second, if we want to stick to the criterion of identity, we may make use of items that may not be characteristic for each country, but at least are commonly purchased everywhere. This means that the expenditure share of such an item is not so low that it has a corresponding high price. In such a case, the item would be essentially *un*characteristic.

Next to identical and common products, there is the category of unique products. These are products that are only present in one of the two countries under comparison. These may be products that are entirely unique, or products that have specific unique characteristics. Matching such products is not possible; if one wants to include them in a comparison, the cost or price of the product in the country where it is not available must be estimated.

Since it is much harder to compare 'like' with 'like' in an international context than in an intertemporal analysis, several methods have been used to match items. Prices of specified items were matched directly if the items were either physically identical, equivalent in quality, equivalent in use, or if the item in one country amounted to the replication of the matched item in the other. In cases where direct matches were not possible, price adjustments proportionate to quality differences were carried out.

In a limited number of cases hedonic regression techniques were used for spatial expenditure comparisons in cases where the quality differences were not easily determined directly. The categories for which this method was used were automobiles and house rents. Prices were regressed in a cross-section analysis on a number of characteristics, like the length, maximum horsepower and number of cylinders in the case of automobiles, and the building year, available facilities and floor area in the case of house rents. The methodology used is very similar to that of Griliches (1961). The application for cars will be discussed in more detail in Section 3.8.

In conclusion the quality problem in international comparisons was explicitly addressed in the different phases of ICP. Moreover, the notion that heterogeneous goods are actually aggregates of homogeneous characteristics is implicitly adopted in the methods that equate different goods by their equivalence in quality or use, and more explicitly in the case of price adjustments due to quality differences and the hedonic regressions. However, the quality problem caused by the trade-off of characteristicity and identity remains one of the main problems of the PPP methodology employed by the United Nations, OECD and Eurostat, and was explicitly addressed by two recent reports that reviewed this methodology, i.e. the 'Castles' report (OECD, 1997) and the 'Ryten' report (United Nations, 1998).

The industry-of-origin approach: unit value ratios (UVRs)

The expenditure approach has been employed more frequently than the industry-of-origin approach, which is partly due to the fact that prices for final goods can be more easily obtained than for net output of industries (Kravis, 1976, p.7). However, for comparisons of relative productivity and output by industry, expenditures on final goods are less suitable. This is because expenditure on final goods excludes intermediate inputs and exports, which are included in production, but includes imports, which are not part of (domestic) production.

One way to get around the lack of information on specified output at the net output level, is the use of unit values. These are derived from dividing the output value in terms of producer prices (ex-factory sales) by produced quantities, which are usually obtained from production censuses. [30] An advantage of unit values as opposed to specification prices, is that the quantities and unit values are consistent with the total value of output. Products are matched between countries, using the unit values as 'prices', which results in unit value ratios. The industry-of-origin approach which adopts unit value ratios has been the subject of several studies at the International Comparisons of Output and Productivity (ICOP) project at the University of Groningen. [31]

Like with the comparison of final product prices, the matching of unit values is not always straightforward, due to the different descriptions of products in the production censuses of different countries. For example, countries may use different quantity units to express production of particular goods. Secondly, information on sales value or quantities is sometimes not disclosed because of reasons of confidentiality. Thirdly, some unique products are produced only in one country and not in another, so they cannot be matched at all. Finally, matching of products is often difficult because of differences in quality.

In one respect, unit value ratios are even more subject to quality problems than purchasing power parities. This is because unit values in practice represent an average price for a mix of product varieties. These different varieties may be available in different proportions in each country, and the proportions are mostly not clear from the censuses of production. So even if the individual varieties would be identical in two countries, their mixes may be subject to the quality problem because the relative proportions in either country are unknown. This is called the product-mix problem. Because the expenditure approach makes use of specification prices of narrowly defined goods, it does not suffer from this particular problem. In addition, unit values have been pointed out as being unreliable price measures. Lichtenberg and Griliches (1989) found that an index based on unit values was a less reliable measure of price change than a producer price index based on specified items.

For the product-mix problem, some adjustments for differences in product mix in comparisons of automobiles were made in several manufacturing studies of ICOP (Maddison and van Ark, 1988; van Ark, 1993). More large-scale adjustment for differences in product mix and quality were made in detailed case studies by the McKinsey Global Institute (1993). [32]

However, despite these shortcomings of the UVR method, this does not necessarily mean that the quality problem is more serious for industry-of-origin studies than for expenditure studies (van Ark, 1993). Firstly, quality differences are likely to be most important in final consumption goods, which are the focus of expenditure PPPs, but less so in many basic, intermediate goods, which are excluded from expenditure PPPs, but make up a large share of manufacturing output. Secondly, unit values relate to a large share of output compared to specification prices. For this reason, unit values are probably more representative for total output than specification prices.
Regardless of the product-mix problem discussed above, both approaches suffer from the fact that products can sometimes not be matched due to differences in 'narrow' product quality. This is analogous to the quality problem inherent in price index numbers over time, as described above. However, adjusting for quality differences is even more difficult for cross-country comparisons than for time series. Whereas quality usually changes rather gradually over time, such differences across countries are not gradual and can be very substantial. For this reason, the hedonic method is often the only viable way to correct for quality differences in cross-sections, since it is often impossible to apply an adjusted version of the matched model method like those discussed above for comparisons over time. Whereas the number of varieties of a particular commodity that can be matched, expressed as a share in total production or sales value, is usually relatively large in the case of price indices, it often happens that no match can be found at all for a particular output category in a cross-country comparison. In such a case, explicit quality adjustments are necessary, of which the hedonic method looks like the best alternative.

2.6 *Some illustrations of the quality problem*

This section provides two illustrations of the nature of quality differences for price comparisons over time and across space. Two industries are taken which are often seen as particularly prone to quality problems, namely the construction sector and the computer manufacturing industry. The construction sector is a typical case of an industry with very heterogeneous products, whereas the computer industry is characterised by products with very rapid quality changes. For both industries there has been a significant amount of work going on to improve producer price indices including the use of hedonic price techniques, in particular in the United States. Below the practices of the statistical agencies in the United States and Germany in compiling producer price indices for the computer industry are compared. The former country is chosen because the use of the hedonic method as an explicit quality adjustment method is most extensive there; Germany is a typical European country where, until very recently, the hedonic method has not been used in official price statistics at all.[33] I also discuss issues in combining PPPs and price indices to obtain estimates of output and productivity for the computer industry.

Quality adjustments of purchasing power parities for comparisons of real output and productivity levels across countries are much scarcer than for time series. The case of the construction industry focuses on comparisons of prices across European countries. Although no use is made of the hedonic method in this case, it provides a good illustration of the problems involved due to quality differences.

The construction sector[34)]

Estimating purchasing power parities for the construction sector is a difficult task since the output is very heterogeneous. Between countries, there is a high variation in designs, the use of materials, and standards for the construction of buildings, infrastructural projects, etc. Eurostat, the statistical agency for the European Union, and the OECD regularly calculate PPPs for the construction sector. These PPP comparisons are based on lists ('Bills of Quantities') that contain a representative collection of identical standard buildings in the countries of the European Union, which consists of a range of building items and activities that have to be performed to construct the building. These lists are updated regularly to preserve representativeness. The Bills of Quantities are completed every PPP round by construction specialists who list local prices for several building items in each country. With these data Eurostat calculates price indices for three types of construction: residential buildings, other buildings and civil works. The prices include VAT and design fees, but corrections for differences in these rates does not lead to very different PPPs (Vermande and van Mulligen, 1999). In effect, the Eurostat methodology is a variant of the matched model method, where an effort is made to compare like with like, but it contains elements of an hedonic price measure as it focuses on characteristics of buildings rather than the building itself. The fundamental difference with the hedonic method, however, is that the prices of these characteristics are listed explicitly instead of being estimated through hedonic regressions.
Price comparisons for the construction sector are also carried out by individual analysts, often from the industry itself,[35)] who employ different methodologies than those of Eurostat. Some analysts compare prices of standard, identical buildings, like Eurostat does, but others compare prices of typical, functionally similar buildings. Both kinds of comparisons have their advantages and disadvantages. In the former case, the buildings can be more easily compared, but they need not be representative for each country. This can be a severe drawback, as it will not reflect the building costs of the 'average' local building in each country. In the case of functionally similar buildings, the buildings are very representative for their respective countries, but since the buildings are now far more heterogeneous, comparability is sacrificed.[36)] Therefore the differences in results from these studies need to be interpreted with some care.

Table 2.1 shows comparative price levels [37)] of the construction sector for four European countries. The first column shows the results from the Eurostat comparison for 1993. The next column shows the result for industrial buildings only from studies using a similar methodology as Eurostat. Finally, column (3) depicts the average result for office, industrial and residential buildings from studies that compare prices of typical, non-identical buildings, which are claimed to be functionally similar. [38)]

Table 2.1
Relative price levels of total construction for several European countries, 1993 (Germany=100)

	Eurostat (1)	Standard [1)] building (2)	Similar [2)] building (3)
Netherlands	97	84	55
United Kingdom	64	71	55
France	77	86	66
Germany	100	100	100

1) Average of price comparisons of standard, identical buildings.
2) Average of price comparisons of typical, functionally similar buildings.

Sources: Column (1) from OECD, Purchasing Power Parities and Real Expenditures (1995); Columns (2) and (3) from PRC Bouwcentrum (1995).

The three different methods give widely differing results, although Germany is the most expensive of the four countries in all comparisons, whereas the United Kingdom always has the lowest comparative price level of the four countries. The differences are largest between the alternative comparisons that use prices of standard, non-representative buildings and the comparisons that employ prices of typical, non-identical buildings. Results for other years, which are not reproduced here, indicate the same, although the Eurostat estimates are usually higher for the Netherlands and the United Kingdom than the other estimates, including those using a similar methodology.

There has been quite some criticism on the methods employed by Eurostat (PRC Bouwcentrum, 1995; Vermande and van Mulligen, 1999). Eurostat asks local construction experts to fill out the 'Bills of Quantities', for which they are required to give or estimate local prices of certain building types. However, these standards buildings are not constructed in all countries concerned, so the experts have to make some informed guesses about the prices that are involved. There is a great likelihood that such estimates include large errors, in particular when – as is the case in the Eurostat comparison – there is only limited international coordination in comparing the estimates (Eurostat, 1999b).

The Eurostat procedure illustrates the difficulties in applying the matched model method in an international price comparison to a sector that is characterised by a large heterogeneity in the output. It is nearly impossible to compare 'like' with 'like', as the matched model method requires, so that imputation is inevitable. Such a price imputation is of course highly subjective, especially if there is no standard method for the imputation procedure.

It is doubtful whether the method of comparing the prices of functionally similar, but not identical buildings, is an option. The advantage of this procedure is that actual prices are compared, but since the compared buildings are usually rather different, some kind of quality adjustment would have to be made to obtain reliable estimates of relative price level. On the other hand, if the value derived from different constructs is comparable in different countries comparing their prices directly may do not much harm. Again we face a trade-off between characteristicity and identity; this question is related to the discussion of identical, common, and unique products discussed in the previous section.

Using hedonic regression analysis might provide such quality adjustments, but considering the large differences between buildings in different countries, it would be hard to specify a set of common characteristics, which are valued in every country. A more practical, but probably even larger problem concerns the huge data requirements for this method, to determine the relative importance of characteristics and their substitutability in different countries.

Another industry which is characterised by cross-country differences in quality of the output, but where quality adjustments for international price level comparisons may be possible, is the car industry, which will be the topic of a more detailed analysis in Chapter 5.

The computer industry

Price indices

The industry where the quality problem is probably most manifest is the computer industry. New and better computer models appear on the market at a fast rate, and the average lifecycle of a computer is only a few months. Set against this background, quality adjustments in price indices for computers are of utmost importance, to prevent serious quality biases from affecting the indices. As already pointed out in Chapter 1, Wyckoff (1995) and a Eurostat taskforce report on computer prices (Eurostat, 1999a), showed that there is a huge discrepancy between price indices for computers of different countries using different methods to deflate computer prices, especially between countries which use hedonic adjustments and countries which do not.

In the case of computers, hedonic deflators in the United States have been used more extensively – though still not comprehensively – than for most other products. Since the 1980s, several papers have been written on this subject in the United States.[39)] This method was adopted by the BEA for deflating computer output in the national accounts (Cartwright, 1986). The BLS introduced hedonic price indices for computers in the PPI and CPI in 1991 and 1998, respectively (Moulton, 2001).

Some countries (e.g., Canada) have followed the U.S. practice of using hedonic price indices for computers. Most European countries have only experimented with this method, and only France and according to a recent press release also

Germany, have very recently begun estimating their own hedonic price indices[40]. Some countries, for example Denmark and Sweden, implemented the U.S. computer price index into their own deflators, adjusted for exchange rate fluctuations. Most methods are very similar to those developed in the United States, and the results of these experiments are also in line with the U.S. results: the price indices for computers decrease substantially over time. However, these results still have an experimental character and indices are usually calculated for only a couple of years. Recently, Eurostat started a project to support statistical agencies of the European Union member countries in constructing hedonic (consumer) price indices for computers.[41]

As a result the computer component of the producer price indices suffers from lack of international comparability. As explained above, this point was extensively elaborated upon in the work of Wyckoff (1995). The same point can also be observed from table 2.2, which shows producer price indices for computer equipment and national accounts deflators for machinery and equipment (which includes computers) for Germany and the United States.[42]

The producer price index for computers is declining in both countries, albeit much faster in the United States than in Germany. The national accounts deflator for machinery and equipment in Germany begins to decline at a later point in time and to a lesser extent than in the United States. Apart from the greater use of hedonic indices in the U.S., the larger share of computer equipment in the production of machinery and equipment in the United States in current prices also accounts for part of the more rapid price decline. This suggests that in Germany the national accounts deflator for machinery and equipment is affected to a much smaller extent by changes in the producer price index for computer equipment.

Table 2.2
National Accounts deflators and Producer Price Indices for Industrial Machinery and Equipment and Computers in the United States and in Germany, 1970–2000 (1992=100)

	United States			Germany [1]		
	NA deflator Industrial Machinery and Equipment (SIC 35)	PPI Domestic Computer Equipment (SIC 3571–3577)	Percentage share in production of computer equipment in SIC 35 (current prices)	NA deflator Machinery and Equipment	PPI Office Machinery and Data-processing Equipment	Percentage share in production of computer equipment in Machiney and Equipment (current prices)
1970	59.5			30.7	161.2	
1975	78.3	1,583.0	8.9	47.3	151.4	
1980	110.1	499.7	14.7	57.9	118.0	6.1
1985	117.6	223.3	25.6	74.5	120.3	11.2
1990	102.6	129.7	23.2	93.1	103.0	7.7
1991	103.0	116.2	22.7	96.4	100.7	7.9
1992	100.0	100.0	24.0	100.0	100.0	6.3
1993	95.4	86.0	23.2	102.0	97.7	6.1
1994	92.0	77.7	23.4	100.7	94.4	5.8
1995	86.5	69.5	24.4	98.1	93.9	5.5
1996	81.2	58.4	25.6	105.8	90.3	5.8
1997	73.4	47.2	26.3	103.6	86.0	6.2
1998	65.8	37.5	26.4	104.2	81.0	7.6
1999	59.6	31.3	26.3	106.6	74.3	7.6
2000	57.7	27.8	28.8		70.8	8.8

[1] Germany refers to former West Germany.

Note: until 1987, Domestic computer equipment in the United States was not disaggregated into four industries, but was aggregated in one industry, 'Electronic computing equipment' (SIC 3573).

Sources:
United States: the national accounts deflator was obtained by dividing output in current prices by output in 1992 chained weights prices (from BEA, Survey of Current Business, various issues). The computer price indices are from Triplett (1996) and the online data base of BLS (http://www.bls.gov). The share of computer equipment in Industrial Machinery and Equipment was calculated by comparing their respective values of shipments in current prices, obtained from U.S. Department of Commerce (1990, 1995), U.S. Census of Manufactures for 1987 and 1992 and unpublished data from their online database (http://www.stat-usa.gov).
Germany: the national accounts deflator was obtained by dividing output in current prices by output in 1991 prices (Statistisches Bundesamt, Volkswirtschaftliche Gesamtrechnungen, various issues). The producer price index was obtained from the online website of the Statistisches Bundesamt (http://www.statistik-bund.de). The share of Data-Processing equipment within Machinery and Equipment was calculated by dividing their respective outputs in current prices (Statistisches Bundesamt, Produzierendes Gewerbe, various issues).

Implications for international productivity comparisons

Although the quality of computers and other IT-equipment changes very fast, international differences in quality changes of a comparable computer may be assumed to be quite small. The quality problem in matched model currency conversion factors (PPPs or UVRs) for computers is therefore probably smaller than in an intertemporal framework. Moreover, the world market for IT-equipment is very competitive, and computers are heavily traded goods. Therefore, purchasing power parities for computers may be not very different from exchange rates.

Usually, it is not possible to calculate currency conversion factors for each year, and price indices are used to update conversion factors extrapolating from benchmark years. But when the procedures to construct these price indices are not the same for the countries compared, as illustrated above, the updated conversion factors will be skewed. This is illustrated by the example below that compares PPP-converted productivity levels for non-electrical machinery and total manufacturing for former West Germany vis-à-vis the United States for 1987 and updated to 1992.

The first row in table 2.3 shows the relative levels of labour productivity in the machinery industry and total manufacturing between West Germany and the United States in 1987. These estimates are based on Van Ark and Pilat (1993), which are derived with the ICOP methodology. The ICOP estimates make use of output and labour input information from the manufacturing censuses and surveys in both countries. Output is converted to U.S. dollars on the basis of unit value ratios, including a quality adjusted UVR for computers, derived from the study by the McKinsey Global Institute (1993).

The regular procedure of updating benchmark comparisons of cross country productivity to later years is by using the real output and labour input series from the national accounts of the two countries. The results of the regular updating procedure from 1987 to 1992 is shown in row 7 of table 2.3. It shows a substantial widening of the productivity gap in non-electrical machinery between 1987 (when German labour productivity was 78.1 per cent of the US level) and 1992 (when the German level was only 64.9 per cent). This is mainly due to the much faster increase of the real value added index obtained from the U.S. National Income and Product Accounts, which includes a hedonic adjustment for computers (see row 3 in table 2.3). By way of experiment the hedonic deflator for computers was eliminated from the US real output index, using the weights shown in table 2.2. The real output index for the USA after this adjustment (row 4) shows a much slower growth, which is closer to the German real output index (row 2 in table 2.3). As a result, row 8 shows that the adjusted productivity level for Germany relative to the U.S. in 1992 is more than 8 percentage points higher than the unadjusted level. At the level of total manufacturing the impact of this experiment on the German/U.S. labour productivity gap is less than 2 percentage points.

Table 2.3
Labour Productivity in Non-electrical Machinery and Total Manufacturing, West Germany Relative to USA, 1987 and 1992, with and without adjustments for hedonic price indices of computers in the USA

		Non-electrical machinery	Total manufacturing
(1)	Value added per hour worked in 1987 (USA=100)[1]	78.1	79.2
	1992/1987 trends (1987=100)		
(2)	Real value added, Germany	109.9	113.6
	Real value added, USA		
(3)	- with hedonic index for computers	121.5	106.0
(4)	- without hedonic index for computers	107.9	103.8
	1992/1987 trends (1987=100)		
(5)	Total hours worked, Germany	97.6	100.1
(6)	Total hours worked, USA	106.2	103.8
	Value added per hour worked in 1992 (USA=100)		
	Extrapolated from 1987		
(7)	- with US hedonic index for computers	64.9	81.8
(8)	- without US hedonic index for computers	73.1	83.5

[1] UVR adjusted for quality of computers.

Sources:
row (1) from Gersbach and van Ark (1994), table 7;
row (2) from Statistisches Bundesamt, Volkswirtschaftliche Gesamtrechnungen, various issues;
row (3) from BEA, Survey of Current Business, various issues;
row (4) as row (3) but taking out price deflator for computers (from Triplett, 1996) using sales value weights from the US Census of Manufactures (see table 2.2), and replacing the computers deflator by the average deflator for the rest of the machinery sector;
row (5) from Statistisches Bundesamt, op. cit., and DIW, Produktionsvolumen und -potential;
row (6) from BEA, Survey of Current Business, various issues, and BLS, Employment and Earnings, various issues.

The latter reflects a more recent analysis by the BEA (Landefeld and Grimm, 2000) who, in response to claims by the German Bundesbank (2000) that the use of hedonic deflators in the U.S. creates a statistical discrepancy between U.S. and German growth rates, show that the impact of the U.S. hedonic price index for computers has a maximum impact of only 0.25 percentage point per year on the real GDP growth rate between 1995 and 1999.
Obviously the counterfactual experiment carried out above should be replaced by a more realistic procedure which includes a hedonic price index for German computers rather than take the one for the US out. Given the fact that the German statistical office recently started using hedonic techniques in their price indices for computers, this may well be possible in the near future. The effect on the comparative productivity levels might then be smaller given the smaller production share of the computer items (PCs) with the most rapid price declines in Germany relative to the United States. However, the present experiment serves to show the sensitivity of international productivity comparisons for the different treatment of computers in the real output series between countries.

2.7 *Summary*

Price comparisons over time and across countries are strongly affected by the statistical treatment of changes in product quality over time and differences in product quality across countries.

As discussed in this chapter, the matched model method (which compares 'like' with 'like') is not adequate to deal with substantial changes or differences in quality. Statistical agencies, the main producers of price indices, employ several methods to deal with relatively small quality changes, the most of which make an implicit assessment of the effect of a difference in quality on price changes. All these methods introduce a quality bias in the price index. This quality bias comes in two forms: the inside the sample bias and the outside the sample bias. The first type of bias occurs when prices of non-identical products are matched. Sometimes a quality adjustment is applied to such a match, but this adjustment is usually ad hoc and depending on the expertise of the price statistician. The second kind of bias occurs when the price changes of matched items are not representative of price changes of unmatched items. This bias is potentially very strong if the share of matched items is relatively low or decreasing, which is usually the case in a fixed base index or for products with a high market turnover.

An alternative to the matched model method is the hedonic method, which makes explicit quality adjustments based on the characteristics of products. This method was pioneered in the first half of the twentieth century, and further developed in the 1960s and 1980 in the United States. Since then, statistical agencies in the U.S. have employed it in several price statistics, and the hedonic method has become a very popular research topic for economists. Statistical agencies in Europe have been much slower in adopting the hedonic method. It was often dismissed because of lack of data or because the hedonic method is too different from current pricing practices. In recent years, statistical agencies in Europe have adapted their views somewhat and have tentatively started experimenting with the hedonic method and even introducing it in their official statistics.

The increased use of the hedonic method gave rise to what later became known as the 'user value – resource cost' debate. This debate was started by critics of the hedonic method, who stated that hedonic price indices and deflators overstated real growth in expenditure or output. According to these critics, price indices and deflators should be based on the costs of products, not on the utility they provide to users, whether these are final or intermediate users. The recent common opinion on this issue seems in line with the statements of Triplett (1983) and Gordon (1990), who put forward that both user value and resource cost are important considerations in price measurement. It is also the point of view that is held in this thesis.

Although most research on quality problems has focused on price indices across time, the quality issue is equally relevant in international price comparisons, and possibly even more. Both the inside the sample and the outside the sample problems are potentially bigger than for price changes. Where quality

changes through time are gradual, they are not in the cross country case. If product specifications are limited, the inside the sample looms large. In addition, the composition of expenditure and output is often widely different in different countries. This product-mix problem gives rise to a large outside the sample problem. In this chapter, the quality problem in international level comparisons has been illustrated with the case of construction. This is a sector with a strongly heterogeneous output, especially in an international context. Both inside the sample and outside the sample problems are large in this case.
But even for those products where reliable cross-country price matches can be made with negligible quality biases, the quality issue can still be relevant. The reason is that for existing benchmark comparisons, price indices or deflators are sometimes used to estimate relative price levels for other years. When the methodology to construct price indices and deflators differs across countries, the reliability of relative price levels which are adjusted with these price indices is seriously affected. This has been illustrated in this chapter for the computer industry. For this industry, there is a major difference in price indices depending on whether or not the hedonic method is used. The U.S. price index for computers, which is based on the hedonic technique, and non-electrical machinery as a whole shows a much more rapid decline than in Germany, where this technique has not been used.
The hedonic method can offer a solution for the quality problem in price indices and international comparisons, provided sufficient information on characteristics can be obtained. This provision also holds for the matched model method, as tight specifications are needed to ensure that prices of (near) identical products are matched.
In the following chapters, the focus is on a more detailed exposition of the hedonic method. Chapter 3 discusses its fundamental similarities and dissimilarities with matched model methods. Chapter 4 contains hedonic applications on price indices for computers in the Netherlands. In Chapter 5 the hedonic method is used to compare output unit value ratios for cars across six countries.

Notes

[1] In the remainder, I will refer to output, with similar implications for expenditure and income.

[2] Likewise, quantity relatives are usually weighted with prices to obtain quantity indices.

[3] The Törnqvist index is the geometric average of price ratios, weighted by their average value shares in both periods:

$$P_T^{01} = \prod_{i=1}^{N} \left(p_i^1 / p_i^0 \right)^{w_i}, \text{ with } \quad w_i = \tfrac{1}{2}\left(p_i^0 q_i^0 \Big/ \sum_{j=1}^{N} p_j^0 q_j^0 + p_i^1 q_i^1 \Big/ \sum_{j=1}^{N} p_j^1 q_j^1 \right)$$

[4] For a description of the difference between the Young and the Laspeyres index, see Chapter 15 of the ILO Manual of the CPI, which is currently in draft. Chapters can be downloaded from http://www.ilo.org/public/english/bureau/stat/guides/cpi/index.htm.

[5] Fisher (1922) already listed several hundreds of different index numbers eighty years ago.

[6] Later, Diewert (1992) developed more tests, and concluded that the Fisher index satisfies twenty 'reasonable' tests. However, he indicated that the selection of which tests to use remains subjective, and that it is not possible to single out one best index number formula with the test approach.

[7] See Diewert (1976), (1987) and (1993) for a concise treatment.

[8] In the Walsh index, price ratios are weighted by the geometric averages of the corresponding quantities:

$$P_W^{01} = \sum_{i=1}^{N} p_i^1 \left(q_i^0 q_i^1\right)^{1/2} \Big/ \sum_{i=1}^{N} p_i^0 \left(q_i^0 q_i^1\right)^{1/2}$$

[9] Sometimes, the following theoretical disadvantage of the chain principle compared to the fixed base principle is pointed out: if for two non-adjacent periods both prices and quantities are equal, than the chain principle will not necessarily lead to a price index of 1. However, it is questionable whether it is appropriate to make a direct match between two periods that are relatively remote in time, as in the time that has passed, many items will have entered and exited the market. See von der Lippe (2001) for a detailed critique of the chaining principle.

[10] This urge to 'know your product' is voiced again nearly a century later by Triplett (forthcoming) with respect to hedonic method, which is the topic of Chapter 3.

[11] It is interesting to note that von Hofsten remarked that at that time, the quality problem was most apparent in goods that declined in quality, but had rising prices, whereas in the present literature the reverse situation is deemed most relevant.

[12] A similar recommendation had been made by Mitchell in 1915, but at that time, a cost-of-living index did not yet have any welfare implications, but was mainly used for the indexation of wages. (Banzhaf, p. 347). Mitchell withdrew his recommendation for a cost-of living index in 1944 on the grounds that there was no satisfactory way to measure it (Banzhaf, p. 356).

[13] This section is to a large extent based on Triplett (forthcoming).

[14] This does not necessarily mean that index numbers are compiled for individual outlets or outlet types. The point of selling is merely used as a controlling variable.

[15] See Boon (1998) for an overview of sampling designs employed by statistical agencies in the European Union and the U.S.

[16] Hence this is also called specification pricing (Gordon, 1990).

[17] The relative lack of detail in the specification of a basic heading is especially apparent in international comparisons, as discussed below.

[18] The weight for this basic heading equals 0.00032, with base year 1995.

[19] In international comparisons, the 'a potato is a potato' rule is usually employed in this respect. This means that the point of purchase is not considered to be a quality determining factor (see United Nations, 1992).

[20] This method is sometimes also referred to as the *class mean method*, as the price index of goods in the same class are imputed for the missing price ratio.

[21] Triplett (1990) presents examples of such quality adjustments made to the price index of cars in the U.S.

[22] Dalén (2002) provides the example of the Austrian CPI for computers.

[23] Although this is Griliches' most famous application of the hedonic method, it was not his first. A few years earlier, he started with hedonic regression analysis of fertiliser prices based on their chemical contents (Griliches, 1958, 1959).

[24] A notion that is familiar to, for example, computer engineers, as illustrated by Triplett (forthcoming).

[25] Moulton (2001) provides an overview of the hedonic price indices currently in use by several U.S. government agencies. Currently, the CPI includes hedonic price indices for computers, audio equipment, camcorders, college textbooks, clothes washers and dryers, DVD players, microwave ovens and refrigerators.

[26] A unit value ratio is a variant of the purchasing power parity.

[27] Gilbert and Kravis (1954), Gilbert and Associates (1958).

[28] Whereas the first ICP round for 1975, around 150 basic headings were used, this number had grown to about 500 for 1982. For the current OECD comparisons, this number is much lower, namely around 150 for consumption expenditures and around 240 for total GDP. Since the 1990s Eurostat carried out PPP studies on an annual basis, with a three-year rolling survey for consumer expenditures.

[29] Especially differences in outlet type and delivery conditions can have major implications for price comparisons between rich and poor countries, as prices of 'identical' products in rich countries generally have a larger intermediate product and consumer amenities component. For a detailed analysis of this point, see Usher (1974).

[30] See, for example, van Ark (1996).

[31] See, for example, Maddison and van Ark (1988), van Ark (1993), Pilat (1994) and Timmer (2000).

[32] See also Gersbach and van Ark (1994).

[33] According to a recent press release, the German Federal statistical office started using hedonic methods in their price indices in 2002. "Hedonic quality adjustments applied for the first time in price statistics", press release of 11 July 2002, www.destatis.de/presse/english/pm2002/p2450051.htm. See also Linz and Eckert (2002).

[34] Most of this section is based on Vermande and van Mulligen (1999).

[35] See Vermande and van Mulligen (1999) for an overview.

[36] In such cases, the assumption is made that the services provided by the building, and therefore its value to users, are the same across countries.

[37] Relative price levels are equal to the PPP divided by the exchange rate.

[38] A more detailed discussion of the studies used for columns (2) and (3) of table 2.1 can be found in PRC Bouwcentrum (1995).

[39] See for example Cole *et al*. (1986), Dulberger (1989) and Berndt and Rappaport (2001).

[40] See Lequiller (2001), Scherrer (2001) and Evans (2002) for the work on France. See Ball and Mehmi (2002) for the most recent work on the United Kingdom and Chapter 4 of this thesis for the Netherlands.

[41] Triplett (2000) and Konijn, Moch and Dalén (2002).

[42] The German data actually refer to data processing equipment, which is the closest classification to computers in the German statistics, although it is somewhat broader than that.

3. The hedonic method

3.1 Introduction

When estimating price changes of heterogeneous goods that are subject to (continuous) quality change, the standard matched model methods usually do not offer a satisfying solution. To be able to construct a price index that takes into account quality changes more systematically, the hedonic method is used. This method was briefly introduced in the previous chapter.

In this chapter the hedonic method is discussed in more detail. Section 3.2 starts with the general theory set up by Rosen (1974), followed by the application of exact price indices on the hedonic method by Feenstra (1995). The estimation of hedonic functions will be discussed in Section 3.3. Issues like the choice of characteristics and the functional form will also be addressed. Section 3.4 describes several methods to derive price indices from the hedonic functions, together with their advantages and drawbacks. Several applications of the hedonic method for price indices of different types of durable goods and construction works are discussed in Section 3.5. Section 3.6 discusses how the hedonic method has been both adopted and criticised by statistical offices. Criticism has not only arisen from statistical agencies, but recently also from economists, on the grounds that given the right circumstances, hedonic indices are unnecessary, and matched model indices will perform as well. The discussion between proponents of this view and defendants of the hedonic method will be outlined in Section 3.7. Although the majority of applications have been in the area of intertemporal price indices, it has sometimes been used in international price comparisons. Section 3.8 discusses some of the major research in the latter area. Section 3.9 summarises.

3.2 Theoretical considerations

The theory of hedonic functions rests on the hedonic hypothesis, which was introduced in Section 2.4: "heterogeneous goods are aggregations of characteristics, and economic behaviour relates to the characteristics" (Triplett, 1987). Characteristics of products are thus at the core in the theory of hedonics, and a hedonic function describes the relation between the price p of an item and the characteristics z_k ($k = 1, ..., K$) contained in this item. Several theories have been developed to provide a theoretical foundation of the hedonic method. Most of these were consumer theory-oriented, but Rosen (1974) established a general equilibrium framework encompassing both consumer utility and production. Since the characteristics of heterogeneous goods are treated as commodities, standard utility and profit functions are extended with characteristics next to homogeneous goods.

A consumer j maximises his utility U in the following way: [1)]

$$\max U_j(q(z) \mid \alpha_j)$$
$$\text{subject to } y = p(z) \qquad (3.1)$$

Where y is total disposable income, $p(z)$ is the hedonic function, and α_j is a vector of the features of the individual consumer which describes his tastes. The goods are aggregated over their characteristics by $q(z)$. The utility function U is unique for each consumer, so the willingness to pay for different values of z is different for every consumer. This willingness to pay can be expressed in a consumer's bid function. If there is a sufficient number of consumers, then the tangencies between the bid functions and the hedonic function $p(z)$ trace out an envelope curve, which defines consumer equilibrium (Rosen, 1974). If consumers have identical tastes, then $p(z)$ identifies the consumer bid functions. [2)]

On the supply side, individual producers have different offer functions, which are derived by the following profit (π) maximising problem:

$$\max \pi = p(z)M(z) - C(M,z) \qquad (3.2)$$

where $M(z)$ is the output of embodied characteristics z by a firm, and $C(M,z)$ the costs it faces. Like on the demand side, the tangent points of the individual offer functions, derived from (3.2) and the hedonic function $p(z)$ span an envelope that is the supply side equilibrium.

The market equilibrium is the collection of points where bid functions of individual consumers and offer functions of individual producers intersect, which are defined by the hedonic function $p(z)$. This function is the market clearing implicit price function. This of course implies that the same characteristics z are relevant to both consumers and producers, i.e., z is associated with both resource costs and user value. Therefore, the selection of which characteristics to include in hedonic regressions is of crucial importance. Also, because the hedonic function is brought about by the market equilibrium, it cannot be used to identify either the demand or supply functions.

Since the bid and offer functions are not necessarily linear, hedonic functions can take on any form in theory. As argued by Triplett (forthcoming), it is therefore incorrect to dismiss any functional form beforehand, which makes it necessary to carry out tests (like the Box-Cox test) to determine which functional form most adequately represents the data at hand.

Although the characteristics of a good are generally not priced separately, the price of a good encompasses the valuation of all the characteristics that together comprise the product, so each characteristic needs to be valued by its 'implicit' price. The hedonic function is then interpreted as a function that disaggregates the price of a good into the implicit prices and the quantities of the characteristics, and it provides estimates of prices for these characteristics. It is because of the fact that the prices for the characteristics must be estimated, these prices are usually termed implicit prices, since they generally cannot be observed themselves.

This method assumes that the characteristics are perfectly separable. Interaction effects between characteristics cannot be handled. In practice, this will not always be the case, which is a reason for Gordon (1990) to prefer a specification for the hedonic function that allows for interaction effects.
When comparing implicit prices with ordinary prices, one can see that there are both differences and similarities between them. Apart from the fact that implicit prices cannot be observed directly, the relations among the prices of characteristics are also more complex than between prices of separate products, because the characteristics are bundled into one product. But there also important similarities between ordinary and implicit prices. First, both measures are an indication of what the seller receives and what the buyer pays irrespective of looking at a set of characteristics or a good. Second, provided that there is competition on the markets, both price index measures are proportional to marginal utilities and costs.

Exact hedonic indices

Rather than combining the utility and profit maximising problems faced by consumers and producers like Rosen does, Feenstra (1995) uses a cost-of-living approach. As discussed in Chapter 2, this approach yields exact price indices. A hedonic exact price index not only takes into account changing prices, but also changing product characteristics. As argued by Lancaster (1971), characteristics rather than heterogeneous goods determine utility and therefore social welfare, so hedonic price indices may provide a better approximation to the true measure of consumer welfare.
Since goods are heterogeneous in their characteristics, preferences of consumers are usually thought to be heterogeneous as well. However, to apply the concept of exact price indices, a single utility function of a representative consumer is needed. Feenstra aggregates individual behaviour to some representative consumer, by assuming that the separate utility functions are identical across consumers with respect to characteristics except for an additive random term. He compares several forms of the hedonic function: a linear specification (equation 3.3), and a log-linear specification (equation 3.4): [3)]

$$p_i^t = \alpha_i^t + \sum_{k=1}^{K} \beta_k^t z_{ik}^t + \varepsilon_i^t \tag{3.3}$$

$$\ln p_i^t = \alpha_i^t + \sum_{k=1}^{K} \beta_k^t z_{ik}^t + \varepsilon_i^t \tag{3.4}$$

where α is the intercept, β_k the regression coefficient for characteristics z_k (k=1, ...,K), and the disturbance terms ε_i^t are independent and identically distributed (i.i.d.). The left hand side of the above equations shows price p of product i (i=1,...,N). Both equations specify a hedonic function in period t.

In this period, the representative consumer faces expenditure function $E(p^t, z^t, U^t)$, with p, z and U again defined as the prices, characteristics, and the utility level of the representative consumer. An exact hedonic price index P_h is defined by the ratio:

$$P_h = \frac{E(p^t, z^t, U^{t-1})}{E(p^{t-1}, z^{t-1}, U^{t-1})} \tag{3.5}$$

If the relation between prices and characteristics is as described by (3.3), Feenstra (Proposition 6) derives the following upper and lower bounds of P_h:

$$\frac{\sum_{i=1}^{N} p_i^t q_i^t}{\sum_{i=1}^{N} \hat{p}_i^{t-1} q_i^t} \le P_h \le \frac{\sum_{i=1}^{N} \hat{p}_i^t q_i^{t-1}}{\sum_{i=1}^{N} p_i^{t-1} q_i^{t-1}} \tag{3.6a}$$

where

$$\hat{p}_i^{t-1} \equiv p_i^{t-1} + \left[\sum_{k=1}^{K} \beta_k^{t-1} (z_{ik}^t - z_{ik}^{t-1}) \right] \tag{3.6b}$$

and

$$\hat{p}_i^t \equiv p_i^t + \left[\sum_{k=1}^{K} \beta_k^t (z_{ik}^{t-1} - z_{ik}^t) \right] \tag{3.6c}$$

The lower and upper bounds in (3.6a) are a Paasche and a Laspeyres index, respectively. The difference with conventional Laspeyres and Paasche index numbers is the explicit quality adjustment in the indices in (3.6a), which are spelled out in equations (3.6b) and (3.6c).

If the hedonic function has a log-linear form like (3.4), these bounds are given by (Feenstra, Proposition 7):

$$\prod_{i=1}^{N} \left(\frac{p_i^t}{\hat{p}_i^{t-1}} \right)^{s_i^t} \le P_h \le \prod_{i=1}^{N} \left(\frac{\hat{p}_i^t}{p_i^{t-1}} \right)^{s_i^{t-1}} \tag{3.7a}$$

where

$$\hat{p}_i^{t-1} \equiv p_i^{t-1} \exp\left[\sum_{k=1}^{K} \beta_k^{t-1} (z_{ik}^t - z_{ik}^{t-1}) \right] \tag{3.7b}$$

and

$$\hat{p}_i^t \equiv p_i^t \exp\left[\sum_{k=1}^{K} \beta_k^t (z_{ik}^{t-1} - z_{ik}^t) \right] \tag{3.7c}$$

The weights s_i^t and s_i^{t-1} in equation (3.7a) are expenditure shares of product *i* in period *t* and period *t*-1, respectively. The geometric averages of the upper and lower bounds in (3.6a) and (3.7a) correspond with Fisher and Törnqvist indices, respectively. Since Fisher and Törnqvist indices are superlative indices, the geometric averages of the upper and lower bounds of (3.6a) and (3.7a) are considered superlative exact hedonic indices (SEHI). If the estimated coefficients of the characteristics (the β's) are the same in both periods, than the upper and lower bounds only differ in the weighting systems.

A standard assumption in the theory of hedonic indices is that there is perfect competition on product markets, and therefore the coefficients, or implicit prices, reflect both user valuations and production costs. However, when competition is imperfect, user value and resource cost do not coincide. This issue was already touched upon in Chapter 2, and was at the basis of the user value vs. resource cost debate which was discussed there. In the case of imperfect competition, producers price their products above marginal costs which results in price mark-ups. User value is still reflected in the implicit prices, but because of the presence of mark-ups, the implicit prices give no clear indication of producer cost. Mark-ups are not observable, and have two effects. Firstly, the bias in the estimates of the coefficients β_k of the characteristics depends on the correlation between mark-ups and characteristics. Secondly mark-ups also determine the difference between marginal costs and the value of characteristics (Feenstra, 1995, p. 647).

Given the specification of individual utility functions and quality-adjusted prices to derive the upper and lower bounds for the SEHI in (3.6a) and (3.7a), Feenstra (Proposition 8) proves that if marginal costs are linear in characteristics, the two mark-up effects mentioned above may offset each other, and price mark-ups do not introduce a bias in the estimates of the implicit prices. But a log-linear specification for marginal costs appears to introduce an upward bias in the value of characteristics from the hedonic regressions, and therefore a downward bias in a log-linear hedonic index when quality is improving (i.e., characteristics are increasing). In this case a linear hedonic function is therefore to be preferred to a log-linear functional form. To test whether there is a bias due to imperfect competition, the superlative indices using the upper and lower bounds of (3.6a) and (3.7a) can be compared. If the Fisher index exceeds the Törnqvist index, then we might expect that there is a downward bias in the log-linear index caused by non-competitive pricing. If the indices are fairly close to each other, then the possible bias is likely not very important.

This practical advantage of linear functions is counteracted by recommendations by Diewert (2001), who argues against linear hedonic functions on theoretical grounds. Diewert restricted the theory developed by Rosen by making two simplifying assumptions. First, each consumer faces the same hedonic utility function. Second, this hedonic utility function is assumed to be separable from other goods. A consumer can trade off between characteristics of a heterogeneous good, independently of his choice of other commodities. On the basis of these assumptions, Diewert dismisses linear hedonic functions, since he finds that in this case, prices are not homogeneous of degree one, which is a standard property in microeconomic theory. Furthermore, Feenstra's preference for a linear specification rests on the case that the mark-up effect *may* offset each other, but in practice this may not be the case at all.

In conclusion, depending on the assumptions that are used, there are some clear advantages as well as drawbacks associated with both the linear and the logarithmic specification. The double logarithmic form has the practical drawback that it is not possible to model characteristics which can have a value equal to zero. A disadvantage of both the linear and the semi-logarithmic specification is that neither allows for multiplicative interactive effects between characteristics, which is sometimes untenable. Another sometimes forgotten drawback of the linear specification is that it does not rule out negative predicted prices.

3.3 *Estimating hedonic functions*

In estimating a hedonic function, there are two issues of major concern, namely the choice of variables and the choice of the functional form. Concerning the characteristics, one has to decide whether these are economically meaningful. Ideally, the characteristics represent what a consumer is interested in when buying the product (user value) and what resources are needed in production (resource cost). As described above, user value and resource costs are very likely not to be identical, mainly due to imperfect competition, which introduces mark-ups in prices. Most characteristics that represent user value also have an associated resource cost, although there are exceptions. For example, the introduction of the second generation birth control pill represented a huge increase in user value with hardly a rise in resource cost. A well known example of a feature that uses resources, but (hardly) brings any user value, is the presence of a catalytic converter in a car. Even though it may bring future discounted utility to some consumers due to reduced emissions of toxic gases, it was primarily introduced because of government regulations (Gordon, 1990).

How should one decide which characteristics to use? Gordon (1990) and Pakes (2002) propose the criterion of user value to decide which characteristics to include. As prices are generally set by producers, and are a reflection of producer cost plus a possible mark-up, resource cost indirectly determine the estimated coefficients of the characteristics.

It is not always straightforward to select characteristics on the basis of user value. Even when it looks obvious that a certain characteristic has user value, it may still be no more than a proxy for what one actually wants to measure. The problem of a proxy variable is that, if it is disproportional to the actual characteristic, quality-adjusted prices will be biased. A good example of a proxy variable is the processor speed in the case of computers. When buying a computer, a consumer is interested in its performance, of which computational speed is one aspect. Computational speed is necessary to be able to run ever more demanding graphics and software. The speed of the processor is only a proxy variable for the computational speed of a computer. There are other factors also influencing computational speed, like size of the working memory, the type of the processor and physical characteristics which are unobservable (at least to the statistician). Arguably, processor speed is a rather *close* proxy for computational speed, but it remains a proxy variable. In fact computational speed or the performance of a computer are proxy variables as well, since what ultimately matters to a consumer is how well the computer is equipped to run particular software applications.[4)] It can sometimes be difficult to determine to what extent a proxy variable is 'close enough' to something that actually represents user value. For this reason, Triplett (forthcoming) argues the importance to 'know your product' for which a hedonic function is estimated.

Even if all characteristics that are included in the hedonic function reflect user value, there still will be other variables which influence prices. Pakes (2002) states that hedonic coefficients cannot be interpreted as the willingness of users to pay for the associated characteristics, since the implicit prices of the characteristics are also determined by market conditions, which are reflected in production costs. Coefficients can therefore take on values which seem strange from a user value point of view.

Pakes cites the example of Cockburn and Anis (1998) who study a certain kind of drugs. Simply stated, there are two variations of a drug, one without (type A) and one with (type B) negative side effects. Type B is used by a small number of patients for whom the type A drug is less effective. Since the mark-up on the type A drug is higher, more drug companies will manufacture versions of it, pushing down the mark-up and price of this drug. The market for the type B drug is not large enough to be profitable for more than one firm, which will set a monopoly price. Therefore, the mark-ups for the type B drug will be higher, resulting in a strong positive coefficient for the negative side effects in the hedonic regression. Although this is in line with economic theory, it is not what one would expect from a user value point of view, since side effects of medicines are judged negatively. Since mark-ups tend to change over time, Pakes argues that hedonic functions should be re-estimated frequently.

A good indicator of the presence of a mark-up is a brand name. Brand names generally also hint at the presence of characteristics which are not easily picked up by the statistician, but are valued by users, like design, vintage, reputation, etc. If mark-ups or non-observable characteristics are disproportionally related to the characteristics that are included, the coefficients of the included characteristics will be biased. In hedonic regressions, brand names are often used as proxies of unobserved characteristics. Coefficients of brand names tend to be positively related with prices, which might indicate that consumers value branded items more than non-branded ones (or items with inferior brands). But the possible introduction of a bias when brand names are included suggests that the statistician should condition on the brand of the product by other means. Estimating separate hedonic functions for different brands may provide a solution, but only when there is no variation in mark-ups and in the relation between observed and unobserved characteris-

tics within individual brands. Ignoring brands altogether is no option, as they generally provide the only option to condition for unobserved variables.
A comparable problem is associated with the point of purchase. Different outlet types offer different kinds of services, such as warranty, opening hours, after-sales service and other amenities. These amenities obviously contain user value, and influence prices, but are not directly related to the product in a physical sense. Therefore, the hedonic function should also, if possible, be conditioned on the outlet where individual items are purchased.
As was illustrated in the previous section, the choice of the functional form is not undisputed in hedonic theory. Triplett (forthcoming) and Pakes (2002) state that empirical considerations should prevail above the theoretical ones. Essentially the data at hand should determine which functional form is the most appropriate. This can be done using the procedures described by Box and Cox (1964) or Davidson and MacKinnon (1981), or by simply comparing mean square errors. In most research, the semi-logarithmic specification of the hedonic function has been the most popular, mainly because of the ease by which hedonic price indices can be derived with this specification.

Summary

The several theoretical problems associated with the hedonic method can be summarized as follows:
1. *Characteristics*. Characteristics should represent both user value and resource cost. In addition, they should be relevant in that they possess user value by themselves, and are not just proxies for something consumers are really interested in.
2. *Excluded variables*. Characteristics which can somehow not be observed, bias the regression results if they are disproportionate to the included variables.
3. *Other price determining variables*. Price mark-ups can bias the hedonic function and make frequent re-estimation of the hedonic function necessary. Mark-ups may also be associated with brand labels which also bias the results. Brand labels and point of sale have obvious effects on prices, but need not necessarily be associated with physical characteristics of a commodity. Ideally hedonic functions should therefore be conditional on these price determining variables in some other way.
4. *Functional form*. There does not seem to be a superior alternative to the hedonic functional form used, although various authors support a specific form based on different theoretical assumptions.

In addition to these theoretical problems, several econometric and practical ones can be noted (Gordon, 1990):
1. *New features*. Sometimes, a new variety of a product contains a completely new characteristic (for example, a DVD player for computers). It is not possible to estimate coefficient for such a characteristics in previous periods;

for each new characteristic, a coefficient β will lack in equations (3.6b) and (3.6c). In such cases, the quality-adjusted prices in equations (3.6b) and (3.6c) will not take into account this new characteristic.[5] A solution is to pool the data of different periods, and include time dummies in the regression.

2. *Multicollinearity*. There can be a relation between the characteristics included in the hedonic function. This is a very common phenomenon, since characteristics tend to increase jointly over time. New computers generally not only have faster processors, but also more hard disk capacity. This need not be a problem if there is no technical relationship between the characteristics, as is the case in the computer example. However, if characteristics are technically related, then this is more problematic. In that case the characteristics will be proxy variables for the same feature that possesses user value. In other cases it may distort regression coefficients. This is the case, for example, with fuel efficiency and engine power. More powerful cars have lower fuel efficiency and vice versa. Although fuel efficiency is a desirable feature for consumers, its coefficient is invariably negative when characteristics on engine performances are included.
3. *Small quantities*. As noted by Griliches (1961, 1971), products which serve only a small part of the market can have disproportionate effects on the regression results, especially if these are unusually cheap or expensive products. Recently, is has been argued[6] that the importance of observations should be reflected in the regression. This prevents items with a low number of sales and deviating values of characteristics from having an unduly influence on the regression results. Carrying out weighted least squares (WLS) rather than ordinary least squares (OLS) seems to provide a rather satisfying solution. Both sales value and quantity sold can be used as weights. Note, however, that this is not equivalent to using weights in price indices, like in Laspeyres and Paasche indices. Applying WLS mainly serves to achieve better estimates of the coefficients, not to achieve weighted indices. Berndt (1991) also argues in favour of weighted least squares regressions, by pointing out that ordinary least squares regressions involve heteroskedasticity.
4. *Accessories*. Sometimes option prices are used for certain accessories. These accessories are mostly removed from the hedonic function, and their option prices are subtracted from the product price. However, the accessory may also affect other characteristics, for which usually no adjustment is made. This is especially relevant for studies that computed hedonic price indices for cars, in which weight was included as a variable. When accessories are removed from a car in the regression, its weight needs to be adjusted accordingly.

