

SOME OBJECTIONS TO HEDONIC INDEXES

Jack E. Triplett
Brookings Institution
Washington, D.C., USA

Note: This document is chapter VII in the author's forthcoming OECD publication:

Triplett, Jack E. Forthcoming. Handbook on hedonic indexes and quality adjustments in price indexes. Directorate for Science, Technology and Industry. Paris: Organization for Economic Co-operation and Development.

It is being posted as a background document (only) for:

SSHRC Conference on Index Number Theory and the Measurement of Prices and Productivity

Vancouver, B.C.

June 30 - July 3, 2004

CHAPTER VII

SOME OBJECTIONS TO HEDONIC INDEXES

Objections to hedonic indexes began with the publication of Griliches' (1961) first article, and they have not abated. Over this 40-plus year period, the major criticism has taken two forms.

One persistent criticism concerns the results: Hedonic indexes are thought to "fall too fast," by some criterion. Denison (1987) was one of the first to voice this criticism. A reader of the draft of this Handbook forwarded a comment from a 2002 Eurostat seminar, which went, roughly: "I would not have paid €15,000 for today's PC ten years ago so I cannot accept that the price has fallen by 90%." Variations on this "falls too fast" theme have appeared in many places.

A second and nearly as persistent criticism concerns the technical properties of hedonic indexes, presumably independently of their results. Ho, Rao, and Tang (2003, footnote 16, page 11) remark: "It [the hedonic technique] has been criticised for lack of theoretical foundation, especially for its functional forms, lack of transparency, and its subjectiveness in selecting the quantities of characteristics." A Committee on National Statistics panel (Schultze and Mackie, eds., 2002, chapter 4) expressed criticisms of hedonic indexes, and urged the U.S. Bureau of Labor Statistics to slow its implementation of hedonic indexes in the U.S. CPI.

This chapter reviews criticisms under the two headings suggested by the preceding paragraphs. Some criticisms of hedonic indexes have been considered in previous chapters of this Handbook. Nevertheless, it seems useful to draw responses to criticisms together in one place, for easier reference, even though some duplication results. Other criticisms that have not yet been addressed are also considered here. The chapter may be incomplete, in ignoring some criticism; if so, it is because I have not been made aware of it, or through omission.

In one of the commonest criticism of hedonic indexes, a speaker says that he has "reservations" about hedonic indexes—or that *others* have expressed "reservations." When pressed to be more specific, the speaker responds that he has reservations, but not (in effect) objections, that he cannot give specifics, he has rather "unease" (a frequently used word) about the method, or that it is others who express "unease" without providing specifics. This form of criticism has often been passed from one person to another, with the result that the origin of the criticism, or what the originator of the criticism had in mind, is impossible to determine. I have encountered this kind of criticism of hedonic indexes many times, and it is a frustration: One cannot respond to a criticism that has no specific content that can be analysed empirically or logically or methodologically. It seems to me that those who make such nebulous criticisms ought to put them into an explicit form that permits analytic discussion, or ought to be challenged to do so. Criticisms are useful in that they may sharpen our thinking, or responses to their salient points may increase understanding, but criticism that takes the form of nonexplicit "reservations" or statements of "unease" contribute neither.

Hulten (2002) noted that, among statisticians and among users, there is a tendency to view old methodology as good methodology because it is known; knowledge of the implications of new methodology is not so extensive, so it is mistrusted (perhaps properly). One suspects that this conservative tendency with respect to methodology explains some of the mistrust of hedonic indexes that is so evidently widespread.

One the other hand, however, it is doubtful that users really understand the matched-model methodology, its implications, and its potential biases, and it is likely that the agencies themselves have not always fully understood the potential biases (these matters are reviewed in Chapter II). Accordingly, even if conservatism with respect to changes in statistical methods has a certain validity, it is doubtful that the “known” properties of the old matched-model methods for treating quality change are correctly valued (because the old methods are in fact not well understood). There are abundant grounds for supposing that the old linking methodologies, correctly understood, should also cause “unease” (see Chapter II). It is the fact of quality change that creates the problem, not new methods for dealing with it.

A. THE CRITICISM THAT HEDONIC INDEXES FALL TOO FAST

Several strands to the “falls too fast” contention exist. Some of these criticisms are general statements, or are meant to be, and some are specific to ICT investment and computer equipment price indexes.

1. General Statements

A long-standing presumption has hedonic indexes *always* falling faster, or rising more slowly, than price indexes constructed with conventional matched-model methods. If this presumption were correct empirically, someone who leans toward conservatism with respect to statistical methods might infer that hedonic indexes were suspect, that they “fall too fast,” where someone who was critical of the matched-model method might take the same putative fact as an indictment of the conventional method. Indeed, a good amount of the past debate about hedonic indexes can be understood in precisely this way—different judgements about the likely relative errors of alternative statistical methods, positions that have been argued under the presumption that hedonic indexes generally decline more, or rise less, than matched-model indexes that are constructed from the same database.

The empirical evidence indicates the presumption is wrong, as a general proposition. Hedonic indexes do not always decline more rapidly than matched-model indexes. For example, the survey in Chapter IV of this Handbook shows that hedonic indexes fall faster for computers, but not for TV’s—refer to tables 4.5 and 4.8. This same conclusion is borne out in other work. Schultze (2002), using information from BLS, shows that after hedonic quality adjustments were added to a number of U.S. CPI series, changed methods did not necessarily result in faster-falling price indexes.

The general criticism, at least in this form, seems poorly informed or based on judgements that are out of date. No general or regular pattern across products emerges from comparisons of matched-model and hedonic indexes that have been computed from the same database. For this reason, the view that hedonic indexes generally “fall too fast” cannot be based on empirical evidence of the actual differences, in practice, between the two methods, when applied across the range of products for which hedonic indexes have been estimated.

2. Computer Price Indexes Fall Too Fast

It may not be a general pattern that applies to all products, but it is certainly true empirically that hedonic indexes for computer equipment fall more rapidly than matched-model indexes. The survey in Chapter IV presents some empirical evidence. In that chapter I concluded that even where matched-model and hedonic indexes were fairly close together, the two different statistical methods produced statistically significant differences.

It is also true that hedonic computer equipment prices fall very rapidly. Moreover, computer price indexes are the subjects where one most often hears the “falls too fast” objection.

a. The Price Indexes

Matched model indexes for computers also fall rapidly if done with matched-model, FR&R (frequently resample and reweight) methods, just not as fast as hedonic indexes that are estimated from the same database. For example, all four matched-model and hedonic computer price indexes estimated by Okamoto and Sato (see table 4.5) decline more than 40 percent per year. The rapid decline in computer prices is not solely an artefact of hedonic methods. At the same time, where we can determine the causes of differences between hedonic and matched model indexes (see the review in Chapter IV), the analysis makes us confident that where they differ, the hedonic is better. Hedonic indexes pick up price changes that matched-model indexes miss, particularly price changes that accompany entries and exits of computers (and other products) when those entry and exit price effects are not instantaneously reflected into price changes for continuing computer models (see the summary in Chapter IV).

If the “falls too fast” contention is made with full knowledge of the research results, therefore, it must rest on some other evidence or on subtle arguments about plausibility. In evaluating these, it is worth bearing in mind that computer prices have been declining at prodigious rates (15 to 30 percent per year, or more) for more than 50 years, though generally computer peripheral equipment and communications equipment prices have fallen somewhat less rapidly (see the review of semiconductor and communications equipment price trends in chapter 10 of Triplett and Bosworth, 2004). Those huge price declines for computer equipment reflect a technological marvel that has no precedent, at least over such an extensive time span. The fact of the technological change is not in serious dispute, but it is clearly worth asking whether that technological change has been measured accurately. Set against the “falls too fast” presumption is Nordhaus (2001), who contended that computing power price decline has been understated in hedonic indexes (his historical series on computing technology extends over 100 years, as he considers pre-computer technology as well).

b. Plausibility

Some criticisms rest on plausibility, which is always an appealing test for an economic measurement. A reader of the draft of this Handbook reported the following conversation:

“Your index says that the price of a PC is 30 percent of its price three years ago, while the nominal price has not changed, it is €1000. I, as a consumer, cannot find a PC worth €300, even if I accept that it has the same functions as three years ago. In other words, I am obliged to buy a computer of €1000, while I don’t want to use the new functionalities which go with it and of which the statisticians tell me that they have lowered the price by 70 percent.”