In summary, attention has to be paid to the right form of the hedonic function (linear or log-linear), the estimation method (OLS or weighted least squares) and above all the choice of variables. It does not suffice to just plug in the variables on which one happens to have data available. Since the data requirements of the hedonic method are huge, the data will generally pose the most restrictions on the hedonic function.

3.4 *Constructing hedonic price indices*

Having estimated a hedonic function, one can calculate price indices in the following four ways (Triplett, 1986, forthcoming):

1. Estimate the price index directly from the regression, also called the *dummy method*. In this way a price index is estimated from a regression using time series, and dummy variables are included for the different periods. The regression coefficients of these time dummies represent the residual price change that cannot be attributed to a change in the quantities of the characteristics, or the change in 'quality'. This method is the one that is used most, in particular by scholars in academia and outside statistical agencies.
2. *Impute a missing price* for a model that does not exist in the period under consideration, but exists in some other period. In this way it is possible to construct a price index with matched model methods, where actual and fitted prices of the same item are matched. This method is mostly applied by statistical agencies.
3. Make an explicit quality adjustment, also called *hedonic quality adjustment*. This is the case, for example, when an old model in the previous period is replaced by a new one in the current period. For each characteristic, the differences in the values, or 'quantities' of characteristics embodied in the old and in the new model is multiplied by its implicit price, to calculate a price adjustment for each characteristic, so that the price index can be adjusted for price changes associated with quality increases.
4. Calculate a '*characteristic price index*'. Since with the hedonic function estimates of implicit prices are calculated, one can compute price index numbers that are defined on the prices and 'quantities' of the characteristics. This method is rarely used by both researchers and statisticians, and will not be discussed further.[7] As a side note, Triplett (1986) notes that this method is conceptually equivalent to the imputation method as long as the correct characteristics are used.

Before turning to a technical discussion of the first three methods, a hypothetical example is introduced that may help to illustrate the difference between the methods presented below.

	Period t-1	Period t	Period t+1
Model #1	p_1^{t-1}		
Model #2	p_2^{t-1}	p_2^{t}	
Model #3	p_3^{t-1}	p_3^{t}	p_3^{t+1}
Model #4		p_4^{t}	p_4^{t+1}
Model #5			p_5^{t+1}

To obtain a price index on the basis of this example, the methods introduced above work as follows:

The dummy method. This method estimates the hedonic index directly from a regression equation. Prices of all items, matched and unmatched, and for all periods are used to determine the regression coefficients. No individual ratios of matched prices between models #2, #3 and #4 are included in the index. Implicitly, all actual prices in all periods are replaced with estimated prices. Price ratios of individual items therefore become equal to the overall index. This is even true for unmatched items like model #1 and #5 at time *t*: the ratio of their estimated 'missing' price and estimated value of their actual price is equal to the overall dummy index as well.

The imputation method. This method uses normal Laspeyres and Paasche formulas for items that can be matched between periods, and extends them with estimated prices for unmatched items. For each period, a separate regression equation is estimated. If we want to calculate a Laspeyres index from periods *t*-1 to period *t*, all models from period *t*-1 (models #1, #2 and #3) are considered. The prices of models #2 and #3 are simply matched. The missing price of model #1 in period *t* is estimated with the regression for period *t*, and this price is matched with the observed price from period *t*-1. The Laspeyres index is then an average of the price ratios of both periods for all three models, weighted by their value shares in period *t*-1. The price ratios for models #2 and #3 are actual price ratios, the one for model #1 is estimated.
The procedure for the quality-adjusted Paasche index is similar. It considers all models from period *t*, which includes model #4, but excludes model #1. Again, the prices of models #2 and #3 are matched. For model #4, the price in period *t*-1 is estimated with the hedonic regression for that period, and this price is matched with the observed price in period *t*. The ratio of these prices is taken as the price change of model #4. The Paasche index is an average of the price ratios for models #2, #3 and #4, weighted by their sales value in period *t*. Again, the price ratios for models #2 and #3 are actual price ratios, whereas the ratio for model #4 is estimated.
Naturally the procedure is then repeated to obtain a price index between period *t* and *t*+1. Both indices can then be chained to obtain a price index from period *t*-1 to period *t*+1.

Hedonic quality adjustment. This approach is rather similar to the imputation method. The difference is that no missing prices are estimated, but rather that the prices of the items for which there is no match in the other period are directly compared, and an explicit quality adjustment is made for the differences in their characteristics. Hence instead of directly estimating the missing prices for models #1 and #4 between periods *t*-1 and *t*, a match between model #1 is period *t* and model #4 is period *t*-1 made with the hedonic quality adjustment method. In this way, again missing prices are imputed, but the imputation is done on the basis of a product match.

What all these variants of the hedonic method have in common and what makes them different from the traditional matched model method is that they include information on new and disappearing items between two periods. The conventional method would only consider the price ratios of models #2 and #3 between periods t-1 and t, and of models #3 and #4 between periods t and t+1. Hence whatever hedonic method is used, it makes greater use of existing data than the matched model method, which is a primary reason to pursue the application of this methodology.

For the sake of simplicity, the semi-logarithmic specification will be used as an example in the technical discussion of the three methods below.

The dummy method

When applying the dummy method, the data for T+1 different periods (t = 0,...,T) are pooled into one regression equation, and dummies are added for these periods to (3.4):

$$\ln p_i^t = \alpha_i^t + \sum_{k=1}^{K} \beta_k^t z_{ik}^t + \sum_{t=1}^{T} \gamma^t T^t + \varepsilon_i^t \tag{3.8}$$

Dummy coefficients T^t with associated regression coefficients γ^t are added for all periods except the reference period (t=0). If one wants to include a dummy for the reference period, the intercept a^t has to be omitted from (3.8). In the hedonic literature, these pooled regressions are always equivalent to a fixed effects model, which assumes that differences across units in different periods can be captured in differences in the intercept. This is a reasonable approach if we can be confident that the differences between units in different periods can be viewed as parametric shifts of the regression function. In the hedonic literature, where most applications are on intertemporal price indices, this assumption is usually implicit and rarely tested. In this thesis, this assumption will be tested in Chapter 5, where the differences between countries are suspected to be more than just shifts of the regressions. In Chapter 4, pooling data across periods is enforced by the data, and no such tests are performed.

The price index for year t is estimated directly from the regression equation (3.8) by taking the antilog of the estimated dummy coefficient γ^t.[8)] The rationale for this is as follows. Suppose that an item appears in two periods, t=0 and t=1. As no dummy is included for period 0, the predicted price of this item in period 0 would be:

$$\ln \hat{p}^0 = \alpha + \sum_{k=1}^{K} \hat{\beta}_k z_k \tag{3.9}$$

and its estimated price in period 1 would be:

$$\ln \hat{p}^1 = \alpha + \sum_{k=1}^{K} \hat{\beta}_k z_k + \hat{\gamma}^1 \tag{3.10}$$

Where $\hat{\gamma}^1$ is the estimated coefficient of the dummy variable for period 1.

Subtracting (3.9) from (3.10), we get:

$$\ln \hat{p}^1 - \ln \hat{p}^0 = \hat{\gamma}^1 \tag{3.11}$$

or

$$\frac{\hat{p}^1}{\hat{p}^0} = e^{\hat{\gamma}^1} \tag{3.12}$$

The main advantage of this method is that it is relatively easy to apply. The two main disadvantages are first, the assumption that the β's remain constant over time, and second, that it gives equal weight to all observations in the index, regardless of the relative importance of each item in total production or expenditure. In statistical practice, weights are generally absent at the level of individual prices. One could argue that the second drawback of the dummy method does not hold anymore, since weights are not available anyway. But even if no quantity weights are available, estimating separate regressions for individual periods still seems preferable. Changing quantities reflect substitution behaviour by buyers, which may reflect shifting preferences. If no quantity weights are available, estimating hedonic regressions for separate periods may provide the only option to represent a change in tastes between these periods. Of course, this implies that hedonic functions provide adequate measures of buyers' preferences, and that the effects of other factors on prices, like mark-ups, are small.

A third, more practical drawback of the dummy method is its lack of transparency. One cannot 'see' what the dummy method does; it is like a black box, where one fills in data and gets a result. The resulting index is not a Laspeyres-type or Paasche-type, but a 'fixed-weight' index, usually called 'regression index'. This is why the dummy method has proven unpopular with statistical agencies, who wish to keep track of what happens in the price index. Another reason why this method is less suitable for the compilation of official price indices, is that time dummies change when another (most recent) period is added to the regression. This problem can be circumvented by pooling data only for consecutive periods, but this would only be possible if the official price indices are based on chain weight indexes, rather than Laspeyres based indexes which is usual practice in price index numbers.

Recently, however, the dummy method has regained some popularity.[9)] One reason is the ease with which it can be applied. Another reason is that the results from the dummy method are actually not all that different from a matched model index. If there is a perfect match between two periods, than the hedonic dummy index will be *exactly* equal to an unweighted geometric matched model index, regardless of which characteristics are used in the regression (Triplett, forthcoming).

Imputing missing prices

The imputation method is a less radical application of the hedonic pricing methodology as price changes of matched products are retained. Only the prices of items that exist in one period only are estimated with the hedonic regressions in the periods where they are not available. This way, 'hedonic imputed' indices consist of two parts: one with actual price ratios representing *m* matched items and one with estimated prices representing unmatched items.
If we have M items, a hedonic imputed Laspeyres index is then calculated as follows:

$$P_L = \frac{\sum_{i=1}^{m} p_i^1 q_i^0 + \sum_{i=m+1}^{M} \hat{p}_i^1 q_i^0}{\sum_{i=1}^{M} p_i^0 q_i^0} \tag{3.13a}$$

which can be rewritten as:

$$P_L = \sum_{i=1}^{m} w_i \frac{p_i^1}{p_i^0} + \sum_{i=m+1}^{M} v_i \frac{\hat{p}_i^1}{p_i^0} \tag{3.13b}$$

with

$$w_i = \frac{p_i^0 q_i^0}{\sum_{i=1}^{M} p_i^0 q_i^0} \tag{3.14a}$$

for the *m* matched items, and

$$v_i = \frac{p_i^0 q_i^0}{\sum_{i=1}^{M} p_i^0 q_i^0} \tag{3.14b}$$

for the *M-m* unmatched items.

That is, in period 0 there are *M* different items of the particular commodity, of which *m* items are available in period 1; this is the first, matched part on the right-hand side of equation (3.13b). These items are unchanged, and the prices of such an item *i* in each period are p_i^0 and p_i^1, respectively. The other *M–m* items have disappeared in period 1, so their prices in period 1 need to be estimated; this is the second, unmatched part on the right-hand side of equation (3.13b). The estimated price of a disappeared item *i* in period 1 is $\hat{p}_i^1$. This price is imputed with the coefficients of the hedonic regression of period 1. Alternatively, one could use the estimates of a pooled regression, if for some reason one decides against estimating separate regressions for different periods. All price ratios are weighted with the relative expenditure shares w_i and v_i in the base period.

The hedonic imputed Paasche index has a similar format as the Laspeyres index:

$$P_P = \frac{\sum_{i=1}^{N} p_i^1 q_i^1}{\sum_{i=1}^{m} p_i^0 q_i^1 + \sum_{i=m+1}^{N} \hat{p}_i^0 q_i^1} \tag{3.15a}$$

which can be rewritten as:

$$P_P = \left[\sum_{i=1}^{m} w_i \frac{p_i^0}{p_i^1} + \sum_{i=m+1}^{N} v_i \frac{\hat{p}_i^0}{p_i^1} \right]^{-1} \tag{3.15b}$$

with

$$w_i = \frac{p_i^1 q_i^1}{\sum_{i=1}^{N} p_i^1 q_i^1} \tag{3.16a}$$

for the m matched items, and

$$v_i = \frac{p_i^1 q_i^1}{\sum_{i=1}^{N} p_i^1 q_i^1} \tag{3.16b}$$

for the N-m unmatched items.

Again, of the N different items of the commodity that are available in period 1, m items also were available in period 0; the first, matched part on the right-hand side of equation (3.15b).[10] As in the Laspeyres index, these items have prices p_i^0 and p_i^1. The remaining N–m items are new, and were not available in period 0. The prices of these items are estimated for period 0, using the hedonic regression for this period; the second, unmatched part on the right-hand side of equation (3.16b). Both terms are weighted with their relative expenditure shares in period 1. The geometric average of the hedonic imputed Laspeyres and Paasche indices given by (3.13a) and (3.15a) yields the hedonic imputed Fisher index.

Depending on the share of unmatched items in total expenditure or output, and the price behaviour of these unmatched items, expanding a pure matched model index with estimates of unobserved price changes can have a drastic effect on the resulting index. It is not possible to say beforehand how much a hedonic imputed index, including estimates of price changes of unmatched items, will differ from a matched model index only, neither which sign this difference will have. The more rapid the quality change and the larger the share of unmatched items, the larger this difference is expected to be. But the data determine the actual differences.

Apart from the fact that the dummy method estimates prices for *all* items, matched and unmatched, an important difference between the imputed method and the dummy method is that the imputation method makes direct use of the regression coefficients for the characteristics in the estimation of the imputed prices. The dummy method only uses these coefficients indirectly, since the hedonic price index is estimated directly from the regression equation, using only the time dummy coefficient. Therefore the level of significance of the coefficients of the characteristics is of less importance when using the dummy method than when using the imputation method.
The imputation method has two advantages compared to the dummy method. First, it is not necessary to estimate one hedonic function for several periods, thereby dropping the assumption that the regression coefficients need to remain constant over time. Second, whereas the dummy method lacks transparency, with the imputation method it is very clear what we are doing. There are prices missing that are needed to construct an index. The hedonic method is used to estimate these prices, and the estimates can be used to fill the holes in the Laspeyres and Paasche indices that were caused by new and disappearing products. This implies that prices of products that can be matched are retained, and are not based on hedonic estimation instead. This variant of the hedonic method is preferred by statistical agencies. Indeed Griliches (1990) lists this method as his preferred interpretation of the hedonic method.

Hedonic quality adjustment

When statisticians witness the disappearance of a particular item, the usual procedure is to look for a 'replacement item' to put in the index. The same happens when an item is undergoing a quality change, resulting in a new item. For purposes of constructing price indices, these 'old' and 'new' items would be considered different items, not a single item that has changed in quality. However, if the first item has disappeared in the second period, than the second item can be considered as a replacement item with which the first one is compared.
The hedonic quality adjustment method makes an estimate of the price difference between the old and new items that is due to the change in quality. What remains is a 'pure' price change. The difference with the imputation method therefore is that the hedonic quality adjustment method makes an estimate of the price differential due to quality changes of *existing*, 'real', items, whereas the imputation method estimates prices of *missing*, 'imaginary', items.
The hedonic quality adjustment method can be illustrated in the following way: [11)]

Observed price change = [Quality-adjusted price change] * [Quality adjustment]

So if an item has increased in quality, the quality-adjusted price change should be smaller than the observed price change. The quality adjustment is the price change that results from a difference in the quality characteristics alone, and is therefore equal to the ratio of the estimated prices of the old item o and the new item n in a period t, would they both have been available in this same period. For a semi-logarithmic regression equation, this hedonic quality adjustment (HQA) is equal to the following:

$$\begin{aligned}
\text{HQA} &= \frac{\hat{p}_n^t}{\hat{p}_o^t} \\
&= \exp[\ln \hat{p}_n^t - \ln \hat{p}_o^t] \\
&= \exp[\sum_{k=1}^{K} \hat{\beta}_k^t z_{nk} - \sum_{k=1}^{K} \hat{\beta}_k^t z_{ok}] \\
&= \exp[\sum_{k=1}^{K} \hat{\beta}_k^t (z_{nk} - z_{ok})] \qquad (3.17)
\end{aligned}$$

where $\hat{\beta}_k^t$ is the estimate of the coefficient of characteristic z_k of a hedonic regression for period t.

With the estimates for HQA, both Laspeyres-type and Paasche-type indices can be calculated. A Laspeyres-type index uses items from the first period as the base, and therefore applies a quality adjustment to the prices of their replacement items (or counterparts) in the second period. If separate hedonic regressions are estimated for individual periods, the regression results from the second period are used to estimate the quality adjustment. A Laspeyres-type price index using a hedonic quality adjustment is then given by:

$$P_L = \frac{\sum_{i=1}^{N} \hat{p}_i^1 q_i^0}{\sum_{i=1}^{N} p_i^0 q_i^0} \qquad (3.18a)$$

with

$$\hat{p}_i^1 \equiv p_i^1 \exp\left[\sum_{k=1}^{K} \beta_k^1 (z_{ik}^0 - z_{ik}^1)\right] \qquad (3.18b),$$

which uses regression estimates β_k^l from period 1 to estimate the quality adjustments. The prices $\hat{p}_i^1$ are prices in period 1 adjusted for differences in quality characteristics. Note that the subscript i refers to a pair of two items, one in each period. In the case of new and disappearing items, this pair consists of two non-identical items, which are each other's replacement items. Of course, if two identical products are matched, than the quality adjustment factor equals unity, and $\hat{p}_i^1$ simply equals p_i^1.

A Paasche-type index can be defined in a similar way:

$$P_P = \frac{\sum_{i=1}^{N} p_i^1 q_i^1}{\sum_{i=1}^{N} \hat{p}_i^0 q_i^1} \tag{3.19a}$$

with

$$\hat{p}_i^0 \equiv p_i^0 \exp\left[\sum_{k=1}^{K} \beta_k^0 (z_{ik}^1 - z_{ik}^0)\right] \tag{3.19b}$$

Of course, a hedonic quality adjustment can only be carried out when the number of items in both periods is the same, and therefore that the number of new items equals the number of disappeared items. [12] If this is not the case, the problem becomes how to pair unmatched items. Instead of selecting individual replacement items, one can also aggregate the items into groups with common features. For example, Silver and Heravi (2002a) aggregate the items over one or more characteristics, which either have discrete values, or are dummies. These characteristics disappear from equations (3.18b) and (3.19b) to calculate hedonic quality adjustments.

Suppose that we aggregate over S summary characteristics, resulting in J different combinations of these characteristics. [13] Within each combination, the weighted mean values of all other $K - S$ characteristics are calculated to compute hedonic quality adjustments within these combinations. The Laspeyres-type index of equation (3.18a) would then change into:

$$P_L = \frac{\sum_{j=1}^{J} \hat{\bar{p}}_j^1 q_j^0}{\sum_{j=1}^{J} \bar{p}_j^0 q_j^0} \tag{3.20a}$$

with

$$\hat{\bar{p}}_j^1 \equiv \bar{p}_j^1 \exp\left[\sum_{k=1}^{K-S} \beta_k^1 (\bar{z}_{jk}^0 - \bar{z}_{jk}^1)\right] \tag{3.20b}$$

where $\bar{p}_j^1$ is the weighted mean price within combination j in period 1, and $\bar{z}_{jk}^0$ and $\bar{z}_{jk}^1$ are the weighted mean values of characteristic z_k within combination j in period 0 and 1, respectively. The quantity q_j^0 is the total quantity sold of all items which fall into combination j. The Paasche index in equation (3.19a) is adjusted similarly.

Aggregating over several characteristics results in less information on quality differences being used. As a consequence, the more observations of individual characteristics are not used for the estimated price, the more the results from a hedonic quality adjustment index converge towards a matched model index.

When aggregating over all characteristics, all items that have no counterpart in the other period will be deleted, and the terms between brackets in equation (3.20b) will become zero, resulting in a hedonic quality adjustment equal to unity. This index will be identical to the matched model index.

For the purpose of constructing official price indices by statistical offices, the number of models is usually equal in both periods. When an item disappears from the index, a replacement item is searched for to take its place in the index. In such a case, an explicit quality adjustment is necessary. However, if no replacement item is found, and the number of models in both periods is not the same, either some observations have to be thrown away, or some models have to be aggregated over their characteristics in some way. The higher the level of aggregation, the more one is calculating ratios of average prices rather than actual prices, which is a drawback of the hedonic quality adjustment method, since statisticians prefer to use actual prices as much as possible.

In general, the hedonic quality adjustment method and the hedonic imputation method lead to fairly similar results. The major difference between these two indices and a hedonic dummy index lies in the fact that explicit weights are used in the two former indices. If no quantity weights are available, and adjacent period regressions are used to estimate the prices, then all three hedonic indices will lead to numerically close results. Another difference is that the imputation method and hedonic quality adjustment method can also be carried out using single period regression, which is by definition not possible for the dummy method. As mentioned above, applying single period regressions may be the only way to allow for shifting preferences and substitution behaviour if no sales data are available. The imputation and hedonic quality adjustment method look therefore better suited to deal with changing tastes.

3.5 Pioneering applications of the hedonic method

Most of the economic literature on the hedonic method has focused on price indices, especially on prices of durable goods. In this section, applications on several products will be discussed. First, I study a product on which much of the pioneering research has focused, i.e. automobiles. Indeed some of the first hedonic applications were for cars, including the work of Court (1939) and Griliches (1961). I then focus on the product group for which probably the most extensive research on quality-adjusted price indices has been carried out, namely computing equipment. The earlier work on hedonic price indices for computing equipment will be discussed below. In Chapters 4 and 5, I will deal with my own applications of hedonic price measurement on computers and cars.

A study that is somewhat different from the mainstream work on the hedonic method is that by Triplett and McDonald (1977). Whereas most studies up to that moment used the dummy method (Griliches' 1961 study being the prime example), they used a hedonic quality adjustment method, which they applied to prices of refrigerators making use of the same data as those used to construct the official price index. This made it possible to compare the implication of different methods in some detail. This was often not true for most other studies, which generally made use of data which were not used in official statistics. Their study will therefore be discussed in more detail below. Finally, hedonic price indices have also been constructed for products other than durable (consumer) goods. A good example is the construction sector, with which the discussion below will close.

Cars

Although the main body of research concerning the hedonic method in recent years has focussed on computers and related equipment, the first hedonic applications were for automobiles. In fact, it was Court (1939) who coined the term 'hedonic' in his research on quality-adjusted price indices for cars.
Unlike the example of computer processors described below, cars do not possess 'obvious' characteristics which can be associated both with user value and resource cost. The main problem of hedonic studies for car prices has always been the choice of appropriate variables in the hedonic function.
The first study to apply hedonic price indices for cars was Court (1939), of which the main purpose was to criticise the existing price index for cars published by the BLS. A simple matching of unit values for three different car types (coupes, two door sedans and four door sedans) revealed that the unit values of these types decreased between 9% and 16% over the period 1925–1935, whereas the official index increased with 45% in the same period.[14)]
So, Court was one of the first to show that price indices can actually decline after an adjustment for changes in product composition.
Among several alternatives, Court compiled an index based on average prices within broad definitions. Court selected product groups of cars, based on weight, wheelbase and horsepower. Cars are divided in the appropriate sets, and unit values of these sets are matched across years, which led to the unit value decrease mentioned above. Court then proposed the hedonic method as a suitable way to calculate a price index, that also takes into account quality changes. The hedonic regressions carried out by Court include the same three variables (weight, wheelbase and horsepower) supplemented with dummy variables for the time periods. The dummy index based on these hedonic regressions shows a price decrease of around 75% for the period 1920–1939. This price decrease is similar to the one Court obtained with an adjusted matched model index based on the same specifications as those used for his regressions. This indicates that explicitly taking differences in characteristics into account is of much importance, regardless of the method of quality adjustment.
Following the work of Court, the hedonic method was not applied for more than twenty years, after which Griliches (1961) revived it. The focus of Griliches' 1961 study was also on cars. The characteristics that Griliches uses are nearly the same as those used by Court: car weight, power and length (rather than wheelbase). In addition Griliches adds some dummy variables

that serve as proxies to other car performance characteristics. Griliches' study has already been discussed in some detail in Section 2.4; suffice it here to say that his hedonic price index for cars (based on the dummy method) is much lower than the official price index published by BLS. During the period 1954–1960, the official index increased with 19.7%. Depending on the hedonic function that Griliches uses, he finds a price increase of only 2.3% (based on a pooled regression for the entire period) or even a price decrease of 4.2% (based on adjacent-year regressions). Triplett (1969) extends Griliches' method for 1960–1965, but surprisingly finds a price increase of as much as 7%. When Triplett applies his hedonic regression estimates to the cars that are actually in the BLS sample, he even finds a price increase of 9.6%, compared with a price *decrease* of 5.2% in the official CPI.

These results are very much at odds with the findings of Griliches and call for explanation. Triplett saw as a possible explanation that for the hedonic regressions, physical characteristics were used rather than actual performance variables. The former are only proxies for the latter. Furthermore the relation between these two types of characteristics may not be constant over time, a fact for which Triplett brings up evidence. But since actual performance variables are hard to quantify in the case of cars, Triplett suspects that the hedonic method will be more problematic for cars than for products which enter the utility function in a more straightforward manner. In a later article (Triplett, 1990) he even places doubts at the credibility of hedonic estimates for car prices.

Probably for this reason, the hedonic method has not been used in a single official price deflator for cars. The results are too ambiguous for the production of reliable price statistics. The only 'official' adoption of the hedonic method for cars has been in the PPP work for the International Comparisons of Prices (ICP) project of the World Bank (Kravis *et al*., 1975, 1978, 1982). This work is discussed in more detail in Section 3.8.

Computers

More than 20 years after the pioneering work of Griliches (1961), discussed above, hedonic regressions were carried out for prices of computer components and peripheral equipment by Cole *et al*. (1986), followed by Dulberger (1989) and Gordon (1990). The first personal computers for commercial use had been introduced several years before, and the rapidly increasing quality of PCs caused large problems in the measurement of prices. Sinclair and Catron (1990) give an overview of the construction of what was called 'an experimental price index for the computer industry'. This index was first produced by the BEA in 1987, shortly after the work of Cole *et al*. (1986) was published in the Survey of Current Business. The goal of that first study was to find a way to measure "the cost improvements embodied in computers and computer peripheral equipment and to develop a methodology for excluding the cost of the improvements from reported prices" (Sinclair and Catron, 1990, p. 16). Stated differently, BEA acknowledged the fact that computers became of higher quality with constant or even decreasing production costs and prices, and that these improvements should be accounted for. The old price indices did not exclude effects of higher quality on prices, and therefore did not reflect pure price changes.

The methodology used by Cole *et al*. and Dulberger is very similar. They construct several types of hedonic price indices, and compare them with conventional matched model indices. The variant of the matched model they compare their hedonic index with is a chained index of matched models, where Paasche indices of two adjacent years are chained. The standard methodology at that time still used a fixed reference period, but because of the speed at which computers in the market are replaced with new models, using a fixed reference period would not be possible.

Cole *et al*. use hedonic regressions based on the premise that differences in the prices of goods mainly reflect differences in the 'volume' of characteristics of goods. For this reason they pay much attention to the issues of the level of aggregation and the selection of characteristics. In fact they are not constructing price indices of computer systems, but rather price indices of individual system components, namely processors, disk drives, printers and general-purpose displays. The reason to develop price indices of components instead of systems is twofold. First, in the studied data set, most purchases were of components. Second, it is easier to find relevant characteristics on the level of components. The focus of the following discussion is on their analysis of quality adjusted prices for processors.

The characteristics employed by Cole *et al*. are selected so that they reflect both user value and resource costs. In the case of processors, these were memory capacity and the speed with which instructions are executed. The former characteristic is measured quite easily. It is simply the random access memory (RAM) in megabytes. For speed, the number of instructions executed per second in millions (MIPS) [15] was selected, adjusted for differences in the job mix of the processors.

Three different hedonic price indices are then computed. The first is a composite index, which uses the regressions to estimate 'missing' prices of new models that were not available in the previous period. This reflects the imputation method discussed above. The second index is a regression index, based on the time dummies in the regressions. Finally, they compute a characteristics price index. Over the period 1972–1984, these hedonic indices show a much faster price decrease than the chained matched model Paasche index they compare their hedonic results with. The average annual price decrease over this period is 8.5% for the matched model method, whereas the decline in hedonic indices vary from 17.6% to 19.2%. Note that the matched model index only compares 'like with like', so there is no inside the sample bias, i.e. existing items in the sample are identical.

Similar results were found for the producer price index (PPI) for the computer industry by Sinclair and Catron (1990). These results indicate that in a period of less than two years, the prices in the computer industry have fallen by about one fifth. Another impressive finding of their study was that the life cycle of a particular type of computer was only about two or three years. If the same time period of five to seven years that is used for re-basing the product sample in other industries included in the PPI program would be used as a reference for the computer industry, this period would include approximately two-and-a-half generations of computers. The 'outside the sample bias' therefore seems to be rather large. This provides another argument why measuring price changes in rapidly changing industries like computers requires different procedures to collect, re-price and adjust the data for quality. In such cases, sampling should be done more frequently, and indices should be chained rather than using a fixed reference.

Shortly after Cole *et al.*, Cartwright (1986) announced that the Bureau of Economic Analysis (BEA) would adopt the hedonic method to deflate the computing equipment components of GNP. The Bureau of Labor Statistics (BLS), however, was much slower in adopting the hedonic method for computers, and only included it in the PPI in 1991, and in the CPI in 1998 (Moulton, 2001).

Refrigerators

As mentioned before, many applications of the hedonic method in the academic literature employ the dummy method to calculate a hedonic index.[16) The accompanying conclusion of many of those studies is that, if the increase of a hedonic index is slower than in the official index, which is usually the case, the official unadjusted BLS index must have an upward bias. This bias is then attributed to insufficient adjustment for quality improvements. However, as Triplett and McDonald (1977) note, this 'error' is not necessarily just a quality bias, since the researchers generally use a different data base than the one that is used to construct the official index. Therefore, the 'quality error' is 'actually an amalgam of quality error, differences in movement between the investigator's price series and the data collected by the BLS, and the effect of different methods of index construction' (Triplett and McDonald, 1977, p. 144).

Triplett and McDonald avoid the non-comparability problem of the price indices by using the same price data as those collected for the Wholesale Price Index (WPI), and performing a hedonic analysis with them. They focus their attention on the refrigerator-freezer manufacturing industry, for the period 1960 to 1972.[17) Apart from using comparable data, another difference with most other investigations is that they make use of a the hedonic quality adjustment method. The hedonic quality adjustment was only applied to refrigerator-freezers that witnessed a change in specifications, and only if there was a difference in the characteristics used in the regressions.[18) The regressions are carried out on *retail list* prices, but the regression results are used to make quality adjustments for *wholesale transaction* prices. This requires the assumption that percentage mark-ups, the difference between wholesale and retail prices, are uncorrelated with characteristics proportions.

In their official index, the BLS also registers the changes in quality of refrigerator freezers, and makes an adjustment using one of the two following methods. BLS staff decides whether the quality changes are minor or major. In the case of a minor quality change, these changes are ignored and the prices are simply matched. This produces a possible quality bias in the index, because the prices of non-identical items are matched. This is known as the "classic example of quality bias" (Triplett and McDonald, 1977; Triplett, forthcoming), and represents the kind of bias many economists assume is predominant in unadjusted price indices. However, for more than half of the minor quality changes in the period studied by Triplett and McDonald, this quality change was also negligible from the perspective of the hedonic method, since there was no change in any characteristic they included in their hedonic regressions.[19] In the case at hand, the 'classic quality error' may be less important than many economists seem to think, a point that is repeated by Triplett (forthcoming) in his analysis of how statistical agencies deal with quality changes.

If the quality change was judged to be a major change, than the link-to-show-no-price-change method, which was discussed in Chapter 2, was used and the entire price change was supposed to be caused by quality changes, with a pure price change of zero. This adjustment will probably have introduced a downward bias in the index, since it is likely that at least some price change cannot be attributed to quality changes. This is especially relevant with the introduction of a new model, which probably will be noted as a major quality change. Again, some of these major quality changes witnessed no change in any characteristic used in the hedonic regressions, so no adjustment was made there.

After constructing several hedonic indices (each with a different mix of the regression results), Triplett and McDonald conclude that, for the entire period, the hedonic indices drop at a somewhat faster rate than the official WPI for refrigerators-freezers, but with some interesting twists. During some sub-periods, the WPI actually decreased *faster* than the hedonic indices, suggesting that the bias is not uniformly upwards.

For the purpose of comparison, Triplett and McDonald also calculate regression indices using the dummy method. As noted above, the regression uses retail prices so the regression indices are actually quality adjusted retail list price indices, whereas the price indices with hedonic quality adjustments are based on wholesale transaction price indices. The two indices are therefore expected to differ, not only because a different method was employed, but because different price data was used as well.

Indeed the difference proved to be quite large (by about a factor two or more), but the signs were identical. Where the hedonic quality adjusted WPI suggested an upward or downward bias in the official WPI, the regression indices estimated the biases to be at least twice as large. Triplett and McDonald offer several possible explanations for these large differences, namely the different weights that both hedonic methods assign to different refrigerator-freezers, the effects of retail and distribution margins, and the differences inherent in the methods themselves. Whatever the main causes, the work of Triplett and McDonald indicates that one should not draw conclusions about the size of a potential quality error based on results from only one hedonic method, especially not if this method uses different (price) data than the biased index.

Construction

As was pointed out in Section 2.6, the construction sector is a sector that is characterised by a large heterogeneity in output. This is true not only for construction as a whole, which includes residential construction, non-residential construction, road building etc., but also for subsectors like residential construction, which is the focus of the current section.

Like for automobiles, it is not a clear-cut case what are relevant characteristics for residential buildings. It is even more difficult to quantify many quality attributes that are relevant to users of houses. Despite these problems, hedonic price deflators for residential buildings have been applied more frequently by statistical agencies than hedonic price indices for cars.

In 1995, Statistics Netherlands (CBS, by its Dutch acronym) needed a new method to construct a consumer price index for construction of houses in the social sector, because the main provider of data stopped reporting. The hedonic method was investigated as a possible new price index, and was applied to a data base that was collected by CBS itself. This data set contained information on several quantifiable characteristics, together with quality information that could be included in a regression analysis in the form of dummy variables. The quantifiable variables were: the number of storeys, the number of building blocks, the number of houses finished in the relevant building project, the volume of the house and the foundation depth.

All quantified characteristics and most of the dummy variables proved to be significant, and the estimated regression index (based on the dummy method) was very similar to the existing index based on the 'old' database. Therefore, a hedonic index was adopted for the CPI for residential buildings in the social sector (van Leeuwen, 1994).

This index was not without criticism, however. The number of observations in the data set dwindled slowly, and with the decline in the share of the social segment in total residential construction the index became less and less representative for total construction of residential buildings. For this reason a new hedonic price index was calculated, this time for total residential construction (Elfering, 2001b). This index also includes more expensive houses, the share of which has sharply increased during the past decade.

For the new hedonic index, yet another data set was used. A drawback of this data set was that it included information on only a limited amount of variables. Of these, only one (the volume of buildings) was a physical characteristic of the building itself. On the other hand, the explanatory power of this variable was fairly large, and including the other variables did not add much explained variance (Elfering, 2001b). [20)]

In neither hedonic indices adopted by CBS, the notion of 'user value' was acknowledged. Prices, which are equal to the total cost of a building project divided by the number of houses in the project, were simply regressed on the data that were available, even if their relation to user value was unclear. For example, the number of buildings that was part of a project proved to have a significant negative price effect. This is to be expected, since economies of scale suggest that fixed cost per unit is lower if the number of units produced is higher. However, economies of scale is neither a physical characteristic of a building, nor is it a performance variable. The user value of this 'characteristic' is negligible, and it is therefore not suitable for inclusion in a hedonic function. The low explanatory power of these variables is acknowledged by Statistics Netherlands, and research is taking place for sources with additional quality characteristics for newly built dwellings.

Likewise the Bureau of the Census in the United States has constructed a hedonic deflator for residential construction from 1963 onwards. The hedonic function of the Bureau includes several physical characteristics, like floor surface, the number of bathrooms and the presence of several amenities, all of which can be associated with both user value and resource cost. The trend in this index is nearly identical to the trend in price per square foot of floor surface, a similar finding as reported by Elfering (2001b). This is possibly problematic, as Pieper (1990) notes, since floor surface appears to be a more homogeneous measure for houses than for apartments (which are also part of residential buildings) and non-residential buildings. Like cars, buildings are very heterogeneous products, with many characteristics which are very difficult to quantify. Apart from data availability, this is the main weakness of the hedonic method for construction. On the other hand, there are major shortcomings to the matched model method as well. In construction, it is generally impossible to compare like with like, a prerequisite for the matched model method (see also Section 2.6). Because of the lack of sufficient data, the applicability of the hedonic method, which seems the only serious alternative, is seriously hampered.

3.6 The hedonic method in official statistics

In 1961, the Stigler Committee recommended the use of the hedonic method by statistical agencies to correct for quality changes. The car study by Griliches discussed above was one of the papers published with the Committee Report. Since then the United States has been at the frontier of the use of hedonic methods in official price statistics. However, up to the end of the 1990s, even in the U.S., there were only three cases in which statistical agencies employed the hedonic method: first, the price index for single-family houses by the Bureau of the Census; second, the National Income and Product Accounts (NIPA) deflator for computer equipment by the BEA; and third, a hedonic adjustment for ageing effects on housing services (rents) in the CPI by the BLS.

Hence despite recommendations and the many demonstrations that hedonic indices are better suited to adjust for quality changes (probably best illustrated by Triplett and McDonald, 1977), most statistical agencies have been very reluctant to include hedonic methods in their procedures of price measurement. Triplett (1990) provides several reasons why statistical agencies were opposed to the use of hedonic methods in price indices. Some of the objections are based on theoretical grounds, but the most significant ones appear to have a practical nature.

For instance, statistical agencies perceived that the hedonic method required them to fundamentally change the philosophy towards measuring changes in prices, which until then was based on observation of existing prices of comparable products. This perception that one was dealing with a fundamentally different approach was mainly based on the fact that most applications of the hedonic method in the economic literature used the dummy method to derive a hedonic price index. No account was taken of weighting procedures, which are very important in the construction of price indices. The question of the appropriate index number form was largely ignored by most researchers. Moreover, most hedonic studies resulted in a price index that was much lower than the relevant official index, which led many scholars to conclude that the difference between the two indices was entirely due to the quality bias in the official, unadjusted index.

That this distinction between matched model price indices and hedonic price indices is wrong was already noted by Griliches (1961, 1964), and replicated by Triplett and McDonald (1977), who adopted several quality adjustments to the official price index, to increase the comparability between their results and the official index.

A related perception was that hedonic price indices were strongly dependent on the measurement procedures adopted. Hedonic price indices were characterised by a lack of robustness. This lack of robustness was to a large extent caused by the fact that many researchers paid little attention to the data they used (e.g. list vs. transaction prices) and the criteria to select characteristics. As Triplett notes, “simply many of the published hedonic studies were not very good” (Triplett, 1990, p. 214). This tainted the reputation of the hedonic method at statistical agencies, many good studies notwithstanding.

A third perception by statistical agencies was that automobiles were considered as a ‘test case’ of the hedonic method. Many hedonic studies in the 1960s and 1970 focused on cars, but they contained too little potential for resolving the measurement problems that arose in price indices for automobiles. Triplett points out that for several reasons, cars do not lend themselves very well for hedonic analysis. Foremost among these reasons is the problem of unobserved variables, which can swamp effects of included characteristics. As it turned out, automobiles represented an unfortunate test case for the hedonic method.

Hence although these perceptions by statistical agencies seem ill-conceived, it was to a large extent based on sometimes poor research that drew false conclusions concerning the difference between a hedonic index and the official index. More recently, however, many of the objections by statistical agencies have dissipated, and even the BLS which was most conservative regarding the hedonic method now makes extensive use of hedonic indices (Moulton, 2001).

As discussed before, some statistical agencies outside the U.S. have recently also implemented hedonic in their official price statistics. Still other statistical offices in Europe are in a advanced stage of research into hedonic indices; see Ball and Mehmi (2002) for the United Kingdom and Chapter 4 of this thesis for the Netherlands. Eurostat recently instigated a project called the European Hedonic Centre, to investigate the feasibility to estimate a common hedonic function for computers for several European countries (Konijn *et al.*, 2002). The amount of research carried out in Europe and the implementation of the hedonic method in official statistics suggests that also European statistical offices are letting go of their initial reserved position regarding the hedonic method. Indeed, Zieschang *et al.* (2001) suggest that there is 'almost certainly support across the majority of national statistical offices for the eventual implementation of characteristics-based hedonic methods for selected commodity groups in their national CPIs'.

3.7 Recent criticisms and the alternative of the high frequency matched model

Besides the optimistic statement at the end of the previous section, there is also recent criticism which goes against the hedonic method, suggesting its popularity seems to be decreasing in the United States. A research panel chaired by Schultze and Mackie (2001) recommends that, although BLS should continue to research and test hedonic methods, a more cautious integration of hedonically adjusted price change estimates in the CPI should be adopted. This recommendation stems from a concern on the perceived credibility of the models that are currently in use. The fact that a large number of econometric issues concerning is still unresolved, is of major concern to the Schultze Commission. The Commission is especially negative about the hedonic dummy method (based on longer time frame), since it relies on the strong assumption that the coefficients of characteristics need to remain constant over time. This conservatism has been strengthened by a research report of the Federal Reserve Board, which calculated matched model price indices for computers with high frequent data, and concluded that given a sufficient amount of matching between adjacent periods, the quality bias of a matched model index will be negligible (Aizcorbe *et al.*, 2000). The hedonic method would in such a case not be necessary.

Indeed, the use of the hedonic method has often been defended on the grounds that matched model indices have a large 'outside the sample' bias, which is caused by the fact that new goods are not included in the index on a timely basis. This is what Silver and Heravi (2002b) call the 'sample degradation' of standard price indices. Such sample degradation is especially apparent in Laspeyres-type indices where periods of observation are few and far between, and the reference weights are kept fixed for too long. But according to Aizcorbe *et al.* (2000) the problem of sample degradation can be largely resolved by reweighting the sample frequently.

When statistical offices apply the matched model method, generally a base year is chosen, and matches are being made for items that exist both in the reference period and the current period. If the commodity in question changes in quality very rapidly, the number of matches will decrease severely over time, until a new reference year is chosen. In such a case, chained indices, where matches are made for adjacent periods will provide a higher number of matches, and fewer items will disappear from the sample because they are too new or too old.

The latter method was adopted by Aizcorbe *et al*. (2000). Since their chained indices are quarterly based, they dubbed their method 'high frequency matched model method'. They analysed prices of personal computers and semiconductors using quarterly data from 1993 to 1999. From this dataset, they constructed chained Fisher indices for matched models, and compared the results with a time dummy index based on the following regression equation:

$$\ln p_i^t = \sum_{i=1}^{N} \iota_i I_i + \sum_{t=1}^{T} \gamma^t T^t + \varepsilon_i^t \tag{3.21}$$

where I_i is a model dummy, that takes on value 1 if the observed price is for model *i*, and T^t a time dummy for the presence of model *i* at time *t*, and may be dubbed 'time-product dummy method'. As such it does not represent a hedonic regression, since it does not make use of a hedonic function including performance characteristics. The regression results were used to calculate a regression index and an index using imputed prices.

The results of the high frequency matched model method and the regression method were very similar. For example, between 1993 and 1999 the Fisher matched price index for desktop computers decreased by 59.9%, whereas the regression index dropped by 60.0%. Another matched index, which used no weights and was a simple geometric mean, fell by 57.8%.

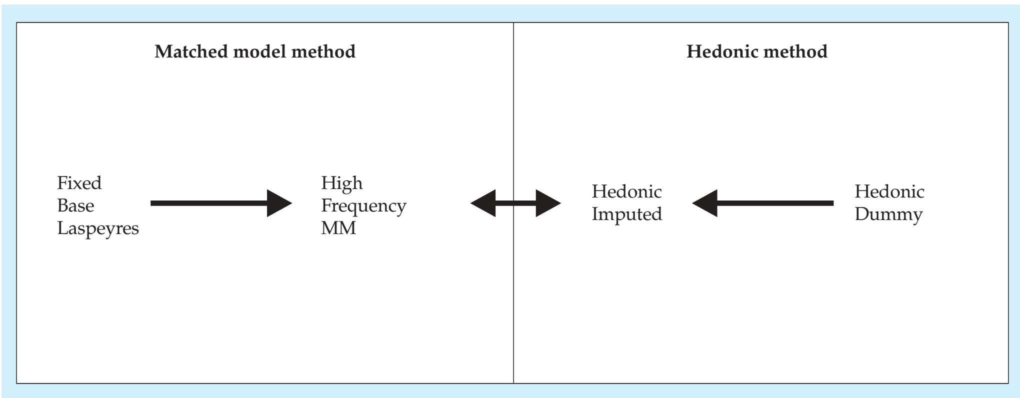
Aizcorbe *et al*. claim that that the small difference between the regression index and the Fisher index is due to the frequency at which the data are collected. Since this is done on a quarterly basis, the value share of computers that cannot be matched (because they disappeared from the market or were just introduced) is very small, i.e., less than 10 percent on average. Hence more than 90% of all revenues could be matched every quarter. Because of the frequent replenishment of the sample, the number of quality changes is small, so they claim that under these circumstances, hedonic price indices do not add much to conventional price index numbers. A caveat, however, is that during some quarters the revenue share of disappearing items was very small, whereas their share in the number of observations was much bigger.

It should be stressed that this result does not necessarily imply that the hedonic method is less useful than first thought. First, the high frequency method is at least as data intensive as the hedonic method. The latter method is sometimes dismissed by statisticians because of the huge and expensive data requirements. However, for the method used by Aizcorbe *et al*., detailed product information is also needed, to make sure that matched items are indeed identical. [21)]

A second point against any matched model method, including the high frequency type, applies to the problems associated with the imputed price change - implicit quality adjustment method, or class mean method, of adjusting for quality changes in matched models, as explained in Section 2.3. Suppose that individual items never change in price, and price changes only occur with the introduction of a new item. Even the high frequency method would not register any price change, and the index would be equal to unity forever. It is very obvious that this will introduce a bias in the index, which will be downward if the generally higher price of the new item is not entirely caused by an increase in quality. This objection is probably more relevant in oligopolistic markets, where buyers have less information about the quality of products than sellers, like the market for new cars. But it may also be true in more competitive markets; this is an empirical matter that needs to be resolved with the data set at hand. The hedonic method cannot be dismissed beforehand even if the share of new and old items (be it share in sales or observations) is relatively small.

Nevertheless, the paper by Aizcorbe *et al*. was received with much enthusiasm by representatives of statistical agencies, who regard the hedonic method with much scepticism. Much of this scepticism originated from the unease about the dummy variant of the hedonic method, which, as described above, is incompatible with standard official price index construction. On the other hand, many researchers who criticise the traditional matched model method, like Silver and Heravi (2002b), have particular difficulties with the fixed base Laspeyres approach, which causes sample degradation. It seems that a common ground needs to be found on the use of high frequency matched model methods and the use of hedonic price imputation. Supporters of the hedonic method tend to have fewer problems with the former approach, whereas critics of the hedonic method can see the merits of the imputation method. In fact, as suggested in a comment by van Ark (in Triplett, 2002) on Silver and Heravi (2002b), there seems to be a movement in statistical practices towards (a combination of) imputed hedonics and high frequency matched models, illustrated in figure 3.1.

Figure 3.1
The movements in statistical practices towards imputed hedonics and high frequency matched models



Source: Derived from van Ark as in Triplett (2002).

One of the aims of the next chapter to assess the difference between the hedonic imputed method and the high frequency matched model method in the case of Dutch computer prices.

3.8 *Applications of the hedonic method across space: international price comparisons*

The bulk of economic research on the hedonic method has focused on comparisons of prices over time. However, there is another field of price comparisons where the quality problem is abundant, namely that of international comparisons of prices. Actually the quality problem is even more imminent in international comparisons. Because of differences in tastes and comparative advantages in production methods, there is a large heterogeneity in the composition of industry output and expenditure patterns, even at the level of individual goods and services. Since it is nearly impossible to compare like with like in an international context, the matched model method has serious limitations for interspatial comparisons. There is a huge 'inside the sample' problem. The 'outside the sample' problem is at least as big, since the value share of matched commodities is generally rather small. [22)]

In cases where matching is hampered due to heterogeneity, the hedonic method might be an appropriate alternative for the conventional matched model method. However, its application in international price comparisons has been limited to this day. Probably the most important cause for this is the lack of appropriate and comparable data. The data intensity of the hedonic method increases in the case of more than one country. For each country involved, data on identical quality characteristics is needed, which is a daunting task. However, because of the inappropriateness of the matched model method in some cases, the hedonic method is one of the alternatives that needs to be further considered.

Despite the data problems associated with the hedonic method, there have been applications in international comparisons. The most obvious efforts are witnessed in the various phases of the International Comparison of Prices (ICP) project (Kravis *et al.*, 1975, 1978, 1982). As described in Chapter 2, the ICP programme goes at great lengths to ensure the comparability of the items in the basic headings for which prices are collected. For this reason the quality problem in this programme is mainly restricted to investment goods and services, which are very hard to compare. Such efforts have been more limited in international comparisons using an industry-of-origin approach. [23)] Recently, Eurostat also started a project to estimate hedonic functions for several member states (Konijn *et al.*, 2002). The focus in this project, however, lies more in the search for a common hedonic function that may be employed for deriving quality-adjusted intertemporal price indices than on international price comparisons.

Efforts to use hedonics to compare quality adjusted relative price levels across countries have also been carried out by individual researchers. The use of scanner data for this purpose is prominent in this field, although such data are more generally used, in addition to being more suitable, for comparisons of prices over time. Examples of the use of scanner data for international price comparisons are Moch and Triplett (2002) and Heravi, Heston and Silver (2002). The second of these has a somewhat broader scope and will be discussed in more detail below.

Probably the first application of the hedonic method to construct quality adjusted PPPs was in ICP (Kravis *et al*., 1975, 1978, 1982). The reason for using a hedonic method was related to the difficulty to match products that have different specifications in different countries. The hedonic method was used to construct PPPs are automobiles and house rents. The following focuses on the methodology used for automobiles. The one for house rents is similar.

The traditional method of price comparison can be easily applied if the same car models are purchased in all countries compared.[24] In this case a straightforward match can be made. However, not in every country the same cars are bought. In this case a solution is to match cars for which model descriptions are as close as possible. But this is usually done on an ad hoc base, and seems rather arbitrary. If, for example, two models are matched on the base of their horsepower, there can still be large differences in other specifications, like engine displacement or weight. An alternative is to classify cars by ranges in different characteristics, and compare unit values of these classes across countries. A similar approach was adopted by Court (1939) in an intertemporal framework; this led to a result that was comparable with the one obtained following a hedonic approach.

Van Ark (1990) used the same approach in an industry-of-origin oriented international comparison of cars, and classified car models on the basis of their engine displacement. This led to substantially different conversion factors than unadjusted ones. However, he compared unit values of cars produced in three countries (France, Germany and the United Kingdom) where the spread in engine displacement of the cars produced is relatively small. For example, in the United States the average car has a much more powerful engine than the average European one. If this country would have been included, an analysis like that of van Ark or Court would be much harder, because of the difficulty to find overlapping classes. Therefore, such a classification method can generally be more easily applied in comparisons across time than across space, because of the larger possible heterogeneity in international comparisons compared to intertemporal comparisons. In such case, the hedonic method may provide a better alternative for quality adjustments based on differences in characteristics.