A similar statement opened this chapter: “I would not have paid €15,000 for today’s PC ten years ago so I cannot accept that the price has fallen by 90%.” This last statement has element of veracity. Extend it further: We now have a 50-year history of performance-corrected computer prices. From this, we can calculate that a computer with the performance of today’s \$2000 PC cost (10,000 x \$2000) in 1953. There were virtually no sales of computers to individuals in 1953, so far as I know. None of us who own computers today would have bought one for \$20 million (in 1953 prices). What is going on here? Is this an argument against the validity or plausibility of the computer price index?

The following sections assess the points made in the above quotations concerning the plausibility of hedonic PC and computer equipment price indexes.

1. *You can't buy a cheap computer.* As a point of fact, you can. The cheapest Dell desktop computer in spring 2004 (at \$339, including applications software, but not a monitor) is close to the €300 example. But to an extent, the point is still valid. Generally, computer makers have used declining production costs (which originate largely in declining semiconductor production costs) as an opportunity to offer more performance per dollar or per euro, and not as an opportunity to cut the price even more by holding the performance constant.

Presumably, computer makers understand their markets. For cars, as well, the current models of the very cheapest cars still offer much more performance, comfort and features than the low-cost models of 20 years or so ago. The number of buyers for minimal transportation autos is low enough that scale economies prevent continued production of the basic transportation models of the past. The same thing is true of computers: If there were sufficient buyers for 100 MHz computers, one would expect they would still be sold. At about the time that 1000 MHz PC's first arrived (at roughly \$2000), I noticed a batch of *new* 100MHz computers offered on the internet, for a little over \$50.00 each. Critics who say they want them either do not know where to find 100MHz computers, or (more probably) there are fewer buyers with preferences for 100MHz computers than the critics think. Or to put it another way, in today's technology the price premium for adding another 100 MHz to the performance of the 100MHz machine is small enough that no buyer would sensibly stop at 100 or 200 MHz, even though they were quite pleased with that level of performance in the past.

For buyers who *really* want a 100MHz computer, its disappearance represents a forced upgrade, which the index should record as a price increase. As explained in Chapter IV, the price effects of exits from the index sample (forced or not) are in fact better estimated with hedonic methods than with conventional matched-model price comparisons, because the hedonic method estimates the quality-corrected price change that the exit implies, where the conventional method invokes the assumption that there is none. Neither method, however, records the value of forced upgrading. Evidence suggests that this source of bias is small, in any case, because the number of buyers who are really forced to upgrade is small.

An adage for technological products is that one can put in value (to the user) faster than cost. When that is the case, the economics of supply and demand result in the disappearance of low-end technology, which does seem to describe the computer market. Indeed, you can't buy yesterday's computer conveniently if you want one, because too few buyers seem to want it to make it economical to produce. That you cannot buy yesterday's computer does not mean that the price index for computers is wrong—if anything, the disappearance of yesterday's computer confirms that prices have dropped greatly.

2. *Performance increases are more than most users need.* This is a variant on the “price falls too fast” argument. The computer price index is performance corrected, but the performance improvements, according to this criticism, are redundant to the needs of users. Alan Blinder, when he was vice chairman of the U.S. Federal Reserve Board, once wrote to me that he was using perhaps 20 percent of the capability of his computer, since he only used it for simple word processing; despite this, the FRB staff replaced it with a new one, he said, which he used at perhaps 15 percent of its capability. The new machine's increased performance gave him little increased usefulness, so *for his purposes* the quality adjustments applied to hedonic computer price indexes overstate the value of the quality change and overstate the decline in computer prices. Many others have followed Blinder's criticism.

A variant appeals to the voracious hardware demands of modern software. Robert Gordon, in his criticism of computer performance measures, has often repeated a witticism from the computer industry: “What Intel giveth, Microsoft taketh away.” Because hardware capacity (increased speed and memory) is becoming cheaper and cheaper, it makes economic sense for software designers to

expropriate part of increasing hardware capabilities for their software. For example, software designers know that because machine memory is increasingly cheap computers will have large memory sizes, so there is little need to economize on memory demands of the software at the expense of other functionalities. In this sense, some computer users may not realize what increased computer performance is doing for them: Graphics capability, for example, is very demanding of hardware capacity. Graphics operate even the screen icons that make mouse control of computer functions feasible, and graphics capability is essential for internet uses. But the package (hardware and software, taken together) may not improve in performance as fast as the measured performance of the hardware, viewed by itself, because some of the increased hardware capacity is employed to accommodate the software's requirements.