Actually, in ICP the estimation of hedonic functions is the first step in calculating quality-adjusted relative price levels. The second step is similar to the alternative matched model approach employed by Court. Cars are classified on the basis of their weight and horsepower. Individual cars of some classes are selected, and the prices of these cars are estimated in each country with the respective hedonic regressions. Quality-adjusted relative price levels are then calculated using different weighting schemes of the various (estimated) price ratios. The application of hedonics by ICP is therefore a mix of the hedonic method and the conventional matching method, although somewhat different from the hedonic imputation method described in Section 3.3.

For the hedonic regressions, it is important to choose the right characteristics that determine prices. To determine which characteristics are relevant for the analysis, a product definition is needed. Kravis *et al*. (1975) distinguish three different product definitions: the universal definition, the binary definition and the national definition. These definitions are used for the selection of specific items in basic headings.

The universal product definition

This definition assumes that the same characteristics are explaining prices in an identical manner in every country concerned. This implies that the marginal prices for each characteristic are the same everywhere. Because the same set of characteristics are equally relevant everywhere, this will result in transitive conversion factors. That is, $PPP^{A/B}$ / $PPP^{C/B}$ = $PPP^{A/C}$. This is a very attractive feature, but the universal definition is rather shaky on empirical grounds. There is no reason that every characteristic will be equally relevant in every country. For example, the marginal price of horsepower will be lower in a country where large, powerful cars are preferred than in a country where mainly smaller and less powerful cars are consumed. Apart from differences in the marginal price of the characteristics, it is not unlikely that the sets of price-determining characteristics are different if they are determined for each country individually.

The binary product definition

Instead of assuming that the same characteristics are equally relevant for every country, the price-determining variables can also be selected for pairs of countries. So, in the case of automobiles, the relevant characteristics in the comparison of countries A and B might be engine displacement and length, whereas comparing prices in countries A and C can involve horsepower and weight. Obviously, this will result in PPPs which are not transitive, since they are not based on the same price-determining variables. The advantage of this definition compared to the universal definition is that the selection of the independent variables takes differences in preferences or country 'characteristicity' into account.

The national product definition

To avoid making assumptions on the relevance of independent variables in different countries, a selection can be made for each individual country. To be able to make price comparisons, a kind of matched model procedure is chosen. The price of a given car model is estimated in each country using the estimated coefficients from the regression. This price can be based on weight and horsepower in one country, and engine displacement and length in another.

Since only one price for each model is estimated in each country, the PPPs are transitive, as long as the prices of the same models are estimated. Because of the resulting transitivity of PPPs and the full account taken of country-specific preferences regarding characteristics, this definition is chosen by Kravis *et al*.

Using the national product definition, Kravis *et al*. carry out separate regressions for each country, and make bilateral comparisons. However, separate regressions have one great disadvantage: because the number of observations per country can be small, the results of the regression may not be very robust. To increase the number of observations in a regression, and thereby getting more reliable coefficients, a pooled regression with dummy variables for the countries is an alternative. However, this method assumes a similar relation between prices and characteristics in every country, and therefore implies moving to the universal product definition. A resort may be the method of flexible pooling. Apart from dummy variables for the countries, additional variables are added, namely the characteristics connected with the country dummies, the interaction effects:

$$\ln p_i = \alpha_i + \sum_{k=1}^{K} \beta_k z_{ik} + \sum_{j=1}^{J} \delta_j D_{ij} + \sum_{k=1}^{K} \sum_{j=1}^{J} \theta_{kj} D_{ij} z_{ik} + \varepsilon_i \tag{3.22}$$

where z is a vector of the price-determining characteristics and D a vector of the country dummies. The θ 's are measures of differences in the effects of the characteristics. If θ_{kj} is not statistically different from zero, then the effect of characteristic k is identical in country j and the base country. The reason that Kravis *et al*. stick with estimating separate country regressions and do not opt for this method, is the computational awkwardness of the latter when the number of countries is getting too big.[25)]
In addition, the regressions follow a WLS procedure, where the number of registrations is used as the weight. In this case, each observation in the regression is weighted according to its relative importance, so that high-price, low-volume models do not influence the regression as much as they would otherwise. Weighted regressions are carried out for each of seven countries.[26)]
Eight different characteristics were included as independent variables, but the mix of selected characteristics was different for each country, depending on the significance of each variable in each country regression.[27)]
The second step is then to select 'common' car models for which prices are estimated. All observations were placed in a grid; on the axes of the grid were eight intervals of horsepower and nine of weight, creating a matrix consisting of 72 cells. Of these cells, ten were selected to provide representatives of the entire overlapping range of weight and horsepower models consumed in the seven countries. One cell consists of the smallest cars with lowest weight and horsepower, whereas another includes the biggest cars with highest weight and horsepower, the rest lying in between. For every cell, one particular car model was chosen to represent it. Preferably these ten cells included observations from as many different countries as possible. Not all cells contained observations for all countries, let alone that the chosen model was available in many countries outside the country where it is manufactured.

For each of the ten cars, the prices are estimated in all countries, using the respective hedonic regressions. However, the cells that are represented by these ten cars are not common to every country. For example, in the case of Japan only cars that are in five mid-range cells are purchased. If no cars from a certain cell are purchased in a country (for example, the five remaining lower- and upper-range cells in the case of Japan), Kravis *et al*. place doubts at the estimated price of the selected car from such a cell, since such a car deviates too strongly from the cars that are commonly purchased in this country. Therefore, they prefer to use only estimated prices of cars from 'common cells' for each country.

For each country, the relative price level was calculated relative to two different base countries: the United States and the United Kingdom. If the United States is chosen as the base country, the resulting relative price levels using base country weights or own country weights are far wider apart than when the United Kingdom is used as the base country. This is not a strange result, as the United Kingdom has observations in all cells, whereas the United States only has observations in seven cells. Moreover, the three cells where there are no U.S. observations are those with the lowest values for horsepower and weight, whereas most other countries have no observations in the cells with the highest values, decreasing the number of overlapping cells to only four or five in most comparisons.

The disadvantage of this mix of hedonic prices and matching methodology is that is still suffers from the major problem of the matched model method, namely the possible lack of common products that can be matched. To avoid the need for overlapping models a multivariate regression was carried out as well. This is known as the country-product-dummy method (CPD), which is also used for price comparisons not concerned with quality differences. A CPD-regression equation is similar to equation (3.21), which was used by Aizcorbe *et al*. (2000):

$$\ln p_i^j = \sum_{i=1}^{N} \iota_i I_i + \sum_{j=1}^{J} \delta^j D^j + \varepsilon_i^j \tag{3.23}$$

where I_i equals one if $i = i'$(i' being the particular item associated with $p_{i'}^{j'}$), and D_j equals one if $j = j'$ (j' being the particular country associated with $p_{i'}^{j'}$). The antilog of the coefficient ι_i is the estimate of the average item i price in the currency of the base country. The antilog of the coefficient δ^j is the estimated country j parity for the heading (consisting of all n items) relative to the base country.

Since Kravis *et al*. also have registration data at their disposal, both an unweighted (OLS) and a weighted (WLS) CPD-regression were carried out. The weighted regression is 'double weighted': each item is weighted with its relative importance in the expenditures of its own country and the relative importance of the given model in each country compared with the total for that model in all seven countries. Both expenditure weights are multiplied, and the products are normalised by the sum of all products. So the 'double weight' ω_i^j equals:

$$\omega_i^j = \frac{w_i^j}{\sum_{i=1}^{N}\sum_{j=1}^{J} w_i^j} \tag{3.24}$$

with

$$w_i^j = \frac{v_i^j}{\sum_{i=1}^{N} v_i^j} * \frac{v_i^j}{\sum_{j=1}^{J} v_i^j} \tag{3.25}$$

where v_i^j is the total expenditure (observed price times number of registrations) of car model *i* in country *j*.

An advantage is that the CPD-method takes into account all observed prices, and does not replace them with estimated ones. On the other hand, no explicit quality adjustment is made. As all car models from all countries are pooled together, this method leads to multilateral relative price levels. The results of the different comparisons are shown in table 3.1; all price levels are relative to the United States.

Table 3.1
ICP automobile price comparisons for seven countries, 1969 (U.S.=100)

	Binary comparisons			Multilateral comparisons (weighted CPD)
	U.S. Weights	Own Weights	Geometric Average	
France	140	101	119	127
West-Germany	163	93	123	104
Hungary	186	177	181	212
Italy	120	85	101	106
Japan	111	89	99	102
United Kingdom	184	104	138	119
United States	100	100	100	100

Source: Kravis et al. (1975), tables 8.5 and 8.8.

As described above, all relative price ratios presented in table 3.1 suffer from several drawbacks. In Chapter 5 of this thesis, a more thorough effort will be made to construct industry purchasing power parities for a set of countries with a very heterogeneous output of cars, based entirely on hedonic regressions.

Since the several rounds of ICP, the use of hedonic methods in official international statistics has been limited and even at ICP, they have been applied only in a few fields. The review reports on the Eurostat PPP programme by Castles (OECD, 1997) and on ICP by Ryten (United Nations, 1998) spend some attention to the problem of incomparability of items within basic headings. Zieschang *et al*. (2001) explicitly suggest that the hedonic approach “should be a part of the package of measures for improving the ICP“. To achieve this, Zieschang *et al*. suggest that national statistical offices, who are making ever more use of hedonic methods, should combine their efforts to collect data on common sets of characteristics, to facilitate the implementation for international comparisons.

Purchasing power parities based on scanner data

Since the calculations of the ICP-project, international price comparisons using hedonics have been rare. The hedonic method has been used most to study price convergence in the European Union, using price reports of the European Commission. Like in ICP, the market that was studied was that of passenger cars. Examples of such studies are Verboven (1996), Goldberg and Verboven (1998) and Gaulier and Haller (2000). The focus of these studies is the issue of price convergence, and they use a hedonic framework to estimate average prices adjusted for quality differences. Although many of the cars included in their samples are sold in all countries, no matched model was used for these cars. The likely reason is that there are some 'holes' in the sample (not every car was sold in every country), and the hedonic method is a relatively easy way to fill in these holes. Another argument for using the hedonic method in a case where the matched model may seem appropriate, is divergence in tastes. If preferences are not identical in different countries, than hedonic estimates based on individual regressions per country will yield different results than parities based on the matched model methodology.
The increased availability of scanner data provides another opportunity to improve international comparisons of prices. An interesting recent study of this sort is by Heravi, Heston and Silver (2003), who use a hedonic framework to calculate quality-adjusted purchasing power parities and relative price ratios for television sets in three countries: France, the United Kingdom and the Netherlands. Their database consists of scanner data, collected with the use of bar codes. An advantage of using scanner data to construct PPPs is that they provide a very rich data source, as scanner data contain detailed information on the transaction price of an item,[28)] the quantity sold, its characteristics, its brand and the type of outlet where it was sold. This contrasts with the normal price collecting procedure, where price collectors observe *list* prices of items that may not be sold at all.
A drawback of scanner data in an international context is contained in the coding. A unique identification code is attached to each individual item. The coding procedure differs between countries, so matching like with like on the basis of codes is not possible. On the other hand, matching may take place on the basis of specifications, if the specifications are detailed enough to ensure like is compared with like. This drawback is less apparent, although not non-existing, in single-country, multi-period scanner data, as individual items are generally given the same unique code in different periods.
Heravi *et al*. consider two different hedonic methodologies to estimate quality-adjusted relative price levels: the dummy variable hedonic method and superlative exact hedonic indices (SEHI), based on the methodology by Feenstra (1995), described in Section 3.2.

The dummy variable hedonic method is a 'standard' hedonic regression of the following form:

$$\ln p_i^j = \alpha_i + \sum_{k=1}^{K_1} \beta_k z_{ik}^j + \sum_{k=K_1+1}^{K} \beta_k z_{ik}^j + \delta^{NL} D_i^{NL} + \delta^F D_i^F + \varepsilon_i^j \tag{3.26}$$

where D^{NL} and D^F are country dummies for the Netherlands and France, respectively; these dummies are similar to time dummies in a hedonic regression pooled over multiple time periods. The United Kingdom serves as the reference country. The intercept term α represents the logarithm of the price of the base item which is the simplest Sony 14" TV set, sold in the United Kingdom, in electrical multiples (the reference outlet type).[29] The characteristics z are divided into two groups: one consisting of K_1 'core' characteristics, which are present in all countries, and another group that consists of characteristics which are not. This distinction is not really necessary, but it is useful for the application of SEHI, where Heravi *et al*. use different combinations of the core characteristics to construct aggregations (cells like those used by Kravis *et al*., 1975), for which quality adjusted price ratios are constructed. The quality adjustments are carried out using the remaining characteristics (non-core characteristics and core characteristics that weren't used in the particular combination of core characteristics).

The main difference between (3.26) and (3.8) is that the time dummies have been replaced by dummies for different countries. Following standard practice, the quality-adjusted PPPs can be derived by taking the antilogs of the δ's.[30] Using the dummy variable hedonic method, Heravi *et al*. apply a weighted least squares regression, to account for the different market shares of each item. This restrict the coefficients of the characteristics to be the same in all countries. To test this, Heravi *et al*. use interaction effects like those in equation (3.22). From these tests, it appears that the coefficients of quite a few variables are not the same for each country. Therefore, pooling the data of separate countries does not seem an adequate way to derive PPPs with this scanner data set.

The superlative exact hedonic indices (SEHI) approach developed by Feenstra (1995) can also be applied in an international context. In this case, a SEHI expresses the costs that a consumer from country A would face in country B relative to the costs he faces in his 'home country' to keep his utility intact, keeping in mind differences in quality between products in both countries. Using a linear specification and replacing time superscripts by country superscripts, equation (3.6a) – which formulates the exact hedonic price index in time perspective – changes to:

$$\frac{\sum_{i=1}^{N} p_i^B q_i^B}{\sum_{i=1}^{N} \hat{p}_i^A q_i^B} \le \frac{E(p^B, z^B, U)}{E(p^A, z^A, U)} \le \frac{\sum_{i=1}^{N} \hat{p}_i^B q_i^A}{\sum_{i=1}^{N} p_i^A q_i^A} \tag{3.27}$$

Using the SEHI, separate regressions are carried out for each country. Cells of similar items are constructed, aggregating the items using six different combinations of core characteristics: screen size only; screen size and flat screen; screen size, flat screen and Nicam stereo sound; brand only; brand and screen size; and brand, screen size and Nicam stereo sound. Within the resulting cells, the average values of other characteristics are used to estimate quality adjusted prices, employing adjusted versions of equations (3.6b) and (3.6c). Some sets of core characteristics take on a larger number of different values than others. [31)]

Different selections of core characteristics will result in different numbers of cells. The more cells are distinguished, the more likely it becomes that in some countries, some cells will contain no observations. In such cases, these cells cannot be used, and the observations that are in them in other countries are deleted. One particular combination employed by Heravi *et al*. results in 600 different cells, which resulted in that for the bilateral country comparisons, between 13% and 22% of all observations were not used, representing between 9% and 18% of the total quantities purchased. This definitively is a drawback of the application of SEHI by Heravi *et al*..

Geometric averages of the indices using base country and current country weights are the resulting superlative indices (Törnqvist in the case of a log-linear specification, Fisher in the current case of a linear specification). The resulting relative price levels of the different combinations of core characteristics in the SEHI approach did not lead to substantially different results. The spread between the highest and the lowest relative price level was never larger than 6%. The price level of France relative to the United Kingdom using the dummy method resulted in figure that was only slightly lower than the SEHI results. The difference between the dummy and the SEHI results in the Netherlands/U.K. comparison was much bigger, however. A possible explanation might be that the television market is France and the United Kingdom are pretty similar, so that taking value shares does not lead to major differences from an unweighted index of price ratios. The results of Heravi *et al*. are summarised in table 3.2.

Table 3.2
Quality-adjusted relative price ratios of televisions in three countries using hedonic methods, June/July 1998 (United Kingdom = 100)

	SEHI results (Fisher)						Dummy method (WLS)
	Screen size (10 cells)	Screen size Flat screen (20 cells)	Screen size Flat screen Nicam stereo (40 cells)	Brand (20 cells)	Brand Screen size (200 cells)	Brand Screen size Outlet type (600 cells)	
Netherlands	90.0	90.5	90.0	94.9	95.0	93.7	58.6
France	86.8	86.8	86.9	89.1	90.4	88.2	84.6

Source: Heravi, Heston and Silver (2003), tables 4 and 6.

3.9 Summary

The analysis of the hedonic pricing techniques has, with some exceptions, shown a divergence in focus between the academic tradition and statistical practice. In theory the hedonic method is not undisputed, and it is not free from practical problems either. Most statistical offices agree that something needs to be done about the quality bias in their price indices. But because of the perceived problems of the hedonic method, some of which were based on misconceptions (Triplett, 1990), traditionally there has been strong resistance to adopting the hedonic method by statistical agencies. Especially the hedonic dummy method, by far the most widely used variant in the academic literature, has met considerable resistance. This is especially the case because the dummy method is far from what is standard practice within the daily production of price statistics.

Two other alternatives of hedonic price measurement, the imputation method and the hedonic quality adjustment method, are much closer to what statisticians are used to; but these are researched far less often in academic circles. They are very suitable to 'patch' the holes in the index sample that are caused by new or disappearing items. It is not clear what the result of this patching on the index will be. If there are few missing prices, the effect can still be large if unmatched items show radically different price behaviour than matched ones. Likewise, a large share of unmatched items does not need to have a profound effect on the index if their price behaviour is similar to that of matched items. The size and sign of the effect is something that can be determined by applying hedonic methods to the data set at hand. Matched model methods provide no way to deal with these observations other than assuming that their price changes are the same as for the other items in the sample. Hedonic and matched model price indices are constructed for a set of scanner data on computers in the Netherlands in Chapter 4.

The matching of identical items is even more complicated in the case of international comparisons. Unless specific price surveys are carried out, such as in the case of the ICP expenditure approach, it is hard to find cross-country price information on identical products. The use of the hedonic method is a useful alternative approach, although comparable data on prices and characteristics that are needed for several countries can be very hard to find. Chapter 5 tries to apply the hedonic method for international comparisons of car prices, based on an industry-of-origin approach.

Notes

[1] Since homogeneous products are products with only one characteristic, they will be treated as characteristics.

[2] Since *p(z)* is an aggregate function across consumers, individual taste functions cannot be identified from *p(z)* if tastes differ.

[3] Here, a semi-logarithmic function is considered. The results for the double logarithmic specification, which also takes logarithms of the characteristics, are similar to those of the semi-logarithmic function.

[4] This is reflected in the fact that computers sold to consumers tend to be more powerful than those bought by businesses, since consumers put more value on graphical capacity and speed, which are needed to play advanced computer games. This requires stronger PCs than for the application of e.g. word processors and spreadsheets, which are of more interest to businesses.

[5] The similar is true for features disappearing from the market. Their coefficients will lack in equations (3.7b) and (3.7c).

[6] Silver (2002), Diewert (2002).

[7] One exception is the use of the characteristics approach as one of the alternatives for the computers price indices in the United States; see Cole et al. (1986) and Dulberger (1989).

[8] In fact this results in a biased estimate. This bias, however, is usually very small, and can be corrected by adding one-half of the coefficient's standard error to the coefficient before exponentiating (Goldberger, 1968).

[9] Diewert (2002), De Haan (2002).

[10] These n items are the same as in the Laspeyres index. If the number of new items is equal to the number of disappeared items, then $M = N$. This of course does not need to be the case.

[11] Again, this applies to a log-linear specification. In the case of a linear hedonic function, it is equal to: [Observed price change] = [Quality-adjusted price change]+[Quality adjustment].

[12] This was assumed the case in equations (3.18a) and (3.19a).

[13] $J = \prod_{s=1}^{S} j_s$, where j_s is the number of different values characteristic s can have.

If all S characteristics are mutually non-exclusive dummies, then $J = 2^S$.

[14] The BLS index was a traditional matched model index, with matches being based on the brand of cars, rather than specifications.

[15] As Triplett (forthcoming) mentions, engineers in those days spoke of "shipping MIPS" rather than of processors or computers, which indicates that this variable is a good proxy for speed.

[16] Early examples are Griliches (1961) and Ohta and Griliches (1976).

[17] The purpose of their study is to provide a quality-adjusted price, and use this for the output index in this industry, to better measure productivity changes. The focus of the present discussion is on their method of price index construction.

[18] Pooled regressions were carried out for three sub-periods.

[19] This does not mean that there will be no quality bias. A change in a non-observable characteristic may have occurred, which cannot be modelled in a hedonic regression. However, if non-observable characteristics introduce a bias in the unadjusted index, this is not a bias that can be eliminated with the hedonic method.

[20] The R^2 of the regressions including only volume was about 0.71; the R^2 of the regression including all five variables was about 0.75.

[21] The authors purchased the database from a private marketing agency. The same is often done for the construction of hedonic price indices.

[22] This is the share in expenditure or gross output, depending on the approach.

[23] Van Ark (1990) is an exception, although he did not make use of hedonics.

[24] The ICP approach is an expenditure approach and therefore the focus is on purchaser prices.

[25] This was a valid argument in 1975, when computational power was much smaller than in today's computers.

[26] The countries studied are France, the Federal Republic of Germany, Hungary, Italy, Japan, the United Kingdom and the United States. For three countries (Kenya, India and Colombia), no regressions were carried out because of too few observations. These countries are left out of the current discussion.

[27] For example, width was used in only one country regression, whereas revolutions per minute (RPM) was included in all regressions but one.

[28] Actually, the price is the weighted average transaction price of a particular item sold at different stores, and therefore a unit value.

[29] This means that no dummies are included for Sony, a 14''-size screen, and electrical multiples. The total number of characteristics available in the dataset is therefore K+3.

[30] Naturally, this will yield PPPs relative to the base country, which is the United Kingdom.

[31] For example, the are twenty different brands, but the core characteristic 'Nicam stereo sound' only takes on values 0 and 1.

4. *Alternative price indices for computers: the case of the Netherlands*

4.1 *Introduction*

Since their introduction for commercial use in the early 1960s, computers have become ever faster and more powerful. This is probably best summarised in what is now known as 'Moore's Law', which states that the computing power of computers doubles every eighteen months. [1)]

Traditionally, the introduction of new goods and rapidly increasing quality of products have posed difficulties for the construction of price indices. Often these phenomena are not sufficiently taken into account. This fact is well illustrated in, for example, Boskin *et al*. (1996) and Wyckoff (1995). [2)]

Given the fast quality change of computers, both the inside the sample and the outside the sample problems (discussed in Chapter 2) are potentially large. Some authors argue that these problems can to a large extent be mitigated by resampling and reweighting the index often (Aizcorbe *et al*., 2000). However, thorough comparisons of hedonic and such high frequency matched model indices based on the same database are not very common.

The present chapter aims to address these points in more detail for the case of a price index for personal computers in the Netherlands. Alternative computer price indices are calculated using two different data sets: one consists of the data from which the official CPI is constructed by Statistics Netherlands (CBS); the other is a scanner data set which was provided by GfK Netherlands, a marketing agency. With these data sets, several matched model indices are calculated and confronted with hedonic indices.

Before I turn to the empirical applications, first some other experiences of statistical offices with hedonic price indices for computers are explored in Section 4.2. The hedonic method has been frequently applied for computers, historically mainly in the United States. Recently statistical offices and researchers in other countries have started experimenting with it, and the ensuing results are discussed here. Section 4.3 discusses the research on the CBS data set. These data are collected from advertisements in computer magazines and websites. Since these advertisements contain rather detailed information on specifications, the data set proved to be suitable for a hedonic analysis. In Section 4.4, the scanner data from GfK are analysed. This data set contains the near 'universe' of all computer sales in the Netherlands from January 1999 to January 2002. Detailed information on prices, specifications and quantities sold in nine different outlet types is given for three types of computer equipment: personal computers (PCs), notebooks and servers. The main purpose of these two sections and this chapter in general is to find out whether the difference between matched model and hedonic indices justifies using the latter. An answer to this question is provided in the concluding Section 4.5.

4.2 *The use of hedonic price indices for computers by statistical agencies outside the U.S.*

Computing equipment is the product class for which the hedonic method has been most frequently applied in recent years, both by researchers and staff of statistical offices. The first major hedonic application on computer prices was Cole *et al*. (1986), and hedonic price deflators were subsequently employed by the BEA (Cartwright, 1986; Sinclair and Catron, 1990). Since then computers have become the archetypical example of the presence of a quality bias in official price statistics, which do not adequately adjust for the rapid quality increases that are witnessed in some products.[3)] The pioneering studies in hedonic price indices for computers in the U.S. were already discussed in Chapter 3. More recent major studies were Berndt *et al.* (1995), Berndt and Rappaport (2001) and Pakes (2002). The focus of this section is on the use of the hedonic method for computers by statistical offices.

Apart from the United States, countries where the hedonic method is applied for the calculation of price indices for computer equipment are Canada, France and very recently Germany. Using data provided by IDC, a marketing agency, Barzyk and MacDonald (2002) explore several methodologies to construct price index numbers for desktop computers in Canada, including a matched model method and a hedonic method. The matched model index is a chained index, where separate indices are calculated for adjacent months. The average monthly matching rate is about 88%, meaning that for each month, 88% of all models can be matched with those in the next month. Since no sales data are available, it is not clear what the market share of the matched models amounts to.

The hedonic index is calculated by chaining the time dummy coefficients of adjacent month regressions. Both the matched model and the hedonic dummy indices decrease rapidly, although the hedonic dummy index drops faster at 4.2% per month, compared with a monthly price decrease of only 2.9% for the matched index. The indices are also compared with the official Canadian price index for computers, which is a hybrid of a matched model method and a hedonic index. For identical computers, prices in adjacent months are matched. If no match for a model in month t can be found in the month t+1, a replacement item is included in the index. A hedonic quality adjustment is used to adjust for the difference in quality, and a 'shadow price' of the replacement item is estimated for month t. The ratio of the actual price of the replacement item at time t+1 and its shadow price at time t is used for the index. This method is close to the hedonic quality adjustment method described in Chapter 3.

The dummy index calculated by Barzyk and MacDonald is very close to the official index. Barzyk and MacDonald prefer the methodology of the official index, as it makes maximum use of actual prices, while minimising the quality error.

In France, a method similar to that of Barzyk and MacDonald is used to estimate a price index for computers. Prices of matched models are used for the index when available. In the case of missing prices, the imputation method is used to estimate them. The French index is therefore a pure example of the imputation method described in Chapter 3 (Bourot, 1997; Evans and Scherrer, 2002).

Like Statistics Canada, the German Federal Statistical Office uses a hedonic quality adjustment to estimate a monthly price index for personal computers (Linz and Eckert, 2002). Prices of identical computers are matched, and hedonic quality adjustments are only applied to pairs of disappeared and new computers. However, two different data sets are used. The first data set, which consists of scanner data similar to those used in Section 4 of this chapter, is used to estimate the hedonic regressions. For the calculation of the index, a second data set is used, which contains price and characteristics from advertisements of mail order companies published on the internet and in trade magazines. There is a lag between the calculation of the regression and the index, so that the index calculation makes use of 'old' regression coefficients. The main reason for the German statistical office to use this two-stage procedure is that it allows to calculate a timely index, without first having to estimate a new hedonic regression.

In other countries, researchers and statistical offices are experimenting with the hedonic method and are still considering the possibility to include it in their official price statistics. In Australia the current price index for computers is based on the U.S. hedonic index, adjusted for exchange rate fluctuations. Recent research suggests that the current index may be severely biased upwards, and a hedonic price index is considered as an alternative.[4)] A similar method using the exchange rate adjusted U.S. price index is applied for the output price index for computers in the national accounts of Denmark.

In 2000, Eurostat set up the 'European hedonic centre', a loose co-operation between E.U. member states to develop hedonic price indices for computers and to obtain more consistent methodologies across Europe (Triplett, 2000). Some of the first results for the United Kingdom and Germany can be found in Ball and Mehmi (2002) and Almus and Moch (2002). A recent overview of the use of hedonic methods by statistical agencies is provided by Almus *et al*. (2002).

Based on the studies discussed here, there seems to be a preference by statistical offices for a mix of the matched model and hedonic methods: prices of matched observations are used when they are available, and the hedonic method is used to fill in the 'gaps' caused by new and disappearing items. The same approach will be advocated in the research presented here on the Dutch computer price index, with a preference for the imputation method described in Chapter 3.

4.3 Alternative price indices using the Dutch CPI data base

The main aim of this section is to investigate to what extent the database of computer prices that is present at Statistics Netherlands can be used to make quality adjustments in the Dutch CPI for computers. This database is used to make the official price index for computers, and is therefore ideally suited for a comparison between the existing index and an hedonic one.[5)] Several variants of the hedonic method will be used to make quality adjustments, namely the dummy method, the imputation method and the hedonic quality adjustment method.

The current practice at Statistics Netherlands

For the construction of the price index for computers, Statistics Netherlands collects data from websites of several computer retailers. This method of data gathering replaced the former method which was in use until February 2001. Data were then collected from advertisements in computer magazines. In most cases, the same retailers were followed before and after February 2001. Nearly all retailers are computer stores that build custom configurations chosen by their clients. The computer price index consists of two parts: systems and components. A system is a 'complete' set, consisting of a computer box, keyboard, mouse, and in most cases a monitor. For the CPI, two kinds of components are used: printers and monitors. In this study, the focus will be on systems only.
Essentially the CBS price index for computers is an unweighted matched model chain index (Elfering, 2001a). Since computers increase in quality very rapidly, each computer system is sold for only a few months. For each retailer in the database, Statistics Netherlands tracks all computer systems in the data set during the period in which they are sold. Within this period, a computer system usually changes only little, so that it is possible to construct a matched model index for this system for this particular retailer. Computers are matched only within retailers, not across. The matching procedure also takes into account all physical characteristics, so that only identical computers are matched.
When a new computer appears, its price ratio relative to the previous period (where it was not available) is imputed by taking the price index for all systems. For example, the price index for all systems in January 2000 was 14.67 (with September 1997=100). A system that was introduced in this month will therefore have a price index of 14.67. When its price has decreased with 10% one month later, its index for February is therefore 0.90 * 14.67 = 13.2. This procedure is known as the class mean method, or 'imputed price - implicit quality adjustment' (IP-IQ) method (Triplett, forthcoming).[6)]

This method is applied for all systems and all retailers in the data set. When a system witnesses a small change in, for example, its hard disk, working memory, or monitor, an option pricing method is used to discount for this change, if possible. For example, if a system comes with a 17″ monitor in one month, but a 15″ monitor was included in the previous month, the prices of 17″ and 15″ monitors sold separately are compared. This price differential is considered as a quality change, and subtracted from the new price. When a system comes with a new processor, however, it is considered an entirely new system, and no quality adjustment will be carried out. It enters the index as a 'new item'. Most quality changes are brought about by a different processor, so the option pricing method is only occasionally used.

For the sake of reliability and representativeness, a system will only be part of the final index if at least three retailers sold such a system during that month. As no single particular configuration is sold by more than one retailer, the criterion for this selection purpose is the processor.[7)] If at least three retailers sell a system with e.g. a 900 MHz Pentium III processor, the price ratios of all systems with this processor will be included in the final index. This data trimming ensures that only 'mature' systems enter the index. Systems that are sold by just one or two retailers are mostly very new or nearly obsolete, and such systems usually witness different price behaviour than other systems. The composite price index is an unweighted arithmetic average of the price ratios across all systems and retailers.

For the hedonic analysis, no such data trimming was carried out. All observations in the advertisement dataset were included. Furthermore only computer systems with particular processor types are included in the CPI. The database that is used here contains information on a lot of computers with other types of processors, but these are excluded from the CPI. On the other hand, some observations that are in the CPI were not available in the advertisements dataset, but these were only a minority. In total, the number of observations used in the hedonic analysis far exceeds the number of observations in the CPI. Only months for which at least 100 observations were available were used for the hedonic regressions, with the result that the present analysis was limited to a period of ten months.

Table 4.1 shows the number of observations by month for the entire database that was used for the hedonic analysis in this section and the total number of observations in the CPI. Although the total number of observations remains fairly constant in this period, the number that is used for the CPI declines steadily. It seems likely that this will increase the outside the sample bias, which will have an adverse effect on the reliability of the index. After June 2000, the total number of available observations declined sharply as well.

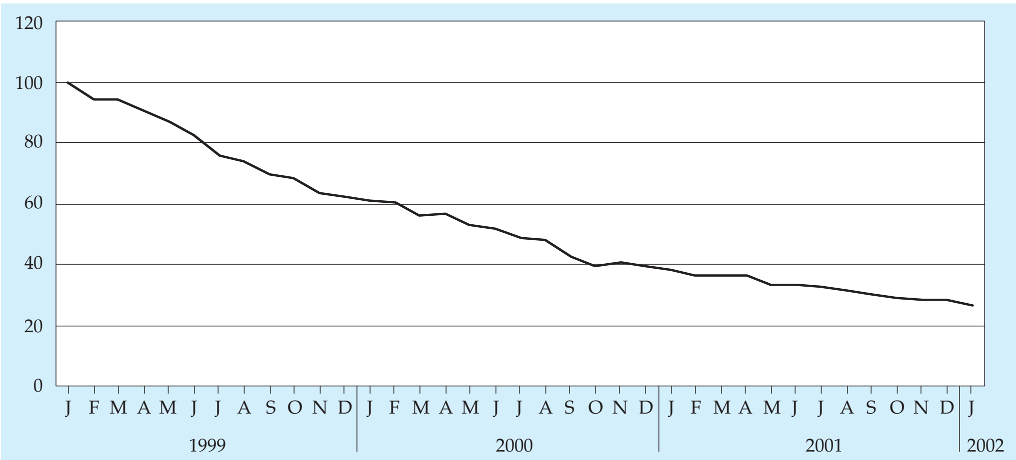
Summing up it is clear that in the consumer price index for computer systems compiled by Statistics Netherlands, only 'like' is compared with 'like'. We can therefore expect a decline in the index, since all systems witness a price decline over their entire existence. As can be seen in figure 4.1, this indeed proves to be the case.

Table 4.1
Number of observations in the total data set and the CPI

	Total data set	CPI
1999		
September	136	43
October	119	35
November	105	39
December	133	31
2000		
January	113	31
February	123	27
March	122	17
April	107	25
May	113	13
June	127	14

Source: CBS.

Figure 4.1
Dutch CPI for computer systems, January 1999–January 2002 (January 1999=100)



Source: CBS.

Apart from explicit adjustments when there are small changes in quality, only prices of identical computer systems are matched, so there is little room for an inside the sample quality bias caused by matching different products in the Dutch CPI for computers. In addition, chaining leads to less 'sample degradation' (Silver and Heravi, 2002b) than in the case where the reference period is held fixed. But the 'outside the sample' bias is also relevant to the extent that the data set used for the CPI is not representative for all computers sold in the Netherlands. Given the declining number of observations that are used for the CPI, this bias is probably substantial.

Data

Although the data set used for the CPI was not collected with the purpose to perform a hedonic analysis, it contains rather detailed information on characteristics of products, and is therefore suitable for such an analysis. Hedonic functions were estimated for the period September 1999 to June 2000. Data for a longer period were available, but outside this short period, the number of observations in the database was too low to perform a reliable regression analysis. Bearing in mind that the main purpose of the current exercise is to illustrate differences between methods, the relatively short period does not need to be a major drawback.
The database consists of advertisements by retailers in computer magazines and catalogues. Although advertisements are not available for all observations in the CPI, the majority of the CPI observations is present in this database. A common feature of all computer retailers studied is that they build systems to order. Advertisements show standard systems, but a customer is free to choose his or her own configuration, which is assembled on the spot. Therefore most systems in the database are 'clones',[8] and standard suggested configurations only. It is therefore not possible to say to what extent these systems are representative of all systems that were sold. Assuming that the standard systems are representative for all actual systems, the relation between price and characteristics is thought to be identical. Some retailers also provide information on the additional cost of a system with a slightly different configuration, especially configurations for which the only difference is another processor. These different configurations were also included in the regressions, to provide better regression estimates, although they were not included in the official CPI. The total number of observations in each month was shown in table 4.1.
The advertisements contain information on a lot of different characteristics of products, but not always the same for each item. What follows is an overview of the information that was given. The data sample is summarised in Appendix A.

Price. Both prices including and excluding VAT were included. For a hedonic regression, it does not matter which one is used, since one is a scaling of the other. Since the official CPI uses prices including taxes, prices gross of taxes (GPRICE) were used here as well.

Time. Pooled hedonic functions were estimated for more than one period, with time dummies T199910 (October 1999) to T200006 (June 2000). September 1999 is the reference period, and no dummy is included for this month. In the adjacent month regressions, a dummy for the second month is used.

Processor and speed. In all cases, the type and the speed in MHz were given. In total, nine different types of processor types are included. The memory size in MB was also always given, and in many cases, but not all, also the memory speed in MHz. The following variables were used in the regressions:
(LN)PSPEED: the (logarithm of) processor speed in MHz
(LN)MEMORY: the (logarithm of) memory size in MB
PD1-PD8: dummies for the processor type. For confidentiality reasons, the names of processors are not disclosed. The most common processor in this period (the sixth type) was chosen as the reference processor, and no dummy variable was included for this processor type.

Storage. The capacity of the hard disk in GB was given in all instances, and sometimes also the type or rotational speed. (LN)HDISK was used as the (logarithm of) the hard disk capacity.

Multimedia. Most configurations come with a CD-ROM player, but in several cases a DVD player and/or a CD-rewriter were part of the configuration, which replace the CD-ROM player. For some systems, the speed of the devices was not given, so these variables (CDROM, DVD, CDREW) are dummy variables. Although DVD players and CD-rewriters replace CD-ROM players, the former two are not mutually exclusive with the latter device, since some configurations contain neither.
Some systems contain a 56k modem, which is represented by the dummy MODEM.

Sound and vision. There is a myriad of different sound cards and graphical cards. Because there are so many different ones, it is too cumbersome and probably not worthwhile to include a dummy for each type. Instead for sound cards, a dummy SOUND was used which takes on value 0 if the configuration has no separate sound card,[9] and equals 1 if a separate sound card is included. Similarly, the dummy variable SPEAKERS registers the presence of a speaker set.
Graphical cards have memory capacities of 8, 16, 32 and 64 MB. In the case of a graphical card that is integrated onto the motherboard, this is set to 4 MB. All configurations that listed a capacity for an integrated graphical card, stated this as 4 MB, which justifies this assumption. (LN)VIDEO is the (logarithm of) the memory capacity of the graphical card in MB.
Most systems, though not all, include a monitor. In all cases, this was either a 15”, 17” or 19” monitor, which were captured by dummy variables MON15, MON17 and MON19. Some monitors had a (Flat) Trinitron screen, represented by the dummy SCREEN.

Operating systems and software. Although software is not a physical characteristic of a computer, it does add user value to the computer, and increased costs to the retailer. However, software companies sometimes pay computer manufacturers to include their software on the computers, actually decreasing the price of a computer when software is included. The actual effect of software

on the price of a computer needs therefore to be investigated with the data set at hand. The number of advertisements that included information on software was limited. Since consumers value computers that are supplied with software, the assumption is that when a system features software, this is mentioned in the advertisement, and vice versa. Dummies were included for three different operating systems (OPD1–OPD3).
The computer systems contained many different software programs other than operating systems; to avoid a large amount of dummy variables, a dummy was simply introduced for whether the system contains additional software (OTHSOFT).

Other specifications. In addition, the following dummy variables were included:
BRAND: equals 1 if the entire system is assembled under a major brand (i.e. all systems that are not 'clones');
NETWORK: equals 1 if the system is a network computer or workstation rather than a personal computer; in the present data set, only branded computers fall into this category;
BOXDESK, BOXMINI, BOXBIG: dummies for the type of computer box: it is either a desktop, a mini tower, a midi tower (the most common one and therefore the reference) or a big tower;
NETCARD: equals 1 if the system comes with a network card; all network computers and workstation contain one, but also some PCs
HARDWARE: equals 1 if the system contains miscellaneous kinds of hardware, which were too infrequent to list separately, like a printer or 120 MB disk drive.

Regressions

To test whether the above-mentioned variables have a significant effect on the price of computers, a pooled regression was run for the entire period. The double logarithmic specification had a somewhat better fit than the semi-logarithmic specification, and both logarithmic forms proved to be better than a linear one. Therefore the double-log specification is used in the current analysis, using the logarithms of gross prices (LNGPRICE) as the dependent variable. All regressions are OLS.
Having chosen a functional form, hedonic regression can now be carried out in different ways: a pooled regression for the entire period, pooled regressions for two adjacent months, and separate regressions for each month. Pooling the data for the entire period requires the strong assumption that the coefficients of the characteristics remain constant over time. Furthermore, because of pooling, observations from one period will influence the resulting hedonic indices for another period, even if they are relatively remote in time. The latter drawback can be avoided by pooling data only for adjacent periods. This will also make the assumption that coefficients remain constant less of a constraint.

Ideally separate regressions would have to be estimated for each individual period. However, if the resulting coefficients are used to impute missing prices, a problem occurs. Since computers are rapidly changing goods, new features appear or become standard equipment, in which case they turn up in the constant term. [10] These features will not be present in the periods prior to their introduction. In the current analysis, computers with DVD players or CD rewriters were not available in every month. If computers with such a characteristic appear in a particular month, but were not available before, their missing prices cannot be estimated accurately in the previous month if one uses single-period regressions. Adjacent period regressions do not have this problem, unless all computers in the second period have a particular characteristic, and no single computer in the first month is equipped with it, a highly unlikely case.
Taking all this into account, regressions were run on data pooled for adjacent months. This avoids the substantial drawbacks of the two alternatives, while not keeping constant the coefficients for too long. Because the data enforces the use of pooled regressions, no test were carried out for fixed effects, which are thus assumed.
After a first round of regressions, a large number of the characteristics listed above were not significant in many or all of the regressions. If one wishes to use the coefficients explicitly to estimate hedonic price indices, this is an awkward result. For this reason, a second round of regressions was run with the characteristics which had significant coefficients in all or nearly all regressions. These variables are PD1, PD4, LNPSPEED, LNMEMORY, LNHDISK, LNVIDEO, MON15, MON17, MON19, CDREW, DVD, SPEAKERS and NETCARD. Apart from SPEAKERS, they all have expected signs. For the construction of price indices, only these variables will therefore be used in the hedonic regressions. The second-round regression results are shown in table 4.2. Significant coefficients are expressed in bold.

Price indices using the CBS database

Hedonic price indices are constructed using the three different methods that were discussed in Chapter 3: the dummy method, the imputation method and the hedonic quality adjustment method. This yields monthly indices, which will be chained, like the CPI, to obtain an index for the entire period. As we use here the dataset that serves for a large part as the basis of the official CPI on computers, the indices are compared with the existing index from Statistics Netherlands. A matched model index for the entire advertisements dataset will be computed as well, where the same method will be used as for the construction of the CPI, i.e. an unweighted arithmetic average of price ratios.

Table 4.2
Coefficients for the adjacent month regressions of consumer prices for PCs, Netherlands, September 1999–June 2000

	September 1999– October 1999	October 1999– November 1999	November 1999– December 1999	December 1999– January 2000	January 2000– February 2000	February 2000– March 2000	March 2000– April 2000	April 2000– May 2000	May 2000– June 2000
Constant	**1.4940**	**1.4236**	**1.9935**	**2.2324**	**2.1472**	**1.4391**	**1.4571**	**2.0791**	**1.8718**
	(0.4517)	*(0.4039)*	*(0.2539)*	*(0.2272)*	*(0.2395)*	*(0.2890)*	*(0.3492)*	*(0.3938)*	*(0.4026)*
T [1]	–0.0227	**0.0530**	–0.0017	**–0.0566**	–0.0117	**–0.0487**	**–0.0455**	**–0.0543**	–0.0180
	(0.0157)	*(0.0141)*	*(0.0103)*	*(0.0082)*	*(0.0078)*	*(0.0099)*	*(0.0103)*	*(0.0118)*	*(0.0120)*
PD1	**–0.2281**	**–0.2701**	**–0.2337**	**–0.2316**	**–0.2488**	**–0.2049**	n.a.	**–0.4004**	**–0.2937**
	(0.0666)	*(0.0641)*	*(0.0391)*	*(0.0465)*	*(0.0603)*	*(0.0739)*	n.a.	*(0.0870)*	*(0.0451)*
PD4	**–0.0999**	**–0.1395**	**–0.1425**	**–0.1206**	**–0.1120**	**–0.1050**	**–0.0981**	**–0.0973**	**–0.0898**
	(0.0178)	*(0.0166)*	*(0.0107)*	*(0.0095)*	*(0.0101)*	*(0.0132)*	*(0.0169)*	*(0.0188)*	*(0.0191)*
LNPSPEED	**0.8837**	**0.8794**	**0.7866**	**0.7427**	**0.7341**	**0.8346**	**0.8178**	**0.7609**	**0.7961**
	(0.0736)	*(0.0654)*	*(0.0397)*	*(0.0358)*	*(0.0382)*	*(0.0460)*	*(0.0534)*	*(0.0610)*	*(0.0636)*
LNMEMORY	**0.1174**	**0.1656**	**0.1759**	**0.1692**	**0.1741**	**0.1713**	**0.1469**	**0.1188**	**0.1444**
	(0.0233)	*(0.0220)*	*(0.0147)*	*(0.0122)*	*(0.0115)*	*(0.0137)*	*(0.0156)*	*(0.0181)*	*(0.0186)*
LNHDISK	**0.0951**	**0.0681**	**0.1202**	**0.1272**	**0.1214**	**0.1488**	**0.1537**	**0.0743**	**0.0456**
	(0.0250)	*(0.0233)*	*(0.0160)*	*(0.0125)*	*(0.0117)*	*(0.0170)*	*(0.0226)*	*(0.0228)*	*(0.0228)*
LNVIDEO	**0.0586**	**0.0415**	–0.0029	0.0125	**0.0416**	**0.0496**	**0.1151**	**0.1319**	**0.0828**
	(0.0253)	*(0.0198)*	*(0.0150)*	*(0.0134)*	*(0.0118)*	*(0.0144)*	*(0.0170)*	*(0.0161)*	*(0.0145)*
MON15	**0.1552**	**0.1420**	**0.1378**	**0.1366**	**0.1420**	**0.1376**	**0.1362**	**0.1403**	**0.1511**
	(0.0205)	*(0.0192)*	*(0.0125)*	*(0.0106)*	*(0.0105)*	*(0.0129)*	*(0.0144)*	*(0.0159)*	*(0.0167)*
MON17	**0.2541**	**0.2244**	**0.2006**	**0.2012**	**0.2177**	**0.2108**	**0.1960**	**0.1989**	**0.2130**
	(0.0196)	*(0.0184)*	*(0.0121)*	*(0.0102)*	*(0.0100)*	*(0.0123)*	*(0.0137)*	*(0.0153)*	*(0.0163)*
MON19	**0.4526**	**0.4214**	**0.3978**	**0.3762**	**0.4885**	**0.3818**	**0.3080**	**0.4149**	**0.5820**
	(0.0591)	*(0.0473)*	*(0.0290)*	*(0.0283)*	*(0.0398)*	*(0.0425)*	*(0.0451)*	*(0.0535)*	*(0.0679)*
CDREW	**0.5750**	**0.4738**	**0.5003**	**0.3565**	**0.3686**	**0.4781**	**0.4966**	**0.4539**	**0.5077**
	(0.0967)	*(0.0853)*	*(0.0593)*	*(0.0455)*	*(0.0429)*	*(0.0568)*	*(0.0617)*	*(0.0678)*	*(0.0764)*
DVD	**0.1124**	**0.1293**	**0.1098**	**0.1532**	**0.1095**	**0.0881**	0.0384	0.0568	**0.0811**
	(0.0428)	*(0.0370)*	*(0.0269)*	*(0.0267)*	*(0.0261)*	*(0.0226)*	*(0.0249)*	*(0.0295)*	*(0.0363)*
SPEAKERS	–0.0442	**–0.0569**	–0.0324	–0.0237	**–0.0382**	**–0.0615**	**–0.1010**	**–0.1193**	**–0.1233**
	(0.0231)	*(0.0216)*	*(0.0167)*	*(0.0152)*	*(0.0146)*	*(0.0184)*	*(0.0203)*	*(0.0212)*	*(0.0192)*
NETCARD	0.0407	0.0213	–0.0316	n.a.	**0.2939**	**0.3315**	**0.4081**	**0.2884**	**0.1420**
	(0.0633)	*(0.0439)*	*(0.0392)*	n.a.	*(0.0316)*	*(0.0302)*	*(0.0353)*	*(0.0324)*	*(0.0297)*
adjusted R^2	0.8192	0.8771	0.9302	0.9360	0.9436	0.9265	0.9237	0.9019	0.8850
SER [2]	0.1219	0.1038	0.0704	0.0619	0.0587	0.0715	0.0771	0.0834	0.0905

[1] Time dummy for the second period in the regression.
[2] Standard error of the regression.

n.a.: not available.
Standard errors are between brackets; coefficients in bold are significant at the 5% level of significance.

A hedonic index using the time dummy is simply derived by taking the antilogs of the time dummy coefficients in table 4.2. With the imputation method the prices of computers that are available in one period, but not in the preceding or following period, are imputed for the period in which they are not available. Since no information on quantities sold is available, we have to rewrite (3.13a) and (3.15a) in the following ways:

$$P_L = \frac{1}{M}\left(\sum_{i=1}^{m} \frac{p_i^1}{p_i^0} + \sum_{i=m+1}^{M} \frac{\hat{p}_i^1}{p_i^0} \right) \tag{4.1}$$

$$P_P = \frac{1}{N}\left(\sum_{i=1}^{m} \frac{p_i^1}{p_i^0} + \sum_{i=m+1}^{N} \frac{p_i^1}{\hat{p}_i^0} \right) \tag{4.2}$$

That is, the Laspeyres and Paasche-type indices are unweighted averages of price ratios of m matched items and M-m, respectively N-m, unmatched items. For the unmatched items, prices are estimated in the period where they are missing. It should be stressed that items are matched if and only if an identical computer model is sold by the same retailer.

With the hedonic quality adjustment (HQA) method, identical computers are matched just like in the imputation method. The remaining computers are divided into two groups: computers that have a 'replacement', offered by the same retailer, with different values for some characteristics, and computers that have no such replacement (either completely 'old' or 'new' computers). A computer is considered a replacement if it does not differ too much from the disappearing computer, but it must be sold at the same retail point. For the first group, a hedonic quality adjustment is applied to the price using the following formula (see also equation 3.18b):

$$\hat{p}_i^1 \equiv p_i^1 \exp\left[\sum_{k=1}^{K} \beta_k (z_{ik}^0 - z_{ik}^1) \right]$$

for the Laspeyres-like index (which uses actual prices from period 0 and adjusted prices for the 'replacement items' from period 1), and (see also equation 3.19b):

$$\hat{p}_i^0 \equiv p_i^0 \exp\left[\sum_{k=1}^{K} \beta_k (z_{ik}^1 - z_{ik}^0) \right]$$

for the Paasche-like index, which uses actual prices from period 1 and adjusted prices for the 'replacement items' from period 0.

For the second group, old and new computers, prices are simply estimated using the imputation method. Therefore, this method is rather a mix of the imputation method and the hedonic quality adjustment method.

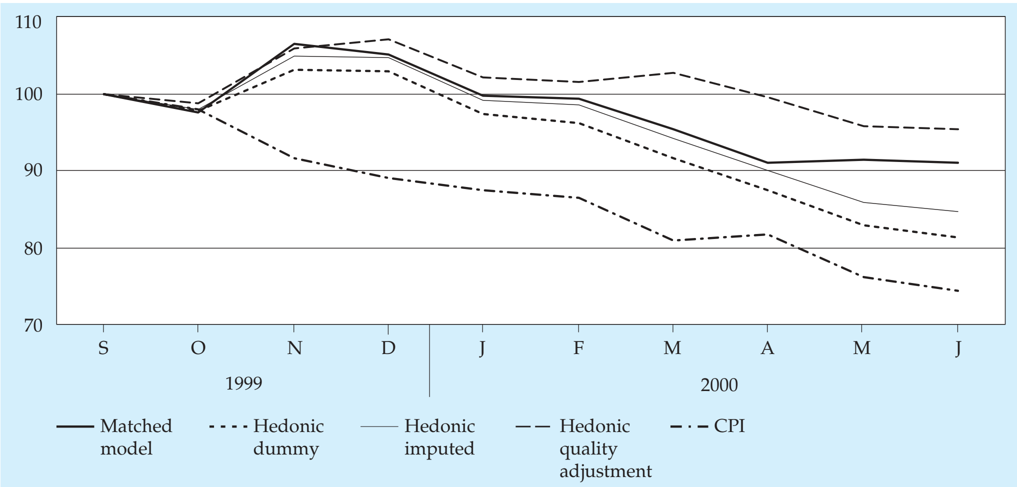
In addition to the three hedonic indices a traditional matched model index was also calculated using the entire database. The number of matched items was therefore larger than the number of observations in the CPI, which only uses systems with particular types of processors, and also deletes observations if systems with a particular processor are sold in less than three stores. The matched index only contains identical computers, offered at the same retail point. Table 4.3 shows the number of matched and unmatched items in each two-month period. The attrition rate is very high, which reflects the rapid model turnover in the computer market. Since no information on quantities is given, we have no clue as to whether this attrition rate also holds in the total value of sales.

Table 4.3
Number of matched, old and new items in the hedonic price indices based on the entire CBS database

Period	Matched	Old	New
September 1999–October 1999	56	80	63
October 1999–November 1999	79	40	28
November 1999–December 1999	38	69	95
December 1999–January 2000	51	82	62
January 2000–February 2000	93	20	30
February 2000–March 2000	36	87	86
March 2000–April 2000	75	47	32
April 2000–May 2000	15	92	98
May 2000–June 2000	12	101	110

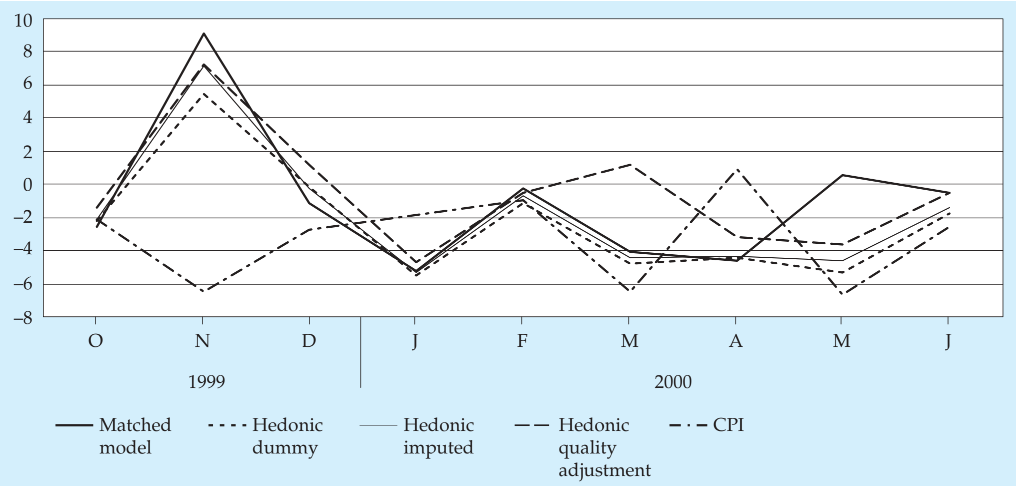
'Old' models are computers that are sold in the first of the two adjacent months, but not in the second. The reverse is true for 'new' models.

Figure 4.2
Hedonic and matched model price indices for computers, September 1999–June 2002 (September 1999=100)



The matched model index and the three hedonic indices are shown in figure 4.2. In the case of the imputation and hedonic quality adjustment methods, the Fisher-like index (equal to the geometric average of the Laspeyres-like and Paasche-like indices) is shown. For comparison, the official CPI is again included. Figure 4.3 shows the relative price change from the previous period for all indices.

Figure 4.3
Relative price change from the previous period (%) for several price indices for computers, October 1999–June 2000



The differences in price changes are the largest between the CPI and the other indices, including the matched model index. The matched model index and the three hedonic indices are fairly close to each other, with the exceptions of March 2000 for the hedonic quality adjustment index and May 2000 for the matched model index. Closer examination of the data revealed that the gap between the HQA index and the other indices in March 2000 is caused by the fact that for March and February, some computers with a difference in processor speed of more than 100 MHz were linked. When these were considered as old and new computers, and their prices were imputed, the gap nearly disappeared completely. This was true even if the their imputed price ratios were estimated using the coefficients of single month regressions rather than the time dummy coefficient from the adjacent-month regression. So the gap cannot be attributed to the fact that the hedonic quality adjustment method employed here explicitly uses the coefficients of the characteristics, whereas the other hedonic indices calculated here do not. This is surprising, since the hedonic quality adjustment method is not expected to give very different results from the imputation method.

The gaps between the CPI and the other four indices in November 1999 and April 2000 are largely the result of the sampling bias that is inherent to the CPI. For the period October-November 1999, some observations that are in the CPI but for which no advertisements were available showed a price decline whereas most other observations in this period increased in price. This occurrence provides a large part of the large gap between the CPI and the other indices in this period. The other matched model index, which uses all observations in the advertisements dataset regardless of processor type, is very close to the hedonic indices. The differences between the CPI and the other indices are therefore likely not caused by a difference in methodologies.
The fact that the matched model index closely tracks the hedonic indices for nearly the entire period suggests that for this data set, the computers included in the matched model index are representative for the non-matched computers. This would indicate that in this case a matched model index does not introduce a significant bias. However, one should note that the current database is far from perfect. Only observations from relatively small retailers (with often only two or three sales points) are included without any indications of the number of computers sold of each model. Whether this conclusion holds when a larger share of the market for personal computers is covered, is the topic of the following section.

4.4 *Alternative price indices using scanner data*

With the increased use of bar code scanning at retail outlets, a wealth of data on retail transactions has come about. These data are called scanner data and are usually collected by marketing agencies like GfK, AC Nielsen, DataQuest and others to perform market research. They are also increasingly being used by economists.[11] As scanner data contain information on prices, quantities sold and in many cases also on specifications, scanner data lend themselves very well for the construction of price index numbers and the application of hedonic regressions.
This section gives an account of research that was performed on a scanner data set on computer sales in the Netherlands. With these data matched model indices and hedonic prices indices are constructed.

Data

The data set provides monthly data on transaction prices, quantities sold and several characteristics of nearly all computers sold in the Netherlands from January 1999 to January 2002. GfK aggregates the data by outlet type, of which nine different types are distinguished: buying groups (BUYING), chain stores (CHAINS), computer stores (CS), department stores and mail order houses (DEPMOH), independent retailers (INDEP), office equipment retailers (OER), photo retailers (PRT), system and software houses (SH) and telecom specialists (TCS). According to the descriptions provided by GfK, two of these outlet types (OER and SH) mainly sell to businesses, whereas the rest supply the bulk of their sales to private consumers. All outlets with a substantial amount of computers sales are fully covered by GfK. In these cases GfK gets the sales data directly from the company's headquarters. For small companies with a limited number of outlets and sales a sample outlet is taken. The total number of sales of this outlet is then multiplied by the number of outlets of the particular company to provide an estimate of its total sales. Since the largest part of computer sales takes place in the bigger stores, the share of these 'extrapolated' sales is fairly small. Therefore the GfK data set provides a near full coverage of the universe of computer sales in the Netherlands.

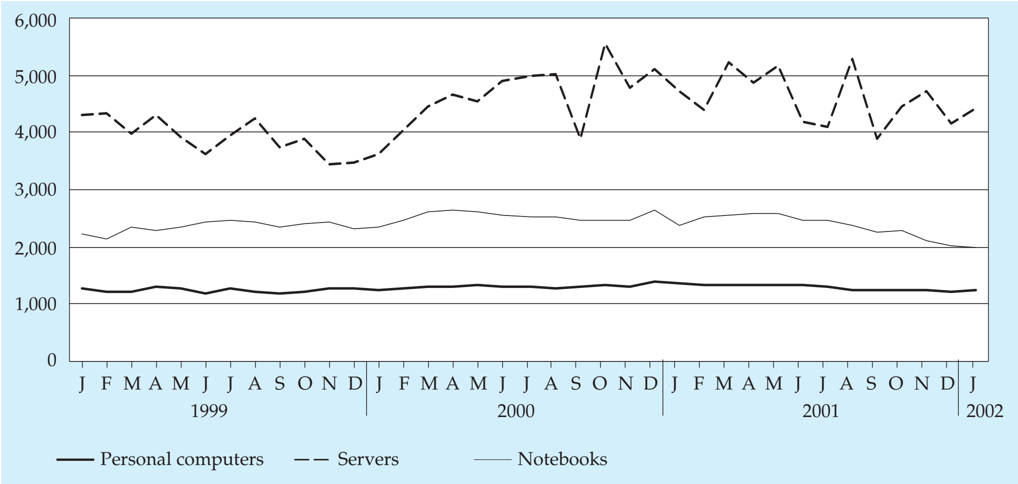
Individual computer models (hereafter called 'items') are given a unique product code based on the bar code. If bar codes correspond to unique computer configurations, we can be sure that a product match based on the product code is exact. For items with the same product code, the characteristics included in the data base were indeed identical. However, the data set only contains a limited set of characteristics, so we can only assume that the other unobserved characteristics are identical as well, to be certain matching is exact. Although they could not provide additional information on unobserved features, GfK ensured that these were identical too. The quality of indices based on these scanner data therefore depends on the plausibility of this statement by GfK.

Computers can be sold at different outlet types. Within an outlet type the price of an item is the average transaction price across all outlets, weighted by quantities sold. The prices in the data set are therefore actually unit values. Since each item is unique, only prices are averaged. Characteristics are the same for each outlet within an outlet type, and the quantities sold can simply be added.

The data set consists of three different kinds of computer equipment: personal computers (towers and desktops), notebooks and servers. Since these are different products, separate hedonic functions and price indices will be estimated for each. A summary of the data is provided in Appendix A. The number of computer sales follows a cyclical pattern that is especially strong in the case of PCs. Computer sales (measured in physical units) peak every March and December (with a relative decline in January and February), whereas the through is between May and August.

Since the main focus of this chapter is on the calculation of price indices, the average prices, or unit values, of computer equipment are shown here by way of comparison. A unit value for computers as presented here contains a strong inside the sample bias, because it explicitly matches computer prices regardless of changes in quality. Figure 4.4 tracks the unit values in euros of the three types of computer equipment. The unit values of notebooks and especially PCs remain fairly constant during the period under consideration, whereas the average price of servers shows a more erratic behaviour.

Figure 4.4
Unit values of PCs, notebooks and servers in euros, the Netherlands, January 1999–January 2002



As pointed out above, the number of characteristics in the data set is limited. The main performance characteristics are processor speed in MHz (SPEED), storage capacity of the hard disk in MB (HDISK) and the memory capacity in MB (MEMORY). These are the only quantifiable, continuous variables which are given for each type of computer. For notebooks there is one additional variable, namely the size of the screen in inches (SCRSIZE). The remainder of the characteristics is qualitative, and can be used in hedonic regressions only as dummy variables. These include the presence of a monitor (MONITOR) and USB port (USB), the screen type in the case of notebooks, the type of processor and the brand of the computer. If the item has no brand (which is the case, for example, when a consumer selects his own configuration), it is labelled as a clone.

The small number of physical characteristics is a disadvantage of the database. No information is available on graphical cards, sound cards, the type of computer, included operation system and application software, and so on. Therefore, we can expect that a relatively large part of the variation in prices cannot be explained by the variation in specifications listed above, unless there is a one-to-one relation between unobserved variables and observed ones, which is not likely. On the contrary, if unobserved variables are disproportionally correlated with included characteristics, then unobserved variables will bias the estimates of the coefficients of the hedonic regressions. This is true whether items with the same product code are exactly identical or not. If they are not, matched model indices will be biased, too. The latter is a potential drawback of the data base, as we have to trust the statement by GfK that matches are exact.

A strong point of the database is that it contains nearly all computer sales during this period, so we know that it is by definition a very good sample. The outside the sample bias is therefore likely small. Furthermore, quantity data is available, so that the relative importance of individual items is known. This allows us to calculate superlative indices which take substitution effects into account.

Hedonic regressions using scanner data

Outlet groups

For each type of computer separate hedonic functions were estimated. Before a decision could be made about the functional form of the regressions, which variables to include and whether or not to pool the data across time, the issue of the outlet type has to be resolved. The outlet where a computer is sold, is not a physical characteristic, but it does tend to influence its price. This is because different outlets have different pricing policies and consumers value services offered by different outlets. The 'a potato is a potato' rule mentioned in Chapter 2 is not used here.
A straightforward solution is to pool the data across outlets and add dummy variables for the outlet types. This is a rather awkward solution since it may bias the estimates of the other characteristics. The relation between characteristics and prices may not be the same across outlets, just as it may differ across time.
A better solution would be to condition for the outlet type and estimate separate regressions for each outlet type. There is, however, one big drawback that comes with this solution, i.e. the number of observations is very low in some cases. To avoid throwing away observations, the data were pooled across different outlet types. To determine how to group the outlet types the following procedure was used. For each equipment type, the data were pooled across all outlet types and months. Thus nine regressions were carried out for each equipment type, using a different outlet type as the base outlet each time. [12)]
If the coefficient of an outlet dummy was not significant, the assumption was made that there is no price differential between this outlet type and the base outlet of the regression, and that these outlets can therefore be pooled. For index number purposes, outlet types OER and SH were never pooled with any of the outlet other types, since the former mainly sell to businesses as opposed to the other types.
This procedure led to a subdivision into five, two and three groups for PCs, notebooks and servers respectively. Table 4.4 gives an overview of the outlet groups. Note that, apart from the distinction between 'consumer outlets' and 'business outlets', there is no a priori expectation on which outlet types will be similar. Instead it is purely an empirical matter.

Table 4.4
Division of outlet types into groups

Personal computers					Notebooks		Servers		
Group 1	Group 2	Group 3	Group 4	Group 5	Group 1	Group 2	Group 1	Group 2	Group 3
BUYING PRT TCS	CHAINS DEPMOH INDEP	CS	OER	SH	BUYING CHAINS CS DEPMOH INDEP PRT TCS	OER SH	CHAINS CS INDEP	OER	SH

Choice of variables

Since the number of quality characteristics in the data set is limited, there is not much room for different selections. All variables which are clearly associated with both user value *and* resource costs, were included in the hedonic functions. These include SPEED, HDISK, MEMORY, MONITOR, and USB. For notebooks, SCRSIZE and dummies for the different available screen types were included as well.

In all regressions two sets of dummy variables remain, i.e., those for processor type and those for brand. Both variables are no clear performance indicators, but are proxies for performance. In the case of processors it is well known that the speed in MHz alone is not a sufficient indicator of its performance. For example, a Pentium processor is considered to be better than a Celeron processor with the same clock speed in MHz. However, any information on other performance characteristics of processors is lacking from the data set, so dummies for the type of processor were included in the regressions. Depending on the type of computer and the month, different processor types were chosen as the base type.

The brand of a computer is a more thorny issue. A brand may indicate all kinds of price determining factors, like price mark-ups, unobserved performance characteristics and so on. Because of the proxy character of this type of variable, without clarity about what is actually proxied, there is a case against including them in the regressions. However, they do appear to have explanatory power, so leaving them out would decrease the fit of the model. For this reason, they were included. [13)]

Functional form and pooling

Like with the CBS data set, the three main functional forms of hedonic regressions are compared, i.e., the linear, semi-logarithmic and double logarithmic specifications. For the data at hand, the double logarithmic specification proved unsuitable. Some items did not include a hard disk, which implies a value of zero for the relevant characteristic. This makes it inappropriate to take logs. This is especially relevant in the case of servers, for which most items are sold without a hard disk. The goodness-of-fit of both the linear and semi-logarithmic specification was tested for regressions of all ten groups listed in table 4.4. These regressions pooled the data across all months. In each case the semi-logarithmic specification proved to have the best fit. This specification was therefore chosen for all hedonic regressions.

To reflect the relative importance in sales of the different items all regressions are weighted least squares (WLS). The physical quantity sold was chosen as the weight in the WLS regressions.

For the same reasons as discussed in the previous section, the data were pooled for adjacent months, resulting in 36 regressions for each of the ten groups of data. An additional reason is that the number of observations in individual months is still fairly low for some groups despite the aggregation of multiple outlet types. For example, with single-month regressions in the case of servers the number of observations would be lower than the number of explanatory variables in several cases. Pooling data for two adjacent months increases the number of observations and therefore reduces the variance of the regression coefficients. Again, no testing on the implied fixed effects was carried out, as the data enforces pooling.

Summing up, for personal computer the following equation is estimated for each combination of outlet group and adjacent months: [14)]

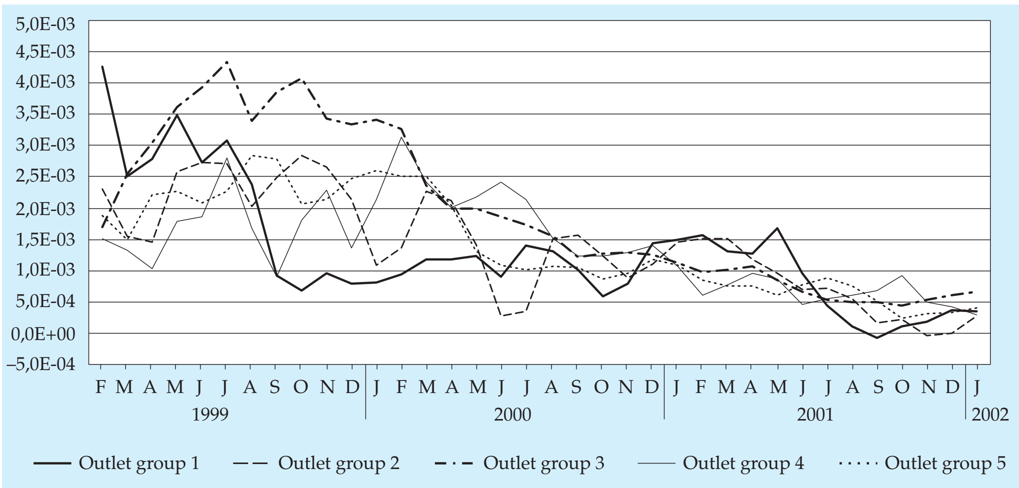
$$\begin{aligned}\ln p_i &= \beta_0 + \beta_1 SPEED_i + \beta_2 HDISK_i + \beta_3 MEMORY_i + \beta_4 MONITOR_i \\ &+ \beta_5 WORKSTAT_i + \beta_6 USB_i + \sum_{j=1}^{30} \gamma_j PTYPE_{ij} + \sum_{k=1}^{55} \delta_k BRAND_{ik} + \tau T_i + \varepsilon_i \end{aligned} \qquad (4.3)$$

Where PTYPE is a dummy for the type of processor of the computer, and BRAND for its brand name. There are 31 different processor types and 56 different brands in the entire database, including generic computers (labelled as ‘clones’). Different processor types are chosen as the reference type to omit from the regression. In the case of BRAND, generic computers are always chosen as the reference.

Stability of coefficients

The division of the entire data set into ten outlet groups and thirty-six adjacent month regressions leads to a total of 360 regressions, with up to 41 independent variables in one regression. Listing all coefficients of each regression here would take up a lot of space; they will be published online as a complement to this thesis.[15)] It is noted that the three main physical characteristics, SPEED, HDISK and MEMORY, are significant in most regressions. These regression results are summarised in figures 4.5 to 4.7. To save space, only regression coefficients are shown for personal computers, which is by far the largest of the three types in terms of quantity sold.

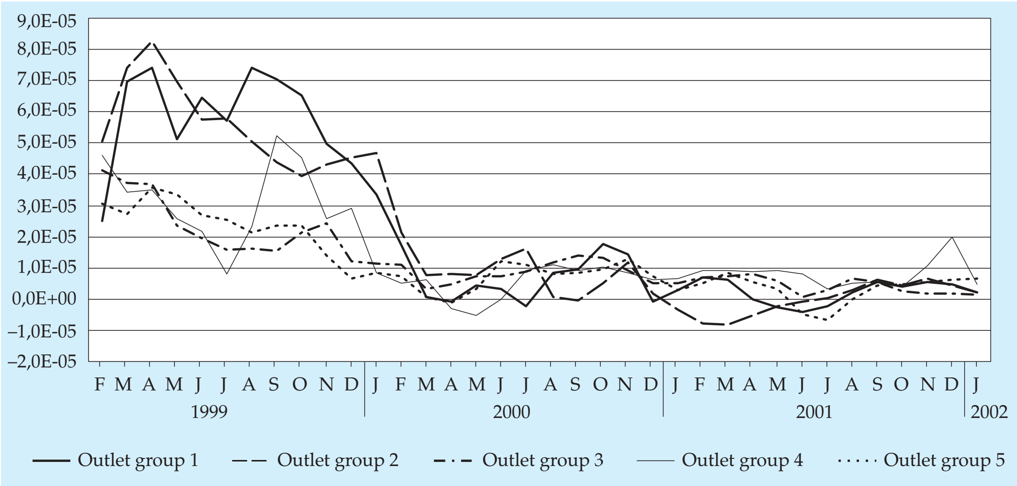
Figure 4.5
Coefficients of variable SPEED, adjacent month regressions January 1999/February 1999 to December 2001/ January 2002, PCs



Note: months on the horizontal axis are the second months in the adjacent month regressions.

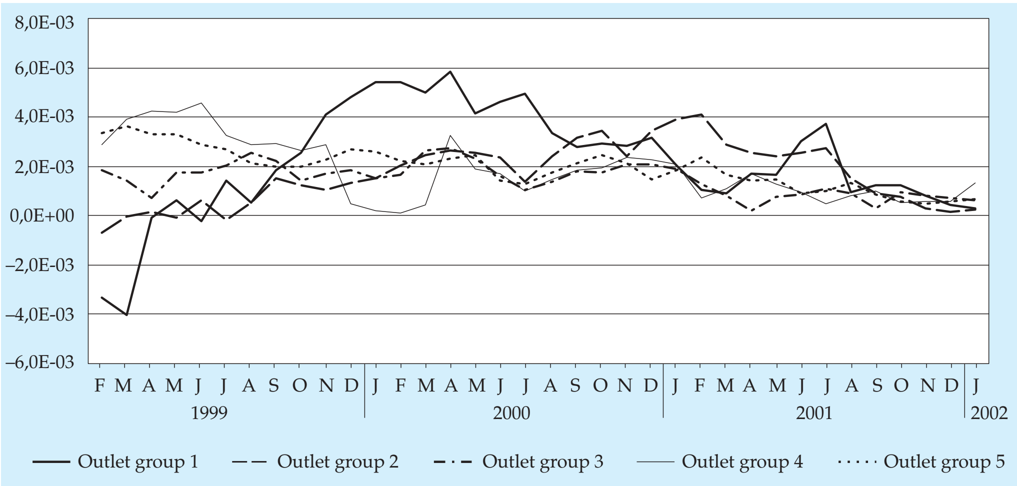
Although the coefficients of the three characteristics are generally significant, figures 4.5 to 4.7 indicate that the coefficients are not constant across outlet groups or time. A downward trend can be witnessed for SPEED and HDISK, and a convergence of the coefficients of the different outlet groups for all three physical characteristics. The downward trend is not surprising, both from a user point of view as from a producer one. The production costs especially of additional units of computation power and hard disk capacity are decreasing at a fast pace, which is reflected in the falling coefficients of these characteristics. On the other hand, buyers' valuation of additional units of these features decreases over time as well because of the fast rate with which computers become more powerful. This is also consistent with the decline in the coefficients of these characteristics. This finding is more or less replicated for notebooks, although the pattern for servers is much more erratic.

Figure 4.6
Coefficients of variable HDISK, adjacent month regressions January 1999/February 1999 to December 2001/ January 2002, PCs



Note: months on the horizontal axis are the second months in the adjacent month regressions.

Figure 4.7
Coefficients of variable MEMORY, adjacent month regressions January 1999/February 1999 to December 2001/ January 2002, PCs



Note: months on the horizontal axis are the second months in the adjacent month regressions.

Note that the falling coefficients are a consequence of choosing a semi-logarithmic specification. In a double logarithmic specification, the coefficients express the cost and value of additional *percentages* of processor speed, memory size or hard disk capacity. The cost and valuation of percentage rather than unit changes of these features are likely to remain more constant over time.

Unmeasured variables may also have affected the coefficients of the characteristics above mentioned. In addition, the relatively low number of characteristics seems to have affected explanatory power of the regressions. This is indicated by the low values for the adjusted R^2 and the high values of the standard error of the regression (SER). The average values for adjusted R^2 for the five PC outlet groups across the adjacent month regressions are 0.75, 0.69, 0.63, 0.62, and 0.57, respectively. The averages of the SERs are 0.88, 0.89, 1.32, 1.08 and 2.01, respectively. These figures indicate that the goodness of fit of the regressions is rather mediocre. In many hedonic regressions, especially for computers, R^2 is close to, and not rarely above, 0.9. The high SERs point at a large spread in the data, which is of course the case given the large number of processor and brand dummies.

These regression estimates suggest that pooling the data, even on a bimonthly basis, is not justified. The reasons that I nevertheless stick to it have a practical nature, and were already mentioned before: the sometimes low number of observations in single months and the problem of characteristics that are available in one month, but not the adjacent one.

Price indices using scanner data

Matched model indices

The main purpose of the project carried out at Statistics Netherlands that was the basis for this chapter is to investigate several alternative methods to construct a price index for computers. Since the standard methodology of Statistics Netherlands and other statistical offices is the matched model method, it is only logical that this method is investigated with the present data as well.

The standard approach of most index numbers is to choose a fixed reference period, and to match prices of products in subsequent periods with prices of the same products in the reference period. If old products disappear or new products are introduced on a frequent basis, the odds are that matching like with like gets increasingly difficult as time progresses. This phenomenon is referred to as ‘sample degradation’ and is a major criticism aimed at conventional price index measurement.

A way to get around this problem is to calculate a chained index, resampling and reweighting every period. The case to do this for computers is very strong, as the average lifecycle of individual items is only a few months on average.

Below prices are matched only for item-outlet combinations, i.e., the price of an item was only matched if it was sold at the same outlet type, to prevent outlet price effects from biasing the index. For each outlet type and computer type, matched model indices were calculated both with a fixed reference period (January 1999) and shifting reference periods, where prices in each period were matched with those of one month earlier. The latter results in a chained index.

For each period, the indices of all outlet types were weighted with their relative sales values:

$$P_L^M = \sum_{g=1}^{G} w_{L,g}^M P_{L,g}^M \tag{4.4}$$

$$P_P^M = \sum_{g=1}^{G} w_{P,g}^M P_{P,g}^M \tag{4.5}$$

with

$$w_{L,g}^M = \frac{\sum_{i=1}^{n_g} p_i^0 q_i^0}{\sum_{g=1}^{G} \sum_{i=1}^{n_g} p_i^0 q_i^0} \tag{4.6}$$

$$w_{P,g}^M = \frac{\sum_{i=1}^{n_g} p_i^0 q_i^1}{\sum_{g=1}^{G} \sum_{i=1}^{n_g} p_i^1 q_i^1} \tag{4.7}$$

where P_L^M and P_P^M are the Laspeyres and Paasche indices for the matched items of all G outlet types, with respective weights $w_{L,g}^M$ and $w_{P,g}^M$, which are shares in total sales value of matched items in outlet type g.

Laspeyres and Paasche indices with a fixed reference period (with January 1999 as the reference period) and chained weights were calculated. The resulting Fisher indices (the geometric average of Laspeyres and Paasche) are shown in figures 4.8 to 4.10. The matching percentages as a share of total expenditure in the base month (the previous month in the chained case, January 1999 for the fixed weight case) and in the current month, remain relatively constant. The matching percentage of the chained indices for computers is for the entire period on average 81.3% of previous period expenditure share and 78.0% of current month expenditure share on average. For notebooks these average shares are somewhat higher at 87.8% and 84.5% for previous month and current month expenditure share, respectively; for servers they are substantially lower, at 65.0% and 65.1% for previous month and current month respectively.

Figure 4.8
Matched model indices with fixed reference period and chained weighting, PCs, all outlet types (January 1999=100)

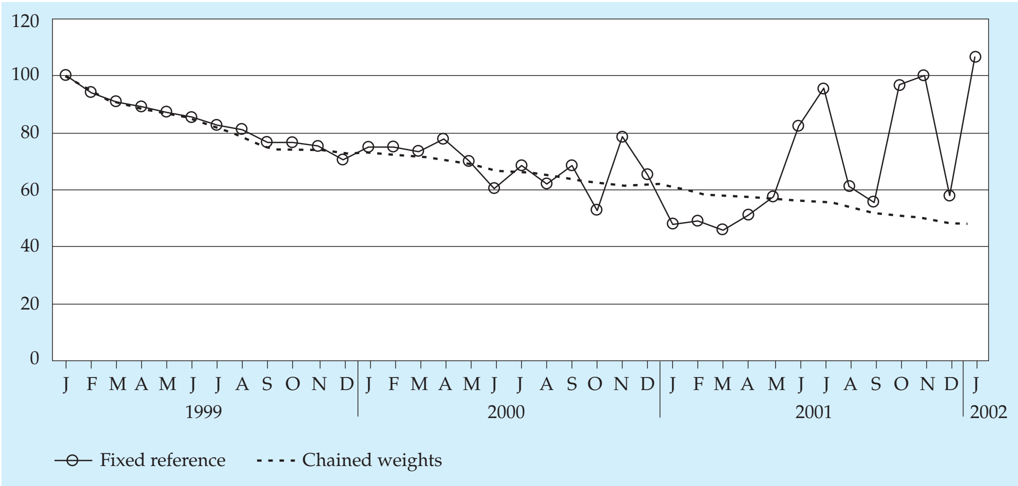
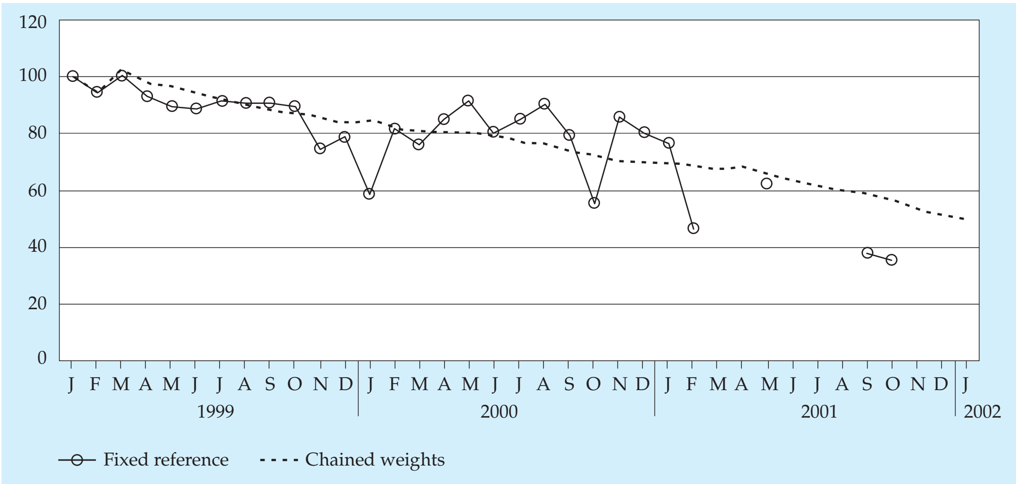
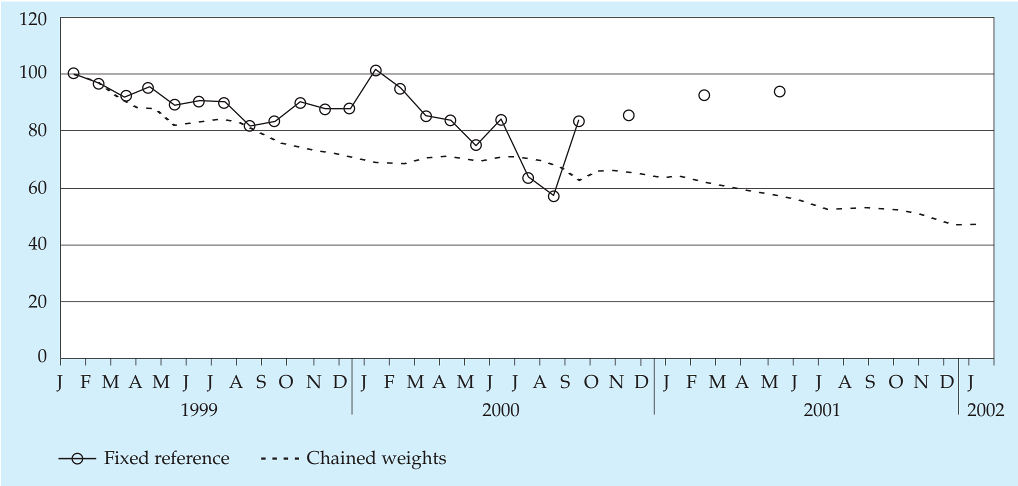


Figure 4.9
Matched model indices with fixed reference period and chained weighting, notebooks, all outlet types (January 1999=100)



For the fixed period index the number of matched observations declines steadily over time. After three years, there remains not a single item for which prices can be matched with the base period in the case of notebooks and servers. For PCs, the share of items for which a match is still possible after three years is negligible; after one year, the expenditure shares of such items has already declined to 24.7% and 3.0% for base and current month respectively. This means that, as expected, sample degradation appears to be very severe. Monthly expenditure shares of matched items are shown in Appendix A.

Figure 4.10
Matched model indices with fixed reference period and chained weighting, servers, all outlet types (January 1999=100)



All this means that the turnover rate is extremely high in the data set used here. For comparison, Pakes (2002) finds an annual turnover rate for computers in the U.S. of 20%;[16] Silver and Heravi (2002a) find a similar annual figure in a scanner data set for appliances in the U.K.
As time passes the fixed weight matched index shows an ever more erratic behaviour, which is caused by the decreasing number of items that can be matched. This gives support to those who claim that the matched model method as applied normally at statistical offices fails in the case of computers, since the index weights are held fixed over a too long period of time.

Hedonic indices

Using the hedonic regression results, three hedonic indices are calculated for each type of equipment: a dummy index and two different imputation indices. A hedonic quality adjustment index is not calculated here, as it is very close to a traditional hedonic imputed index. The time consuming task of calculating such an index is probably not worth the effort.
The hedonic dummy indices that are presented here are chained indices using antilogs of the coefficients of the time dummies of the adjacent month regressions. This is a fixed weight index, since observations are not weighted by their relative shares in sales. The regression results are based on weighted least squares regressions, where quantities are used as weights. Weights are therefore used only implicitly.
Indices based on the imputation method are analysed here in a bit more detail. The standard way in which this method is applied was discussed in Section 3.4. Missing prices are estimated with a hedonic function and these prices are matched with actual prices. This procedure has its drawback, however, when the explanatory power of the hedonic regression is low.

The regressions calculated here show relatively low values for R^2, and given the presence of quite a few unobserved variables, the coefficients of the observed variables are likely to be biased. This can have major implications for the residuals of actual prices. If residuals are biased, then the matching of actual with estimated prices will take over this bias. To prevent the residuals from entering the index, actual prices can be replaced with fitted estimated prices as well. This procedure is known as 'double imputation' (De Haan and Opperdoes, 2002). The results presented here include both traditional, 'single imputed', and double imputed hedonic indices.

As I am estimating adjacent month regressions rather than single month regressions, the ratio of estimated prices in the double imputed index equals the antilog of the coefficient τ of the time dummy in equation (4.3), and makes no direct use of the coefficients of the other explanatory variables. Rewriting (3.13b) and (3.15b), the hedonic double imputed Laspeyres and Paasche indices can be written as:

$$P_L = w^{0M} P_L^M + (1 - w^{0M}) e^{\hat{\tau}} \tag{4.8}$$

$$P_P = \left[\frac{w^{1M}}{P_P^M} + \frac{(1 - w^{1M})}{e^{\hat{\tau}}} \right]^{-1} \tag{4.9}$$

where w^{0M} and $(1 - w^{0M})$ are the base month expenditure shares of matched and disappearing items, respectively; w^{1M} and $(1 - w^{1M})$ are current month expenditure shares of matched and new items. P_L^M and P_P^M are Laspeyres and Paasche indices for matched items only. The double imputed hedonic indices are therefore weighted averages of matched model indices and hedonic dummy indices.

The double imputation hedonic method is somewhat related to the IP-IQ method described in Chapter 2, as the price change for new and old models is set equal to the overall price change of all models. It will therefore lie closer to a matched model index than a single imputed hedonic index.

Following De Haan and Opperdoes (2002), the Fisher index, which is the geometric average of the Laspeyres and Paasche indices, can be rewritten in the following way:

$$P_F = \left(P_L P_P \right)^{½} \qquad \Leftrightarrow$$

$$P_F = \left(\frac{P_L}{P_L^M} P_L^M \, \frac{P_P}{P_P^M} P_P^M \right)^{½} \qquad \Leftrightarrow$$

$$P_F = \left(\frac{P_L}{P_L^M} \right)^{½} \left(\frac{P_P}{P_P^M} \right)^{½} \left(P_L^M P_P^M \right)^{½} \tag{4.10}$$

Next, define:

$$\lambda = \left(\frac{P_L}{P_L^M} \right)^{1/2}$$

$$\pi = \left(\frac{P_P}{P_P^M} \right)^{1/2}$$

and rewrite (4.10) as:

$$P_F = \lambda \pi P_F^M \tag{4.11}$$

where $P_F^M = \left(P_L^M P_P^M \right)^{1/2}$ is the Fisher index of matched items only.

The factors λ and π can be interpreted as the effects of disappeared items on the Laspeyres index and new items on the Paasche index respectively. If the shares of disappearing and new items are small (i.e. w^{0M} and w^{1M} are close to one), then P_F^M is a close approximation of P_F. This is true in the case of two periods, but if the Fisher indices are combined into a chained index, small differences between P_F^M and P_F can have substantial effects in the long run (like, say, thirty-six periods).

However, we cannot say a priori whether P_F^M will overstate or understate P_F. This depends on the net effect of λ and π combined. Whether $\lambda\pi > 1$ or $\lambda\pi < 1$ is an empirical matter, and will be investigated here. However, economic theory may provide some insight in the sizes of λ and π. In a competitive and transparent market, where consumers have full information, we expect to find that the Laspeyres price index of matched items is smaller than the index of disappearing items, and so $\lambda > 1$. Demand for disappeared items has fallen to zero, which may be caused by their obsolescence. This may imply that the imputed prices of these items are too high as compared to items with continued sales. On the other hand, retailers sometimes offer old computer models at discount prices to clear shelves, and make room for newer models. [17)] If this effect is stronger, this will result in a value of $\lambda < 1$.

Likewise, the economics of new goods implies that new models are likely to have high base period prices, had they been available. In this respect the concept of Hicks' reservation price is sometimes put forward. This is the price that sets the demand for a product just equal to zero (Hicks, 1940; Triplett, forthcoming). Imputed prices of new goods are therefore likely to be 'too low' as compared to items that were available previously (i.e. lower than the reservation price). This leads to the expectation that the Paasche index of matched items is larger than the index of new items, and so $\pi < 1$. But the effect of new models can also go in the other direction. When introduced, prices of new computers sometimes contain a premium, which is based on their newness and exclusiveness, leading to a value of π which is greater than 1.

Before I turn to the comparison between the CPI and the price indices based on the scanner data, I first pay some attention to the differences between the matched model indices and hedonic indices that were estimated with the GfK data set. For all three types of computer equipment, four indices are calculated for each outlet group: a matched model index, a hedonic dummy index (based on the time dummies from the adjacent month regressions) and two hedonic imputation indices (based on the imputation methods described above). In all cases, bimonthly Fisher indices were chained over the entire period. The indices from the individual outlet types were aggregated with the expenditure shares of the outlet types to derive aggregate indices for the three types of computer equipment.[18] The resulting indices are represented in figures 4.11 to 4.13. To conserve space, table 4.5 only shows the average monthly price changes of all indices, including the average effects of old and new items. Monthly indices are included in Appendix A.

Table 4.5
Average monthly price changes (%) for different indices and effects of old and new models, January 1999–January 2002, all equipment types

	Personal computers	Notebooks	Servers
Price change:			
Matched model	–2.0	–1.9	–2.1
Hedonic dummy	–3.2	–2.4	–2.6
Hedonic single imputed	–2.5	–2.2	–2.4
Hedonic double imputed	–2.3	–2.0	–2.3
Effect of old models:			
Hedonic single imputed	–0.5	–0.2	–1.0
Hedonic double imputed	–0.1	–0.1	–0.1
Effect of new models:			
Hedonic single imputed	0.0	–0.1	0.6
Hedonic double imputed	–0.1	0.0	–0.2

Figure 4.11
Hedonic and matched model Fisher indices for PCs, the Netherlands, all outlets (January 1999=100)

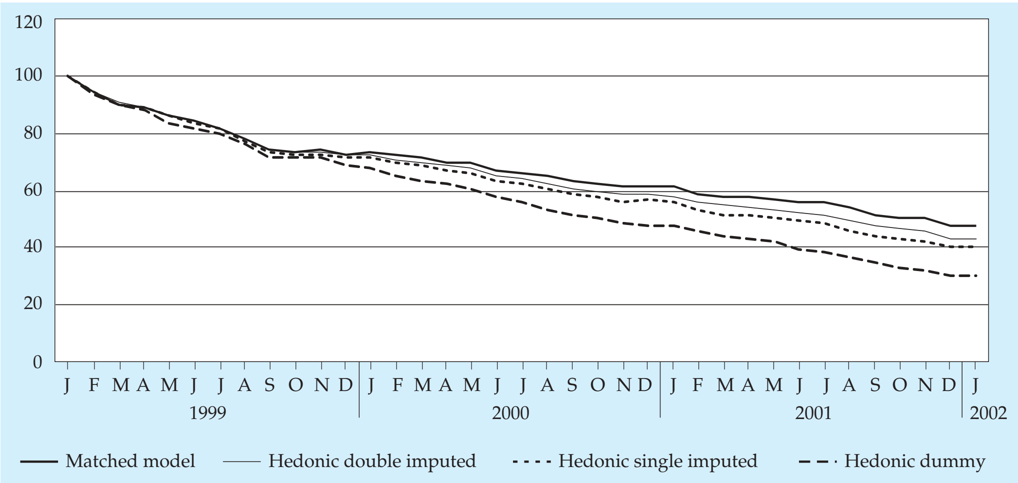


Figure 4.12
Hedonic and matched model Fisher indices for notebooks, the Netherlands, all outlets (January 1999=100)

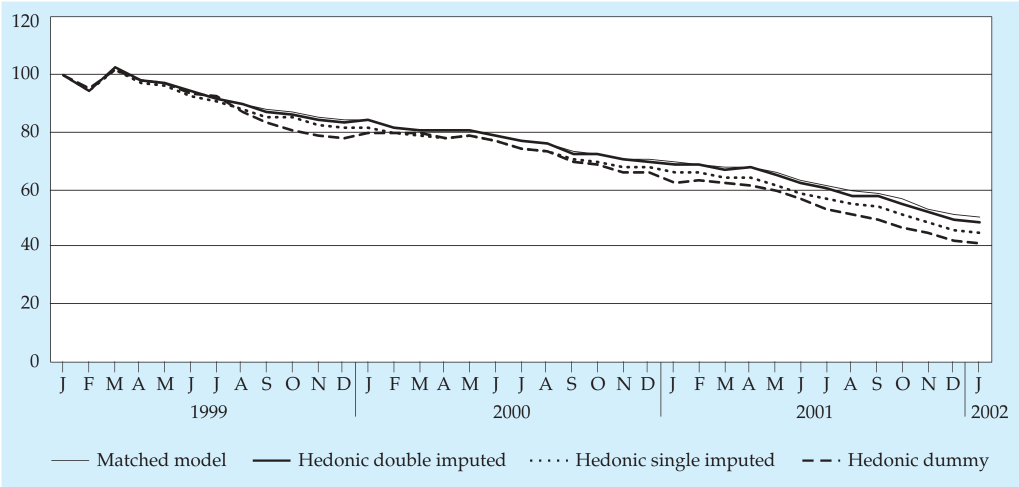
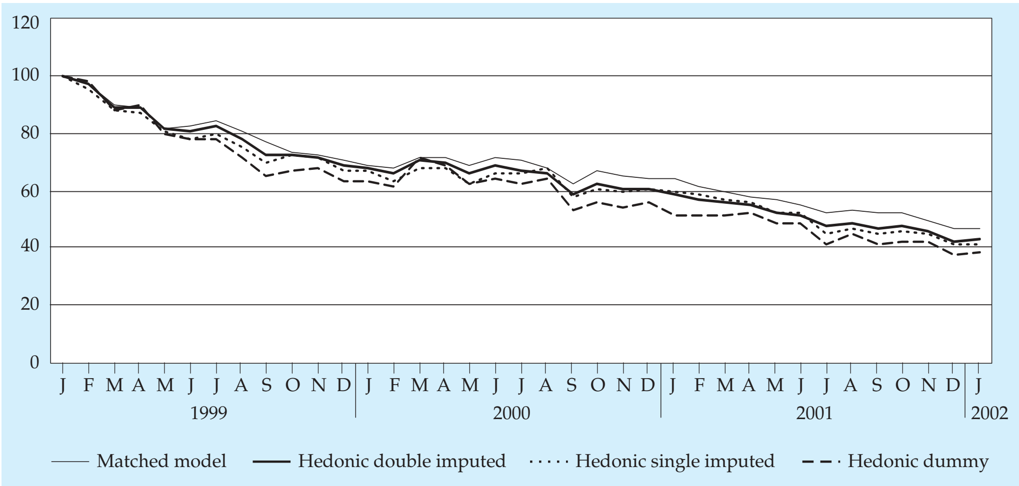


Figure 4.13
Hedonic and matched model Fisher indices for servers, the Netherlands, all outlets (January 1999=100)



The fact that the matched model indices are above the hedonic indices is not surprising on theoretical grounds. The current matched model procedure leaves out unmatched observations, and therefore implicitly assumes that the price change of unmatched items equals the price change of matched items. As discussed by Triplett (forthcoming), this 'imputed price change, implicit quality adjustment' (IP-IQ) method will introduce an upward bias in the price index if prices are falling, which is the case here. These results are in line with Barzyk and MacDonald (2002) and Evans and Scherrer (2002).

The differences between the hedonic imputed indices and the matched model indices are relatively small on a month-to-month basis. The differences are larger for PCs than for notebooks and servers. As expected, the single imputed hedonic index is below the double imputed hedonic index. The dummy index is in all cases below the other three. This reflects the importance of using weights in indices. As all regressions are WLS with quantities as weights, the dummy index is only implicitly weighted. The other indices use expenditure shares as weights in a Laspeyres-Paasche-Fisher framework, and therefore make proper use of weights. The lower dummy index suggests that computers with a rapid price decline are over-represented in the dummy index. This is confirmed when unweighted versions of the other three indices are computed. Although not shown here, the unweighted matched model indices are substantially below the weighted matched model, and the unweighted hedonic imputed indices nearly coincide with the hedonic dummy indices. Given the differences between the several indices shown in table 4.5, one can conclude that the effect of weighting (the difference between the hedonic imputed indices and the dummy indices) is roughly equal to or exceeds the effect of using hedonics or not (the difference between the hedonic imputed indices and the matched model index). Note that both effects are in a different direction. Both effects, however, pale in comparison to the effect of using a chained principle with shifting reference periods rather than a fixed base period, as shown in figures 4.8 to 4.10.

A large share of matched items can explain the similarity between the hedonic imputed indices on the one hand and the matched model index on the other. In such a case, the effect of old and new items will be minimal. However, in the current data base, 15–35% of all observations is not matched, a substantial amount. This suggests that matched items give a reasonable representation of the entire market, although this conclusion is somewhat weaker for personal computers than for notebooks and servers.

In all cases but one, the separate effects of old and new items is negative, and both effects are usually small. These effects are of course larger in an absolute sense for the single imputed hedonic indices, as the price change of the new and old items in the double imputed hedonic index is imputed from all computers rather than just old and new ones. The results of the single imputed hedonic index for servers are somewhat different. The effects of new and old items are relatively large, but of opposite sign. This may be caused by the distinct properties of servers, which serve different purposes than personal computers and notebooks. The difference between the matched model index and the single imputed hedonic index is nearly entirely caused by computers that exit from the market. A likely explanation is that old and obsolescent computers are dumped for bargain prices to clear shelves.

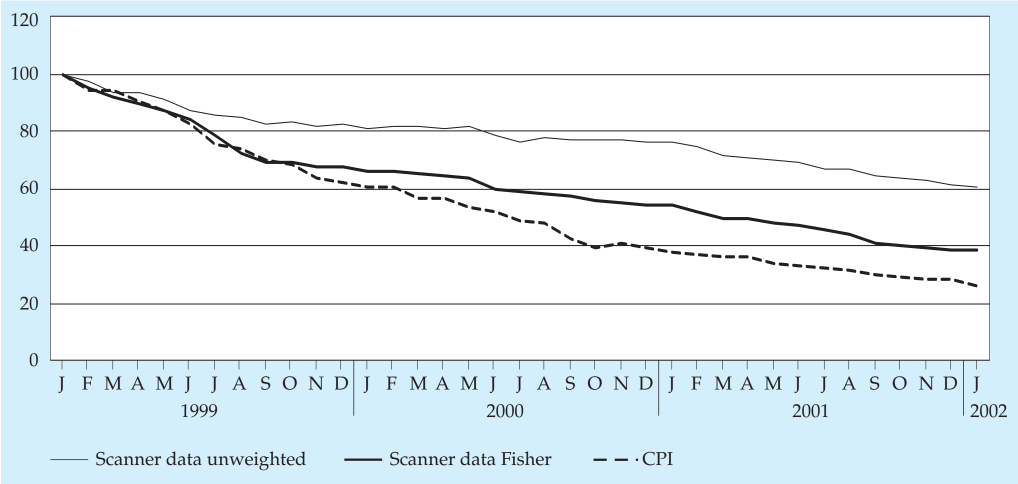
Although the month-to-month differences between the indices are small in most cases, these differences cumulate to larger gaps over longer periods of time. This may justify using hedonic methods, especially for PCs, although their benefit is dwarfed by that of using a chained index.