A further variant depends on efficiency of the computer network. Computers interact, and the efficiency of a computer system requires compatibility among all the nodes. System efficiency, therefore, may demand more capability at individual nodes than is required for a particular user at that node, which is probably one reason that the FRB staff insisted that Blinder have a more powerful machine than his word processing strictly required.

Ultimately, though, one needs to ask why users buy ever more powerful computers, if they really do not use their capabilities. Though games and graphics uses drive much of the demand for performance that is built into the really high end PC, faster computers with larger memories and more features occupy the bulk of the market. It might be, as critics contend, that users do not really need the performance of the computers they are buying. But so what? They still buy them. Is there any precedent in economic statistics for subtracting from an industry's output the performance it produces that is not used by buyers of the industry's product?¹ I observe that an astonishing number of Ferraris and other high-performance cars are parked on London streets and used to motor into the city, where they must use only one or two of their six gears, and a tiny fraction of their performance. Do we deduct from auto manufacturers' output the portion of Ferrari performance that is not used by Ferrari owners? If computer buyers purchase increased PC performance, that is what they choose to buy, and it is output of computer manufacturers, just as surely as the output of Ferraris used to go grocery shopping is part of the output of the auto industry.

It is possible to buy cheaper, lower performance computers than the typical PC computer sold today. Buyers don't. Thus, they don't behave the way this critique presumes.

3. Performance measures used in hedonic functions are incomplete or inadequate measures of computer performance. Though most critics do not make this argument, they could and it might be used to buttress their other contentions.

The adequacy of performance measures used in empirical work on PC's was considered at length in Chapter V. As explained there, existing measures are problematic, to an extent. But the direction of bias to hedonic price indexes from inadequate performance measures it is not entirely clear. For example, MHz as a measure of computer speed is clearly inferior to a direct measure of work done, for the reasons discussed in Chapter V; but the only comparison of alternatives to date (Chwelos, 2003) finds little bias.

McCarthy (1997) contended that hardware performance measures (such as MHz) that are used in most PC hedonic functions also serve as proxies for the improvement in the software that increasingly has been bundled into computers. McCarthy posited that the software portions of the package were

¹ If the increased performance is not treated as quality change, it will not show up in industry output measures that are formed, as is usually the case, by deflation by an output price index.

declining in price more slowly than the hardware components.² Under those circumstances, MHz, memory size and similar measures of hardware capability overstate the performance change of the hardware/software package, which causes the price index for the PC bundle to decline too rapidly. McCarthy's contention is correct so far as it goes (see the section on proxy variable bias in Chapter VI). But existing measures of computer performance also make no allowance for the inclusion of new software packages. When word processing software and photo editing software were first added "free" to the price of the PC, no adjustment was made for their inclusion in most empirical studies, which biases the PC index in the other direction.

Bias from inappropriate or inadequate performance measures poses a potential problem for existing empirical work on computer prices. But the magnitude of the bias is unknown. Surely a one or two variable PC hedonic function is subject to substantial bias from omitted variables. But it is not so clear that the bias is large in hedonic indexes that are based on more adequate hedonic functions, and the direction of bias, if any, has not been determined.

The same objections, incidentally, could be raised against conventional matched-model indexes, because they normally control for quality change in computers using the same variables that are used in the hedonic function. If the variables are inadequate for hedonic indexes, they are inadequate to exactly the same extent when used as the pricing specification for matched-model comparisons (see Chapter II).

4. *We are forced to buy the improved technology because of monopolistic behaviour on the part of computer and software producers.* On the industrial organization aspects of the computer industries, there is little market power for PC's. It is easy to assemble PC's from purchased components and therefore relatively easy to enter the PC market. Market power is more of a factor in the production of microprocessor chips and of PC software.

Aizcorbe, Oliner, and Sichel (2003) make a nice estimate of the influence of the market power markup on semiconductor price changes. They estimate that markup changes account for only a small proportion of the price decline. Most of the decline in semiconductor prices reflects technological change, not changes in the markups in response to oligopolistic rivalry.