A reason for concern is the low explanatory power of the hedonic regressions that are estimated with (4.3). No information on a number of possible important characteristics is available, which results in collinear coefficients of the included variables, a bad fit, and a large variance of the time dummy coefficients. A better data set with more quality characteristics might increase the significance of the difference between a matched model index and a hedonic imputed index. However, it is not clear whether such a data base exists for computers as yet. Alternative large data sets, such as those used by Barzyk and MacDonald (2002) and Evans and Scherrer (2002), which contain a lot more information on quality characteristics, are based on manufacturer catalogues, and do not provide information on sales data. Such data sets may be used to test the difference between unweighted matched model indices and hedonic imputed indices, but not for superlative indices which include sales data.

Confrontation with the CPI

Up to this point, the focus of this chapter has been a comparison between hedonic and matched model indices in the CBS and scanner data bases. Here, I will make a comparison between the two datasets, to judge the accuracy of the Dutch CPI for computers when confronted with the dataset comprising the near universe of computer sales. For this purpose, two adjustments are made: first, the indices that were shown in the previous section are aggregates of all outlet groups, both business-to-consumer and business-to-business outlet types. Since the CPI, which contains only PCs, by definition only applies to sales to private consumers, only price indices which apply to 'consumer outlets' will be compared with the CPI. For PCs, these are the outlets in groups 1 to 3 (see table 4.4). Second, because the CPI is an unweighted (arithmetic) index, an unweighted index using the scanner data is constructed as well.
Figure 4.14 compares the official CPI with two matched model indices for consumer outlets in the scanner data set, a Fisher index and an unweighted arithmetic price index. The differences between the several indices are very large. The average monthly price changes of the unweighted scanner data index, the Fisher scanner data index and the CPI are –1.4%, –2.6% and –3.7%, respectively.
These gaps between the CPI and the scanner data indices of about one and two percentage points *per month* are very large. If we consider the matched model index using the scanner data for all consumer outlets as the 'true' price index for computers in the Netherlands, the Dutch CPI for computers contains a substantial *downward* bias, contrary to popular belief concerning official price index numbers for computer equipment. If the CPI would be calculated using a geometric average of price ratios rather than an arithmetic average, this downward bias would be even larger, given the well-known upward drift in chained arithmetic indices (Diewert, 1995).
This bias may be attributed to three different factors: methodology, sampling and weighting.

Figure 4.14
Official CPI for computers vs. scanner data matched model indices, the Netherlands, all consumer outlets (January 1999=100)

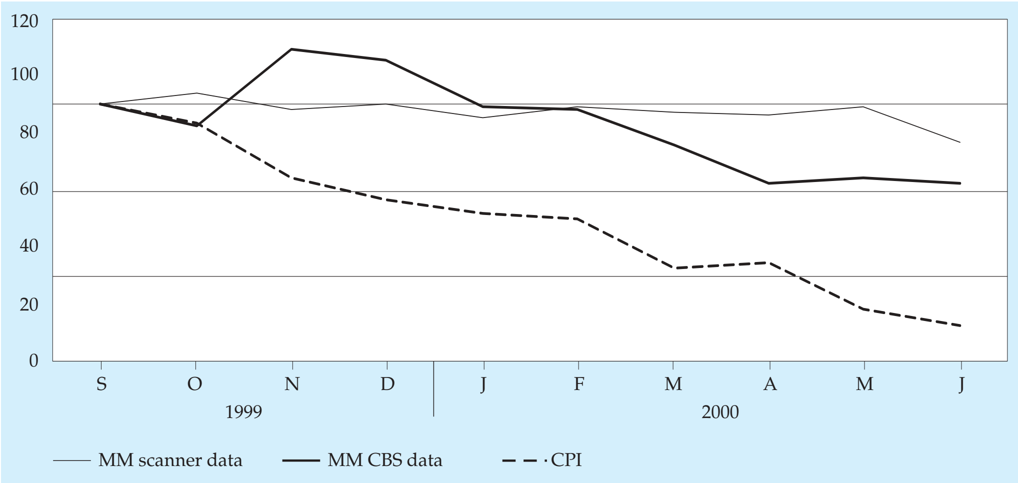


The method that is used to construct the CPI cannot explain the large difference. The CPI is calculated with a matched model methodology, where only identical computers are matched, similar to how the unweighted matched model index on the scanner data set was constructed.

Sampling and weighting issues are more likely candidates for the widening gap between the matched model indices and the CPI. On sampling, the only difference between the CPI and the unweighted scanner data index is that the former uses a limited sample of observations. The CPI contains only several dozen observations each month, and this number is even in decline, whereas the scanner data indices presented in figure 4.14 contain on average 460 matches per month, with the number of matches increasing through time. This suggests that the price sample of the CPI does not give a good representation of the Dutch PC market. Additionally each computer is given the same weight in the index. This is a second drawback of the official index, but one that is standard practice in CPI measurement in general, namely that at the level of price collection of individual items, which together comprise 'basic headings', no weighting scheme is applied. On the other hand, a geometric unweighted price index on the scanner data (not shown in figure 4.14) nearly coincides with the scanner data Fisher index. For a geometric average index, which is more adequate than an arithmetic one, weighting therefore seems not an important issue for the universe of computer sales, where the law of large numbers holds. But in a sample of several dozen observations, single observations may have unduly influence, so weighting remains important there.

This conclusion concerning sampling and weighting is confirmed when we compare the results of the scanner data analysis with those of the exercise using the CBS data base. Figure 4.15 combines three indices from figures 4.2 and 4.14, for the period where all three are known (September 1999–June 2000): the official CPI, the (unweighted) matched model index using the entire CBS data base on computer prices, and the unweighted matched model index using the scanner data set for all consumer outlets.

Figure 4.15
Official CPI for computers vs. matched model indices using the CBS and scanner data base



The unweighted matched model index based on the CBS data set (labelled 'MM CBS data'), which uses more observations than the CPI, appears to lie closer to the unweighted scanner data index than the official CPI. However, this is to a large extent caused by the strange behaviour of the 'MM CBS data' index in October-November 1999, discussed in the previous section. Excluding this period, both indices based on the CBS data set on average differ from the unweighted scanner data index by about the same amount. This suggests that simply extending the official CPI with observations that are presently not used provides no quick solution to the downward bias in the CPI. Drawing a better sample of computers, possibly including sales data, therefore still seems necessary.

4.5 *Summary*

Using a database on computer prices and characteristics that is employed by Statistics Netherlands to construct the CPI for computers, several hedonic indices were calculated, together with a matched model index. The results seem to indicate that for this data set, high frequency matched model indices introduce a relatively modest quality bias in the price index for computers. The same conclusion was reached with the use of a scanner data set, although there are substantial differences between the matched model indices using either data set.

Therefore, it seems that the largest drawback of the indices (both hedonic and matched model) calculated with the data collected by CBS is the data set itself. The first flaw of the data is that weights are not provided. No information on the sales volume is given for the models in the sample, so that popular items and infrequently sold configurations are treated equally. But since index number construction at the lowest level of aggregation is usually unweighted, one could direct this criticism at the entire CPI. A second point of concern is that the sample for the CPI can be improved. Together these two flaws lead to a substantial difference of about one percentage point per month between the unweighted official CPI and a Fisher index based on a scanner data set, and a difference of more than two percentage points per month between the CPI and an unweighted scanner data index. Since the scanner data set used here contains virtually all computer sales in the Netherlands, we can assume that a Fisher index for this data set equals the actual matched model price index for computers, and that therefore the CPI has a downward bias of one percentage point per month, barring quality biases.

The difference between matched model indices and hedonic (imputed) indices is less clear cut. On the one hand, these are smaller than sometimes is found. On the other hand, the differences may be small, but they are not insubstantial. Concerning hedonic indices, a reasonable conclusion based on the results found in this chapter would be that if data on characteristics are available, hedonic methods should be pursued, and the differences between hedonic indices and matched model indices should be evaluated. The fact that these differences are modest in the case of computers, a product with a fast rate of quality change, might indicate that for other products with slower quality changes the differences between both basic methods are negligible. But this is not a forgone conclusion, and needs to be determined empirically.

The results presented here point out that a more important point than the question of using hedonics or not is that of using a chained index principle rather than a fixed base. Resampling may not be necessary on a monthly basis for every product, but holding the reference period fixed for several years is definitively not an option. Furthermore, efforts should be made to draw a good sample, if possible with sales information. The chained index principle is already employed in the Dutch CPI for computers, but this is not the case everywhere and for each product.

The advantage of a high frequency matched model index compared with the hedonic method is that price statisticians are very familiar with the matched model procedure. Furthermore, although no *explicit* quality adjustment is carried out with this methodology, it entails an *implicit* quality adjustment, as pointed out in Chapter 2. This again calls for attention with respect to the sampling of observations, so that the possible outside the sample bias caused by the class mean method remains minimal. However, when a good sample is drawn with the purpose for constructing a high frequency matched model index, a hedonic index is usually possible as well. After all, information on characteristics is used for the matching of observations, and this information may as well be used for hedonic regressions.

Summing up, the main conclusion of this chapter is that for the construction of good and unbiased price indices for computers or in fact for any other product, using a chained index principle with frequent resampling is absolutely necessary. Using information on expenditure shares, or hedonic methods to make explicit quality adjustments can provide even better estimates of price change, but this is of less immediate importance.

Having said that, a point has to be made concerning the relevance of research into price indices for computers. In private conversations with statisticians, it has been argued that a lot of fuss has been made about these price indices, whereas the ultimate share of computers in the total CPI is small. In the Netherlands the weight of computers in the CPI is less than 0.4%, so biases in the price index for computers are hardly reflected in the total CPI. A similar point can probably be made for other countries.

But this is not true for computers as investment goods, where their share is much bigger than in the CPI. The present chapter only studied consumer price indices, but of course the quality problem in computers is no less relevant in producer price indices or output deflators. When output deflators for computers are not adjusted for quality change, implicitly or explicitly, real output measures of computer producing and real productivity measures of computer using industries are seriously defunct, as brought forward by Wyckoff (1995), who compared different deflators for computer equipment across the OECD. A practical problem with output deflators and producer price indices lies in the fact that sufficient data can be harder to obtain than for consumer prices, although the experiences of the U.S. statistical agencies point out that it can actually be done.

Looking at the relevance of this research in the case of consumer prices, it can be stated that this lies in the nature of computers: quality change in this commodity type is very fast, probably the fastest in any type of durable good. Necessary conditions for the correctness of a matched model index are that sampling is done on a frequent basis, and that price indices are chained and do not refer to a fixed base period. Based on the research presented here, for computers, monthly sampling and chaining go a long way to provide better estimates of actual price change. Taking into account the difference between hedonic and matched model indices presented here, this may not be sufficient when there is (rapid) quality change. The challenge for statistical offices to collect good data and compare hedonic with matched model indices remains.

Notes

[1] For a recent discussion of Moore's Law, see Mann (2000).
[2] Strictly speaking, a good with a different quality would also count as a new good. Here a 'new good' is defined as a completely new product, like the first VCR or cellular phone. Goods with different quality are other (usually better) variations of an existing good, like a PC with a 1,500 MHz processor vs. a PC with a 1,200 MHz processor. These can be considered as new models rather than new goods.
[3] Triplett (forthcoming) argues that this conclusion is often drawn too easily. Researchers who find a larger price increase in their hedonic price indices than in official statistics usually attribute the difference to a quality bias in the latter, although the difference may also be caused by the use of different data, the difference between geometric means and arithmetic means etcetera.
[4] Lim and McKenzie (2001), McKenzie (2002).
[5] This approach is in line with the work by Triplett and McDonald (1977) on refrigerators in the U.S.
[6] See Chapter 2 for a more detailed description of this method.
[7] Nearly all retailers in the data set build their own custom configurations, which are therefore sold by one retailer by definition.
[8] This means that the system as a whole has no brand; it is not a finished product delivered by a computer manufacturer like Compaq.
[9] In these cases a simple sound card is integrated onto the motherboard.
[10] The latter happened with CD-ROM players a few years ago, and with USB-ports more recently.
[11] Examples are Ioannidis and Silver (1999), Silver and Heravi (2001, 2002a, 2002b) and Heravi, Heston and Silver (2003).
[12] For servers, only five regressions were run as servers are only sold in five different outlet types.
[13] See Section 3.3 for a more detailed discussion of the issue of brands.
[14] For the sake of simplicity, this equation shows the OLS equation. Actual estimation is carried out with WLS, where the weights are based on the quantity sold of each item. The equations for notebooks and servers are nearly identical, as described above.
[15] The relevant URL is http://www.cbs.nl/proefschriften.
[16] Note that this is a percentage of the number of observations rather than total sales.
[17] If discounts take place, their effects should appear in the time dummy when adjacent period regressions are estimated, as discounts are a pure deflationary effect.
[18] Also the hedonic dummy indices were aggregated across outlets using expenditure shares, although the individual dummy indices for the different outlets are unweighted indices.

5. *Quality-adjusted industry PPPs for automobiles*

5.1 Introduction

In the previous chapter, the focus was on quality adjustments in time-series based price index numbers. Throughout the literature, time series of price indices have been the main focus of quality corrections. Consumer price indices have received most attention, although producer price indices were also looked into (Gordon, 1990, being the prime example). However, the quality problem is not confined to the field of (consumer) price indices over time. When constructing measures of comparative price levels across countries, currency conversion factors like purchasing power parities (PPPs) or unit value ratios (UVRs) are needed, for which one encounters the similar problems. The characteristics of price indices and those currency conversion factors are very similar: both are ratios of prices, where in the case of price indices prices in different periods are compared, whereas PPPs and UVRs compare prices between countries. When the quality of a product that is purchased or manufactured in region *A* is not identical to its counterpart in region *B*, a quality bias is introduced in the PPP or UVR for this product.

In the case of price indices, the hedonic method is frequently used to make an explicit quality adjustment. As Triplett (2000) mentions, hedonic functions can also be used to measure PPPs. Because of the similarities between price indices and conversion factors, it seems therefore worthwhile to investigate to what extent this method can be used in international comparisons of prices. Two major efforts to apply the hedonic method in international comparisons stem from the ICP project, namely for cars and house rents, of which the former have been described in more detail in Section 3.8.

In international comparisons, cars are subject to both major inside the sample and outside the sample problems. The inside the sample problem is obvious: cars produced in different countries are not identical, so making identical matches is impossible. To tackle this problem, cars may be divided into classes based on size and/or engine performance. This solution was chosen by e.g. Kravis *et al*. (1975) and van Ark (1990). Cars within such classes are assumed not to differ too much, and the average price or unit value of cars within these classes are matched to provide currency conversion factors which are then weighted at the corresponding expenditure or production shares of each class.

A strong outside the sample problem appears when some or many classes are empty in either country in the comparison. This problem is especially relevant when the output mix differs strongly between countries, for example in comparisons between the United States and European countries. Cars produced and purchased in the United States are generally much bigger and more powerful than in Europe.

In intertemporal studies of price change, this output mix problem is very small, as the output mix of a single country generally remains fairly stable over short periods of time, but the inside the sample problem is much more relevant in intertemporal price indices. Like with computers, the bulk of the applications of the hedonic method to solve the inside the sample problem for cars has been for time series. The two pioneering studies on quality adjusted price indices for cars go quite far back in time, namely those of Court (1939) and Griliches (1961), which were already discussed in Section 3.5. Other major studies that built upon the framework provided by Court and Griliches are Otha and Griliches (1976), Gordon (1990) and Raff and Trajtenberg (1997). Indeed, cars are easily the second most popular item for which the hedonic method has been carried out, after computers.

Some hedonic applications for international price comparisons for cars have been carried out as well, but they focus on expenditures and consumer prices of cars that are sold in the respective markets. Kravis *et al*. (1975) is the most well known example (see Section 3.8). Studies by Goldberg and Verboven (1998) and Gaulier and Haller (2000), which are focused on issues like price dispersion and convergence in the automobile markets, include a common set of cars that are sold in each country in their analysis.

The focus of the present chapter is more production oriented, in line with the studies on industry output and productivity of the ICOP (International Comparisons of Output and Productivity) project at Groningen University.[1)] For each country only car models that are produced domestically are included in the analysis. Because this chapter studies countries with very dissimilar output mixes, there is a large amount of variation in the characteristics of representative cars in each country. Since producer prices of individual car models are not publicly available, retail list prices were used as the dependent variable in the hedonic regressions. The resulting quality adjusted prices were subsequently adjusted for distribution margins to derive a close approximation of quality adjusted industry PPPs. The countries that are studied are the United States, Germany, France, the United Kingdom, Italy and Japan. The period studied is 1995 to 1999.

Section 5.2 gives an overview of the quality issue for automobiles, and discusses relevant issues and previous research. In Section 5.3, the data and the methods that are used in this analysis are discussed. The results are presented in Section 5.4, and Section 5.5 concludes this chapter.

5.2 *Hedonic price indices for cars*

Price index numbers

Quality adjustments in price indices for cars using the hedonic method have been fairly common since Griliches' 1961 article. Before I turn to the issues that are relevant for international comparisons, I first pay attention to some general issues.

As pointed out in Chapter 3, the hedonic hypothesis basically assumes that a commodity can be viewed as a bundle of characteristics for which implicit prices can be derived. These characteristics represent utility to the user and usually cost to the producer as well. An important general point, that is especially relevant in the case of cars, is the distinction between characteristics that actually represent user value, and characteristics that are merely proxies for the former. Usually the second type of characteristics is far easier to measure than the first type. In the case of cars, actual performance characteristics include safety, reliability, durability, engine performances, space and so on. It is hard if not impossible to obtain direct measures of these performance variables. Physical characteristics, which represent these features, are often much easier to measure. These include maximum horsepower, engine displacement, size dimensions, the presence of airbags, and so on. Hedonic analysis on car prices therefore assumes that there is a direct relation between the latter variables and the former. This assumption has to be made for any type of commodity for which one wishes to apply the hedonic method, but given the complexity of cars, it is very outspoken in this case.

Although there is a relation between performance variables and physical characteristics, it is not a one to one relation, and sometimes they may even go in opposite directions. Some features related to, for example, design or prestige may add to user value but may not be related to measured physical characteristics or even production costs. Other examples concern fuel economy and engine performance, as already mentioned in Chapter 3. These are variables that are both valued positively by consumers, but the coefficients of hedonic regressions of these characteristics usually have opposite signs. This brings us to another problem that plagues hedonic regressions for cars, that is, multicollinearity. Many physical characteristics are technically related, which results in unreliable estimates of the regression coefficients.

The function that links the performance variables to the physical characteristics may also change through time, which makes the choice of some physical characteristics less obvious. Gordon (1990) gives the example of the 'full-sized' cars of General Motors in 1977. Decreases in the 'volumes' of several physical characteristics (weight, overall length, width and height) were accompanied with increases in associated performance variables, namely interior size dimensions, including trunk capacity. A standard hedonic regression would register this as a quality decrease, while an actual quality increase has taken place. This stresses the importance of using actual performance variables, although this may often be infeasible in practice.

The brand effect that was dealt with in Chapter 3 is especially apparent in the case of cars. Car manufacturers price their products well above marginal costs, but this mark-up is different for each car model. Ohta and Griliches (1976) call this brand effect a 'putative make effect', since it is not based on physical characteristics and does not enter the cost function of the producer. Such a make effect is reflected in non-physical characteristics, like prestige, reputation and service availability. These are not costless, but the costs do not depend closely on the volume of output (Ohta and Griliches, p. 331). Unlike price effects of physical characteristics (both observed and non-observed), such price effects may be smaller in the market for used cars, which they investigate.

The two studies mentioned here, Gordon (1990) and Ohta and Griliches (1976), are major studies on hedonic price indices for the United States based on consumer (list) prices. Both estimate hedonic regressions are for adjacent years, using somewhat different data. The main physical characteristics are weight, length and maximum horsepower. Both studies also use make dummies and dummy variables for other quality characteristics.[2)] Optional equipment was excluded from the price of a car, to make the observations more comparable. However, Ohta and Griliches do not make corresponding adjustments in the weight of the car. Gordon, who warns for this bias, estimates the effects of removing optional equipment on weight, and adjusts correspondingly.

The resulting hedonic price indices for the United States, which were derived using the dummy method, are not very different between Gordon and Ohta and Griliches. The average annual price changes for 1955–1971, which is the period analysed in both studies, equal 0.90% according to Ohta and Griliches and 0.81% according to Gordon. Differences in year-to-year changes are also small. The same is also true for the official matched model CPI for new cars. Although there are some differences in sub-periods, the overall price change in this period is comparable with the hedonic indices of Gordon and Ohta and Griliches. But for the period from 1974 to 1983 the CPI increases much more slowly than Gordon's hedonic index. This discrepancy can be explained entirely "by the allowance in the CPI for quality improvements taking the form of government-mandated safety and anti-pollution devices, in contrast to the hedonic index that makes no such allowance" (Gordon, 1990, p. 352). If the hedonic index is adjusted for these measures and changes in fuel economy, it nearly coincides with the official CPI between 1958 and 1983.

Ohta and Griliches also pay attention to the important issues of make effects and the dichotomy between performance variables vis-à-vis physical characteristics. Concerning the former, they compare the results of the hedonic regressions for new and used car markets for four different years. Make (or brand) effects in the new and used markets differ slightly, but since the sample sizes are generally large, they conclude that "make effects observed in the new car market persist in the used car market at roughly similar orders of magnitude" (Ohta and Griliches, 1976, p. 359). Brand effects, or 'putative make effects', as Ohta and Griliches dub them, are therefore not only valued in new cars.

Ohta and Griliches also test the 'two-stage hypothesis' that physical characteristics produce performance variables as outputs, which in turn should generate user value, meaning that physical characteristics are therefore good proxies for user value. This hypothesis is tested by using data on rating information from Consumer Reports. This magazine rates various performance variables of individual automobiles, on which Ohta and Griliches perform hedonic regressions to identify which variables are most significant in explaining variation in car prices. The two stage hypothesis is accepted if physical characteristics do not explain much of the residuals from hedonic regressions using performance variables, and if performance variables are explained well by physical characteristics. Ohta and Griliches find that both conditions are satisfied for the data at hand, so that the two stage hypothesis appears to hold in the current case.

The results found by Gordon and Ohta and Griliches indicate that the U.S. CPI for new cars is relatively unbiased, and if it is biased, this bias cannot be tackled by calculating hedonic indices instead. These results are not in accordance with the earlier findings of Griliches (1961, 1964) discussed earlier. Gordon criticises Griliches' earlier results on the ground of the use of unsuitable data, which show different changes in physical characteristics than the extensive data set used by Gordon. According to Raff and Trajtenberg (1997), who carried out a hedonic analysis of U.S. car prices in the period 1906 to 1940, the major quality-adjusted price decreases were witnessed in the years from 1906 to 1920, during which passenger cars were relatively new consumer products. Since then the market has matured and quality adjusted price indices have not fallen that much more rapidly than unadjusted indices.

More recently hedonic applications on car prices have also been carried out for European countries. De Haan and Opperdoes (2002) and Harhoff (2002) estimated hedonic functions for new cars in the Netherlands and Germany respectively.[3] It turns out that for both countries, the hedonic indices do not differ very much from the official CPI for cars or matched model indices on the same data sets.[4] This is in accordance with what Gordon finds on the U.S. CPI for new cars for a sixteen-year period, and the assertion made by Raff and Trajtenberg.

Summarising it appears that cars are too complex products, and hedonic analysis is therefore not the most suitable approach for dealing with the quality issue.[5] Indeed the studies cited above find that the differences between a matched model CPI and hedonic indices both in the United States and several European countries are very small on year-to-year differences, and the accumulated differences stay minor for relatively long periods of time. This suggests that matched model methods may provide sufficient estimates of year-to-year price changes of new cars. The question remains whether the same can be said of for international price comparisons.

International price comparisons

In international comparisons of relative price levels, there are two approaches to calculate currency conversion factors: the expenditure approach and the industry-of-origin approach. The former is used in the International Comparisons Project (ICP),[6] where consumer prices of products are compared across countries, which results in purchasing power parities (PPPs). For example, if an average car in the U.S. costs on average $20,000, and an average German car DM 40,000, than the purchasing power of one dollar in Germany would be two marks. The industry-of-origin approach provides a currency conversion factor at the producer level. In this case, unit values of products (which are equal to producer price value divided by the quantity produced) that are actually produced in each country are compared, resulting in unit value ratios (UVRs). This is the method applied in the framework of the International Comparisons of Output and Productivity (ICOP) project.[6]
Regardless of the approach, the usual way to calculate conversion factors is based on the matched model method. As described in Section 2.5, prices of products are compared under the assumption that the goods in both countries are identical. Taking cars, this means that the weighted average price (or unit value) of cars in country *A* is compared to the weighted average price (or unit value) of cars in country *B*. Usually shares in sales value or production volume are used as weights, depending on the approach.
In the case of cars, a more disaggregated price comparison can be achieved when using the expenditure approach than with the industry-of-origin approach. Market prices of individual car models are available from various data sources (see below), but producer unit values (requiring both producer price values and quantities) can generally only be calculated for all passenger cars combined on the basis of industrial production statistics.[8]
Another advantage of using expenditure prices for cars is that there is substantial overlap in cars that can be purchased in different countries. This is especially the case within Europe, where roughly the same range of cars and car models is available in most countries. Because of this overlap, it is possible to compare prices of most available car models. Overlap in samples between countries is less for produced cars. There is much concentration of the production of models within Europe, so that a specific type of cars that is produced in one country is usually not produced in other countries. The matched model method then breaks down, and a higher level of aggregation is used to obtain a unit value ratio, assuming that the average unit value of a car is comparable between countries. To deal with this problem, various studies have aimed to refine the product matches. For example, van Ark (1990) distinguished car production in France and the United Kingdom into five categories based on engine displacement. He then used weights on the quantities of models produced from industry sources and assumed that items within each category were more homogenous than for all produced cars produced together. This method in fact is a hybrid between the matched model method and the hedonic method. Despite a large spread in the mix of cars purchased, Kravis *et al*. (1975) also adopted this approach for a wider range of countries. An all-out

application of the hedonic method without resorting to car classes may provide a solution if output mixes differ considerably, but it has to be applied with caution.

An industry-of-origin study that did not employ the hedonic method at all was McKinsey Global Insitute (1993). For the construction of quality-adjusted industry PPPs, two adjustments are made in this study. First, akin to Kravis *et al*. (1975) and van Ark (1990), cars are subdivided into five market segments. Within these classes, a further explicit quality adjustment is made by assessing the price differentials a consumer, who has unrestricted access to foreign products, would be willing to pay for a car of the same category based on his perception of quality differences. This price differential is interpreted as the quality difference between cars of a particular segment. For the Japan/U.S. comparison prices of 'standardised' cars were compared and averaged. [9] For the Germany/U.S. comparison interviews with local experts were used to determine explicit quality adjustments.

5.3 Data and methodological issues

The focus of this section is on quality-adjusted comparisons of producer prices of cars in six major car manufacturing countries (the United States, Germany, France, the United Kingdom, Italy and Japan) using the hedonic method. This requires prices, quantities and specifications of individual car models.

Unlike characteristics and quantities produced, producer prices are usually not available on the level of individual models of produced automobiles. I therefore first had to work on information on produced quantities of specific models in combination with their retail prices (net of taxes). As a second step I adjusted the retail prices to producer price level (see below).

Several sources provide information on list prices and characteristics of cars that are available on the market. From these sources, I only included observations for cars from locally produced brands in the analysis. For each country, cars from the following brands were included:

- Germany: Audi, Volkswagen, Opel, Mercedes, BMW and Ford;
- France: Citroën, Peugeot and Renault;
- United Kingdom: Ford, Honda, Rover, Nissan, Vauxhall (Opel) and Toyota; [10]
- Italy: Alfa Romeo, Fiat and Lancia;
- United States: Chrysler/Plymouth, Dodge, Eagle, Ford, Lincoln/Mercury, Buick, Cadillac, Chevrolet, Oldsmobile, Pontiac and Saturn;
- Japan: Toyota, Nissan, Mazda, Mitsubishi, Honda, Isuzu, Fuji Heavy, Daihatsu and Suzuki.

In all countries these brands have large market shares in domestic car sales. In Japan, these brands represent nearly 100% of domestic production, whereas it is between 55% and 70% in the other countries. But for the purpose of producer price analysis we are ultimately only interested in cars produced rather than sold domestically.

Data on the quantities and (retail) prices produced cars and models produced in the United States come from the *Market Data Book*, which is an annual publication by Automotive News. This source contains information on the annual production of cars by model, characteristics by model and retail prices by type. A model is defined here as a production line of a particular manufacturer, e.g., a Ford Taurus. Such a model can in turn consist of multiple types, each with a different price. When this is the case, the most basic type is chosen as representative for all types of this model. One single observation therefore consists of the number produced of one model, and the characteristics and the retail price of the most basic type of this model. This raises the issue of representativeness or characteristicity, namely whether this particular type is best suited to represent the model.[11)] Unfortunately, the number of characteristics is fairly limited. Those included were engine displacement, engine power, weight, and size dimensions (length, width and height).

The basic source for Japan, the Japanese *Motor Vehicle Guidebook*, is primarily a glossary for consumers on cars that are available on the Japanese market. Data on quantity and type of cars produced is therefore not available. However, the Japanese car market is a strongly protected market, and foreign cars are rarely available. As a matter of fact all *Guidebook*s that were used for this study only included cars from Japanese manufacturers. All cars listed are therefore assumed to be produced domestically. Since the information in the Guidebooks is by type rather than model, the number of observations for Japan is much higher than for any other country, especially the European ones. But because no production information is available, it is not possible to construct weighted conversion factors.

Data on the number of cars produced in the European Union is available from *World Automotive Statistics* and the Motorsat-website. But the data on prices and characteristics for the European countries are obtained from another source provided by the European Commission. The European Commission requires European car manufacturers to provide records of their most selling models each year, including characteristics and list prices they charge in each member country of the European Union. From this data source, I only obtained observations of cars from locally produced brands in each country.

Since the price and quantity data for the European countries do not come from the same source, the observations on prices and quantities had to be matched. The data from the European Commission lists prices and specifications of several models for each brand. For example, in the case of Germany, four different types are listed for the year 1997: the Astra 1.6, the Corsa 1.2, the Omega GL 2.0 and the Vectra GL 1.6. The assumption is made that the type that is listed in the data from the European Commission is representative for all other types within a particular model (e.g. the Vectra GL 1.6 represents the entire Vectra-line). The database on the number of cars produced only distinguishes between different models, like the Opel Vectra and the Opel Astra. Combining the two databases on quantities (and characteristics) and prices implies that the entire production of all Opel Vectras is linked with the price and characteristics of the Vectra GL 1.6 from the EC-database.

Like with the United States, the characteristicity issue is relevant here. In the EC-database, only the types that are most frequently sold are included. In this example, this means that of all Opel Vectra types, the GL 1.6 is the mostly sold, and therefore probably the mostly produced model. Table 5.1 illustrate the linking of the two databases for Germany in 1997. The columns labelled Production line and Quantity are from Motorsat, the remaining columns from the European Commission. The inclusion of the mostly sold types causes the number of observations for the European cars to be very low in each year, so that individual hedonic regressions for separate countries in single years are not feasible.

Although the approach used here is essentially representative of the industry-of-origin approach, the prices in table 5.1 as well as those for other countries are not producer prices but retail prices, net of taxes. The use of transport and distribution margins for the individual countries, which can be derived from the input-output tables at the aggregate level of motor vehicle production may provide a partial solution for this problem, which is applied here. Since we make use of an industry-of-origin approach in combination with expenditure prices adjusted for distribution margins, the resulting conversion factors are termed *industry-based expenditure PPPs*. [12)]

Table 5.1
All observed cars for Germany, 1997

Model	Type	Quantity (1,000s)	Retail price (DM)	Engine (cc)	Power (HP)	Length (m)	Width (m)	Height (m)
Audi A3	A3 1.6E	128	27,261	1,595	101	4.15	1.74	1.42
Audi A4	A4 2.0E	289	35,991	1,781	125	4.48	1.74	1.41
Audi A6	A6 2.4E	122	48,826	2,393	165	4.80	1.81	1.45
Audi A8	A8 2.8E	16	73,913	2,771	193	5.03	1.88	1.44
BMW 3	316i 4dr	315	33,565	1,596	102	4.44	1.70	1.40
BMW 5	520i 4dr	233	48,478	1,991	150	4.78	1.80	1.43
BMW 7	735I 4D	49	83,913	3,498	235	4.99	1.86	1.43
Ford Escort	1.6 CLX/LX	289	22,501	1,597	90	4.14	1.70	1.40
Ford Fiesta	1.25	258	17,318	1,242	75	3.83	1.64	1.38
Ford Scorpio	2.3 16 v GHIA	20	33,308	2,295	147	4.83	1.76	1.40
Mercedes C	C 180	228	39,600	1,799	122	4.49	1.72	1.42
Mercedes E	E 200	232	49,500	1,998	136	4.80	1.80	1.43
Opel Astra	1.6	489	26,052	1,598	75	4.05	1.70	1.41
Opel Corsa	1.2	167	17,090	1,195	45	3.73	1.61	1.42
Opel Omega	GL 2.0	114	45,548	1,998	116	4.79	1.79	1.46
Opel Vectra	GL 1.6 16V	283	34,352	1,598	101	4.48	1.71	1.43
Volkswagen Golf	CL 2D 60PS 5G	519	19,070	1,390	60	4.02	1.70	1.43
Volkswagen Passat	CL 90PS 5G	335	31,131	1,781	101	4.68	1.74	1.46
Volkswagen Polo	60 PS	178	16,920	1,390	60	3.72	1.66	1.42
Volkswagen Vento	CL 90PS 5G	38	24,037	1,781	90	4.38	1.70	1.43

Sources: The columns labelled Model and Quantity are from Motorsat (http://perso.club-internet.fr/motorsat/), the remaining columns form the European Commission (http://europa.eu.int/comm/competition/car_sector/price_diffs/).

Cars are very complex products of which the prices are determined by a large amount of characteristics. Some characteristics are only present in cars of a particular manufacturer or even a specific model or type, which makes quality adjustments relative to cars from other manufacturers difficult. On the one hand, taking too few characteristics into account in the hedonic regression can leave a remaining quality bias in the price ratio. On the other hand, as many of the characteristics are correlated with each other, the use of all of them would result in many insignificant coefficients. With the data at hand, however, the latter seems no real problem. Since the data on characteristics for the countries come from different sources, there are only a few matching characteristics which are available for all six countries. These include engine displacement (ENGINE), engine power (POWER), weight (WEIGHT), length (LENGTH), width (WIDHT) and height (HEIGHT).

For the selection of characteristics, the same criterion is used as was done by Gordon (1990) and in Chapter 4 of this thesis, that is, characteristics have to be associated with both user value and production costs. The characteristics mentioned above can be divided into three groups: engine performances, size indicators and weight. The two former have a meaning from the buyer's perspective: engine performance and the amount of space in a car are relevant to most buyers, albeit as proxies. This is not the case for weight. Although virtually all characteristics are reflected in the weight of a car, weight itself is not something buyers desire in itself (Gordon, 1990). This is probably the reason why the coefficient of weight, when included, always has a large explanatory power in determining car prices (Griliches, 1961).

This leaves five potential characteristics. Although engine displacement and engine power are not perfectly collinear, there exists a strong technical link between the two. If one is used in the regression, including the other will not lead to a significant increase in explanatory power of the regression. Therefore hedonic regressions were run with six different combinations of characteristics: ENGINE with either LENGTH, WIDTH or HEIGHT; and POWER with either LENGTH, WIDTH and HEIGHT respectively. Carrying out regressions with either POWER or ENGINE and two or more size indicators did not lead to major increases in explanatory power, and were therefore not further used. The same was true when all characteristics were used simultaneously, but this also leads to strong multicollinearity, which resulted in coefficients with unexpected signs.

The unweighted average values of all characteristics are shown in table 5.2. It is clear from this table that there are large differences in characteristics between cars from the respective countries. Cars from Japan and especially the United States have much more horse power than European cars. The unadjusted average price level of the United States would therefore too high compared with the other countries, resulting in a PPP that underestimates the purchasing power of the dollar.

Table 5.3 shows the averages of prices and characteristics weighted by the number of cars produced. As the information on production is limited, not all years are available for each country and Japan is missing altogether. The differences between tables 5.2 and 5.3 are as expected: larger and more powerful cars are over-represented in the average if the number of cars produced is not used as a weight. But as this is the case for each country, it is not clear beforehand how large the difference between weighted and unweighted PPPs will be. I will look into this in more detail in the next section.

Table 5.2
Average prices and characteristics of all countries, unweighted, 1995–1999

	Retail price[1] (own currency)	Engine (cc)	Power (HP)	Length (m)	Width (m)	Height (m)	Number of observations
1995							
United States	19,824	3,010	164	4.81	1.80	1.36	63
Germany	32,877	1,783	104	4.39	1.72	1.41	18
France	85,971	1,472	80	4.05	1.68	1.42	12
United Kingdom	11,025	1,635	95	4.21	1.69	1.41	18
Italy	24,244	1,671	102	4.21	1.70	1.43	10
Japan	2,266	1,920	147	4.35	1.68	1.39	407
1996							
United States	20,403	2,940	165	4.80	1.80	1.37	67
Germany	32,354	1,811	108	4.40	1.72	1.42	18
France	96,789	1,547	88	4.11	1.70	1.46	14
United Kingdom	12,110	1,683	103	4.28	1.69	1.41	17
Italy	24,688	1,610	100	4.18	1.71	1.43	11
Japan	2,259	1,917	147	4.35	1.68	1.39	381
1997							
United States	20,360	2,851	164	4.76	1.79	1.37	59
Germany	36,419	1,864	114	4.43	1.74	1.42	20
France	101,120	1,605	92	4.17	1.72	1.47	13
United Kingdom	12,041	1,667	100	4.22	1.69	1.42	15
Italy	26,178	1,601	107	4.23	1.73	1.43	9
Japan	2,313	1,937	153	4.36	1.69	1.39	356
1998							
United States	21,553	2,870	171	4.75	1.80	1.37	57
Germany	37,860	1,852	116	4.42	1.74	1.43	17
France	97,869	1,571	92	4.18	1.72	1.47	14
United Kingdom	11,937	1,655	100	4.20	1.70	1.43	17
Italy	25,459	1,597	106	4.22	1.73	1.43	9
Japan	2,304	1,933	155	4.39	1.70	1.40	263
1999							
United States	21,724	2,888	173	4.75	1.80	1.38	52
Germany	38,035	1,838	116	4.41	1.74	1.43	18
France	87,657	1,508	88	4.13	1.71	1.47	14
United Kingdom	12,568	1,667	102	4.22	1.70	1.45	15
Italy	26,817	1,602	110	4.22	1.73	1.43	9
Japan	2,341	1,893	155	4.35	1.69	1.40	252

[1] Retail prices of Japan and Italy are in 1,000s.

Sources: For European countries, see table 5.1; U.S. data are from Automotive News, data for Japan from JAMA.

Table 5.3
Average prices and characteristics of all countries except Japan, weighted by quantity produced, 1995–1999

	Retail price[1] (own currency)	Engine (cc)	Power (HP)	Length (m)	Width (m)	Height (m)
1995						
United States	16,292	2,617	141	4.74	1.78	1.37
1996						
United States	16,627	2,558	143	4.73	1.78	1.38
1997						
United States	17,072	2,546	145	4.73	1.78	1.38
Germany	31,123	1,671	98	4.31	1.72	1.42
Italy	17,720	1,265	70	3.92	1.68	1.44
1998						
United States	17,841	2,597	149	4.75	1.78	1.39
Germany	31,116	1,676	100	4.34	1.73	1.43
France	81,829	1,422	78	4.05	1.68	1.41
United Kingdom	10,413	1,514	94	4.18	1.69	1.43
Italy	19,195	1,326	77	3.95	1.68	1.44
1999						
United States	17,670	2,606	152	4.74	1.79	1.39
Germany	31,594	1,686	102	4.35	1.73	1.43
France	76,210	1,395	77	4.04	1.68	1.42
United Kingdom	10,145	1,479	90	4.15	1.68	1.43
Italy	19,709	1,360	83	3.95	1.68	1.44

[1] Retail prices of Italy are in 1,000s.

Sources: see table 5.2.

Estimating separate regressions for each country-year combination would be desirable, but this is not possible as the number of observations for the European countries is very limited in individual years (see last column of table 5.2). This runs from 9 in Italy (1997) to 20 in Germany (1997). In contrast, the lowest number of observations for the United States and Japan in 1999 was 52 and 252 models respectively.

Pooling the data is therefore necessary. The data were pooled both across countries and across years. Dummies were included for each country except the United States (which served as the reference country) and for each year except the first year. Pooling the data across countries assumes that the relation between prices and characteristics is the same in every country for a given year. These assumptions are obviously not realistic, and even more problematic than the hypothesis that coefficients are stable across time for a given country. Whereas the changes in quality through time are gradual, this is not the case with quality differences in an international context. Both assumptions therefore need to be tested.

Hedonic OLS regressions were run with the logarithm of prices (LOGPRICE) as the dependent variable. Both semi-logarithmic regressions and double-logarithmic regressions were run. In the latter case LOGPRICE was regressed on the logarithms of the characteristics. The semi-logarithmic regressions had a better fit in all cases, both expressed in adjusted R^2 and standard error of the regression. Hence the results reported here relate to the semi-logarithmic regressions.[13)]

Since I pooled the data in two different ways, I performed standard F-tests to test whether either way can be rejected on statistical grounds. To test whether the relation between prices and characteristics is the same across years, I use the five year-regressions that pool the data across countries. In this way, all coefficients (i.e. those of the characteristics and the five country dummies) are restricted to be the same in each year. To test the hypothesis that the relation between prices and characteristics was the same across countries, I used the six country-regressions that pool the data across years. The coefficients (i.e. those of the characteristics and four year dummies, 1995 being the base year) are then restricted to be the same across countries. Both test were done for all six combinations of characteristics. Table 5.4 shows the values of these F-tests for each of the six specifications of the hedonic model.

Table 5.4
F-values for hypotheses of identical coefficients across countries and years

Characteristics used		Coefficients are identical in every	
		country	year
ENGINE	LENGTH	19.9	1.0
ENGINE	WIDTH	26.2	1.7
ENGINE	HEIGHT	31.7	0.7
POWER	LENGTH	2.9	0.7
POWER	WIDTH	4.0	1.9
POWER	HEIGHT	3.5	0.6

As the critical value for the F-test is 1.39 at the 5%-level of significance, the null hypothesis that the relation between prices and characteristics is identical in all countries is rejected in all cases. The international hedonic function hypothesis which states that there is a common hedonic function across countries, as for example claimed for computers in Triplett (2000) clearly does not hold for cars. This leads to the conclusion that we cannot pool the data across countries, and so have to run separate regressions for each country. The second hypothesis, i.e. that coefficients are constant through time assuming they are the same in every country, is only rejected in two cases. Pooling the data from all five years for each country thus seems less restrictive.

The next issue to consider is which combination of characteristics provides the best fit. To determine this, the explanatory power of all combinations are compared in table 5.5.

Table 5.5
Explanatory powers of different specifications of the hedonic model

Country	Regression using coefficients:					
	ENGINE LENGTH	ENGINE WIDTH	ENGINE HEIGHT	POWER LENGTH	POWER WIDTH	POWER HEIGHT
United States						
adjusted R^2	0.7251	0.7374	0.7181	0.8246	0.8245	0.8146
SER[1)]	0.2240	0.2189	0.2268	0.1789	0.1790	0.1839
Germany						
adjusted R^2	0.9298	0.8993	0.8129	0.9549	0.9356	0.9073
SER[1)]	0.1209	0.1448	0.1973	0.0969	0.1157	0.1389
France						
adjusted R^2	0.9332	0.9263	0.9186	0.8998	0.9147	0.8487
SER[1)]	0.1007	0.1057	0.1111	0.1233	0.1138	0.1515
United Kingdom						
adjusted R^2	0.8831	0.8638	0.8530	0.8673	0.8873	0.8807
SER[1)]	0.1333	0.1439	0.1495	0.1420	0.1309	0.1347
Italy						
adjusted R^2	0.9384	0.9296	0.9220	0.9409	0.8968	0.8530
SER[1)]	0.0925	0.0989	0.1041	0.0906	0.1198	0.1429
Japan						
adjusted R^2	0.8545	0.8543	0.8613	0.8477	0.8315	0.8103
SER[1)]	0.1964	0.1966	0.1918	0.2013	0.2117	0.2247

[1)] Standard error of the regression.
Data for each country are pooled across years.

Table 5.5 shows that the combination of POWER and LENGTH provides the best fit in three of the six countries (United States, Germany and Italy). None of the other specifications has the best fit for more than one country. As a result the POWER-LENGTH specification is chosen for the hedonic regressions. This is in accordance with economic and technical intuition: horsepower is usually a better indicator of engine performance than its basic displacement, and should therefore be better reflected in price. Concerning the size indicators, length has the highest variation and has therefore the biggest effect on the amount of space in a car.
For the regressions including POWER and LENGTH, the results are shown in table 5.6. POWER and LENGTH had significant coefficients in all cases, except for the regression of the United Kingdom, where LENGTH is insignificant. [14)]

Table 5.6
Coefficients of country regressions with POWER and LENGTH as explanatory variables

	United States	Germany	France	United Kingdom	Italy	Japan
Constant	7.69	7.49	8.88	8.06	7.15	5.46
t-value	48.6	41.3	36.2	30.6	29.2	97.9
POWER	0.01	0.01	0.01	0.01	0.01	0.01
t-value	28.9	12.7	5.70	14.3	6.10	49.4
LENGTH	0.25	0.51	0.46	0.02	0.57	0.31
t-value	7.30	10.2	5.80	0.30	7.80	20.6
T1996	0.03	–0.03	0.05	0.01	0.03	0.00
t-value	0.99	–0.81	1.09	0.21	0.70	0.14
T1997	0.04	0.01	0.05	0.04	0.04	–0.01
t-value	1.29	0.33	0.96	0.71	1.04	–0.83
T1998	0.06	0.04	0.01	0.02	0.03	–0.03
t-value	1.83	1.09	0.20	0.43	0.66	–2.07
T1999	0.05	0.04	–0.05	0.03	0.06	–0.02
t-value	1.61	1.35	–1.07	0.56	1.52	–0.97

The results from the regressions using POWER and LENGTH will be used in the next section to derive the industry PPPs.

5.4 *Applying the hedonic estimates to derive PPPs*

Hedonic regressions can be used in several ways to construct price indices or conversion factors. The three mostly used methods are the dummy method, the imputation method and the hedonic quality adjustment (see Chapter 3).

The dummy method can only be applied if we pool the data across countries. The results in the previous section showed that this is statistically unsound. PPPs derived with this method will therefore only be included by way of comparison.

The imputation method is generally used to estimate prices of unmatched items that are not available in one of the two countries. This approach works if there is a substantial amount of matched observations. Because of the large differences in product mixes, there are only unmatched observations in international comparisons of car prices. Therefore this method is not useful for computing industry PPPs for cars.

With the hedonic quality adjustment, the price differential of a good in two countries is divided in two parts: a 'true' price differential and a price differential that is due to differences in characteristics. To calculate these price differentials, observations are matched like with the normal matching procedure and an explicit quality adjustment for each matched pair is made. Obviously not all items can be matched, as the number of observations differs between countries. Therefore the observations are usually aggregated over one or more characteristics, as pointed out in Chapter 3. Preferably the characteristics over which this aggregation takes place are dummy variables or other variables with a limited number of values, so that the number of aggregations stays manageable. In the present case the only possible aggregation is to aggregate over all cars in each country. Since it is the only feasible method in the present case, the hedonic quality adjustment method is adopted here to derive quality adjusted currency conversion factors.

For each country, the average value of price, engine power and length of all cars is calculated, using the shares of models in the total production volume, as shown in table 5.1. The only country for which production volumes were available for all years is the United States. For Germany and Italy, quantities are only given for the years 1997–1999, and for France and the United Kingdom for 1998–1999. For Japan, no quantity data is available on the level of individual models at all.

Using the hedonic quality adjustment, the PPP for passenger cars of countries A and B are calculated in two ways. First, one can obtain the ratio of the estimated price of an 'average car' from country A in country B, $\hat{p}_B$, relative to the average car price in country A, p_A. If country A is the reference country, this yields the 'base country PPP'. Second, one can obtain the ratio of the average car price in country B, p_B, relative to the estimated price of an average car from country B in country A, $\hat{p}_A$. If country A is the reference country, this yields the 'own country PPP'.

The estimated prices $\hat{p}_A$ and $\hat{p}_B$ are calculated in the following way:

$$\hat{p}_A = p_A \exp[\beta_{POWER}(POWER_B - POWER_A) + \beta_{LENGTH}(LENGTH_B - LENGTH_A)]$$

$$\hat{p}_B = p_B \exp[\beta_{POWER}(POWER_A - POWER_B) + \beta_{LENGTH}(LENGTH_A - LENGTH_B)]$$

Where $POWER_A$ and $POWER_B$ are the average power of cars in countries A and B respectively, and $LENGTH_A$ and $LENGTH_B$ are the average length in countries A and B. The β's are the coefficients from the hedonic regression. If we use the coefficients from the country-pooled regressions, the Laspeyres indices ($\hat{p}_B / p_A$) and Paasche indices ($p_B / \hat{p}_A$) are identical. If, as was indicated as the preferred method above, regressions are estimated for each country separately, the coefficients from the regression of country A are used in computing $\hat{p}_A$, and the coefficients from the regression of country B are used to calculate $\hat{p}_B$. [15] This will yield different results for the two indices.

Because were are using consumer prices while using a producer perspective, we have to adjust the currency conversion factors for the difference between producer and consumer prices. Part of the difference between producer prices and purchaser prices is explained by transport and trade services (Hooper and Vrankovich, 1995). Both transport and trade margins can be obtained from input-output tables, and these are used here to adjust the consumer price based PPPs to achieve closer approximations of industry PPPs from a producer point of view. [16]

Since American cars are on average more powerful and larger than cars from the other countries (see tables 5.2 and 5.3), and since these characteristics have a positive effect on prices, we expect the relative value of the U.S. dollar based on unadjusted prices to be biased downwards. Table 5.7 presents PPPs based on unweighted average prices and values of characteristics. Here we only show the conversion factors of the European currencies and the Japanese yen vis-à-vis the U.S. dollar. [17)]

Table 5.7
PPPs of five currencies relative to the U.S. dollar for the car industry, unadjusted and quality adjusted using hedonic methods, unweighted, 1995–99

PPP	Unadjusted	Dummy method	Hedonic quality adjustment			
			Pooled regression	Country regressions		
				US base	Own base	Average[2)]
1995						
DM/$	1.80	2.88	2.86	3.14	2.77	2.95
FF/$	4.72	9.64	9.50	11.7	8.96	10.2
£/$	0.59	1.06	1.04	1.29	1.00	1.13
IL/$[1)]	1.28	2.25	2.17	2.41	2.08	2.24
¥/$	124	150	158	156	152	154
1996						
DM/$	1.72	2.69	2.62	2.9	2.57	2.73
FF/$	5.16	9.86	9.63	11.8	9.27	10.4
£/$	0.63	1.05	1.03	1.27	1.00	1.13
IL/$[1)]	1.26	2.20	2.17	2.44	2.09	2.26
¥/$	120	146	153	152	147	149
1997						
DM/$	1.93	2.75	2.78	3.02	2.73	2.87
FF/$	5.36	9.68	9.45	11.4	9.19	10.2
£/$	0.62	1.05	1.04	1.29	1.01	1.14
IL/$[1)]	1.33	2.19	2.12	2.36	2.07	2.21
¥/$	122	139	145	146	142	144
1998						
DM/$	1.89	2.81	2.85	3.06	2.76	2.90
FF/$	4.90	9.46	9.12	10.7	8.62	9.60
£/$	0.58	1.05	1.04	1.30	0.98	1.13
IL/$[1)]	1.22	2.18	2.09	2.25	1.98	2.11
¥/$	115	137	143	140	136	138
1999						
DM/$	1.89	2.83	2.86	3.11	2.80	2.95
FF/$	4.38	8.65	8.37	10.3	8.09	9.10
£/$	0.61	1.05	1.06	1.37	1.03	1.18
IL/$[1)]	1.29	2.22	2.12	2.35	2.06	2.20
¥/$	117	138	145	145	141	143

[1)] In thousands of Lira per dollar.
[2)] Geometric average of U.S. base and own base PPP.