In software, on the other hand, this is a big problem, as antitrust cases on both sides of the Atlantic against the leading software supplier bear out. The essence of the antitrust cases involved alleged behaviour that forced packages on unwilling buyers, who were ultimately the consumers. However, as the antitrust cases abundantly show, one can find economists on either side. Whether there has been welfare loss from slowing innovation is in dispute. There is little evidence of a price effect, but some for forced quantity and, one suspects, effects from spurious "upgrading." The evidence at this point is mixed, and insufficient to make an assessment for the bias, if any, in PC price measures that include software in the bundle.

5. *"Whatever: I wouldn't have paid the price suggested by extrapolating the hedonic index into the past."* This objection, though no doubt coupled with some of the points above, is logically separate from them. It is a more subtle point that requires a complex response.

As the textbooks show, the market demand for any product is made up of the aggregation of individual demands. In the case of new products such as computers, sales are small at a relatively high price (such as existed in the past). A famous "forecast" in the early 1950s predicted that world-wide

² On prices for PC component software, see Prud'homme and Yu (2002), Abel, Berndt, and White (2003), and White, Abel, Berndt, and Monroe (2004). The latter reports price declines for PC operating system and productivity suite software of around 13-18 percent per year, where PC hardware prices have been falling at 20-30 percent per year, or even more, which is consistent with McCarthy's speculation.

demand for computers would amount to only a small number of machines, and this might have been correct at computer prices of the time. For most buyers, and certainly for individuals, computer prices were too high to consider buying one. For those individuals, demands were zero.

Hicks (1940) showed a way to analyse the demand for new products. The “reservation price” is the highest price an individual would pay for the new product, or the minimum price that would keep the individual out of the market. The price index for the new good that did not exist in the previous period could be found by comparing the current price (that was paid) to the reservation price (the price that resulted in zero demand) in the previous period. Hausman (1997) famously used Hicks’ approach to estimate price indexes for new products.

Somewhat overlooked at times is the applicability of Hicks’ approach to the demands for individuals. Consider individual A (the speaker in the sentence quoted at the beginning of this section). He indicated that his demand for a computer (with specifications equal to the current one) was zero when the price of that PC was €15,000, so his reservation price is lower than €15,000. Though he has not given his reservation price, suppose it is €2000. He certainly would buy one when the price fell to €1000.

Then we can apply Hicks’ rule to construct *a price index for individual A*: The decline in PC prices from €15,000 to €2000 is irrelevant for individual A, because he was not in the market for a computer at any price above €2000. But the decline from the point where a PC cost €2000 to its €1000 cost is relevant.

Individuals differ in taste, so each has a different reservation price. Early computer buyers have higher reservation prices, so they own computers when the rest of us are still computerless. Increasing market demand for computers occurs as computer prices fall to or below the reservation prices of more and more potential buyers. As the price of computers falls, more and more individuals and firms buy computers. The market demand for computers, of course, is the aggregation of the individual demands of those potential buyers who have different reservation prices.

Similarly, aggregate price indexes are aggregations, in principle, of individual price indexes (see the discussion in Pollak, 1975, for the consumer price index, or cost-of-living index, but the point applies to other price indexes as well). That the aggregate price index for PC’s, over the whole history of PC’s, does not follow individual A’s own price index for PC’s is an expected result because individual A’s own price index for computers is undefined above his reservation price. For individual B, who was willing to pay €15,000 for a PC with today’s specification, the aggregate price index from the €15,000 cost point to the present is relevant, but not perhaps the index for earlier years that extrapolates the price back to €20 million. On the other hand, the firm that was willing to pay €20 million for an early computer has a massive gain from being able to buy one (or many) for €15,000 or €1,000.

The aggregate price index will not represent the price experience for buyers who were not in the market for PC’s. This does not, however, invalidate the aggregate hedonic price index as a measure of what actually happened to computer prices. The aggregate price index measures the price declines that potential buyers face, not necessarily what each of them would have been willing to pay at points where they were not computer buyers. As prices fall, as measured by the hedonic index, more and more buyers for computers exist, so for more and more of them the aggregate price index also matches their personal price index. This, as noted at the outset, is a subtle point in interpreting aggregate price indexes.³

³ The price index literature also contains references to consumer surplus, the gains to buyers who would have been willing to pay more than the current price to obtain a computer but who do not have to. I ignore that for present purposes, but the above discussion is obviously closely related.