The column 'unadjusted' represents the unadjusted PPPs, which are simply the ratios of average prices. The column 'dummy method' presents the results using the dummy method by way of comparison. These PPPs are derived from the regression where data are pooled over countries rather than years. The next four columns contain the results using hedonic quality adjustments. In the column 'pooled regression', the coefficients from the country-pooled regressions were used. In the columns 'U.S. base' and 'own base' the coefficients from the regression for the United States and the compared country were used, with the last column stating the geometric averages of these two columns. Given the discussion above, the PPPs from this column 'average' represent our preferred measures.

As was to be expected, the differences between the unadjusted conversion factors and the quality-adjusted industry-based expenditure PPPs are large. In nearly all cases the quality adjusted PPPs of each national currency relative to the U.S. dollar are 20%–100% higher than the unadjusted PPPs. The differences between the various hedonic applications are relatively small, i.e., a few percent at most. This indicates that the issue of how we apply the hedonic method is of less importance than whether we apply it.

These assertions are confirmed in table 5.8, which shows the results of weighted averages for prices and characteristics. The differences between the unweighted and weighted PPPs are caused by differences between unweighted and weighted average prices and characteristics. Referring to table 5.2 and 5.3, in all cases the weighted averages of prices and characteristics are lower than the unweighted ones. This is to be expected, as cars with a higher POWER/LENGTH combination of characteristics, which are larger and more expensive, have smaller production runs than smaller and medium- sized cars.

Although the differences between weighted and unweighted averages are not the same for each country and each year, tables 5.2 and 5.3 show that the size of these differences, with some exceptions, is roughly in the same order of magnitude for all countries. Since PPPs are ratios of average prices, and both weighted and unweighted hedonic quality-adjusted PPPs are derived with the same regression equations, the differences between unweighted and weighted PPPs are expected not to be too large. [18)]

This expectation is confirmed in table 5.8. Although there are some differences, especially in the unadjusted PPPs, the difference between our preferred unweighted and weighted PPPs (labelled 'average' in tables 5.7 and 5.8) does not exceed 7%. For Germany and Italy, there is no clear direction in the differences between weighted and unweighted PPPs. For France, weighted PPPs are in all cases higher than unweighted PPPs, whereas the reverse is true for British cars. In all cases, the difference between unadjusted PPPs and the preferred hedonic PPPs far exceeds the difference between different hedonic PPPs and between weighted and unweighted PPPs. This indicates that the effect of using hedonic quality adjustments to industry PPPS for passenger cars outweighs the effect of using quantity information.

Table 5.8
PPPs of four currencies relative to the U.S. dollar for the car industry, unadjusted and quality adjusted using hedonic methods, weighted by quantity produced, 1997–99

	Unadjusted	Dummy method	Hedonic quality adjustment			
			Pooled regression	Country regressions		
				US base	Own base	Average[2]
1997						
DM/$	1.96	2.75	2.85	3.15	2.80	2.97
IL/$[1]	1.07	2.19	2.04	2.43	1.97	2.19
1998						
DM/$	1.88	2.81	2.82	3.06	2.71	2.88
FF/$	4.95	9.46	9.26	10.9	8.64	9.71
£/$	0.62	1.05	1.01	1.15	0.95	1.05
IL/$[1]	1.11	2.18	2.17	2.47	2.01	2.23
1999						
DM/$	1.94	2.83	2.85	3.14	2.79	2.96
FF/$	4.68	8.65	8.64	10.6	8.32	9.38
£/$	0.61	1.05	1.02	1.22	0.98	1.10
IL/$[1]	1.16	2.22	2.13	2.52	2.05	2.27

[1] In thousands of Lira per dollar.
[2] Geometric average of U.S. base and own base PPP.

5.5 *Summary*

The issue of quality differences that is apparent in price index numbers is at least as relevant for international comparisons. Because the composition of production of some goods is strongly heterogeneous across countries, a quality adjustment is necessary in those cases. The hedonic method, which is applied more frequently in the field of price indices, may also provide a solution in the case of international price comparisons.

A good example of goods for which price comparisons are very complex, are cars. Whereas the same cars are to a certain extent sold in more countries, most models are only produced in a limited number of countries. Comparing the car production of these countries therefore justifies an adjustment for differences in the output mixes.

In calculating quality-adjusted price indices for automobiles, the hedonic method is not undisputed. Cars may be too complex for this method, since there is a vast array of price determining characteristics, many of which appear only in certain models or types, and most of which are correlated with each other. In price indices, there may be sufficient overlap in the year-to-year availability of cars to use alternative methods, but this is not the case for international comparisons.

I therefore experimented with the hedonic method, and compare the hedonic results with an unadjusted measure of price ratios. Choosing the United States as the reference country, unadjusted conversion factor proved to be biased downwards by as much as 50%. The size of this bias was only to a small extent changed by different applications of the hedonic method, although separate country regressions are to be preferred to regressions with pooled data for all countries. The obvious reason of this huge differential is the fact that American cars are on average much larger, both in size and engine performances, than European cars.
The hedonic method may not be the best way to apply to indices of car prices, but this chapter certainly indicates that unadjusted *international* price ratios can introduce large biases. This makes the case for the hedonic method in international comparisons of industries with large differences in output mixes and/or industries that produce very heterogeneous goods rather strong.

Notes

1) See van Ark (1993, pp. 34–38) and van Ark and Timmer (2001) for a discussion.
2) Whereas both use dummies for engine type and in some instances vintage, Ohta and Griliches employ a dummy for body type, and Gordon includes dummies for trim level.
3) De Haan and Opperdoes studied the period 1990–1999, while Harhoff examined the years 1986–1997. Similar research on the same data for the Netherlands was performed by Bode and van Dalen (2001) and van der Grient and Oei (2002).
4) Because of this, the German Federal Statistical Office announced in a press release that it would not employ the hedonic method for their CPI for cars (http://www.destatis.de/presse/englisch/pm2003/p0580051.htm).
5) See also Triplett (1969; 1990) as well as a comment on Raff and Trajtenberg (1997).
6) Kravis *et al*. (1975, 1978, 1982).
7) Van Ark (1993); Van Ark and Timmer (2001).
8) An exception is made in the case of Japan, where passenger cars a distinguished into three categories based on engine displacement.
9) McKinsey does not provide details on how these 'standardised' cars are defined.
10) Although there are no British domestic brands anymore (Rover was acquired by BMW), all these manufacturers have car plants in the United Kingdom. The same is true for Germany in the case of Opel, which is owned by General Motors.
11) See also Section 2.5.
12) The major remaining problem is of course that retail prices also reflect pricing strategies of car manufacturers, which may not be eliminated by the adjustment for transport and distribution margins.
13) Detailed regression results are shown in Appendix B.

[14] To save space, only the regression results with POWER and LENGTH are shown here. The results from the other regressions are shown in Appendix B. POWER and LENGTH are also the main physical characteristics that were employed by Gordon (1990) and Ohta and Griliches (1976). Another physical characteristic they both use is weight. The inclusion of weight as an explanatory variable improves the fit of the regressions somewhat. It strengthens the effects of multicollinearity, resulting in a negative coefficient of LENGTH. As I will make use of the regression coefficients for the construction of PPPs, the reliability of coefficients is of some importance. Including WEIGHT has adverse effects in this respect, so it was left out of the analysis. In addition, the economic case to include WEIGHT is weak: it has only indirect user value and production cost, and is not something that is valued by the buyer as such.

[15] As mentioned above these regressions are pooled over the years, since the number of observations for the European countries is too small in individual years.

[16] The margins used here are shown in Appendix B.

[17] Bilateral PPPs for all countries are given in Appendix B.

[18] The PPPs labelled 'dummy method' in tables 5.7 and 5.8 are the same in both tables, as the dummy index is a fixed weights index (Feenstra, 1995).

6. *Summary and conclusions*

The main topic of this thesis has been the quality issue in price comparisons through time and across space. In studying this topic, this thesis has tried to answer the following questions: first, what is the impact of differences in quality on intertemporal and international price comparisons? Second, what are the theoretical and practical advantages and drawbacks of the hedonic method compared to the matched model method to deal with this 'quality issue'?
Below I briefly summarize the main conclusions from the preceding chapters. I will provide an assessment of how quality issues will affect the future of price measurement practice and the use of hedonics, and discuss the implications for measurement practices in construction of time series price indices and currency conversion factors.

The quality issue: 'inside the sample bias' and 'outside the sample bias'

Starting with the first question, which is the topic of Chapter 2, it is clear that quality differences occur for many goods and services. In a dynamic economy, new products appear continuously, and nearly as many disappear. Likewise existing products are continuously improved, and often provided at lower nominal prices given the improvement in quality characteristics. For accountants of economic activity, these changes pose a large problem on how to measure price changes which are controlled for quality improvement. In the case of international comparisons, the heterogeneity issue is even stronger. Not only are differences between similar goods larger across countries, the overlap in output mixes and expenditure packages is also smaller than in a single-country intertemporal case.
The quality issue works in two ways, i.e., as an 'inside the sample bias' and 'outside the sample bias'. The outside the sample bias relates to the entry and exit of products in the sample. The problem of neglecting new and disappearing products leads to non-representativeness of the price index sample and creates a bias the price measurement. When trying to compare 'like with like' either over time or across countries, the incorporation of new goods in a price index at the moment of introduction is not possible. New goods do not have a counterpart in a previous period, so no price change for such a good can be measured for its first phase of existence. Moreover most statistical offices only renew their price samples every five or even ten years, so that relatively new goods are only picked up after they have been around for several years. [1] The often dramatic price changes that new goods witness in their first years of existence are therefore often not measured at all. Although new goods generally do not have large shares in total expenditure, a bias is introduced never-

theless. A similar problem occurs when varieties ('new models') of existing products are introduced. Since the standard method of price index calculation is comparing prices of nearly identical products only, new models are often neglected in a similar way.

Like new products, products that disappear from the market also pose a problem when the price sample is kept fixed for a long period. When a product stops being available, its price can no longer be observed and a hole appears in the price index. Statistical offices sometimes try to fill these holes by estimating what the price of this product would have been if it still existed, occasionally on the basis of consulting firms that produced those items.

The best solution to diminish the outside the sample bias is to frequently update the sample of the price index, including the prices and weights of new products as soon as they become available, and deleting disappeared ones from the index as quickly as possible.

The second kind of bias is analogously called the 'inside the sample bias'. This bias occurs when prices of non-identical products are matched. This is especially apparent when a new variety of an existing product replaces an old variety. In the search for a replacement item, the statistician often links the prices of the two varieties, making an implicit or an (ad hoc) explicit adjustment for differences in specifications of the products. Various adjustment procedures are discussed in Chapter 2, all of which lead to biased estimates of price changes. The direction of the bias depends on the size and direction of both quality change and nominal price change.

When taken together these two problems combined can have a significant effect on the measurement of price change. The Advisory Report to the Senate Finance Committee (Boskin *et al.*, 1996) estimated that the new goods and quality bias resulted in an annual upward bias of the CPI in the U.S. of 0.6 percentage points. Similar studies in other countries showed comparable results. It is difficult to assess the plausibility of these 'net' estimates, as the exact magnitude of this bias is determined by 'gross' biases in thousands of individual sub-indices and their weights in the final index. Such biases may be offsetting each other to an important extent, and these offsetting effects may change over time and across countries. Regardless of the bias in the final index, an improper index methodology can have a large effect in the case of individual goods and services. Computers provide a well researched and very relevant example, but the problem of wrong index numbers is not confined to this product alone. Indeed not all prices of goods and services are upwardly biased, and some prices (e.g., services in a non-market environment) may in fact show a significant downward bias when it comes to a careful adjustment of quality characteristics (Triplett, forthcoming).

What has been said on price index numbers is equally relevant in comparisons of prices across countries, which are carried out to derive currency conversion factors such as purchasing power parities or unit value ratios. In both the expenditure approach and the industry-of-origin approach, prices of products are matched on the assumption that they are identical across countries.

Here we also encounter the two biases mentioned above. In particular the outside the sample bias is big in international comparisons of prices, especially when comparisons are made from the producer perspective. Some products (like aircraft or automobiles) are simply not produced in every country. Especially specialization on the basis of comparative advantages strengthens outside the sample biases. [2] Also increased differentiation in the product value chain, in many cases within vertically integrated multinational firms, can significantly affect adequate price comparisons across countries. Furthermore data sources from different countries can sometimes not be matched, or information is not available because of confidentiality reasons. Together this leads to a large percentage of output or expenditure that is not matched in international price comparisons. Since there is not a natural way to link countries to each others as with periods of time, quality differences between countries cannot be assumed to be as gradual as quality changes over time. Hence the outside the sample bias cannot be easily remedied, not even with larger samples of goods and services. [3]

Because of the lack of gradual differences in quality of goods and services and the lack of an obvious scaling of countries, the inside the sample bias is also more severe in international comparisons than in price index numbers. In case of the expenditure approach, which uses specification pricing, it may still be feasible to match items which are closely similar across countries, but such matched items may not be the most representative for both countries' expenditure patterns. The industry-of-origin approach, which uses unit values, is based on average 'prices' of aggregates of goods. Not only quality differences between individual products are then important, but also the composition of the aggregates for which unit values are calculated. The importance of this output mix problem was pointed out in Chapter 5. For international comparisons of output and productivity at industry level, the expenditure approach does not suffice because expenditure price comparisons only cover final products or services. In order to cover output of intermediate products, which is a substantial part of total output, industry of origin comparisons remain necessary, and the need to address the inside the sample bias therefore remains high on the agenda.

The hedonic method and its uses

I now turn to the second question addressed in this thesis. As the matched model method is in most fields still the default option of price measurement, I will discuss here the relative merits and drawbacks of the hedonic method.

Based on the premise that buyers are interested in consuming the *characteristics* of finished goods and services rather than the goods and services themselves, the hedonic method applies regression techniques to make an explicit quality adjustment on price indices or ratios. Because actual consumer behaviour obviously not only relates to final products, but also or rather to their characteristics, this premise makes the theoretical case for the hedonic method very strong. Especially if one is interested in cost-of-living-indices, which hold consumer utility intact, the hedonic hypothesis seems inevitable.

Nevertheless, there have been theoretical objections against the hedonic method. The expansion of applications of the hedonic method in the U.S. has given rise to the revival of an old debate that has become known as the 'resource cost – user value' debate. Although in a sense somewhat of a separate issue, this debate has major theoretical implications for the use of hedonics. The core of the debate is the conflict between those who think that only variables that reflect a change in production costs should be taken into account when measuring price changes, and those who think that ultimately the value that users attach to a product (or a certain characteristic) is what counts. Triplett (1983) pointed out that both views have their merits depending on the approach of the price measurement. But from an economic point of view, user value and resource cost must converge. First, producers who do not take user value into account in their production process will lose market share to competitors that provide products and services with higher user value. Second, if in a free market economy the same user value is delivered at higher production cost than a competitor there will ultimately be a loss of business. Therefore this thesis holds the view that price measures should reflect both user value and producer costs, and that characteristics used for the estimation of hedonic functions should also reflect both.

Apart from theoretical considerations, practical ones have to be assessed as well, given the fact that the construction of price indices is of a highly empirical nature. Obviously, the greatest strength of the hedonic method compared with the matched model method is that it allows the statistician or researcher to calculate a price index or price ratio when the number of matched items is low or even zero. The hedonic method is therefore much better suited to deal with strong inside and outside the sample biases than the matched model method.

But as pointed out in Chapter 3, the hedonic method also has several practical drawbacks. One which is raised most often, is the lack of sufficient data on individual items and their characteristics. However, as discussed, this argument is a fallacy. The data problem for the hedonic method is not any bigger than for the matched model method. Products can only be matched if the analyst is sufficiently certain that they are (nearly) identical, which requires the same information on the characteristics of the product available. In fact the application of the hedonic method by statistical offices may not be that more expensive or data intensive after all, since much information on characteristics of goods is collected anyway for the purpose of matching. With the availability of detailed data on prices including characteristics (e.g., scanner data) such information is increasingly available. Summing up, the 'data argument' cannot be used to dismiss the hedonic method in favour of the conventional matched model method.

Relevant practical problems that need to be addressed when applying the hedonic method were discussed in more detail in Chapter 3. The bottom line of the assessment in that chapter is that the correct specification of a hedonic model is the crux to the construction of quality-adjusted price indices and currency conversion factors. In practice this may prove hard, because the choice of which characteristics to include is not straightforward. Furthermore, many 'desired' characteristics may prove unobservable or not quantifiable. The difficulty of specifying an adequate hedonic function will likely be the biggest obstacle to the practical application of the hedonic method.

This thesis includes two original empirical applications of the hedonic method. In Chapter 4, the hedonic method is applied on advertisement and scanner data for computers in the Netherlands. Price indices declined somewhat more than matched model indices, confirming the hypothesis that an index of a product with rapidly changing characteristics, like computers, is upwardly biased when unadjusted for quality changes. Although the difference between hedonic and matched model indices was not small, the effect of using an explicit quality adjustment by means of hedonics is already substantially reduced when using a chained index method rather than a fixed base. In the case of chaining, resampling takes care of much of the quality problem. Monthly chaining may not be necessary for every product, but yearly resampling of the price sample is a likely minimum. In this respect, in dealing with quality issues one has more to gain from frequent resampling than from the implementation of hedonic quality adjustment methods as such. Another important matter is the extensive use of weights for the individual items in the index. These conclusions were confirmed when the (high frequency) matched model and hedonic methods were applied on the database for the construction of the Dutch CPI for computers.

Although frequent resampling and the use of expenditure weights would therefore tremendously improve price indices, this can not always be realised in practice. Statistical offices have limited budgets for the collection of price data, and encounter special difficulties with the collection of expenditure shares. The hedonic method may provide a partial solution when the price sample is inadequate, especially when the number of matches is low and sampling is infrequent, but it does not provide a full and satisfactory remedy.

For international comparisons of producer prices of cars, which are studied in Chapter 5, there appears to be no overlap in items that can be matched. In different countries, different cars are produced, so applying the matched model method on an individual product basis is not possible. Using the hedonic method shows that average price ratios, unadjusted for quality differences, cause a severe understating of the real output value of the automobile industry for countries where on average larger and powerful cars are produced vis-à-vis the real output value in countries which produce smaller cars. However, the hedonic function for cars suffers from one major problem: the number of available characteristics is quite limited when set against the huge amount of price-determining characteristics of such complex products. In fact the available characteristics for cars (such as size and weight) are all proxies of the characteristics that actually matter.

Although the correct specification of hedonic functions for complex products like cars is very difficult and may be even impossible, we should not dismiss the results beforehand and stick with the unadjusted matched model results. Paying close attention to how we specify a hedonic function remains important, but one must accept the limited availability of data. By choosing the best possible specification, one should determine whether it is good enough, and judge the results by comparing them with matched model results. In some cases matching products may not be possible at all and one can then either has to rely on price relatives for related products in other industries or use the hedonic method, despite its limitations.
Because both the inside the sample and outside the sample biases are much bigger in international price comparisons than for price indices over time, the case for hedonic quality adjustments in the former case is in fact quite strong. However, the specification of a common hedonic function, as is possible for different periods (provided they are not too far apart), is much more difficult to apply across countries. The relation between prices and characteristics is determined by preferences and production processes, which are far less similar between countries than between relatively close periods within a single country. Only in very special cases with internationally standardised products the estimation of a common international hedonic function might be possible. Again computers may be a good example, as illustrated by Moch and Triplett (2002). For the data on cars used in Chapter 5, the relation between prices and characteristics differed too much between countries to employ a common hedonic function. The inadequacy of a common hedonic model in international price comparisons also precludes the use of the hedonic dummy method, which requires the pooling of data across countries.

The future of hedonics

Although the popularity of the hedonic method is still large under many scholars, it has come under recent attack from an American expert panel on price measurement (Schultze and Mackie, 2001). The Schultze Committee states that many econometric issues surrounding the hedonic method have still not adequately been resolved. Because the hedonic method has, in the Committee's view, not fully been developed, it states that the class of goods to which the hedonic method may be justifiably applied is narrow. However, much of the criticism and conservatism against the hedonic method expressed in this report is directed towards a variant of the hedonic method that is not advocated by many scholars, namely the dummy method.
The work by Schultze and Mackie ignores a lot of recent research on hedonic indices, as appears from a recent criticism by Diewert (2002). Many statistical offices have increased their use of the hedonic method compared with their previous methodologies. The views stated by Schultze and Mackie on hedonics may be shared by some statisticians and even economists, they are certainly not the *communis opinio* regarding the hedonic method.

Instead the debate on the hedonic method appears to be moving in favour of the strong advocates of hedonics, like Diewert (2002), Pakes (2002), Silver (2003) and Triplett (forthcoming).

Since this thesis puts the hedonic method as an appropriate tool for quality adjustment in price measurement, what is the best way forward in both statistical practice and academic application of hedonic price indices? As stated above, theoretical considerations are of importance when specifying a hedonic model. The choice of characteristics, and how to deal with proxies for unobservable characteristics are important issues to deal with for the correct specification of a hedonic model.

Provided such theoretical issues can be resolved, an important point is when to use the hedonic method as an instrument for quality adjustment. A distinction is made here between price indices across time and international comparisons. As noted above, for price index measurement, most can be gained by using a chained index principle with frequent resampling. The effect of hedonics in comparison with a the chain principle is modest. For price index numbers, hedonics therefore does need to be a high priority.

For international price comparisons, the need for the hedonic method is bigger. There is no international equivalent to a high frequency matched model index as an alternative method. Besides using price relatives from related products, as is common practice in current ICOP studies, hedonics is the only option especially in those fields where the overlap in production and expenditure of the same items are small. Collecting necessary data for international price comparisons is not an easy task, and using the hedonic method only makes it harder. In most cases it requires industry specific databases from, for example, multinational companies, that produce in more than one countries or marketing agencies or industry associations with offices in various countries. Without quality information, simply matching prices of products based on a similar functional use becomes an educated guess at best. This is especially true for complex products with many characteristics.

Finally, given the increased importance of the services sector in the economies of advanced countries, more attention needs to be paid to price measurement in services. Although this thesis focused on examining the quality issue in two types of (durable) goods, all of the above is equally relevant for services. As the measurement problems are huge, and the difficulty to distinguish prices, quantities and quality characteristics as set out in Chapter 1, the use of hedonic measurement may be even more relevant here. Services are often complex in terms of possessing a broad range of relevant characteristics, distinguishing them from other items. This is a problem for price measurement in general, be it hedonic or conventional measures. If we somehow can resolve these issues and enough information is available for matching prices, there is no reason why we cannot apply the hedonic method. As pointed out above, the only real drawback of the hedonic method compared with the matched model method is the difficulty of specifying an adequate hedonic function. As with physical goods, data issues are similar for both methods.

The measurement problems in the field of services are not easily overestimated. These are still huge, and preclude a rapid adoption of the hedonic method in services. But for physical goods, where the output concept is much clearer, the largest pay-off from using the hedonic method lies in international price comparisons.

Notes

[1] More frequent resampling is presently considered by many statistical offices, and has, for example, been implemented in the Dutch CPI in 2003.
[2] See, for example Dornbusch *et al*. (1977).
[3] Linking countries with the use of optimal spanning trees may resolve this problem to some extent. See Hill (1999).

Appendix tables

Table of contents

Appendix A. Appendix tables to Chapter 4

Appendix B. Appendix tables to Chapter 5

Appendix A. Appendix tables to Chapter 4

Table A.1
Summary of the CBS advertisement data base by month

	Number of observations	Average price (euro)	Average speed (MHz)	Average hard disk size (GB)	Average memory (MB)
1999					
September	136	1,256	464	10.1	92
October	119	1,234	470	9.9	88
November	105	1,375	481	10.7	91
December	133	1,442	492	13.9	92
2000					
January	113	1,399	503	14.9	93
February	123	1,419	510	15.0	91
March	122	1,486	552	16.2	94
April	107	1,430	559	16.5	94
May	113	1,415	595	16.3	91
June	127	1,362	597	16.0	88

Table A.2
Summary of the GfK data base by month, PCs

	Number of obser-vations	Quantity (1,000s)	Value (mln. euro)	Average price (euro)	Average speed (MHz)	Average hard disk (GB)	Average memory (MB)
1999							
January	1,212	90.7	116.3	1,282	338	4.8	54
February	1,132	92.2	112.5	1,220	342	4.8	55
March	1,010	115.2	139.5	1,211	354	5.1	58
April	962	85.8	111.1	1,295	370	5.4	62
May	1,040	69.9	88.8	1,270	376	5.8	64
June	1,064	68.9	82.6	1,199	378	5.7	62
July	1,178	72.5	91.8	1,266	397	6.2	65
August	1,130	75.0	91.3	1,217	408	6.2	68
September	1,158	78.1	92.8	1,188	421	6.8	70
October	1,150	84.2	102.0	1,211	437	7.2	70
November	1,261	91.8	117.3	1,278	453	7.9	72
December	1,261	114.6	147.4	1,286	468	8.2	74
2000							
January	1,225	89.9	112.6	1,253	468	8.3	72
February	1,198	90.7	115.2	1,270	488	9.1	77
March	1,227	115.5	151.9	1,315	510	10.1	80
April	1,132	79.2	103.8	1,311	518	10.2	79
May	1,088	65.8	87.6	1,331	535	10.6	80
June	1,069	68.0	88.0	1,294	557	11.1	83
July	1,084	62.7	81.1	1,293	569	11.5	82
August	1,147	72.2	92.6	1,283	601	12.0	86
September	1,205	77.0	100.0	1,299	626	12.7	90
October	1,214	84.1	112.7	1,340	672	13.9	88
November	1,275	94.7	124.1	1,310	691	14.2	91
December	1,355	112.1	155.7	1,389	715	15.4	97
2001							
January	1,271	80.0	109.7	1,371	715	15.3	95
February	1,224	80.8	107.1	1,325	730	15.6	99
March	1,259	99.3	133.0	1,339	777	16.7	103
April	1,189	75.4	101.0	1,340	774	17.2	102
May	1,205	60.9	82.3	1,351	814	18.2	116
June	1,235	69.2	92.4	1,335	877	21.1	135
July	1,139	57.4	75.5	1,315	904	20.6	129
August	1,164	60.6	75.2	1,241	898	20.5	123
September	1,200	69.1	86.1	1,246	1,007	22.8	142
October	1,254	73.2	90.9	1,242	1,073	24.8	155
November	1,308	87.6	108.1	1,234	1,138	26.3	164
December	1,361	106.2	129.9	1,223	1,252	30.1	188
2002							
January	1,389	83.8	104.9	1,252	1,256	30.3	189

Table A.3
Summary of the GfK data base by month, notebooks

	Number of observations	Quantity (1,000s)	Value (mln. euro)	Average price (euro)	Average speed (MHz)	Average hard disk size (GB)	Average memory (MB)
1999							
January	308	13.0	29.2	2,242	251	4.0	38
February	312	16.1	34.5	2,137	262	4.1	38
March	281	16.4	38.4	2,340	273	4.3	43
April	289	14.4	33.1	2,301	278	4.3	46
May	312	11.8	27.9	2,359	292	4.5	51
June	316	13.5	32.8	2,427	302	4.7	54
July	390	14.2	35.0	2,466	315	4.7	56
August	343	16.3	40.0	2,448	334	4.9	57
September	313	16.7	39.3	2,360	345	5.0	57
October	350	15.5	37.2	2,398	351	5.2	59
November	408	19.0	46.1	2,434	363	5.4	60
December	416	23.3	53.9	2,317	368	5.3	57
2000							
January	456	18.3	42.7	2,335	375	5.3	57
February	450	19.2	47.5	2,469	395	5.7	60
March	500	24.8	65.3	2,628	415	6.0	61
April	514	19.2	50.5	2,637	428	6.1	63
May	483	17.0	44.2	2,599	440	6.2	61
June	487	17.7	45.4	2,562	458	6.3	63
July	500	17.0	43.0	2,537	474	6.5	65
August	510	19.1	48.1	2,516	494	6.7	65
September	526	18.7	45.8	2,451	521	7.0	66
October	527	22.5	55.5	2,466	541	7.1	66
November	619	24.2	59.6	2,457	558	7.4	69
December	630	28.4	74.9	2,633	583	8.1	74
2001							
January	617	20.6	49.2	2,391	570	8.1	73
February	549	20.5	51.7	2,527	583	8.1	75
March	569	24.0	61.6	2,569	617	8.5	81
April	522	19.3	49.9	2,585	623	9.3	87
May	547	14.8	38.4	2,590	650	10.2	89
June	515	17.8	43.8	2,463	668	10.2	92
July	484	16.0	39.6	2,471	704	11.6	98
August	562	16.7	39.4	2,363	716	11.9	103
September	561	18.1	41.1	2,266	732	12.3	110
October	594	21.6	49.2	2,282	751	13.0	133
November	587	21.8	46.3	2,122	785	14.1	136
December	608	24.4	49.0	2,007	838	15.2	150
2002							
January	687	20.4	40.7	1,999	855	16.3	160

Table A.4
Summary of the GfK data base by month, servers

	Number of observations	Quantity (1,000s)	Value (mln. euro)	Average price (euro)	Average speed (MHz)	Average hard disk size (GB)	Average memory (MB)
1999							
January	131	3.5	15.1	4,310	353	3.4	102
February	135	3.2	13.8	4,325	378	4.2	113
March	121	3.5	13.8	3,984	378	4.1	108
April	134	3.2	14.0	4,320	392	2.9	115
May	152	2.8	11.1	3,929	402	3.2	121
June	150	3.3	12.0	3,615	399	4.7	115
July	170	3.7	14.6	3,938	420	4.1	120
August	188	3.7	15.8	4,243	438	5.7	145
September	193	4.0	15.0	3,729	458	6.3	149
October	183	4.8	18.7	3,888	516	5.8	147
November	189	5.3	18.2	3,447	543	4.0	133
December	207	7.1	24.6	3,487	549	4.4	145
2000							
January	167	4.5	16.4	3,630	526	5.8	129
February	173	4.2	17.1	4,044	501	6.3	148
March	177	5.4	24.2	4,470	516	4.8	158
April	173	3.2	14.8	4,676	528	6.4	184
May	143	2.7	12.2	4,535	550	4.0	169
June	152	2.9	14.3	4,903	574	4.3	182
July	150	2.7	13.5	4,992	594	6.2	171
August	170	3.0	15.1	5,026	641	7.8	159
September	159	2.8	10.9	3,883	636	6.6	139
October	150	3.4	18.8	5,554	669	6.0	263
November	153	3.9	18.5	4,785	700	4.3	182
December	162	4.6	23.3	5,108	716	6.2	193
2001							
January	167	3.7	17.4	4,711	720	7.4	197
February	163	3.7	16.3	4,400	747	11.4	198
March	154	4.9	25.5	5,216	785	7.5	248
April	138	2.9	14.2	4,862	782	11.8	229
May	139	2.4	12.2	5,174	811	16.2	255
June	122	3.2	13.3	4,183	865	6.4	162
July	108	3.0	12.3	4,102	906	7.1	200
August	127	3.3	17.4	5,284	944	18.8	263
September	114	2.8	10.8	3,883	948	7.6	247
October	117	3.1	13.8	4,448	956	12.9	236
November	116	3.1	14.6	4,737	921	9.0	269
December	127	3.4	14.2	4,167	957	10.8	258
2002							
January	117	3.1	13.6	4,426	999	18.2	338

Table A.5
Expenditure shares (%) of matched items in the current and reference months, PCs

	Chained weights		Fixed weights	
	Previous month	Current month	January 1999	Current month
1999				
February	78.6	79.1	78.6	79.1
March	80.9	80.7	69.6	71.2
April	77.0	72.8	65.0	56.9
May	82.6	75.4	61.2	50.8
June	80.5	82.3	54.9	47.8
July	83.1	80.0	50.9	36.1
August	79.8	78.7	43.6	23.7
September	86.4	76.2	37.7	17.8
October	79.7	80.5	31.9	10.8
November	88.1	73.7	31.3	5.8
December	85.0	81.2	27.1	3.6
2000				
January	79.8	81.4	24.7	3.0
February	75.1	70.4	11.3	1.2
March	77.1	69.9	12.4	1.7
April	79.2	82.1	8.3	1.6
May	80.3	77.1	5.9	1.8
June	81.4	80.1	7.8	0.9
July	84.0	79.8	6.1	1.5
August	84.5	79.0	5.3	0.7
September	75.9	70.5	6.7	0.5
October	76.4	76.0	5.6	0.2
November	84.2	78.0	2.4	1.3
December	85.3	76.9	2.2	0.5
2001				
January	78.6	79.2	3.9	0.7
February	80.3	82.8	2.0	0.7
March	81.6	79.8	2.7	0.1
April	85.7	80.2	2.7	0.1
May	83.1	79.9	0.3	0.1
June	80.5	74.2	2.3	0.2
July	78.5	82.5	0.2	0.3
August	78.8	79.4	2.4	0.3
September	78.3	80.4	2.7	0.3
October	81.8	77.5	0.5	0.1
November	82.7	74.7	0.3	0.0
December	83.5	75.7	0.5	0.0
2002				
January	87.7	80.2	0.0	0.0
Average	81.3	78.0		

Table A.6
Expenditure shares (%) of matched items in the current and reference months, notebooks

	Chained weights		Fixed weights	
	Previous month	Current month	January 1999	Current month
1999				
February	89.7	87.5	89.7	87.5
March	88.4	85.4	82.9	69.7
April	90.9	91.2	77.3	58.3
May	90.8	82.4	65.5	37.0
June	83.9	90.9	64.6	31.8
July	93.2	77.4	54.4	19.7
August	86.3	84.6	46.7	11.8
September	91.9	92.2	34.3	8.9
October	86.5	85.1	31.0	4.3
November	93.6	86.4	16.2	2.9
December	93.1	89.5	24.2	3.0
2000				
January	88.0	88.5	22.3	1.9
February	85.5	81.9	14.4	0.4
March	90.0	84.0	8.0	1.3
April	91.9	86.6	7.5	0.2
May	86.7	86.1	2.2	0.1
June	90.6	89.6	9.5	0.4
July	87.3	86.9	2.5	0.4
August	88.2	79.9	1.9	0.1
September	85.7	82.7	7.1	0.4
October	90.4	87.2	4.4	0.1
November	90.6	76.4	2.7	0.2
December	86.2	83.7	1.8	0.0
2001				
January	84.2	78.0	0.2	0.0
February	80.8	77.0	10.1	0.0
March	77.5	78.7	0.0	0.0
April	87.4	85.4	0.0	0.0
May	85.2	82.8	0.6	0.0
June	85.6	82.7	0.0	0.0
July	80.4	87.4	0.0	0.0
August	89.5	81.2	0.0	0.0
September	87.5	83.2	0.6	0.2
October	90.6	84.2	1.2	0.0
November	81.2	81.1	0.0	0.0
December	87.5	87.4	0.2	0.0
2002				
January	92.9	85.7	1.4	0.0
Average	81.3	78.0		

Table A.7
Expenditure shares (%) of matched items in the current and reference months, servers

Month	Chained weights		Fixed weights	
	Previous month	Current month	January 1999	Current month
1999				
February	69.9	67.6	74.8	74.8
March	65.7	62.4	62.9	64.4
April	74.2	79.9	63.9	63.8
May	72.6	69.8	47.2	55.1
June	68.6	64.2	44.2	43.9
July	68.5	66.3	39.4	42.1
August	68.2	70.9	25.5	23.7
September	71.2	70.6	27.1	19.1
October	60.0	58.8	22.8	9.5
November	61.3	74.5	16.7	9.5
December	78.9	71.3	9.6	3.8
2000				
January	62.7	72.4	7.2	5.3
February	72.1	71.9	3.3	6.0
March	70.7	79.8	3.1	8.2
April	75.5	70.6	0.7	3.4
May	66.6	73.5	0.8	0.8
June	78.4	66.8	1.5	0.3
July	70.9	64.5	0.6	0.1
August	58.6	53.2	0.6	0.0
September	57.7	58.0	0.8	0.2
October	64.5	58.8	0.0	0.0
November	49.7	65.7	10.8	2.1
December	70.6	57.5	0.0	0.0
2001				
January	60.9	59.9	0.0	0.0
February	61.4	66.3	0.3	1.0
March	69.2	68.2	0.0	0.0
April	66.0	74.4	0.0	0.0
May	65.7	51.3	0.1	0.0
June	37.8	66.2	0.0	0.0
July	54.9	66.8	0.0	0.0
August	61.7	40.9	0.0	0.0
September	52.2	65.1	0.0	0.0
October	64.5	58.0	0.0	0.0
November	62.4	63.1	0.0	0.0
December	62.2	59.2	0.0	0.0
2002				
January	62.7	56.0	0.0	0.0
Average	81.3	78.0		

Table A.8
Matched model and hedonic indices, including effects of new and old models, PCs

	Matched model	Hedonic dummy	Hedonic single imputed			Hedonic double imputed		
			Index	Effect of		Index	Effect of	
				old models (λ)	new models (π)		old models (λ)	new models (π)
1999								
January	100.0	100.0	100.0			100.0		
February	94.1	93.8	94.1	1.006	0.994	94.0	1.000	0.999
March	89.9	90.0	90.1	0.997	1.005	90.4	1.001	1.000
April	88.9	88.3	89.2	0.993	1.007	89.2	0.999	0.998
May	86.5	83.8	85.9	0.992	0.998	86.3	0.997	0.998
June	84.2	81.4	83.6	0.996	1.004	84.0	1.000	0.999
July	81.4	79.5	81.2	0.997	1.008	81.3	1.000	1.001
August	78.2	75.9	77.3	0.990	1.001	77.9	0.999	0.999
September	74.3	71.6	73.2	0.995	1.002	73.9	0.999	0.999
October	73.7	71.1	72.7	0.997	1.003	73.3	0.999	1.001
November	73.7	71.2	72.8	0.999	1.002	73.4	1.000	1.001
December	72.5	68.5	71.3	0.998	0.998	71.9	0.998	0.998
2000								
January	72.9	67.5	71.4	0.995	1.000	72.0	0.998	0.998
February	72.1	64.8	69.6	0.990	0.996	70.6	0.996	0.996
March	71.4	63.4	68.5	0.995	0.999	69.7	0.997	1.000
April	70.1	62.3	67.1	0.998	1.000	68.4	1.000	1.000
May	69.5	60.9	66.2	0.997	0.998	67.7	0.998	0.998
June	67.0	57.5	63.2	0.995	0.995	64.9	0.998	0.997
July	66.0	55.5	62.0	0.995	1.002	63.7	0.998	0.998
August	65.0	53.5	60.6	0.995	0.996	62.4	0.998	0.998
September	63.3	51.3	58.6	0.996	0.997	60.6	0.997	0.999
October	62.2	50.0	57.6	0.999	1.002	59.4	0.998	0.999
November	61.6	48.2	56.3	0.991	0.995	58.5	0.998	0.998
December	61.7	47.9	56.5	1.000	1.002	58.5	0.999	1.000
2001								
January	61.1	47.7	56.0	0.998	1.002	58.0	1.000	1.000
February	58.9	45.4	53.4	0.990	1.000	55.8	0.998	0.999
March	57.7	44.1	51.7	0.990	0.998	54.6	0.998	1.000
April	57.6	43.5	51.7	1.000	1.002	54.3	0.999	0.999
May	56.8	41.9	50.5	0.994	0.995	53.4	0.998	0.998
June	56.2	39.4	49.1	0.988	0.995	52.2	0.995	0.994
July	55.7	38.2	48.3	0.991	1.001	51.5	0.998	0.998
August	54.1	36.4	46.2	0.993	0.994	49.8	0.999	0.997
September	51.6	34.7	44.2	0.999	1.002	47.5	0.999	0.999
October	50.8	33.2	43.1	0.993	1.000	46.4	0.997	0.997
November	49.9	32.4	42.4	0.997	1.003	45.5	1.000	0.999
December	47.8	30.2	40.3	0.993	1.001	43.3	0.997	0.997
2002								
January	47.7	30.7	40.2	0.998	1.000	43.4	1.001	1.001

Table A.9
Matched model and hedonic indices, including effects of new and old models, notebooks

	Matched model	Hedonic dummy	Hedonic single imputed			Hedonic double imputed		
			Index	Effect of		Index	Effect of	
				old models (λ)	new models (π)		old models (λ)	new models (π)
1999								
January	100.0	100.0	100.0			100.0		
February	94.5	94.9	94.5	0.996	1.004	94.5	1.000	1.001
March	102.7	101.4	102.1	0.995	0.999	102.6	0.999	0.999
April	97.7	97.9	97.2	0.998	1.003	97.7	1.001	1.001
May	96.8	97.1	96.1	0.997	1.000	96.7	1.001	0.999
June	94.1	93.6	92.8	0.995	0.998	93.9	0.999	1.000
July	91.7	92.3	90.7	1.000	1.002	91.8	1.000	1.002
August	89.8	87.1	87.8	0.992	0.997	89.4	0.997	0.998
September	88.0	83.1	85.6	0.998	0.997	87.4	0.999	0.999
October	87.5	80.4	84.9	0.996	1.001	86.6	0.998	0.999
November	85.0	79.2	82.6	0.999	1.003	84.3	1.000	1.002
December	84.0	78.1	81.4	0.998	1.000	83.3	1.000	1.001
2000								
January	84.6	79.4	82.0	0.997	1.002	83.9	1.000	1.001
February	81.9	79.3	79.8	1.001	1.004	81.7	1.002	1.003
March	81.0	79.7	78.8	0.999	0.999	81.1	1.000	1.003
April	80.4	77.9	78.1	1.000	0.998	80.3	0.999	0.999
May	80.7	78.5	78.6	1.001	1.002	80.6	1.000	1.001
June	79.1	77.3	77.1	1.000	1.001	79.1	1.000	1.000
July	76.8	73.9	74.5	0.997	0.999	76.6	0.999	1.000
August	76.3	72.9	73.7	1.000	0.996	76.1	0.999	0.999
September	73.0	69.9	70.5	1.001	0.998	72.8	1.001	0.999
October	72.7	68.6	69.9	1.000	0.997	72.3	0.999	0.999
November	70.5	66.1	67.9	0.998	1.002	70.1	1.000	0.998
December	70.1	66.0	67.4	0.998	1.000	69.7	1.001	1.000
2001								
January	69.5	62.3	65.6	0.997	0.986	68.5	0.997	0.994
February	68.9	63.6	65.5	1.002	1.005	68.3	1.002	1.005
March	67.6	62.0	63.9	0.995	0.998	66.9	0.999	0.999
April	68.2	61.5	64.1	0.999	0.996	67.3	0.999	0.998
May	65.6	59.5	61.8	1.000	1.002	64.8	1.000	1.001
June	63.1	56.3	59.1	1.000	0.994	62.2	0.999	0.998
July	61.3	53.3	57.1	0.999	0.996	60.2	0.998	0.998
August	59.3	51.1	55.1	1.001	0.998	58.1	0.999	1.000
September	59.0	49.9	54.1	0.996	0.989	57.7	0.998	1.000
October	56.5	46.9	51.5	0.998	0.996	55.2	0.999	0.999
November	53.2	44.5	48.3	0.998	0.999	52.0	1.000	1.002
December	51.1	41.9	46.1	0.999	0.994	49.9	0.999	0.998
2002								
January	50.2	41.4	45.2	0.999	0.999	49.0	1.000	1.001

Table A.10
Matched model and hedonic indices, including effects of new and old models, servers

	Matched model	Hedonic dummy	Hedonic single imputed			Hedonic double imputed		
			Index	Effect of		Index	Effect of	
				old models (λ)	new models (π)		old models (λ)	new models (π)
1999								
January	100.0	100.0	100.0			100.0		
February	97.0	98.4	95.0	0.989	0.991	97.4	1.007	0.997
March	89.3	88.2	88.1	0.998	1.009	88.8	1.000	0.990
April	88.6	89.4	87.2	0.985	1.013	88.4	1.003	1.001
May	81.9	80.0	80.7	0.986	1.015	81.2	0.994	0.999
June	82.5	77.5	77.9	0.971	0.986	80.6	0.997	0.989
July	84.6	78.2	79.4	0.990	1.006	82.2	0.998	0.997
August	80.9	71.5	74.9	0.981	1.004	77.5	0.992	0.994
September	77.0	64.7	69.5	0.977	0.999	72.4	0.988	0.993
October	73.7	67.3	72.3	1.025	1.059	72.1	1.017	1.023
November	72.7	67.6	71.1	1.001	0.996	71.3	1.003	1.000
December	70.5	63.2	67.0	0.989	0.983	68.6	0.997	0.996
2000								
January	69.0	63.3	67.0	1.003	1.017	67.5	0.999	1.005
February	68.2	61.2	63.6	0.974	0.987	66.4	0.998	0.997
March	71.2	71.4	67.6	1.004	1.014	70.8	1.011	1.012
April	71.2	68.3	67.3	0.990	1.007	70.0	0.996	0.993
May	69.0	61.9	62.6	0.961	0.996	66.3	0.983	0.993
June	71.0	64.5	65.8	0.993	1.029	68.4	0.998	1.006
July	70.8	62.7	65.9	0.991	1.015	67.2	0.992	0.994
August	68.1	64.3	67.4	1.040	1.022	66.1	1.015	1.007
September	62.4	53.1	58.1	0.952	0.988	58.9	0.985	0.988
October	66.9	55.8	60.8	0.987	0.989	62.3	0.994	0.993
November	65.3	54.0	59.9	0.999	1.010	60.5	0.997	0.997
December	63.8	56.0	60.6	1.013	1.024	60.3	1.011	1.009
2001								
January	64.1	50.9	59.8	0.988	0.994	58.4	0.986	0.978
February	61.5	51.0	58.7	1.003	1.020	56.7	1.007	1.004
March	59.9	51.7	56.7	0.996	0.996	55.8	1.004	1.006
April	58.1	52.6	56.0	1.002	1.016	54.8	1.012	1.002
May	57.2	48.4	52.5	0.955	0.997	52.5	0.987	0.983
June	54.7	48.4	52.1	1.018	1.020	51.5	1.016	1.011
July	52.5	41.1	45.1	0.952	0.946	47.3	0.977	0.978
August	52.8	44.8	46.9	1.003	1.031	48.9	1.011	1.019
September	52.6	41.3	45.3	0.966	1.004	47.1	0.981	0.988
October	52.1	42.4	46.2	1.015	1.013	47.2	1.011	0.999
November	49.0	41.8	44.7	1.002	1.027	45.5	1.015	1.010
December	46.9	37.9	40.8	0.962	0.992	42.4	0.985	0.989
2002								
January	47.2	38.4	41.0	0.991	1.008	42.6	1.002	0.998

Appendix B. Appendix tables to Chapter 5

Table B.1
Results from a semi-logarithmic regression, pooled across countries, 1995

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	**7.2521**	**7.2171**	**8.4222**	**7.3502**	**7.1589**	**7.5369**
	(0.1207)	*(0.2224)*	*(0.2357)*	*(0.1060)*	*(0.1957)*	*(0.2170)*
ENGINE	**0.0004**	**0.0004**	**0.0005**			
	(0.0000)	*(0.0000)*	*(0.0000)*			
POWER				**0.0054**	**0.0060**	**0.0074**
				(0.0002)	*(0.0002)*	*(0.0002)*
LENGTH	**0.2849**			**0.3235**		
	(0.0323)			*(0.0259)*		
WIDHT		**0.6914**			**0.9163**	
		(0.1464)			*(0.1218)*	
HEIGHT			–0.1303			**0.7702**
			(0.1659)			*(0.1540)*
Germany	**1.1086**	**1.1130**	**1.1497**	**0.9749**	**0.9527**	**0.9179**
	(0.0604)	*(0.0637)*	*(0.0652)*	*(0.0528)*	*(0.0571)*	*(0.0591)*
France	**2.2982**	**2.2474**	**2.2825**	**2.1817**	**2.0967**	**2.0524**
	(0.0714)	*(0.0751)*	*(0.0771)*	*(0.0629)*	*(0.0677)*	*(0.0700)*
UK	**0.1373**	**0.1202**	**0.1461**	–0.0031	–0.0496	–0.1023
	(0.0607)	*(0.0639)*	*(0.0657)*	*(0.0533)*	*(0.0574)*	*(0.0593)*
Italy	**0.9299**	**0.9000**	**0.9368**	**0.7695**	**0.7048**	**0.6438**
	(0.0763)	*(0.0803)*	*(0.0826)*	*(0.0672)*	*(0.0725)*	*(0.0751)*
Japan	**–1.6687**	**–1.6585**	**–1.6610**	**–1.9755**	**–2.0027**	**–2.1085**
	(0.0322)	*(0.0338)*	*(0.0348)*	*(0.0281)*	*(0.0311)*	*(0.0297)*
adjusted R^2	0.9637	0.9599	0.9582	0.9711	0.9661	0.9641

Notes:
Standard errors are between brackets; coefficients in bold are significant at the 5% level of significance;
Coefficients in (1) are from a regression with ENGINE and LENGTH as independent variables;
Coefficients in (2) are from a regression with ENGINE and WIDTH as independent variables;
Coefficients in (3) are from a regression with ENGINE and HEIGHT as independent variables;
Coefficients in (4) are from a regression with POWER and LENGTH as independent variables;
Coefficients in (5) are from a regression with POWER and WIDTH as independent variables;
Coefficients in (6) are from a regression with POWER and HEIGHT as independent variables.