3. Summary

The “falls too fast” arguments are not compelling. The size of computer price declines provokes, understandably, astonishment, and strains (it seems) credulity. Yet, the anecdotes that have been spun as criticism do not stand up very well to analysis. Computer prices have been falling at 20-30 percent per year for a very long time, and the great improvement in computer performance is not in doubt among computer professionals. It is implausible that very much of the 10 or 20 or 50 year improvement in computer capabilities and the consequent decline in the price of computer performance is disguised by forced changes or lack of free choice among computer buyers and so forth.

Technical improvements to hedonic indexes can be made—improvements in the variables in hedonic functions, for example—and will no doubt lead to revisions of hedonic indexes. But there is little reason to suppose that that such revisions will eliminate or reduce appreciably the remarkable story of the computer’s 50-year price decline.

B. TECHNICAL CRITICISMS OF HEDONIC INDEXES

1. General Criticisms

Ho, Rao, and Tang (2003) repeat what they have heard said about hedonic indexes, and they accurately summarise a persistent form of technical criticism. Their list includes theoretical appropriateness of the functional forms, transparency and reproducibility, and the origin and choice of the characteristics variables.

a. Functional Form

First, there is the criticism of an alleged lack of theoretical foundation with respect to choice of hedonic functional form. As Chapter VI and also the theoretical appendix explains, this longstanding criticism is misconceived. Rosen (1974) showed conclusively why in theory the choice of functional form is entirely an empirical matter, it is not determined on theoretical grounds. Significant contributions since Rosen (including Pakes, 2003, and the work with collaborators he cites there) has not changed this conclusion. Even though one still sometimes sees erroneous statements that one or the other of the functional forms commonly used in hedonic research is theoretically unacceptable, the authors of such statements apparently do not understand the appropriate theory.⁴ In Triplett (2004), I considered the curious fact that this functional form criticism has arisen for product hedonic functions (such as for computers) but it has not arisen in the parallel literature in labour economics on human capital or wage hedonic functions.

b. Transparency and Reducibility

It has often been said that hedonic indexes lack transparency and reducibility. This contention, expressed by statistical agencies in the past, has been revived by the recent Committee on National Statistics report (Schultze and Mackie, eds., 2002): Are the quality adjustments that hedonic indexes provide objective and reproducible? Some observers imply that investigator effects create too much dispersion in hedonic indexes to rely on them in official statistics, and that there is too much subjectiveness in choosing the characteristics.

⁴ Or they ignore the salient parts of it, in favor of special cases that do not provide guidance for empirical research. An example is Diewert (2003).

The transparency criticism of hedonic methods implies the converse—that matched-model methods for handling quality change are transparent and reproducible. This is a dreadful misunderstanding. The opposite is more nearly the case—it is hedonic methods that are more transparent and more capable of reproduction, certainly outside statistical agencies, which is where it matters.⁵

Traditional methods are not transparent. If an agency decides to use the IP-IQ method (see Chapter II) when a quality change arises in a price index sample, then the price index operation itself is more transparent—to it—than the dummy variable form of the hedonic method (e.g., equation 4.1). The same thing is true of the direct comparison method. But when does an agency decide to use implicit quality adjustment (IP-IQ)? When does it use direct comparison or, in cases where it is still done, link-to-show-no-price-change? Those decisions, which are not transparent to outsiders, contribute to opaqueness in the conventional method.

Neither are the quality adjustment procedures themselves transparent. The various forms of “linking” that are traditionally done when quality changes are encountered may not be liked, but agencies understand what they do, and how what they do relates to the rest of the index calculations. Judgmental quality adjustments may be arbitrary, but they are transparent, at least to the agencies that do them. Observers outside statistical agencies, however, cannot see what is done within them. The total calculation of the index using conventional methods is decidedly not transparent to outsiders.

Moreover, the implications of the use of some of these methods have not always been understood even by the agencies themselves. Their properties are not transparent (see the discussion in Chapter II), whether or not the procedures themselves are transparent.⁶

The conventional method is also not as reproducible as sometimes assumed. Reproducibility of a statistical method requires that capable people, confronted with similar problems, resolve them in ways that produce quantitatively similar results. As noted earlier, Hoffmann (1998) describes a virtual experiment in the German CPI price index for washing machines. In Germany, separate consumer price indexes are computed for each German Land and quality adjustments are performed independently in each of these Lander indexes. Because quality change arises from particular events in the samples, there is no assurance that each Land’s price index compiler faced exactly the same quality adjustment problem. Nevertheless, over a fifteen-year period, those independently adjusted washing machine price indexes gave a very large range of estimates of price change—from more than 10% decline in the index that fell the most to just over a 30% increase in the highest (see Figure 2.4 in Chapter II). Chapter II discusses similar “virtual experiments” on computer price indexes across OECD and European countries that have yielded similar variances among matched-model methods implemented in different statistical systems—see Wyckoff (1995) and Eurostat Task Force (1999), and also the Dalén (2002) and Ribe (2002) studies. Additional studies within Eurostat for the HICP are exemplary.

Accordingly, the transparency and reproducibility argument does not provide strong support for the traditional method. Mechanical or computational transparency might be greater in conventional methods; but overall transparency is not greater, because the statistical procedures that come into play before quality adjustments are carried out are not themselves transparent. The effects of these decisions have mostly not been quantified and are not factored into calculation of index number variances, even in

⁵ I mean no criticism of Ho, Rao, and Tang (2003) here, they are correctly reporting what has often been said.

⁶ Two changes in agency practice illustrate what I have in mind. Armknecht (1996) remarked that the former rule followed when quality changes were encountered in the U.S. CPI was: “When in doubt, link it out [with the IP-IQ method]”. This implies, correctly, that the implications of the IP-IQ method had formerly not fully been understood within BLS. A second: Eurostat banned use of the link-to-show-no-price-change method in the HICP indexes, because of its severe bias. That the method was widely used in Europe and is still used elsewhere suggests that its properties are not fully understood.

countries where price index sampling variances are routinely published. Agencies think, or behave, as if their methods were reproducible; but actually, all they know is one set of decisions were made with one result, not that the same result would be reached if the decision process were in fact replicated.

Griliches (1990, pages 190-91) said it well: “The fact that [hedonic research] is difficult to do, and that an actual empirical implementation calls for much judgement on the part of the analyst and hence exposes him to the charge of subjectivity, is still the most telling objection today. The fact that the standard procedures also involve much judgement...is usually well hidden behind the official facade of the statistical establishment.”

One might, indeed, prefer a method whose statistical properties can be examined (the hedonic price index) over one whose statistical properties are less easy to describe. Especially when an agency does not use probability sampling in selecting items and outlets, the statistical properties and the overall transparency of conventional methods are not their strong points.

2. The CNSTAT Panel Report

The report of the U.S. Committee on National Statistics (CNSTAT) Panel on Conceptual, Measurement, and Other Statistical Issues in Developing Cost-of-Living Indexes (Schultze and Mackie, eds., 2002, hereafter “Panel”) contains a chapter on quality change and hedonic indexes (its chapter 4). The report has been read in a number of ways. Its interpretation was clarified by a series of presentations at the Brookings Institution in February, 2002⁷, at the Bureau of Labor Statistics in June 2002, at the National Bureau of Economic Research in July, 2002, at the American Economic Association’s annual meetings in January, 2003, and in a subsequent article by its chairman (Schultze, 2003).

On the one hand, the report contains positive statements about hedonic methods (all pages from Schultze and Mackie, eds., 2002):

- “Hedonics currently offers the most promising technique for explicitly adjusting observed prices to account for changing product quality.” (page 122)
- “In a well-specified equation, coefficients on the explanatory variables [of the hedonic function] reveal the marginal relationship between the product characteristics and price at various values of [the characteristics].... Then, as long as the set of observable characteristics includes all characteristics that matter to consumers and the equation is properly specified, these results can be used to correct for product quality change.” (pages 123-124)
- “The successful use of hedonic methods rests on a modeler’s ability to identify and measure quality-determining characteristics and specify an equation that effectively links them to the prices of different models or varieties.” (page 124)

Accordingly, one interpretation of the Committee’s chapter is that it was intended as a compendium of good practices. Though the language of the report sometimes suggests that members of the committee thought their points were new, most of the prescriptions in this part of the committee’s report are well established in the hedonic literature. Indeed, they are consistent with the analysis of the Chapters II-VI of this Handbook, and the Committee cites the “Advance” version of the Handbook. Looked at in that way, the Committee’s section on hedonic indexes would probably be endorsed by most

⁷ The presentations were made at a Brookings Institution Workshop on Economic Measurement and can be accessed online at: <http://www.brookings.edu/es/research/projects/productivity/20020201.htm>

hedonic researchers. On the other hand, the chapter also contains language that is inconsistent with this interpretation, and apparently very few readers of the chapter have so interpreted it.