Table B.2
Results from a double logarithmic regression, pooled across countries, 1995

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	**1.7876**	–0.1101	**1.9054**	**4.6564**	**4.8480**	**4.4974**
	(0.1991)	*(0.2650)*	*(0.2407)*	*(0.1342)*	*(0.1145)*	*(0.1385)*
ENGINE	**1.1215**	**1.5167**	**0.9874**			
	(0.0561)	*(0.0558)*	*(0.0233)*			
POWER				**0.8802**	**1.0173**	**0.9888**
				(0.0395)	*(0.0403)*	*(0.0214)*
LENGTH	**–0.5633**			**0.4463**		
	(0.2131)			*(0.1535)*		
WIDHT		**–3.6068**			–0.3112	
		(0.3508)			*(0.2596)*	
HEIGHT			0.2005			**1.0072**
			(0.2379)			*(0.2238)*
Germany	**0.9873**	**1.0530**	**0.9689**	**0.9780**	**0.9878**	**0.9514**
	(0.0640)	*(0.0591)*	*(0.0648)*	*(0.0608)*	*(0.0613)*	*(0.0605)*
France	**2.1218**	**2.2276**	**2.1238**	**2.1721**	**2.1673**	**2.1266**
	(0.0758)	*(0.0701)*	*(0.0767)*	*(0.0722)*	*(0.0728)*	*(0.0717)*
UK	0.0064	0.0720	–0.0029	–0.0289	–0.0348	–0.0653
	(0.0644)	*(0.0594)*	*(0.0652)*	*(0.0610)*	*(0.0614)*	*(0.0606)*
Italy	**0.7564**	**0.8345**	**0.7508**	**0.7256**	**0.7144**	**0.6671**
	(0.0811)	*(0.0747)*	*(0.0824)*	*(0.0771)*	*(0.0775)*	*(0.0767)*
Japan	**–1.7595**	**–1.7747**	**–1.7661**	**–2.0266**	**–2.0731**	**–2.0738**
	(0.0337)	*(0.0309)*	*(0.0341)*	*(0.0324)*	*(0.0334)*	*(0.0304)*
adjusted R^2	0.9582	0.9649	0.9577	0.9622	0.9617	0.9631

Note: see Table B.1.

Table B.3
Results from a semi-logarithmic regression, pooled across countries, 1996

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	**7.3121**	**7.2340**	**8.4499**	**7.3653**	**7.0654**	**7.7971**
	(0.1245)	*(0.2335)*	*(0.2235)*	*(0.1084)*	*(0.2035)*	*(0.2126)*
ENGINE	**0.0004**	**0.0004**	**0.0005**			
	(0.0000)	*(0.0000)*	*(0.0000)*			
POWER				**0.0051**	**0.0056**	**0.0070**
				(0.0002)	*(0.0002)*	*(0.0002)*
LENGTH	**0.2973**			**0.3384**		
	(0.0332)			*(0.0264)*		
WIDHT		**0.7507**			**1.0230**	
		(0.1530)			*(0.1263)*	
HEIGHT			–0.0676			**0.6427**
			(0.1615)			*(0.1509)*
Germany	**1.0197**	**1.0172**	**1.0516**	**0.9073**	**0.8786**	**0.8464**
	(0.0602)	*(0.0636)*	*(0.0654)*	*(0.0532)*	*(0.0576)*	*(0.0606)*
France	**2.3014**	**2.2433**	**2.2842**	**2.2044**	**2.1104**	**2.0545**
	(0.0671)	*(0.0709)*	*(0.0740)*	*(0.0596)*	*(0.0641)*	*(0.0683)*
UK	0.1143	0.1043	0.1256	–0.0143	–0.0515	–0.1032
	(0.0618)	*(0.0652)*	*(0.0671)*	*(0.0546)*	*(0.0591)*	*(0.0619)*
Italy	**0.8826**	**0.8348**	**0.8750**	**0.7450**	**0.6589**	**0.6188**
	(0.0736)	*(0.0778)*	*(0.0800)*	*(0.0654)*	*(0.0704)*	*(0.0741)*
Japan	**–1.7371**	**–1.7284**	**–1.7357**	**–2.0049**	**–2.0277**	**–2.1412**
	(0.0316)	*(0.0334)*	*(0.0344)*	*(0.0279)*	*(0.0309)*	*(0.0301)*
adjusted R^2	0.9659	0.9623	0.9605	0.9727	0.9679	0.9649

Note: see Table B.1.

Table B.4
Results from a double logarithmic regression, pooled across countries, 1996

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	**1.9422**	0.1629	**2.0196**	**4.7278**	**4.9745**	**4.6789**
	(0.2087)	*(0.2782)*	*(0.2102)*	*(0.1376)*	*(0.1171)*	*(0.1384)*
ENGINE	**1.0733**	**1.4689**	**0.9811**			
	(0.0579)	*(0.0585)*	*(0.0245)*			
POWER				**0.8497**	**0.9726**	**0.9689**
				(0.0398)	*(0.0403)*	*(0.0221)*
LENGTH	–0.3849			**0.5172**		
	(0.2175)			*(0.1553)*		
WIDHT		**–3.3305**			–0.0992	
		(0.3678)			*(0.2630)*	
HEIGHT			0.1602			**0.8261**
			(0.2275)			*(0.2134)*
Germany	**0.9196**	**0.9855**	**0.9066**	**0.9003**	**0.9045**	**0.8739**
	(0.0638)	*(0.0597)*	*(0.0644)*	*(0.0598)*	*(0.0605)*	*(0.0600)*
France	**2.1539**	**2.2608**	**2.1492**	**2.1912**	**2.1808**	**2.1285**
	(0.0711)	*(0.0669)*	*(0.0727)*	*(0.0669)*	*(0.06780*	*(0.0677)*
UK	0.0157	0.0703	0.0060	–0.0556	–0.0645	–0.0876
	(0.0656)	*(0.0612)*	*(0.0661)*	*(0.0613)*	*(0.0619)*	*(0.0613)*
Italy	**0.7481**	**0.8532**	**0.7439**	**0.7239**	**0.7103**	**0.6757**
	(0.0781)	*(0.0735)*	*(0.0790)*	*(0.0734)*	*(0.0742)*	*(0.0735)*
Japan	**–1.8114**	**–1.8262**	**–1.8164**	**–2.0532**	**–2.0923**	**–2.1037**
	(0.0332)	*(0.0309)*	*(0.0336)*	*(0.0314)*	*(0.0325)*	*(0.0299)*
adjusted R^2	0.9611	0.9663	0.9608	0.9656	0.9648	0.9659

Note: see Table B.1.

Table B.5
Results from a semi-logarithmic regression, pooled across countries, 1997

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	**7.3861**	**7.0861**	**8.9180**	**7.5552**	**7.0874**	**7.9871**
	(0.1272)	*(0.2408)*	*(0.2440)*	*(0.1102)*	*(0.2026)*	*(0.2212)*
ENGINE	**0.0004**	**0.0004**	**0.0005**			
	(0.0000)	*(0.0000)*	*(0.0000)*			
POWER				**0.0054**	**0.0057**	**0.0070**
				(0.0002)	*(0.0002)*	*(0.0001)*
LENGTH	**0.2692**			**0.2893**		
	(0.0336)			*(0.0268)*		
WIDHT		**0.8386**			**1.0098**	
		(0.1579)			*(0.1259)*	
HEIGHT			**–0.4260**			**0.5097**
			(0.1773)			*(0.1585)*
Germany	**1.0624**	**1.0463**	**1.1121**	**0.9389**	**0.9089**	**0.8901**
	(0.0565)	*(0.0590)*	*(0.0609)*	*(0.0486)*	*(0.0509)*	*(0.0544)*
France	**2.2927**	**2.2256**	**2.3229**	**2.1940**	**2.1082**	**2.0781**
	(0.0671)	*(0.0701)*	*(0.0739)*	*(0.0581)*	*(0.0606)*	*(0.0659)*
UK	0.1162	0.0857	0.1297	–0.0047	–0.0481	–0.0913
	(0.0633)	*(0.0657)*	*(0.0679)*	*(0.0548)*	*(0.0571)*	*(0.0607)*
Italy	**0.9350**	**0.8801**	**0.9643**	**0.7496**	**0.6727**	**0.6503**
	(0.0778)	*(0.0813)*	*(0.0834)*	*(0.0673)*	*(0.0702)*	*(0.0747)*
Japan	**–1.7472**	**–1.7396**	**–1.7378**	**–2.0476**	**–2.0537**	**–2.1561**
	(0.0319)	*(0.0331)*	*(0.0342)*	*(0.0277)*	*(0.0297)*	*(0.0291)*
adjusted R^2	0.9686	0.9663	0.9647	0.9761	0.9737	0.9707

Note: see Table B.1.

Table B.6
Results from a double logarithmic regression, pooled across countries, 1997

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	**1.5863**	–0.1773	**1.9281**	**4.8468**	**4.9652**	**4.8103**
	(0.2608)	*(0.2866)*	*(0.2091)*	*(0.1408)*	*(0.1129)*	*(0.1361)*
ENGINE	**1.1859**	**1.5270**	**1.0192**			
	(0.0570)	*(0.0612)*	*(0.0248)*			
POWER				**0.9158**	**0.9745**	**0.9651**
				(0.0393)	*(0.0400)*	*(0.0215)*
LENGTH	**–0.7121**			0.2309		
	(0.2175)			*(0.1612)*		
WIDHT		**–3.4989**			–0.0937	
		(0.3898)			*(0.2736)*	
HEIGHT			–0.4377			**0.4743**
			(0.2482)			*(0.2320)*
Germany	**0.9788**	**1.0571**	**0.9847**	**0.9394**	**0.9419**	**0.9218**
	(0.0594)	*(0.0562)*	*(0.0607)*	*(0.0558)*	*(0.0561)*	*(0.0564)*
France	**2.1567**	**2.2956**	**2.1969**	**2.1889**	**2.1878**	**2.1520**
	(0.0705)	*(0.0674)*	*(0.0734)*	*(0.0666)*	*(0.0673)*	*(0.0683)*
UK	0.0130	0.0788	0.0292	–0.0451	–0.0504	–0.0664
	(0.0666)	*(0.0626)*	*(0.0678)*	*(0.0627)*	*(0.0627)*	*(0.0629)*
Italy	**0.8231**	**0.9628**	**0.8402**	**0.7086**	**0.7032**	**0.6807**
	(0.0819)	*(0.0781)*	*(0.0835)*	*(0.0771)*	*(0.0772)*	*(0.0774)*
Japan	**–1.8070**	**–1.8209**	**–1.8030**	**–2.0897**	**–2.1093**	**–2.1122**
	(0.0333)	*(0.0311)*	*(0.0339)*	*(0.0319)*	*(0.0328)*	*(0.0303)*
adjusted R^2	0.9648	0.9693	0.9642	0.9686	0.9685	0.9688

Note: see Table B.1.

Table B.7
Results from a semi-logarithmic regression, pooled across countries, 1998

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	**7.0679**	**5.2413**	**9.0920**	**7.2686**	**5.1785**	**8.0566**
	(0.1628)	*(0.4200)*	*(0.2704)*	*(0.1406)*	*(0.3125)*	*(0.2482)*
ENGINE	**0.0004**	**0.0004**	**0.0005**			
	(0.0000)	*(0.0000)*	*(0.0000)*			
POWER				**0.0053**	**0.0048**	**0.0069**
				(0.0002)	*(0.0002)*	*(0.0002)*
LENGTH	**0.3505**			**0.3579**		
	(0.0415)			*(0.0336)*		
WIDHT		**1.9985**			**2.1545**	
		(0.2656)			*(0.1890)*	
HEIGHT			**–0.5187**			**0.4702**
			(0.1925)			*(0.1743)*
Germany	**1.0724**	**1.0319**	**1.1078**	**0.9600**	**0.9351**	**0.9001**
	(0.0626)	*(0.0640)*	*(0.0685)*	*(0.0536)*	*(0.0526)*	*(0.0611)*
France	**2.2662**	**2.1736**	**2.2755**	**2.1705**	**2.0933**	**2.0456**
	(0.0684)	*(0.0698)*	*(0.0760)*	*(0.0589)*	*(0.0574)*	*(0.0675)*
UK	0.0968	0.0598	0.0865	–0.0070	–0.0242	–0.1205
	(0.0635)	*(0.0644)*	*(0.0694)*	*(0.0546)*	*(0.0535)*	*(0.0617)*
Italy	**0.9016**	**0.8084**	**0.9033**	**0.7442**	**0.6708**	**0.6308**
	(0.0811)	*(0.0830)*	*(0.0884)*	*(0.0698)*	*(0.0682)*	*(0.0789)*
Japan	**–1.7810**	**–1.7495**	**–1.7810**	**–2.0629**	**–1.9957**	**–2.1795**
	(0.0349)	*(0.0359)*	*(0.0380)*	*(0.0298)*	*(0.0315)*	*(0.0319)*
adjusted R^2	0.9706	0.9695	0.9655	0.9778	0.9785	0.9715

Note: see Table B.1.

Table B.8
Results from a double logarithmic regression, pooled across countries, 1998

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	**0.7866**	**0.6019**	**0.9676**	**4.3815**	**4.4646**	**4.6141**
	(0.2414)	*(0.3042)*	*(0.2574)*	*(0.1876)*	*(0.1407)*	*(0.1666)*
ENGINE	**1.1762**	**1.2245**	**1.1464**			
	(0.0576)	*(0.0763)*	*(0.0301)*			
POWER				**0.8875**	**0.7846**	**1.0036**
				(0.0413)	*(0.0428)*	*(0.0259)*
LENGTH	–0.1235			**0.6327**		
	(0.2315)			*(0.1925)*		
WIDHT		–0.6637			**2.4357**	
		(0.6252)			*(0.4017)*	
HEIGHT			–0.4360			**0.5116**
			(0.2502)			*(0.2439)*
Germany	**1.0217**	**1.0299**	**1.0358**	**0.9556**	**0.9430**	**0.9348**
	(0.0620)	*(0.0625)*	*(0.0623)*	*(0.0599)*	*(0.0580)*	*(0.0609)*
France	**2.1865**	**2.2006**	**2.2137**	**2.1749**	**2.1350**	**2.1313**
	(0.0677)	*(0.0686)*	*(0.0690)*	*(0.0567)*	*(0.0636)*	*(0.0674)*
UK	0.0252	0.0286	0.0421	–0.0379	–0.0331	–0.0752
	(0.0628)	*(0.0627)*	*(0.0631)*	*(0.0608)*	*(0.0587)*	*(0.0615)*
Italy	**0.8217**	**0.8368**	**0.8383**	**0.7117**	**0.6796**	**0.6718**
	(0.0804)	*(0.0815)*	*(0.0806)*	*(0.0779)*	*(0.0751)*	*(0.0787)*
Japan	**–1.8226**	**–1.8296**	**–1.8169**	**–2.0976**	**–2.0314**	**–2.1411**
	(0.0343)	*(0.0348)*	*(0.0340)*	*(0.0333)*	*(0.0349)*	*(0.0319)*
adjusted R^2	0.9708	0.9709	0.9710	0.9724	0.9741	0.9719

Note: see Table B.1.

Table B.9
Results from a semi-logarithmic regression, pooled across countries, 1999

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	**7.3912**	**5.8620**	**8.7731**	**7.5106**	**5.7858**	**7.8761**
	(0.1554)	*(0.3703)*	*(0.2581)*	*(0.1374)*	*(0.3013)*	*(0.2402)*
ENGINE	**0.0004**	**0.0004**	**0.0005**			
	(0.0000)	*(0.0000)*	*(0.0000)*			
POWER				**0.0054**	**0.0049**	**0.0069**
				(0.0002)	*(0.0002)*	*(0.0002)*
LENGTH	**0.2675**			**0.3031**		
	(0.0401)			*(0.0330)*		
WIDHT		**1.6164**			**1.8057**	
		(0.2363)			*(0.1838)*	
HEIGHT			–0.2870			**0.5958**
			(0.1806)			*(0.1664)*
Germany	**1.0838**	**1.0494**	**1.1105**	**0.9596**	**0.9360**	**0.9123**
	(0.0623)	*(0.0625)*	*(0.0663)*	*(0.0548)*	*(0.0540)*	*(0.0601)*
France	**2.1783**	**2.1055**	**2.1727**	**2.0756**	**2.0080**	**1.9584**
	(0.0695)	*(0.0695)*	*(0.0748)*	*(0.0615)*	*(0.0601)*	*(0.0675)*
UK	0.1055	0.0789	0.1027	–0.0102	–0.0254	–0.1073
	(0.0671)	*(0.0669)*	*(0.0717)*	*(0.0594)*	*(0.0583)*	*(0.0649)*
Italy	**0.9430**	**0.8675**	**0.9420**	**0.7584**	**0.6950**	**0.6589**
	(0.0819)	*(0.0822)*	*(0.0870)*	*(0.0725)*	*(0.0710)*	*(0.0789)*
Japan	**–1.7550**	**–1.7266**	**–1.7596**	**–2.0587**	**–1.9972**	**–2.1677**
	(0.0365)	*(0.0368)*	*(0.0387)*	*(0.0321)*	*(0.0340)*	*(0.0331)*
adjusted R^2	0.9707	0.9709	0.9672	0.9765	0.9771	0.9719

Note: see Table B.1.

Table B.10
Results from a double logarithmic regression, pooled across countries, 1999

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	**1.4370**	**0.9816**	**1.5822**	**4.8242**	**4.8566**	**4.7288**
	(0.2268)	*(0.3058)*	*(0.2532)*	*(0.1823)*	*(0.1367)*	*(0.1698)*
ENGINE	**1.2166**	**1.2527**	**1.0545**			
	(0.0583)	*(0.0779)*	*(0.0288)*			
POWER				**0.9006**	**0.7994**	**0.9699**
				(0.0430)	*(0.0471)*	*(0.0257)*
LENGTH	**–0.7470**			0.3024		
	(0.2374)			*(0.1941)*		
WIDHT		**–1.6922**			**1.6261**	
		(0.6283)			*(0.4183)*	
HEIGHT			–0.0866			**0.6655**
			(0.2482)			*(0.2461)*
Germany	**1.0073**	**1.0231**	**0.9986**	**0.9446**	**0.9339**	**0.9273**
	(0.0625)	*(0.0634)*	*(0.0636)*	*(0.0620)*	*(0.0609)*	*(0.0618)*
France	**2.0572**	**2.0984**	**2.0711**	**2.0736**	**2.0467**	**2.0371**
	(0.0696)	*(0.0707)*	*(0.0716)*	*(0.0695)*	*(0.0683)*	*(0.0697)*
UK	0.0077	0.0199	0.0141	–0.0531	–0.0488	–0.0825
	(0.0673)	*(0.0676)*	*(0.0688)*	*(0.0669)*	*(0.0656)*	*(0.0667)*
Italy	**0.8183**	**0.8594**	**0.8244**	**0.6861**	**0.6709**	**0.6554**
	(0.0822)	*(0.0837)*	*(0.0837)*	*(0.0818)*	*(0.0799)*	*(0.0881)*
Japan	**–1.8074**	**–1.8271**	**–1.8089**	**–2.0930**	**–2.0378**	**–2.1189**
	(0.0634)	*(0.0371)*	*(0.0370)*	*(0.0364)*	*(0.0387)*	*(0.0342)*
adjusted R^2	0.9700	0.9698	0.9692	0.9702	0.9712	0.9706

Note: see Table B.1.

Table B.11
Results from a semi-logarithmic regression, pooled across years, United States

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	**8.1849**	**6.5329**	**8.1618**	**7.6933**	**6.4962**	**7.2204**
	(0.2100)	*(0.4601)*	*(0.3316)*	*(0.1584)*	*(0.3217)*	*(0.2747)*
ENGINE	**0.0003**	**0.0002**	**0.0003**			
	(0.0000)	*(0.0000)*	*(0.0000)*			
POWER				**0.0053**	**0.0046**	**0.0062**
				(0.0000)	*(0.0002)*	*(0.0002)*
LENGTH	**0.1702**			**0.2548**		
	(0.0473)			*(0.0351)*		
WIDHT		**1.4596**			**1.4116**	
		(0.2796)			*(0.1946)*	
HEIGHT			**0.5531**			**1.1467**
			(0.2383)			*(0.1957)*
T1996	0.0520	0.0502	0.0498	0.0312	0.0320	0.0235
	(0.0393)	*(0.0384)*	*(0.0398)*	*(0.0314)*	*(0.0314)*	*(0.0323)*
T1997	0.0795	**0.0792**	0.0727	0.0420	0.0448	0.0236
	(0.0406)	*(0.0397)*	*(0.0411)*	*(0.0325)*	*(0.0325)*	*(0.0333)*
T1998	**0.1258**	**0.1159**	**0.1135**	0.0601	0.0566	0.0269
	(0.0410)	*(0.0400)*	*(0.0415)*	*(0.0328)*	*(0.0328)*	*(0.0337)*
T1999	**0.1284**	**0.1149**	**0.1128**	0.0542	0.0484	0.0131
	(0.0420)	*(0.0410)*	*(0.0427)*	*(0.0337)*	*(0.0336)*	*(0.0347)*
adjusted R^2	0.7251	0.7374	0.7181	0.8246	0.8245	0.8147

Note: see Table B.1.

Table B.12
Results from a double logarithmic regression, pooled across years, United States

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	**2.1931**	**2.5595**	**2.2628**	**3.3947**	**3.7219**	**3.6272**
	(0.3167)	*(0.3193)*	*(0.2990)*	*(0.2330)*	*(0.1594)*	*(0.1858)*
ENGINE	**0.9820**	**0.7861**	**0.9663**			
	(0.0481)	*(0.0743)*	*(0.0353)*			
POWER				**1.1534**	**1.0509**	
				(0.0387)	*(0.0539)*	
LENGTH	–0.1168			**0.3721**		**1.2037**
	(0.2568)			*(0.1787)*		*(0.0319)*
WIDHT		**1.7076**			**1.3163**	
		(0.6180)			*(0.3882)*	
HEIGHT			–0.4156			0.3160
			(0.3123)			*(0.2447)*
T1996	0.0517	0.0511	0.0525	0.0269	0.0285	0.0247
	(0.0400)	*(0.0395)*	*(0.0399)*	*(0.0310)*	*(0.0306)*	*(0.0311)*
T1997	0.0734	0.0766	0.0752	0.0344	0.0386	0.0294
	(0.0413)	*(0.0408)*	*(0.0412)*	*(0.0320)*	*(0.0317)*	*(0.0321)*
T1998	**0.1156**	**0.1146**	**0.1197**	0.0414	0.0444	0.0324
	(0.0417)	*(0.0411)*	*(0.0416)*	*(0.0324)*	*(0.0320)*	*(0.0323)*
T1999	**0.1156**	**0.1126**	**0.1214**	0.0334	0.0353	0.0227
	(0.0427)	*(0.0422)*	*(0.0427)*	*(0.0332)*	*(0.0328)*	*(0.0334)*
adjusted R^2	0.7157	0.7228	0.7173	0.8295	0.8336	0.8280

Note: see Table B.1.

Table B.13
Results from a semi-logarithmic regression, pooled across years, Germany

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	**6.7674**	**2.2779**	**8.0296**	**7.4891**	**4.6019**	**6.8368**
	(0.1891)	*(0.7815)*	*(1.2858)*	*(0.1813)*	*(0.6798)*	*(0.8905)*
ENGINE	**0.0004**	**0.0003**	**0.0008**			
	(0.0000)	*(0.0001)*	*(0.0000)*			
POWER				**0.0057**	**0.0056**	**0.0094**
				(0.0004)	*(0.0006)*	*(0.0003)*
LENGTH	**0.6590**			**0.5078**		
	(0.0555)			*(0.0496)*		
WIDHT		**4.3803**			**2.9841**	
		(0.5131)			*(0.4299)*	
HEIGHT			0.6487			**1.7696**
			(0.9251)			*(0.6354)*
T1996	–0.0129	–0.0371	–0.0248	–0.0263	–0.0439	–0.0501
	(0.0403)	*(0.0483)*	*(0.0660)*	*(0.0323)*	*(0.0386)*	*(0.0465)*
T1997	0.0376	–0.0183	0.0230	0.0106	–0.0290	–0.0269
	(0.0393)	*(0.0474)*	*(0.0646)*	*(0.0317)*	*(0.0378)*	*(0.0455)*
T1998	0.0777	–0.0057	0.0564	0.0359	–0.0226	–0.0210
	(0.0409)	*(0.0497)*	*(0.0677)*	*(0.0331)*	*(0.0394)*	*(0.0477)*
T1999	**0.0907**	0.0047	0.0750	0.0439	–0.0156	–0.0124
	(0.0403)	*(0.0491)*	*(0.0667)*	*(0.0326)*	*(0.0389)*	*(0.0470)*
adjusted R^2	0.9298	0.8993	0.8129	0.9549	0.9356	0.9073

Note: see Table B.1.

Table B.14
Results from a double logarithmic regression, pooled across years, Germany

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.3448	1.2584	**–1.5465**	**4.4536**	**4.5956**	**4.6341**
	(0.4771)	*(0.6483)*	*(0.5475)*	*(0.2258)*	*(0.1774)*	*(0.3070)*
ENGINE	**0.8694**	**0.7555**	**1.6201**			
	(0.1105)	*(0.1523)*	*(0.0746)*			
POWER				**0.7443**	**0.6810**	**1.0827**
				(0.0817)	*(0.0737)*	*(0.0381)*
LENGTH	**2.3571**			**1.6665**		
	(0.3110)			*(0.3414)*		
WIDHT		**6.3256**			**4.8275**	
		(1.0515)			*(0.7555)*	
HEIGHT			–0.6589			**2.1234**
			(1.1538)			*(0.8762)*
T1996	–0.0129	–0.0336	–0.0171	–0.0378	–0.0519	–0.0620
	(0.0442)	*(0.0480)*	*(0.0574)*	*(0.0415)*	*(0.0384)*	*(0.0454)*
T1997	0.0386	–0.0087	0.0333	0.0024	–0.0318	–0.0317
	(0.0431)	*(0.0472)*	*(0.0562)*	*(0.0407)*	*(0.0376)*	*(0.0445)*
T1998	0.0861	0.0149	0.0860	0.0229	–0.0268	–0.0281
	(0.0448)	*(0.0498)*	*(0.0589)*	*(0.0427)*	*(0.0392)*	*(0.0466)*
T1999	**0.0984**	0.0256	0.1028	0.0274	–0.0228	–0.0237
	(0.0442)	*(0.0494)*	*(0.0580)*	*(0.0421)*	*(0.0386)*	*(0.0459)*
adjusted R^2	0.9156	0.9007	0.8584	0.9262	0.9363	0.9115

Note: see Table B.1.

Table B.15
Results from a semi-logarithmic regression, pooled across years, France

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	**9.0466**	**7.4754**	**9.8687**	**8.8844**	**5.6764**	**9.9265**
	(0.1893)	*(0.8421)*	*(0.1821)*	*(0.2547)*	*(0.6525)*	*(0.2542)*
ENGINE	**0.0008**	**0.0007**	**0.0011**			
	(0.0001)	*(0.0002)*	*(0.0001)*			
POWER				**0.0066**	**0.0036**	**0.0117**
				(0.0012)	*(0.0014)*	*(0.0009)*
LENGTH	**0.2725**			**0.4595**		
	(0.0728)			*(0.0793)*		
WIDHT		**1.6907**			**3.1630**	
		(0.6328)			*(0.4478)*	
HEIGHT			–0.1419			0.2890
			(0.1611)			*(0.2053)*
T1996	0.0569	0.0454	0.0537	0.0530	0.0360	0.0287
	(0.0397)	*(0.0417)*	*(0.0440)*	*(0.0488)*	*(0.0449)*	*(0.0599)*
T1997	0.0407	0.0169	0.0347	0.0480	0.0036	0.0293
	(0.0406)	*(0.0430)*	*(0.0449)*	*(0.0498)*	*(0.0461)*	*(0.0612)*
T1998	0.0353	0.0131	0.0445	0.0096	–0.0239	–0.0023
	(0.0398)	*(0.0431)*	*(0.0440)*	*(0.0489)*	*(0.0453)*	*(0.0601)*
T1999	–0.0159	–0.0406	–0.0005	–0.0523	–0.0872	–0.0676
	(0.0397)	*(0.0436)*	*(0.0441)*	*(0.0487)*	*(0.0451)*	*(0.0600)*
adjusted R^2	0.9332	0.9263	0.9186	0.8998	0.9147	0.8487

Note: see Table B.1.

Table B.16
Results from a double logarithmic regression, pooled across years, France

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	1.0702	2.3797	–0.3270	**6.1722**	**6.9058**	**6.4854**
	(0.6636)	*(1.3175)*	*(0.5606)*	*(0.2141)*	*(0.2245)*	*(0.2376)*
ENGINE	**1.2198**	**1.0321**	**1.5918**			
	(0.1451)	*(0.2623)*	*(0.0846)*			
POWER				**0.7744**	**0.4907**	**1.0579**
				(0.1258)	*(0.1352)*	*(0.0650)*
LENGTH	**0.9614**			**1.2476**		
	(0.3253)			*(0.3910)*		
WIDHT		**2.6948**			**4.3257**	
		(1.1846)			*(0.8462)*	*(0.2595)*
HEIGHT			0.0915			**0.5464**
			(0.2439)			
T1996	0.0589	0.0487	0.0528	0.0520	0.0392	0.0335
	(0.0418)	*(0.0430)*	*(0.0449)*	*(0.0485)*	*(0.0437)*	*(0.0506)*
T1997	0.0450	0.0223	0.0381	0.0498	0.0125	0.0349
	(0.0428)	*(0.0445)*	*(0.0458)*	*(0.0494)*	*(0.0449)*	*(0.0516)*
T1998	0.0362	0.0151	0.0411	0.0128	–0.0163	0.0041
	(0.0419)	*(0.0446)*	*(0.0449)*	*(0.0486)*	*(0.0441)*	*(0.0507)*
T1999	–0.0145	–0.0379	–0.0058	–0.0521	–0.0808	–0.0639
	(0.0419)	*(0.0454)*	*(0.0450)*	*(0.0484)*	*(0.0439)*	*(0.0506)*
adjusted R^2	0.9258	0.9218	0.9152	0.9010	0.9193	0.8921

Note: see Table B.1.

Table B.17
Results from a semi-logarithmic regression, pooled across years, United Kingdom

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	**7.3018**	**6.0610**	**8.7219**	**8.0647**	**5.7855**	**7.6148**
	(0.2176)	*(0.7629)*	*(0.2513)*	*(0.2634)*	*(0.6499)*	*(0.1906)*
ENGINE	**0.0005**	**0.0004**	**0.0006**			
	(0.0000)	*(0.0000)*	*(0.0000)*			
POWER				**0.0111**	**0.0091**	**0.0103**
				(0.0008)	*(0.0008)*	*(0.0006)*
LENGTH	**0.2764**			0.0243		
	(0.0577)			*(0.0746)*		
WIDHT		**1.4436**			**1.5290**	
		(0.4871)			*(0.4191)*	
HEIGHT			–0.3440			**0.4475**
			(0.2083)			*(0.1532)*
T1996	0.0544	0.0607	0.0647	0.0103	0.0144	0.0184
	(0.0449)	*(0.0485)*	*(0.0504)*	*(0.0479)*	*(0.0442)*	*(0.0455)*
T1997	0.0716	0.0635	0.0691	0.0353	0.0352	0.0373
	(0.0464)	*(0.0501)*	*(0.0520)*	*(0.0495)*	*(0.0456)*	*(0.0468)*
T1998	0.0629	0.0421	0.0645	0.0207	0.0098	0.0153
	(0.0448)	*(0.0488)*	*(0.0504)*	*(0.0479)*	*(0.0441)*	*(0.0453)*
T1999	0.0858	0.0676	0.0941	0.0275	0.0195	0.0195
	(0.0463)	*(0.0504)*	*(0.0523)*	*(0.0496)*	*(0.0456)*	*(0.0470)*
adjusted R^2	0.8832	0.8635	0.8529	0.8676	0.8874	0.8810

Note: see Table B.1.

Table B.18
Results from a double logarithmic regression, pooled across years, United Kingdom

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.0667	0.2081	0.2430	**4.1416**	**3.7090**	**4.4994**
	(0.3458)	*(0.3824)*	*(0.4078)*	*(0.3794)*	*(0.2938)*	*(0.2612)*
ENGINE	**1.1828**	**1.1773**	**1.2135**			
	(0.0689)	*(0.0903)*	*(0.0617)*			
POWER				**1.0729**	**0.7695**	**0.9437**
				(0.1136)	*(0.0954)*	*(0.0638)*
LENGTH	0.3266			0.1806		
	(0.2572)			*(0.4441)*		
WIDHT		0.7058			**3.9445**	
		(0.7809)			*(0.8631)*	
HEIGHT			0.1941			**1.4021**
			(0.2312)			*(0.2504)*
T1996	0.0513	0.0531	0.0556	–0.0058	0.0047	0.0097
	(0.0421)	*(0.0422)*	*(0.0423)*	*(0.0633)*	*(0.0561)*	*(0.0533)*
T1997	0.0638	0.0618	0.0633	0.0152	0.0205	0.0205
	(0.0434)	*(0.0436)*	*(0.0436)*	*(0.0655)*	*(0.0578)*	*(0.0549)*
T1998	0.0512	0.0456	0.0481	0.0048	–0.0072	–0.0046
	(0.0420)	*(0.0425)*	*(0.0423)*	*(0.0634)*	*(0.0559)*	*(0.0531)*
T1999	0.0835	0.0786	0.0802	0.0155	0.0071	–0.0012
	(0.0434)	*(0.0439)*	*(0.0438)*	*(0.0656)*	*(0.0579)*	*(0.0550)*
adjusted R^2	0.8977	0.8966	0.8964	0.7695	0.8193	0.8371

Note: see Table B.1.

Table B.19
Results from a semi-logarithmic regression, pooled across years, Italy

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	**7.2806**	**6.9655**	**6.6386**	**7.1463**	**6.5129**	**9.3074**
	(0.2718)	*(0.5093)*	*(1.2297)*	*(0.2448)*	*(0.6044)*	*(1.6580)*
ENGINE	**0.0007**	**0.0009**	**0.0011**			
	(0.0001)	*(0.0001)*	*(0.0000)*			
POWER				**0.0047**	**0.0075**	**0.0099**
				(0.0008)	*(0.0008)*	*(0.0006)*
LENGTH	**0.3812**			**0.5714**		
	(0.1056)			*(0.0731)*		
WIDHT		**0.8869**			**1.6203**	
		(0.3544)			*(0.3881)*	
HEIGHT			1.0935			–0.1983
			(0.8461)			*(0.1505)*
T1996	0.0550	0.0485	0.0716	0.0278	–0.0025	0.0142
	(0.0406)	*(0.0438)*	*(0.0459)*	*(0.0396)*	*(0.0525)*	*(0.0628)*
T1997	**0.1186**	**0.1186**	**0.1556**	0.0432	–0.0048	0.0228
	(0.0436)	*(0.0474)*	*(0.0481)*	*(0.0418)*	*(0.0555)*	*(0.0658)*
T1998	**0.0986**	**0.0964**	**0.1359**	0.0276	–0.0219	0.0125
	(0.0436)	*(0.0477)*	*(0.0481)*	*(0.0417)*	*(0.0557)*	*(0.0658)*
T1999	**0.1509**	**0.1463**	**0.1845**	0.0638	0.0006	0.0236
	(0.0434)	*(0.0476)*	*(0.0480)*	*(0.0421)*	*(0.0555)*	*(0.0659)*
adjusted R^2	0.9384	0.9296	0.9220	0.9409	0.8968	0.8530

Note: see Table B.1.

Table B.20
Results from a double logarithmic regression, pooled across years, Italy

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.1222	–1.3055	**–3.0171**	**4.8263**	**5.5229**	**5.1198**
	(0.8989)	*(0.7042)*	*(0.7952)*	*(0.2567)*	*(0.2527)*	*(0.6679)*
ENGINE	**1.0608**	**1.4599**	**1.6479**			
	(0.2108)	*(0.1326)*	*(0.0749)*			
POWER				**0.4723**	**0.7650**	**0.9486**
				(0.0945)	*(0.0849)*	*(0.0549)*
LENGTH	**1.4381**			**2.1302**		
	(0.5138)			*(0.3891)*		
WIDHT		1.0133			**1.9316**	
		(0.7058)			*(0.7632)*	
HEIGHT			2.3846			1.6421
			(1.2824)			*(1.5863)*
T1996	0.0641	0.0633	0.0859	0.0366	0.0150	0.0364
	(0.0461)	*(0.0497)*	*(0.0484)*	*(0.0460)*	*(0.0565)*	*(0.0600)*
T1997	**0.1142**	**0.1220**	**0.1507**	0.0415	0.0061	0.0298
	(0.0495)	*(0.0537)*	*(0.0507)*	*(0.0484)*	*(0.0597)*	*(0.0629)*
T1998	0.0898	0.0946	**0.1242**	0.0238	–0.0128	0.0151
	(0.0493)	*(0.0539)*	*(0.0507)*	*(0.0484)*	*(0.0600)*	*(0.0628)*
T1999	**0.1403**	**0.1419**	**0.1703**	0.0530	–0.0022	0.0145
	(0.0490)	*(0.0537)*	*(0.0507)*	*(0.0490)*	*(0.0596)*	*(0.0630)*
adjusted R^2	0.9209	0.9103	0.9131	0.9205	0.8810	0.8659

Note: see Table B.1.

Table B.21
Results from a semi-logarithmic regression, pooled across years, Japan

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	**6.2933**	**6.5336**	**7.4370**	**5.4588**	**4.9985**	**6.0808**
	(0.0653)	*(0.1158)*	*(0.1125)*	*(0.0564)*	*(0.1062)*	*(0.1354)*
ENGINE	**0.0006**	**0.0006**	**0.0006**			
	(0.0000)	*(0.0000)*	*(0.0000)*			
POWER				**0.0053**	**0.0057**	**0.0070**
				(0.0001)	*(0.0001)*	*(0.0001)*
LENGTH	0.0367			**0.3080**		
	(0.0189)			*(0.0152)*		
WIDHT		–0.0802			**1.0398**	
		(0.0786)			*(0.0698)*	
HEIGHT			**–0.7306**			**0.3412**
			(0.0798)			*(0.0951)*
T1996	0.0094	0.0097	0.0145	0.0021	–0.0009	–0.0027
	(0.0141)	*(0.0141)*	*(0.0137)*	*(0.0145)*	*(0.0152)*	*(0.0161)*
T1997	0.0195	0.0200	0.0203	–0.0122	–0.0166	–0.0204
	(0.0143)	*(0.0143)*	*(0.0140)*	*(0.0147)*	*(0.0155)*	*(0.0164)*
T1998	0.0141	0.0169	0.0218	**–0.0330**	**–0.0476**	**–0.0396**
	(0.0156)	*(0.0157)*	*(0.0153)*	*(0.0161)*	*(0.0169)*	*(0.0179)*
T1999	**0.0436**	**0.0453**	**0.0528**	–0.0158	**–0.0340**	**–0.0363**
	(0.0158)	*(0.0159)*	*(0.0155)*	*(0.0163)*	*(0.0171)*	*(0.0182)*
adjusted R^2	0.8545	0.8543	0.8613	0.8463	0.8306	0.8094

Note: see Table B.1.

Table B.22
Results from a double logarithmic regression, pooled across years, Japan

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	**–0.3964**	**–2.2896**	**0.2324**	**2.7757**	**2.9898**	**2.7258**
	(0.1036)	*(0.1383)*	*(0.1127)*	*(0.0718)*	*(0.0593)*	*(0.0817)*
ENGINE	**1.3361**	**1.6256**	**1.0014**			
	(0.0332)	*(0.0307)*	*(0.0130)*			
POWER				**0.8506**	**0.9360**	**0.9527**
				(0.0213)	*(0.0212)*	*(0.0120)*
LENGTH	**–1.3678**			**0.4449**		
	(0.1273)			*(0.0881)*		
WIDHT		**–4.3967**			0.0402	
		(0.2024)			*(0.1500)*	
HEIGHT			**–0.4036**			**0.6217**
			(0.1369)			*(0.1358)*
T1996	0.0041	0.0082	0.0063	–0.0001	–0.0006	–0.0037
	(0.0165)	*(0.0151)*	*(0.0171)*	*(0.0166)*	*(0.0167)*	*(0.0166)*
T1997	0.0151	0.0205	0.0146	–0.0118	–0.0139	–0.0150
	(0.0168)	*(0.0153)*	*(0.0174)*	*(0.0169)*	*(0.0170)*	*(0.0169)*
T1998	–0.0031	**0.0354**	–0.0058	–0.0333	–0.0339	–0.0384
	(0.0184)	*(0.0168)*	*(0.0190)*	*(0.0184)*	*(0.0186)*	*(0.0185)*
T1999	**0.0444**	**0.0844**	**0.0428**	–0.0076	–0.0110	–0.0168
	(0.0186)	*(0.0171)*	*(0.0192)*	*(0.0187)*	*(0.0189)*	*(0.0188)*
adjusted R^2	0.7993	0.8333	0.7863	0.7973	0.7941	0.7967

Note: see Table B.1.

Table B.23
Trade and transport margins in the transport equipment industry, six countries, 1995–1999

	1995	1996	1997	1998	1999
United States	0.14	0.14	0.13	0.13	0.14
Germany	0.05	0.05	0.05	0.05	0.05
France	0.05	0.05	0.05	0.05	0.05
United Kingdom	0.07	0.07	0.07	0.07	0.07
Italy	0.09	0.09	0.09	0.09	0.09
Japan	0.05	0.05	0.05	0.05	0.05

Note: Not all years in the above table are given in the OECD I/O data base. For years with missing values, the values of the nearest years were imputed.

Source: OECD (2002), Preliminary Input/Output Tables.

Table B.24
Bilateral PPPs for the car industries of six countries, unadjusted and quality-adjusted using hedonic methods, unweighted, 1995

	Unadjusted	Dummy method	Hedonic Quality adjustment			
			Pooled regression	Country regressions		
				Base country	Own country	Average[1]
DM/$	1.80	2.88	2.86	3.14	2.77	2.95
FF/$	4.72	9.64	9.50	11.7	8.96	0.22
£/$	0.59	1.06	1.04	1.29	1.00	1.13
IL/$	1,277	2,254	2,174	2,414	2,076	2,239
¥/$	124	150	158	156	153	155
FF/DM	2.62	3.35	3.32	3.57	3.55	3.56
£/DM	0.33	0.37	0.36	0.36	0.38	0.37
IL/DM	709	783	759	792	785	788
¥/DM	68.8	52.2	55.1	55.3	54.9	55.1
£/FF	0.13	0.11	0.11	0.11	0.11	0.11
IL/FF	271	234	229	223	218	221
¥/FF	26.2	15.6	16.6	16.8	14.7	15.7
IL/£	2,158	2,125	2,084	2,093	2,011	2,052
¥/£	209	142	151	153	118	134
¥/IL	0.10	0.07	0.07	0.07	0.07	0.07

[1] Geometric average of base country and own country.

Table B.25
Bilateral PPPs for the car industries of six countries, unadjusted and quality-adjusted using hedonic methods, unweighted, 1996

	Unadjusted	Dummy method	Hedonic Quality adjustment		
			Pooled regression	Country regressions		
				Base country	Own country	Average[1]
DM/$	1.72	2.69	2.62	2.90	2.57	2.73
FF/$	5.16	9.86	9.63	11.77	9.27	10.4
£/$	0.63	1.05	1.03	1.27	1.00	1.13
IL/$	1,263	2,199	2,166	2,442	2,091	2,260
¥/$	120	146	153	152	148	150
FF/DM	3.00	3.67	3.67	3.92	3.90	3.91
£/DM	0.37	0.39	0.39	0.39	0.40	0.40
IL/DM	734	818	826	868	862	865
¥/DM	69.9	54.4	58.4	57.8	57.6	57.7
£/FF	0.12	0.11	0.11	0.10	0.10	0.10
IL/FF	245	223	225	222	219	221
¥/FF	23.3	14.8	15.9	15.8	14.1	14.9
IL/£	2,000	2,096	2,097	2,143	2,067	2,105
¥/£	191	140	148	147	116	131
¥/IL	0.10	0.07	0.07	0.07	0.07	0.07

[1] Geometric average of base country and own country.

Table B.26
Bilateral PPPs for the car industries of six countries, unadjusted and quality-adjusted using hedonic methods, unweighted, 1997

	Unadjusted	Dummy method	Hedonic Quality adjustment		
			Pooled regression	Country regressions		
				Base country	Own country	Average[1]
DM/$	1.93	2.75	2.78	3.02	2.73	2.87
FF/$	5.36	9.68	9.45	11.4	9.19	10.2
£/$	0.62	1.05	1.04	1.29	1.01	1.14
IL/$	1,331	2,191	2,120	2,363	2,070	2,212
¥/$	122	139	145	146	143	145
FF/DM	2.78	3.52	3.40	3.66	3.62	3.64
£/DM	0.32	0.38	0.37	0.38	0.39	0.39
IL/DM	691	796	764	804	799	802
¥/DM	63.4	50.4	52.3	52.7	52.6	52.6
£/FF	0.12	0.11	0.11	0.11	0.11	0.11
IL/FF	248	226	224	223	218	221
¥/FF	22.8	14.3	15.4	15.5	13.8	14.6
IL/£	2,133	2,086	2,048	2,052	1,977	2,014
¥/£	196	132	140	141	108	123
¥/IL	0.09	0.06	0.07	0.07	0.07	0.07

[1] Geometric average of base country and own country.

Table B.27
Bilateral PPPs for the car industries of six countries, unadjusted and quality-adjusted using hedonic methods, unweighted, 1998

PPP	Unadjusted	Dummy method	Hedonic Quality adjustment			
			Pooled regression	Country regressions		
				Base country	Own country	Average[1]
DM/$	1.89	2.81	2.85	3.06	2.76	2.90
FF/$	4.90	9.46	9.12	10.7	8.62	9.60
£/$	0.58	1.05	1.04	1.30	0.98	1.13
IL/$	1,224	2,181	2,087	2,246	1,980	2,109
¥/$	115	137	143	140	137	139
FF/DM	2.59	3.36	3.20	3.38	3.34	3.36
£/DM	0.31	0.37	0.36	0.37	0.38	0.37
IL/DM	647	775	732	757	756	757
¥/DM	60.8	48.6	50.0	49.9	49.5	49.7
£/FF	0.12	0.11	0.11	0.11	0.11	0.11
IL/FF	250	231	229	229	224	226
¥/FF	23.5	14.5	15.6	15.8	14.1	14.9
IL/£	2,093	2,080	2,015	2,014	1,963	1,988
¥/£	197	130	138	139	106	122
¥/IL	0.09	0.06	0.07	0.07	0.07	0.07

[1] Geometric average of base country and own country.

Table B.28
Bilateral PPPs for the car industries of six countries, unadjusted and quality-adjusted using hedonic methods, unweighted, 1999

PPP	Unadjusted	Dummy method	Hedonic Quality adjustment			
			Pooled regression	Country regressions		
				Base country	Own country	Average[1]
DM/$	1.89	2.83	2.86	3.11	2.80	2.95
FF/$	4.38	8.65	8.37	10.2	8.09	9.10
£/$	0.61	1.05	1.06	1.37	1.03	1.18
IL/$	1,285	2,223	2,122	2,347	2,062	2,200
¥/$	117	138	145	145	142	144
FF/DM	2.31	3.06	2.93	3.17	3.13	3.15
£/DM	0.32	0.37	0.37	0.38	0.39	0.38
IL/DM	678	787	743	780	775	777
¥/DM	61.5	48.8	50.8	51.0	51.0	51.0
£/FF	0.14	0.12	0.13	0.12	0.12	0.12
IL/FF	294	257	254	251	243	247
¥/FF	26.6	16.0	17.3	17.4	15.4	16.4
IL/£	2,094	2,116	2,008	2,019	1,921	1,969
¥/£	190	131	137	138	105	120
¥/IL	0.09	0.06	0.07	0.07	0.07	0.07

[1] Geometric average of base country and own country.

Table B.29
Bilateral PPPs for the car industries of six countries, unadjusted and quality-adjusted using hedonic methods, weighted by quantity produced, 1997

	Unadjusted	Dummy method	Hedonic Quality adjustment			
			Pooled regression	Country regressions		
				Base country	Own country	Average [1]
DM/$	1.96	2.75	2.85	3.15	2.80	2.97
IL/$	1,075	2.19	2.04	2.43	1.97	2.19
IL/DM	548	796	716	784	786	785

[1] Geometric average of base country and own country.

Table B.30
Bilateral PPPs for the car industries of six countries, unadjusted and quality-adjusted using hedonic methods, weighted by quantity produced, 1998

	Unadjusted	Dummy method	Hedonic Quality adjustment			
			Pooled regression	Country regressions		
				Base country	Own country	Average [1]
DM/$	1.88	2.81	2.82	3.06	2.71	2.88
FF/$	4.95	9.46	9.26	10.9	8.64	9.71
£/$	0.62	1.05	1.01	1.15	0.95	1.05
IL/$	1,115	2.18	2,175	2,466	2,008	2,226
FF/DM	2.64	3.36	3.28	3.48	3.45	3.46
£/DM	0.33	0.37	0.36	0.35	0.37	0.36
IL/DM	593	775	770	824	823	824
£/FF	0.12	0.11	0.24	0.10	0.11	0.10
IL/FF	225	231	235	239	238	238
IL/£	1,809	2,080	2,158	2,243	2,215	2,229

[1] Geometric average of base country and own country.

Table B.31
Bilateral PPPs for the car industries of six countries, unadjusted and quality-adjusted using hedonic methods, weighted by quantity produced, 1999

	Unadjusted	Dummy method	Hedonic Quality adjustment			
			Pooled regression	Country regressions		
				Base country	Own country	Average [1]
DM/$	1.94	2.83	2.85	3.14	2.79	2.96
FF/$	4.68	8.65	8.64	10.6	8.32	9.38
£/$	0.61	1.05	1.02	1.22	0.98	1.10
IL/$	1,161	2.22	2,134	2,522	2,047	2,272
FF/DM	2.42	3.06	3.03	3.27	3.25	3.26
£/DM	0.31	0.37	0.36	0.36	0.37	0.37
IL/DM	600	787	748	822	815	819
£/FF	0.13	0.12	0.25	0.11	0.11	0.11
IL/FF	248	257	247	254	249	251
IL/£	1,906	2,116	2,101	2,208	2,065	2,135

[1] Geometric average of base country and own country.

References

Abraham, K.G. (1997), "The CPI Commission: Discussion", *American Economic Review*, May, Vol. 87, pp. 94–95.

Aizcorbe, A., C. Corrado and M. Doms (2000), *Constructing price and quantity indexes for high technology goods*, Federal Reserve Board, Washington.

Almus, M. and D. Moch (2002), *Hedonic Price Indices for PCs. Results from Germany*, Presented at the conference on Price Indices and the Measurement of Quality Changes, ZEW, Mannheim, 25–26 April.

Almus, M., T. Eckert, and D. Moch (2002), *Hedonic Methods in Official Price Statistics*, Presented at the conference on Price Indices and the Measurement of Quality Changes, ZEW, Mannheim, 25–26 April.

Ark, B. van (1990), "Manufacturing Productivity Levels in France and the United Kingdom", *National Institute Economic Review*, August, pp. 62–77.

Ark, B. van (1993), *International Comparisons of Output and Productivity*, Monograph Series No. 1, Groningen Growth and Development Centre, Groningen.

Ark, B. van (1996), "Issues in Measurement and International Comparison of Productivity – An Overview", in OECD, *Industry Productivity. International Comparison and Measurement Issues*, OECD Proceedings, Paris.

Ark, B. van, and D. Pilat (1993), "Productivity Levels in Germany, Japan, and the United States: Differences and Causes", *Brookings Papers on Economic Activity: Microeconomics 2*, Washington D.C., December.

Ark, B. van and M.P. Timmer (2001), *PPPs and International Productivity Comparisons: Bottlenecks and New Directions*, Presented at Joint Worldbank-OECD Seminar on Purchasing Power Parities. Recent Advances in Methods and Applications, Worldbank, Washington, D.C., 30 January–2 February.

Ark, B. van, R. Inklaar and R. McGuckin (2002), ""Changing Gear". Productivity, ICT and Services Industries: Europe and the United States", *Research Memorandum GD-60*, Groningen Growth and Development Centre, Groningen.

Automotive News (1999), *1999 Market Data Book*, Detroit (also issues for 1995–1998).

Ball, A. and S. Mehmi (2002), *Implementing Hedonic Methods for PCs: The UK Experience*, Presented at the Symposium on the use of the hedonic method by statistical agencies, 25 October, Statistics Netherlands, Voorburg.

Banzhaf, H.S. (2001), "Quantifying the Qualitative: Quality-Adjusted Price Indexes in the United States, 1915–61" in J.L. Klein and M.S. Morgan (eds.), *The age of economic measurement*, History of Political Economy vol. 33, Duke University Press, Durham.

Barzyk, F. and M. MacDonald (2002), *The treatment of quality change for computer price indexes – A review of current and proposed practices*. Presented at the conference on Price Indices and the Measurement of Quality Changes, ZEW, Mannheim, 25–26 April.

Berndt, E.R. (1991), *The practice of econometrics: classic and contemporary*, Addison-Wesley, Reading, MA.

Berndt, E.R., Z. Griliches and N.J. Rappaport (1995), "Econometric Estimates of Price Indexes for Personal Computers in the 1990's", *Journal of Econometrics*, vol. 68, pp. 243–268.

Berndt, E.R. and N.J. Rappaport (2001), *Price and Quality of Desktop and Mobile Personal Comoputers: A Quarter Century Historical Overview*, Paper presented at the Symposium on Hedonic Methods In Price Statistics, Statistisches Bundesamt, Wiesbaden, 16 June.

Bode, B. and J. van Dalen (2001), "Kwaliteit-gecorrigeerde prijsontwikkelingen van nieuwe auto's, 1990–1999", *Research report*, Erasmus University Rotterdam.

Boon, M. (1998), "Sampling Designs in Compiling Consumer Price Indices: Current Practices at EU Statistical Institutes", *Research in Official Statistics*, vol. 1, pp. 39–52.

Boskin, M.J., E.R. Dulberger, R.J. Gordon, Z. Griliches and D.W. Jorgenson (1996), *Toward a more accurate measure of the cost of living*, Advisory Commission to study the Consumer Price Index. Interim report to the Senate Finance Committee, December 4.

Boskin, M.J., E.R. Dulberger, R.J. Gordon, Z. Griliches and D.W. Jorgenson (1998), *Consumer Prices, the Consumer Price Index and the Cost of Living*, Journal of Economic Perspectives, vol. 12, no. 1, pp. 3–26.

Bourot, L. (1997), "Indice des micro-ordinateurs et des imprimantes: bilan d'une rénovation", *Internal working paper no. 571/E312*, INSEE, Paris.

Box, G.E.P. and D.R. Cox (1964), "An Analysis of Transformations", Journal of the Royal Statistical Society, Series B, vol. 26, no. 2, pp. 21–243.

Brachinger, H.W., B. Schips and W. Stier (1999), *Expertise on the Relevance of the "Boskin-Report" for the Swiss Consumer Price Index*, University of Fribourg, mimeo.

Bureau of Labor Statistics (1997), *Measurement Issues in the Consumer Price Index*, Response to the U.S. Congress, Joint Economic Committee, June.

Cartwright, D.W. (1986), "Improved Deflation of Purchases of Computers", *Survey of Current Business*, March, pp. 7–9.

Cockburn, I.M. and A.H. Anis (1998), "Hedonic Analysis of Arthritis Drugs", *NBER Working Paper 6574*, NBER, Cambridge, MA.

Cole, R., Y.C. Chen, J.A. Barquin-Stollerman, E.R. Dulberger, N. Helvecian and J.H. Hodge (1986), "Quality Adjusted Price Indices for Computer Processors and Selected Peripheral Equipment", *Survey of Current Business*, January, pp. 41–50.

Court, A.T. (1939), Hedonic Price Indexes with Automotive Examples, *The Dynamics of Automobile Demand*, The General Motors Corporation , New York, pp. 99–117.

Dalén, J. (2002), *Personal computers in different HICPs*. Paper presented at the Brookings Workshop on Economic Measurement: Hedonic Indexes: Too Fast? Too Slow? Or Just Right?, Brookings Institution, Washington, D.C., 1 February.

Davidson, R. and J.G. MacKinnon (1981), "Several Tests for Model Specification in the Presence of Alternative Hypotheses", *Econometrica*, vol. 49, no. 3, pp. 781–793.

Denison, E.F. (1989), *Estimates of productivity change by industry: an evaluation and an alternative*, Brookings Institution, Washington, D.C.

Deutsche Bundesbank (2000), *Monthly Report August 2000*, Frankfurt am Main.

Deutsche Bundesbank (2001), *Monthly Report May 2001*, Frankfurt am Main.

Diewert, W.E. (1976), "Exact and superlative index numbers", *Journal of Econometrics*, vol. 4, pp. 115–145.

Diewert, W.E. (1987), "Index Numbers", in J. Eatwell, M. Milgate and P. Newman (eds.), *The New Palgrave: A Dictionary of Economics, Volume 2*, Macmillan, London, pp. 767–780.

Diewert, W.E. (1992), "Fisher Ideal Output, Input and Productivity Indices Revisited", *Journal of Productivity Analysis*, vol. 3, pp. 211–248.

Diewert, W.E. (1993), "The Early History of Price Index Research", in W.E. Diewert and A.O. Nakamura (eds.), *Essays in Index Number Theory, Volume 1*, North-Holland, Amsterdam, pp. 33–65.

Diewert, W.E. (1995), "Axiomatic and Economic Approaches to Elementary Price Indexes", *Discussion Paper 95–01*, University of British Columbia, Vancouver.

Diewert, W.E. (1999), "Axiomatic and Economic Approaches to International Comparisons", in A. Heston and R.E. Lipsey (eds.), *International and Interarea Comparisons of Income, Output and Prices*, Studies in Income and Wealth vol. 61, NBER and University of Chicago Press, Chicago, pp. 13–107.

Diewert, W.E. (2001), "Hedonic Regressions: A Consumer Theory Approach", *Discussion Paper 01–12*, University of British Columbia, Vancouver.

Diewert, W.E. (2002), "Hedonic Regressions: A Review of Some Unresolved Issues", *mimeo*, University of British Columbia, Vancouver.

DIW (Deutsches Institut für Wirtschaftsforschung) (1991), *Produktionsvolumen und -potential. Produktionsfaktoren des Bergbaus und des Verarbeitendes Gewerbe*, Berlin.

Dornbusch, R., S. Fisher and P. Samuelson (1977), "Comparative Advantage, Trade and Payments in a Ricardian Model with a Continuum of Goods", *American Economic Review*, vol. 67, pp. 823–839.

Dulberger, E.R. (1989), "The Application of a Hedonic Model to a Quality-Adjusted Price Index For Computer Processors", in D.W. Jorgenson and R. Landau (eds.), *Technology and Capital Formation*, MIT Press, Cambridge.

Elfering, S. (2001a), *Procesbeschrijving van het consumentprijsindexcijfer en het producentenprijsindexcijfer*, mimeo, Statistics Netherlands, Voorburg.

Elfering, S. (2001b), *Het prijsindexcijfer van de bouwkosten van nieuwe woningen*, mimeo, Statistics Netherlands, Voorburg.

Eurostat (1995), *Short Term Indicators in Construction: Methodological Handbook for the Construction Sector, version 4*, Luxembourg.

Eurostat (1999a), *Volume Measures for Computers and Software*, Report of the Task Force, Luxembourg.

Eurostat (1999b), *Volume Measures for Construction*, Report of the Task Force, Luxembourg.

Eurostat (2001), *Compendium of HICP reference documents*, Luxembourg.

Evans, R. (2002), *INSEE's Adoption of Market Intelligence Data for its Hedonic Computer Manufacturing Price Index*, Presented at the Symposium on the use of the hedonic method by statistical agencies, 25 October, Statistics Netherlands, Voorburg.

Evans, R. and P. Scherrer (2002), *Hedonic methods for computers in industrial price statistics: The French experience*, Presented at the conference on Price Indices and the Measurement of Quality Changes, ZEW, Mannheim, 25–26 April 2002.

Feenstra, R.C. (1995), "Exact Hedonic Price Indices", *Review of Economics and Statistics*, vol. 77, pp. 634–54.

Fisher, I. (1922), *The Making of Index Numbers*, Houghton Mifflin, Boston.

Folkertsma, C. (1998), "De vertekening van de prijsindex", *ESB*, 6 March 1998, pp. 189–191.

Freudenberg, M. and D. Ünal-Kesenci (1994), "French and German Productivity Levels in Manufacturing: A Comparison Based on the Industry-of-Origin Method", *Document de Travail 94-10*, CEPII, Paris.

Gaulier, G. and S. Haller (2000), "The Convergence of Automobile Prices in the European Union: An Empirical Analysis for the Period 1993–1999", *Working Paper no. 14*, CEPII, Paris.

Gersbach, H. and B. van Ark (1994), "Micro Foundations for International Productivity Comparisons", *Research Memorandum GD-11*, Groningen Growth and Development Centre, Groningen.

Gerschenkron, A. (1962), *Economic Backwardness in Historical Perspective*, Harvard University Press, Cambridge, Mass.

Gilbert, M. (1961), "Quality Changes and Index Numbers", in *Economic Development and Cultural Change*, Vol. 9, no. 3, pp. 287–294.

Gilbert, M. and I.B. Kravis (1954), *An International Comparison of National Products and the Purchasing Power of Currencies*, OECD, Paris.

Gilbert, M. and Associates (1958), *Comparative National Products and Price Levels*, OECD, Paris.

Goldberg, P.K. and F. Verboven (1998), "The Evolution of Price Dispersion in the European Car Market", *NBER Working Paper 6818*, NBER, Cambridge, MA.

Goldberger, A.S. (1968), "The Interpretation and Estimation of Cobb-Douglas Functions.", *Econometrica,* vol. 35, pp. 464–472.

Gordon, R.J. (1990), *The Measurement of Durable Goods Prices*, University of Chicago Press, Chicago.

Grient, H.A. van der, and M.H. Oei (2002), "CPI van nieuwe auto's; een hedonisch alternatief", *mimeo*, Statistics Netherlands, Voorburg.

Griliches, Z. (1961), "Hedonic Price Indices for Automobiles: An Econometric Analysis of Quality Change", in *The Price Statistics of the Federal Government*, General Series, no. 73, pp. 137–196, NBER, New York.

Griliches, Z. (1964), "Notes on the Measurement of Price and Quality Changes", in *Models of Income Determination*, pp. 381–418, NBER, New York.

Griliches, Z. (1971), "Introduction: Hedonic Price Indices Revisited" in Z. Griliches (ed.), *Price Indices and Quality Change: Studies in New Methods of Measurement*, Cambridge, Harvard University Press.

Griliches, Z. (1990), Hedonic Price Indexes and the Measurement of Capital and Productivity: Some Historical Reflections", in E.R. Berndt and J.E. Triplett (eds), *Fifty Years of Economic Measurement*, University of Chicago Press, Chicago, pp. 185–202.

Haan, J. de (2002), *Time Dummy Approaches to Hedonic Price Measurement*, research paper, Statistics Netherlands, Voorburg.

Haan, J. de and L. Hoven (2001), *CBS-Onderzoeksprogramma ter verbetering van de Consumentenprijsindex*, Statistics Netherlands, Voorburg.

Haan, J. de and E. Opperdoes (2002), *Hedonic and matched model Fisher price indices for cars*, Working paper, Statistics Netherlands, Voorburg.

Harhoff, D. (2002), *Quality-Adjusted Price Indices for Cars in Germany*, Presented at the conference on Price Indices and the Measurement of Quality Changes, ZEW, Mannheim, 25–26 April 2002.

Heravi, S., A. Heston and M. Silver (2003), "Using Scanner Data to Estimate Country Price Parities: A Hedonic Regression Approach", *Review of Income and Wealth*, vol. 49, no. 1, pp. 1–21.

Hicks, J. R. (1940), "The Valuation of the Social Income" *Economica* N.S. vol. 7, no. 2, pp. 105–124.

Hill, P. (1984), "Real Gross Product in OECD Countries and Associated Purchasing Power Parities", *Working Paper No. 17*, OECD, Paris.

Hoffmann, J. (1998) *Problems of Inflation Measurement in Germany*, Discussion Paper 1/98, Deutsche Bundesbank, Frankfurt.

Hofsten, E. von (1952) *Price Indexes and Quality Changes*. George Allen & Unwin Ltd., London.

Hooper, P. and E. Vrankovich (1995), "International Comparisons of the Levels of Unit Labor Costs in Manufacturing", *International Finance Discussion Papers*, No. 527, Board of Governors of the Federal Reserve system, Washington, D.C.

Ioannidis, C. and M. Silver (1999), "Estimating Hedonic Indices: An Application to UK Television Sets", *Journal of Economics, Zeitschrift für Nationalökonomie*, vol. 69, no. 1, pp. 71–94.

Japanese Automobile Manufacturers Association (1999), *Motor Vehicle Guidebook*, Tokyo (also issues for 1995–1998).

Konijn, P., D. Moch and J. Dalén (2002), *Searching for the European Hedonic Function*, Presented at the Symposium on the use of the hedonic method by statistical agencies, 25 October, Statistics Netherlands, Voorburg.

Kravis, I.B. (1986), "The Three Phases of the International Comparison Project", in *The World Bank Research Observer*, vol. 1, pp. 3–26.

Kravis, I.B., Z. Kenessy, A. Heston and R. Summers (1975), *A System of International Comparisons of Gross Product and Purchasing Power*, Johns Hopkins University Press, Baltimore.

Kravis, I.B., A. Heston and R. Summers (1978), *International Comparisons of Real Product and Purchasing Power*, Johns Hopkins University Press, Baltimore.

Kravis, I.B., A. Heston and R. Summers (1982), *World Product and Income; International Comparisons of Real Gross Product*, Johns Hopkins University Press, Baltimore.

Lancaster, K. (1971), *Consumer Demand: A New Approach*, Columbia University Press, New York

Landefeld, J.S. and B.T. Grimm (2000), "A Note on the Impact of Hedonics and Computers on Real GDP", *Survey of Current Business*, December, pp. 17–22.

Lebow, D.E. and J.B. Rudd (2001), *Measurement Error in the Consumer Price Index: Where Do We Stand?*, Federal Reserve Board, Washington, D.C.

Leeuwen, G. van (1994), *A hedonic approach to output price indices for construction*, research paper, Statistics Netherlands, Voorburg.

Lequiller, F.(2001), "The new economy and the measurement of GDP growth"; *INSEE Working Paper no G 2001/01,* Paris.

Lichtenberg, F.R. and Z. Griliches (1989), "Errors of Measurement in Output Deflators", *Journal of Business & Economic Statistics*, vol. 7 no. 1, pp. 1–9.

Lim, P.P. and R. McKenzie (2001), "Hedonic Price Analysis for Personal Computers in Australia: An Alternative Approach to Quality Adjustments in the Australian Prices Indexes", *mimeo*, Australian bureau of statistics, Canberra.

Linz, S. and G. Eckert (2002), Zur Einführung hedonischer Methoden in die Preisstatistik, *Wirtschaft und Statistik*, October, pp. 857–863.

Lippe, P. von der (2001), *Chain Indices. A Study in Price Index Theory*, Statistisches Bundesamt, Wiesbaden.

Maddison, A., and B. van Ark (1988), *Comparison of Real Output in Manufacturing*, World Bank Working Papers, WPS 5, Washington D.C., June.

Mann, C. (2000), "The End of Moore's Law?", *Technology Review*, vol. 103, no. 3, pp. 42–48.

McKenzie, R. (2002), *Hedonic Price Indexes for Computers in Australia*, Presented at the conference on Price Indices and the Measurement of Quality Changes, ZEW, Mannheim, 25–26 April 2002.

McKinsey Global Institute (1993), *Manufacturing Productivity*, Washington D.C., October.

Minne, B. and H. Noordman (1999), "Kwaliteitsstijging Nederlandse export in 1988–1996?", *Research Memorandum 151*, Bureau of Economic and Policy analysis, The Hague.

Moch, D. and J.E. Triplett (2002), *International Comparisons of Hedonic Price Indexes for Computers: A Preliminary Examination*, Presented at the 27th General Conference of the International Association for Research in Income and Wealth, Djurhamn, 18–24 August.

Moulton, B.R. (2001), *The Expanding Role of Hedonic Methods in the Official Statistics of the United States*, Presented at the OECD Meeting of National Accounts Expert, 9–12 October, Paris.

Moulton, B.R. and K.E. Moses (1997), "Adressing the Quality Change Issue in the Consumer Price Index", in *Brookings Paper on Economic Activity, Macroeconomics 1*, pp. 305–366.

Mulligen, P.H. van (1999), *Measuring Quality of Production: International Comparisons of Manufacturing*, University of Groningen, mimeo.

OECD (1992), *Purchasing Power Parities and Real Expenditures, EKS Results, Volume 1*, Paris.

OECD (1995), *Purchasing Power Parities and Real Expenditures, EKS Results, Volume 1*, Paris.

OECD (1997), *Review of the OECD-Eurostat PPP Program*, Paris.

OECD (2002), *Preliminary Input/Output Database*, Paris.

Ohta, M. and Z. Griliches (1976), "Automobile Prices Revisited: Extensions of the Hedonic Hypothesis", in M. Terleckyj (ed.), *Household Production and Consumption. Studies in Income and Wealth vol. 40*, Columbia University Press, New York.

Oulton, N. (1998), "The Implications of the Boskin Report", in *National Institute Economic Review*, National Institute of Economic and Social Research, London, pp. 89–98.

Paige, D. and G. Bombach (1959), *A Comparison of National Output and Productivity*, OEEC, Paris.

Pakes, A. (2002), *A reconsideration of hedonic price indices with an application to PCs*. Paper presented at the Workshop on Economic Measurement: Hedonic Indexes: Too Fast? Too Slow? Or Just Right?, Brookings Institution, Washington, D.C., 1 February.

Pieper, P.E. (1990), "The Measurement of Construction Prices: Retrospect and Prospect", in E.R. Brendt and J.E. Triplett (eds.), *Fifty Years of Economic Measurement*, University of Chicago Press, Chicago, pp. 239–271.

Pilat, D. (1994), *The Economics of Rapid Growth. The Experience of Japan and Korea*, Edward Elgar, Aldershot U.K. and Brookfield U.S.

PRC Bouwcentrum (1995), *Een vergelijking van Internationale Bouwkostenvergelijkingen*, Bodegraven (summary in English available).

Raff, D.M.G. and M. Trajtenberg (1997), "Quality-Adjusted Prices for the American Automobile Industry: 1906–1940", in T.F. Bresnahan and R.J. Gordon (eds.), *The Economics of New Goods*, Studies in Income and Wealth, vol. 58, University of Chicago Press, Chicago, pp. 71–107.

Rosen, S. (1974), "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition.", *Journal of Political Economy*, vol. 82, no. 1, pp. 34–55.

Scherrer, P. (2001), *Hedonic Methods in Price Statistics: The French Experience*, Presented at the Symposium on Hedonic Methods In Price Statistics, Statistisches Bundesamt, Wiesbaden, 16 June.

Schreyer, P. (2002), "Computer price indices and international growth and productivity comparisons", *Review of Income and Wealth*, vol. 48(1), pp. 15–32.

Schultze, C.L. and C. Mackie, eds. (2001), *At what price? Conceptualizing and measuring cost-of-living and price indexes*, Panel on conceptual, measurement and other statistical issues in developing cost-of-living indexes, National Academy Press, Washington D.C.

Silver, M. and S. Heravi (2001), "Scanner data and the measurement of inflation", *The Economic Journal*, June, pp. 384–405.

Silver, M. and Heravi, S. (2002a), "The Measurement of Quality-Adjusted Price Changes", in M. Shapiro and R. Feenstra (eds.), *Scanner Data and Price Indexes*, National Bureau of Economic Research, Studies in Income and Wealth, vol. 61, University of Chicago Press, Chicago.

Silver, M. and Heravi, S. (2002b), "Why the CPI Matched Models Method May Fail Us: Results from an Hedonic and Matched Experiment Using Scanner Data", *European Central Bank Working Paper No. 144*, Frankfurt.

Silver, M. (2003), "The Use of Weights in Hedonic Regressions: the Measurement of Quality-Adjusted Price Changes", *mimeo*, Cardiff University.

Sinclair, J. and B. Catron (1990), "An experimental price index for the computer industry", *Monthly Labor Review*, October, pp. 16–24.

Statistisches Bundesamt (1991), *Volkswirtschaftliche Gesamtrechnungen, Revidierte Ergebnisse, 1950 bis 1990 (Fachserie 18, Reihe S. 15)*, Wiesbaden.

Statistisches Bundesamt (1995), *Preise (Fachserie 17, Reihe 4)*, Wiesbaden.

Statistisches Bundesamt (1997), *Volkswirtschaftliche Gesamtrechnungen, Hauptbericht 1996, Konten und Standardtabellen (Fachserie 18, Reihe 1.3)*, Wiesbaden.

Statistische Bundesamt (2000), *Produzierendes Gewerbe (Fachserie 4, Reihe 3.1)*, Wiesbaden (also previous issues).

Stigler, G., D.S. Brady, E.F. Denison, I.B. Kravis, P.J. MacCarthy, A. Rees, R. Ruggles and B.N. Swerling (1961), *The Price Statistics of the Federal Government*, NBER, New York.

Timmer, M.P. (2000), *The Dynamics of Asian Manufacturing. A Comparative Perpective in the Late Twentieth Century*, Edward Elgar, Cheltenham.

Triplett, J.E. (1969), "Automobiles and Hedonic Quality Measurement", *Journal of Political Economy*, vol. 77, pp. 408–417.

Triplett, J.E. (1983), "Concepts of Quality in Input and Output Price Measures: A Resolution of the User Value-Resource Cost Debate", in M.F. Foss (ed.), *The U.S. National Income and Product Accounts: Selected Topics*, NBER, University of Chicago Press, Chicago.

Triplett, J.E. (1986), "The Economic Interpretation of Hedonic Methods", in *Survey of Current Business*, January, pp. 36–40.

Triplett, J.E. (1987), "Hedonic functions and hedonic indices", in J. Eatwell, M. Milgate and P. Newman (eds.), *The New Palgrave: A Dictionary of Economics, Volume 2*, Macmillan, London, pp. 630–634.

Triplett, J.E. (1990), "Hedonic Methods in Statistical Agency Environments: An Intellectual Biopsy", in E.R. Berndt and J.E. Triplett (eds.), *Fifty Years of Economic Measurement*, University of Chicago Press, Chicago, pp. 207–237.

Triplett, J.E. (1996), "High Tech Industry Productivity and Hedonic Price Indices", in OECD, *Industry Productivity. International Comparison and Measurement Issues*, OECD Proceedings, Paris.

Triplett, J.E. (2000), *Internationally Comparable Price Indices for Computers and High Tech Products: A Research Proposal*, mimeo, Brookings Institution, Washington, D.C.

Triplett, J.E. (2002), *Summary of the Brookings Workshop on Economic Measurement: Hedonic Prices Indices: Too Fast? Too Slow? Or Just Right?*, Brookings Institution, Washington, D.C., 1 February.

Triplett, J.E. (forthcoming), *Handbook on quality adjustment of price indices for information and communication technology products*, draft, OECD, Paris.

Triplett, J.E. and R.J. McDonald (1977), "Assessing the Quality Error in Output Measures: The Case of Refrigerators", *The Review of Income and Wealth*, vol. 23, no. 2, pp. 137–156.

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

United Nations (1994), *World Comparisons of Real Gross Domestic Product and Purchasing Power, 1985*, New York.

United Nations (1998), *Evaluation of the International Comparison Programme*, New York.

U.S. Bureau of Economic Analysis (1986), *Supplement to Survey of Current Business*, Washington D.C.

U.S. Bureau of Economic Analysis (1993), *Survey of Current Business*, May, Washington, D.C.

U.S. Bureau of Economic Analysis (1997), *Survey of Current Business*, November, Washington, D.C.

U.S. Bureau of Labor Statistics (1988), *Employment and Earnings*, various monthly issues, Washington, D.C.

U.S. Bureau of Labor Statistics (1993), *Employment and Earnings*, various monthly issues, Washington, D.C.

U.S. Bureau of Labor Statistics (2000), "Price index research in the coming decades", *Monthly Labor Review*, September, pp. 31–36.

U.S. Department of Commerce (1990), *U.S. Census of Manufactures 1987*, Washington, D.C.

U.S. Department of Commerce (1995), *U.S. Census of Manufactures 1992*, Washington, D.C.

Verboven, F. (1996), "International Price Discrimination in the European Car Market", *Rand Journal of Economics*, vol. 27, no. 2, pp. 240–268.

Vermande, H.M. and P.H. van Mulligen (1998), "Construction Costs in the Netherlands in an International Context", *Construction Management & Economics*, vol. 17, pp. 269–283.

Ward, M. (1985), *Purchasing Power Parities and Real Expenditures in the OECD*, OECD, Paris.

Waugh, F.V. (1928), "Quality Factors Influencing Vegetable Prices", *Journal of Farm Economics*, vol. 10 no. 2, pp. 185–196.

Wyckoff, A.W. (1995), "The impact of computer prices on international comparisons of labour productivity", *Economics of Innovation and New Technology*, vol. 3, pp. 277–293.

Zieschang, K.D., P.A. Armknecht and D. Smith (2001), *Integrating inter-area and international price comparisons with consumer price index compilation*, Presented at Joint Worldbank-OECD Seminar on Purchasing Power Parities. Recent Advances in Methods and Applications, Worldbank, Washington D.C., 30 January–2 February.

Nederlandse samenvatting (Summary in Dutch)

Het hoofdonderwerp van deze dissertatie is het kwaliteitsprobleem in prijsvergelijkingen door de tijd en tussen landen. Bij het bestuderen van dit onderwerp is er in dit proefschrift een poging gedaan om de volgende vragen te beantwoorden: ten eerste, wat is het gevolg van kwaliteitsverschillen op intertemporele en internationale prijsvergelijkingen? Ten tweede, wat zijn de theoretische en praktische voor- en nadelen van de hedonische methode ten opzichte van de conventionele methoden om met dit 'kwaliteitsprobleem' om te gaan?

Hieronder vat ik kort de belangrijkste conclusies van dit proefschrift samen. Ik zal uiteenzetten hoe kwaliteitsproblemen de toekomst van het maken van prijsvergelijkingen en het gebruik van de hedonische methode zullen beïnvloeden. Tevens zal ik de implicaties bespreken voor de praktijk van het maken van tijdreeksen van prijsindices and conversiefactoren voor munteenheden.

Het kwaliteitsprobleem: 'vertekening binnen de steekproef' en 'vertekening buiten de steekproef'

Te beginnen met de eerste vraag, het onderwerp van Hoofdstuk 2, is het duidelijk dat kwaliteitsverschillen optreden voor veel goederen en diensten. In een dynamische economie verschijnen voortdurend nieuwe producten, en bijna evenveel verdwijnen. Evenzo worden bestaande producten voortdurend verbeterd, en vaak aangeboden tegen lagere nominale prijzen, gegeven de verbetering van de kwaliteitskarakteristieken. Voor boekhouders van economische activiteiten introduceren deze veranderingen een probleem hoe prijsveranderingen te meten, die rekening houden met kwaliteitsverbetering. In het geval van internationale vergelijkingen is de kwestie van heterogeniteit nog sterker. Verschillen tussen vergelijkbare producten zijn niet alleen groter tussen landen, ook de overlap in productiesamenstellingen en uitgavenpatronen is kleiner dan in het geval van een intertemporele prijsindex voor een enkel land.

Het kwaliteitsprobleem werkt op twee manieren, namelijk als een 'vertekening binnen de steekproef' en een 'vertekening buiten de steekproef'. De vertekening buiten de steekproef heeft betrekking op de introductie en het verdwijnen van goederen in en uit de steekproef. Het negeren van nieuwe en verdwijnende producten leidt ertoe dat de steekproef van de prijsindex niet representatief wordt, en veroorzaakt zo een vertekening in de prijsmeting. Als dezelfde producten vergeleken moeten worden over de tijd of tussen landen, is het niet mogelijk nieuwe producten op te nemen op het moment van hun introductie. Nieuwe goederen hebben geen equivalent in de vorige periode, dus de prijsverandering van zo'n goed kan in de eerste fase van zijn bestaan niet

gemeten worden. Bovendien vernieuwen de meeste statistische bureaus hun prijssteekproeven slechts eens per vijf of zelfs tien jaar, zodat relatief nieuwe goederen pas opgepikt worden als ze al enkele jaren bestaan.[1] De vaak dramatische prijsveranderingen die nieuwe goederen ondergaan in de eerste jaren van hun bestaan woorden zo meestal niet gemeten. Hoewel het aandeel van nieuwe goederen in de totale bestedingen doorgaans niet groot is, wordt desalniettemin een vertekening geïntroduceerd. Een vergelijkbaar probleem treedt op als variëteiten van bestaande producten ('nieuwe modellen') worden geïntroduceerd. Aangezien met de standaardmethode van prijsindexberekeningen alleen prijzen van bijna identieke producten worden vergeleken, worden nieuwe modellen vaak op vergelijkbare wijze genegeerd.
Net als nieuwe producten veroorzaken ook producten die van de markt verdwijnen een probleem als de prijssteekproef te lang wordt vastgehouden. Als een product niet meer verkrijgbaar is, kan de prijs ervan niet langer waargenomen worden en ontstaat er een gat in de index. Statistische bureaus proberen soms deze gaten te vullen door een schatting te maken van de prijs van het product als het nog verkrijgbaar was geweest, soms op basis van het raadplegen van de bedrijven die de producten produceerden.
De beste oplossing om de vertekening buiten de steekproef te verkleinen is het regelmatig bijwerken van de steekproef voor de prijsindex, en daarbij prijzen en gewichten van nieuwe producten op te nemen zodra ze verschijnen, en verdwenen producten zo snel mogelijk te verwijderen.
De tweede soort van vertekening wordt op analoge wijze de 'vertekening binnen de steekproef' genoemd. Deze vertekening treedt op als de prijzen van niet-identieke producten worden vergeleken. Dit is met name relevant als een nieuwe variëteit een oude vervangt. Op zoek naar een vervangingsproduct koppelt de statisticus vaak de prijzen van twee variëteiten van hetzelfde product, en maakt daarbij een impliciete of (ad hoc) expliciete aanname over de specificatieverschillen van de producten. In Hoofdstuk 2 zijn verschillende aanpassingsmethodes besproken, die alle leiden tot vertekende schattingen van prijsveranderingen. De richting van de vertekening hangt af van de omvang en de richting van zowel de kwaliteitsverandering als de prijsverandering.
Gecombineerd hebben deze twee problemen een significant effect op het meten van prijsveranderingen. Het 'Advisory Report to the Senate Finance Committee' (Boskin *et al.*, 1996) schatte dat de vertekening door nieuwe goederen en kwaliteitsveranderingen resulteerden in een jaarlijkse opwaartse vertekening in de CPI van de V.S. van 0,6 procentpunt. Soortgelijke studies in andere landen lieten vergelijkbare resultaten zien. Het is moeilijk om de geloofwaardigheid van deze 'netto' schattingen te beoordelen, omdat de exacte omvang van deze vertekening bepaald wordt door 'bruto' vertekeningen in duizenden individuele subindices en hun gewichten in de uiteindelijke index. Zulke vertekeningen kunnen elkaar in belangrijke mate neutraliseren, en deze neutraliserende effecten kunnen veranderen over de tijd en tussen landen. Ongeacht de vertekening in de uiteindelijke index kan een ongeschikte indexmethodologie een groot effect hebben bij individuele goederen en diensten. Computers vormen een veel onderzocht en relevant voorbeeld, maar het pro-

bleem van verkeerde indexcijfers is niet voorbehouden aan dit product alleen. Niet alle prijzen van goederen en diensten vertonen een opwaartse vertekening, en sommige prijzen (bv. in de niet-commerciële dienstverlening) kunnen in feite een significante neerwaartse vertekening vertonen als kwaliteitsverschillen nauwkeurig onderzocht worden (Triplett, te verschijnen).

Wat over prijsindices is gezegd is even relevant in prijsvergelijkingen tussen landen, die worden uitgevoerd om conversiefactoren voor munteenheden af te leiden, zoals koopkrachtpariteiten en eenheidswaardeverhoudingen. Zowel in de bestedingsbenadering als de industrie-van-oorsprong benadering worden prijzen van producten gekoppeld onder de aanname dat ze identiek zijn in verschillende landen.

Hier komen we ook de twee vertekeningen tegen die hierboven genoemd werden. Met name de vertekening buiten de steekproef is groot in internationale prijsvergelijkingen, vooral als vergelijkingen gedaan worden vanuit een producentenperspectief. Sommige producten (zoals vliegtuigen of auto's) worden simpelweg niet in ieder land geproduceerd. Vooral specialisatie op basis van comparatieve voordelen versterkt de vertekening buiten de steekproef. [2] Tevens kan een toenemende differentiatie in de productwaardeketen, in veel gevallen binnen verticaal geïntegreerde multinationale bedrijven, een significant effect hebben op adequate prijsvergelijkingen tussen landen. Bovendien kunnen databronnen uit verschillende landen vaak niet gekoppeld worden, of is informatie niet beschikbaar vanwege geheimhoudingsplicht. Gecombineerd leidt dit tot een groot percentage van de productie of bestedingen dat niet gekoppeld wordt in internationale vergelijkingen. Omdat er geen natuurlijke wijze is om landen met elkaar te verbinden zoals met verschillende tijdsperiodes, kan men niet aannemen dat kwaliteitsverschillen tussen landen even gradueel zijn als over tijd. Daarom kan de vertekening buiten de steekproef niet eenvoudig opgelost worden, zelfs niet met grote steekproeven van goederen en diensten. [3]

Vanwege het gebrek aan graduele verschillen in de kwaliteit van goederen en diensten het ontbreken van een duidelijke inschaling van landen is de vertekening binnen de steekproef ook ernstiger in internationale vergelijkingen dan in prijsindices. In het geval van de bestedingsbenadering, die gebruik maakt van specificaties, kan het nog mogelijk zijn artikelen te koppelen die redelijk vergelijkbaar zijn tussen landen, maar zulke gekoppelde artikelen hoeven niet het meest representatief te zijn voor de bestedingspatronen van beide landen. De industrie-van-oorsprong benadering, die gebruik maakt van eenheidswaarden, is gebaseerd op 'prijzen' van aggregaten van goederen. Niet alleen kwaliteitsverschillen tussen individuele zijn in dit geval belangrijk, maar ook de samenstelling van de aggregaten waarvoor eenheidswaarden worden berekend. Het belang van dit probleem van de productie-samenstelling werd aangestipt in Hoofdstuk 5. Voor internationale vergelijkingen van productie op bedrijfstakniveau voldoet de bestedingsbenadering niet, omdat prijsvergelijkingen voor bestedingen alleen finale goederen en diensten omvatten. Om de productie van intermediaire producten af te dekken, die een substantieel onderdeel van de totale productie vormen, blijven industrie-van-oorsprong vergelijkingen nodig, en de noodzaak om de vertekening binnen de steekproef te ondervangen blijft daarom hoge prioriteit houden.

De hedonische methode en het gebruik ervan

Ik zal nu aandacht besteden aan de tweede vraag van dit proefschrift. Omdat de 'matched model' methode op de meeste terreinen nog steeds de standaardkeuze voor prijsmeting is, zal ik hier de relatieve verdiensten en nadelen van de hedonische methode behandelen.

De hedonische methode, die gebaseerd is op de vooronderstelling dat gebruikers geïnteresseerd zijn in de karakteristieken van goederen en diensten in plaats van in de goederen en diensten zelf, past regressietechnieken toe om een expliciete kwaliteitsaanpassing op prijsindices of prijsratio's te maken. Omdat daadwerkelijk consumentengedrag duidelijk niet alleen betrekking heeft op finale producten maar ook of eerder op hun karakteristieken, verschaft deze vooronderstelling een sterk argument voor de hedonische methode. Met name wanneer men geïnteresseerd is in indices voor de kosten van levensonderhoud, die het nut constant houden, lijkt de hedonische methode onvermijdelijk.

Desalniettemin zijn er ook bezwaren geuit tegen de hedonische methode. De toename in het gebruik van de hedonische methode in de V.S. gaf aanleiding tot een opleving van een oud debat dat bekend is geworden als het 'productiekosten vs. gebruikerswaarde'-debat. Hoewel het in feite een enigszins andere kwestie is, heeft dit debat grote gevolgen voor het gebruik van de hedonische methode. De kern van het debat is het conflict tussen hen die denken dat alleen met variabelen die een verandering in productiekosten weergeven rekening gehouden moet worden, en degenen die denken dat het uiteindelijk de waarde die gebruikers aan een bepaald product hechten (of een zekere karakteristiek) is die telt. Triplett (1983), geeft aan dat beide gezichtpunten hun waarde hebben, afhankelijk van de benadering van de prijsmeting. Maar vanuit een economisch perspectief moeten productiekosten en gebruikerswaarde convergeren. Ten eerste zullen producenten die geen rekening houden met gebruikerswaarde in hun productieproces marktaandeel verliezen aan concurrenten die producten met een hogere gebruikerswaarde leveren. Ten tweede, als in een vrijemarkteconomie een producent dezelfde gebruikerswaarde levert tegen hogere productiekosten dan die van een concurrent, zal hij uiteindelijk verlies lijden. Daarom wordt in dit proefschrift de mening gehanteerd dat prijsmaatstaven zowel gebruikerswaarde als productiekosten moeten weergeven, en dat karakteristieken die gebruikt worden voor het schatten van hedonische modellen ook beide moeten weergeven.

Behalve theoretische overwegingen moet ook aan praktische punten aandacht besteed worden, gegeven het feit dat het maken van prijsindices een sterk empirische zaak is. De grootste kracht van de hedonische methode is duidelijk dat het de statisticus of onderzoeker toestaat een index of ratio te berekenen als het aantal gekoppelde artikelen laag is, of zelfs nul. De hedonische methode is daardoor veel geschikter om de vertekeningen binnen en buiten de steekproef op te lossen dan de 'matched model' methode.

In Hoofdstuk 3 werd al aangegeven dat de hedonische methode ook enkele praktische nadelen heeft. Het nadeel dat het meest wordt aangevoerd is het gebrek aan gegevens over individuele artikelen en hun karakteristieken. Zoals besproken in dit proefschrift, is dit argument een misvatting. Het dataprobleem van de hedonische methode is niet groter dan voor de 'matched model' methode. Producten kunnen alleen gekoppeld worden als de onderzoeker voldoende zeker is dat ze (nagenoeg) identiek zijn, wat dezelfde informatie over productkarakteristieken vereist. In feite zou de toepassing van de hedonische methode door statistische bureaus helemaal niet veel duurder of data-intensiever hoeven zijn, aangezien veel informatie over eigenschappen toch al verzameld wordt om prijzen te kunnen koppelen. Met de beschikbaarheid van gedetailleerde gegevens over prijzen en karakteristieken, zoals bijvoorbeeld scanner data, is dergelijke informatie in een toenemende mate verkrijgbaar. Samenvattend kan gesteld worden dat het 'data argument' niet gebruikt kan worden om de hedonische methode te verwerpen ten gunste van de conventionele 'matched model' methode.

Relevante praktische problemen waar aandacht aan besteed dient te worden bij toepassing van de hedonische methode zijn in detail behandeld in Hoofdstuk 3. De uiteindelijke conclusie van dat hoofdstuk is dat een juiste specificatie van een hedonisch model de kern is van het maken van kwaliteitsgecorrigeerde prijsindices en conversiefactoren van munteenheden. Dit kan in de praktijk lastig zijn, omdat de keuze welke karakteristieken op te nemen niet vanzelfsprekend is. Bovendien kunnen veel 'gewenste' karakteristieken niet waarneembaar of kwantificeerbaar blijken te zijn. De problemen bij het opstellen van een adequate hedonische functie zullen waarschijnlijk het grootste obstakel vormen bij de praktische toepassing van de hedonische methode.

Deze dissertatie bevat twee originele empirische toepassingen van de hedonische methode. In Hoofdstuk 4 wordt deze methode toegepast op advertentie- en scannerdata van computers in Nederland. De prijsindices daalden enigszins sneller dan 'matched model' indices, wat de hypothese bevestigt dat een prijsindex van een product met snel veranderende kwaliteit, zoals computers, een opwaartse vertekening heeft als het niet aangepast wordt voor kwaliteitsveranderingen. Hoewel het verschil tussen de hedonische en 'matched model' indices niet klein was, wordt het effect van een expliciete kwaliteitsaanpassing door middel van de hedonische methode al substantieel gereduceerd wanneer er gebruik gemaakt wordt van een kettingindex in plaats van een index met een vaste basisperiode. In het geval van een kettingindex wordt een groot gedeelte van het kwaliteitsprobleem opgelost door het frequenter trekken van de steekproef. Maandelijks ketenen hoeft niet noodzakelijk te zijn voor ieder product, maar het jaarlijks trekken van een nieuwe steekproef is een waarschijnlijk minimum. In dit opzicht valt er bij het bestrijden van het kwaliteitsprobleem meer te winnen van het frequent trekken van steekproeven dan van het gebruik van een hedonische kwaliteitsaanpassing. Een andere belangrijke kwestie is het gebruik van gewichten voor individuele artikelen in de index. Deze conclusies werden bevestigd bij het toepassen van de (hoge frequentie) 'matched model' en de hedonische methode op de database die gebruikt wordt voor het maken van de Nederlandse CPI voor computers.

Hoewel het regelmatig trekken van een steekproef en het gebruik van bestedingsgewichten prijsindices in grote mate zou verbeteren, kan dit in de praktijk niet altijd gerealiseerd worden. Statistische bureaus hebben beperkte budgetten voor de verzameling van prijsgegevens en hebben extra moeite met

het inwinnen van bestedingsaandelen. De hedonische methode kan tot op zekere hoogte een oplossing zijn als de prijssteekproef inadequaat is, vooral wanneer het aantal gekoppelde artikelen laag is en de steekproeven niet regelmatig getrokken worden, maar het is geen volledige en bevredigende oplossing.
In internationale prijsvergelijkingen van producentenprijzen van auto's, het onderwerp van Hoofdstuk 5, blijkt er geen overlap te bestaan in artikelen die gekoppeld kunnen worden. In verschillende landen worden verschillende auto's geproduceerd, dus het toepassen van de 'matched model' methode op basis van individuele producten is niet mogelijk. Het gebruik van de hedonische methode laat zien dat gemiddelde prijsratio's die niet zijn aangepast voor kwaliteitsverschillen een forse onderschatting veroorzaken van de reële productiewaarde van de auto-industrieën van landen waar gemiddeld grotere en krachtigere auto's worden geproduceerd ten opzichte van landen waar kleinere auto's worden geproduceerd. De hedonische functie voor auto's heeft echter een groot probleem: het aantal beschikbare karakteristieken is beperkt vergeleken met de enorme hoeveelheid prijsbepalende eigenschappen van dergelijke complexe producten. In feite zijn de beschikbare specificaties voor auto's (zoals omvang en gewicht) alle slechts benaderingen van de karakteristieken die er werkelijk toe doen.
Hoewel een juiste specificatie van hedonische functies voor complexe producten als auto's erg moeilijk en misschien onmogelijk is, moeten de resultaten niet op voorhand verworpen worden en aan de onaangepaste 'matched model' resultaten vastgehouden worden. Het blijft belangrijk zorgvuldig aandacht te besteden aan het specificeren van een hedonische functie, maar de beperkte beschikbaarheid van data moet geaccepteerd worden. Bij het kiezen van de best mogelijke specificatie moet besloten worden of deze goed genoeg is, en de resultaten beoordelen door ze te vergelijken met 'matched model' resultaten. Soms kan het koppelen van producten onmogelijk blijken en in zulke gevallen moet men vertrouwen op prijsverhoudingen van verwante producten in andere industrieën of de hedonische methode gebruiken, ondanks de tekortkomingen.
Omdat zowel de vertekening binnen als buiten de steekproef veel groter is in internationale prijsvergelijkingen dan in intertemporele prijsindices is het pleidooi voor hedonische kwaliteitsaanpassingen daar in feite erg sterk. Echter, de specificatie van een gemeenschappelijk hedonisch model, zoals mogelijk voor verschillende periodes (op voorwaarde dat ze niet te ver uit elkaar liggen) is veel moeilijker toe te passen over verschillende landen. De relatie tussen prijzen en karakteristieken wordt bepaald door voorkeuren en productieprocessen, die veel minder vergelijkbaar zijn tussen verschillende landen dan tussen relatief nabije periodes in een enkel land. Slechts in zeer bijzondere gevallen met internationaal gestandaardiseerde producten kan het schatten van een gemeenschappelijke internationale hedonische functie mogelijk blijken. Wederom vormen computers een mogelijk goed voorbeeld, zoals geïllustreerd in Moch en Triplett (2002). Voor de gegevens over auto's die gebruikt werden in Hoofdstuk 5, is de relatie tussen prijzen en karakteristieken te verschillend in verschillende landen voor een gemeenschappelijke hedonische functie.

De ongeschiktheid van een gemeenschappelijk hedonisch model in internationale prijsvergelijkingen sluit ook het gebruik van de hedonische dummy methode uit, die vereist dat gegevens van verschillende landen samengevoegd worden.

De toekomst van de hedonische methode

Hoewel de hedonische methode nog steeds erg populair is bij wetenschappers, is er recentelijk kritiek op geuit door een Amerikaans panel van experts op prijsmeting (Schultze en Mackie, 2002). De Schultze Commissie stelt dat veel econometrische kwesties rondom de hedonische methode nog steeds niet afdoende zijn opgelost. Omdat de hedonische methode volgens de Commissie nog niet volledig is ontwikkeld, stelt zij dat de klasse van goederen waarop het gebruik van de hedonische methode gerechtvaardigd is, beperkt is. Veel van de kritiek en de terughoudendheid ten aanzien van de hedonische methode die in dit rapport wordt geuit is echter gericht op een variant van de hedonische methode die niet door veel wetenschappers wordt bepleit, namelijk de dummy methode.
Het werk van Schultze en Mackie negeert veel recent onderzoek naar hedonische indices, zoals blijkt uit een recente kritiek van Diewert (2002). Veel statistische bureaus hebben het gebruik van de hedonische methode uitgebreid ten opzichte van andere methodes. De meningen verkondigd door Schultze en Mackie over de hedonische methode worden wellicht gedeeld door sommige statistici en zelfs economen, maar zij vormen zeker niet de gangbare mening wat betreft deze methode. Integendeel, het debat over de hedonische methode lijkt te verschuiven ten gunste van uitgesproken proponenten van de methode, zoals Diewert (2002), Pakes (2002), Silver (2003) en Triplett (te verschijnen).
Nu dit proefschrift de hedonische methode naar voren schuift als geschikt instrument voor kwaliteitsaanpassing in prijsmeting, wat is dan de beste route voorwaarts, zowel in de statistische praktijk als in de wetenschappelijke toepassing van hedonische prijsindices? Zoals hierboven vermeld zijn theoretische overwegingen belangrijk. De keuze van karakteristieken, en hoe om te gaan met benaderingen van niet-waarneembare karakteristieken zijn belangrijke kwesties bij de juiste specificatie van een hedonisch model.
Op voorwaarde dat zulke kwesties opgelost kunnen worden is de vraag wanneer de hedonische methode te gebruiken voor kwaliteitsaanpassing een belangrijk punt. Hier wordt onderscheid gemaakt tussen prijsindices over de tijd en internationale vergelijkingen. Zoals hierboven opgemerkt valt bij prijsindices de meeste winst te boeken bij het gebruik van een kettingindexprincipe met regelmatige steekproeftrekkingen. Het effect van de hedonische methode in vergelijking met het kettingprincipe is bescheiden. Wat betreft prijsindices hoeft de hedonische methode daarom geen hoge prioriteit te hebben.
In internationale prijsvergelijkingen is de noodzaak van de hedonische methode groter. Er is geen internationaal equivalent van het vaker trekken van steekproeven voor de 'matched model' methode als alternatief. Naast het gebruik van prijsverhoudingen van verwante producten, wat veel gedaan wordt

in huidige studies van ICOP, is de hedonische methode de enige optie, met name daar waar de overlap in productie en bestedingen van dezelfde artikelen klein is. Het verzamelen van de noodzakelijke gegevens voor internationale prijsvergelijkingen is geen eenvoudige opgave, en het gebruik van de hedonische methode maakt dit alleen maar moeilijker. In de meeste gevallen vereist het specifieke databases per industrie, bijvoorbeeld afkomstig van multinationals die in meer dan een land produceren, of marktonderzoeksbureaus of bedrijfstakorganisaties met vestigingen in meerdere landen. Zonder informatie over kwaliteit wordt het simpelweg koppelen van prijzen van producten op basis van een gemeenschappelijk gebruik op zijn best een geïnformeerde gok. Dit is vooral het geval bij complexe producten met veel kwaliteitskarakteristieken.

Tenslotte moet meer aandacht geschonken worden aan prijsmeting van diensten, gegeven het toegenomen belang van de dienstensectoren in de ontwikkelde economieën. Hoewel dit proefschrift zich geconcentreerd heeft op het kwaliteitsprobleem in twee soorten (duurzame) goederen, is al het hierboven genoemde even relevant voor diensten. Met de enorme meetproblemen, en de moeilijkheid om prijzen, hoeveelheden en kwaliteits-karakteristieken te onderscheiden die in Hoofdstuk 1 uiteen gezet zijn, is het gebruik van de hedonische methode wellicht nog relevanter in het geval van diensten. Diensten zijn vaak complex in de zin dat ze een breed scala aan relevante karakteristieken bezitten die ze van andere producten onderscheiden. Dit is een probleem voor het meten van prijzen in het algemeen, hedonisch of traditioneel. Als deze problemen opgelost kunnen worden en er voldoende informatie beschikbaar is voor het koppelen van prijzen en producten, is er geen reden waarom de hedonische methode niet toegepast zou kunnen worden. Zoals hierboven uiteengezet is, is de het enige echte nadeel van de hedonische methode in vergelijking met de 'matched model' methode de moeilijkheid in het specificeren van een adequate hedonische functie. Net als met fysieke goederen zijn dataproblemen vergelijkbaar voor beide methoden.

De meetproblemen op het gebied van diensten zijn moeilijk te overschatten. Deze zijn nog steeds enorm, en sluiten een snelle toepassing van de hedonische methode in diensten uit. Maar voor fysieke goederen, waar het productconcept veel helderder is, ligt de grootste toegevoegde waarde van de hedonische methode vooral in internationale prijsvergelijkingen.

Noten

1) Het frequenter trekken van een nieuwe steekproef wordt momenteel door veel statistische bureaus overwogen en is bijvoorbeeld ingevoerd in het Nederlandse consumentenprijsindexcijfer (CPI) in 2003.

2) Zie bijvoorbeeld Dornbusch *et al*. (1977).

3) Het verbinden van landen met behulp van 'optimal spanning trees' kan tot op zekere hoogte een oplossing van dit probleem betekenen. Zie Hill (1999).