Another way of interpreting the Committee's report is that it was critical of the BLS implementation of hedonic indexes, rather than of the method itself. For example, the committee expresses disapproval of hedonic functions whose variables are disproportionately or largely brand dummy variables (true of some BLS hedonic functions),⁸ it criticizes the BLS for instituting hedonic quality adjustments (the hedonic quality adjustment method is described in Chapter III, above) without plans for updating the hedonic functions,⁹ and it expresses concern that too little research had been done on the stability of coefficients in the hedonic functions estimated by the BLS. Those seem valid reservations against the BLS' particular implementations, but do not necessarily imply a defect in the hedonic technique—quite the contrary, the Panel's points suggest BLS deviations from best practice (see for example the discussion about keeping the coefficients up to date in Chapters III and IV of this Handbook). Against that interpretation of the report, however, is the language “the panel is not convinced that anyone could have done this work better than the BLS” (page 134), an odd assessment since other countries' statistical agencies have implemented hedonic indexes with explicit plans for, e.g., updating coefficients regularly, and researchers regularly update coefficients for hedonic studies.

A third interpretation of the Committee's views is possible. When the CNSTAT committee's staff director, Christopher Mackie, presented the hedonic portions of the report at the Brookings Workshop on Economic Measurement in February, 2002¹⁰, he seemed to interpret the Panel's conclusions as a set of empirical impossibility theorems. Mackie's interpretation is consistent with some of the language in the report: “The long list of unresolved issues discussed in this chapter explains why even some proponents of hedonics advocate a less aggressive expansion of its use in the CPI than BLS appears to be taking” (p.144)].¹¹ Many readers of the report have so interpreted it. At subsequent meetings, particularly at the NBER 2002 Summer Institute, the principal author of the Committee's chapter 4 (Richard Schmalensee) indicated that Mackie's interpretation was not the appropriate one to give to the committee's report, and Schultze (2003) makes the same point. Some notably infelicitous language (for example, the “long list of unresolved issues” boils down to a small number of mostly standard matters) and some carelessness with the hedonic literature¹² makes this understandably a bit hard for readers to comprehend from the report alone.

Assessment. A review of the Panel's chapter 4 suggests the following main critical points about hedonic indexes.

⁸ “If one assumes that brand, in itself, does not lead to higher valuations by consumers, one must believe that it is an acceptable proxy for unmeasured quality characteristics. Incidence of repairs might be one such example.... Moreover, brands are repositioned in terms of relative quality from time to time, and reputations sometimes change in response to advertising campaigns, so that brand dummy coefficients may be inherently unstable” (Schultze and Mackie, eds., 2002, page 133).

⁹ Some BLS hedonic functions (TV's, clothing, and computers) are regularly updated, but others apparently are not. The Panel's criticism has seemingly been accepted by BLS, at least for some of its hedonic functions.

¹⁰ See his outline notes on the Brookings website at:

http://brook.edu/dybdocroot/es/research/projects/productivity/workshops/20020201_schultze.pdf

¹¹ No citations support this statement, and I can identify no “proponent” whose views can be so summarised.

¹² Van Mulligen (2003, page 154) remarks: “The work by [the Panel] ignores a lot of recent research on hedonic indices...”; he goes on to cite as a source corroborating this view a paper by one of the Panel members.

- “In practice, the critical question is whether one can reliably estimate functions that capture the relationship between market price and characteristics that confront individual consumers” (page 124).¹³
- The Panel refers to “difficult econometric problems that plague all hedonic analysis—e.g., identifying appropriate functional forms and relevant product characteristics...” (page 126).
- “Theory provides little guidance to help determine the appropriate functional form for hedonic equations” (page 142).¹⁴
- “For many classes of goods—and perhaps especially services—it can be extremely difficult to identify which characteristics are actually associated with price” (page 142).
- The Panel refers to “the general problem of price data that reflect nonobservable seller attributes...”¹⁵ and “the data requirements and the operational difficulty of producing [some types of hedonic indexes] on a high-frequency, up-to-date schedule” (page 128).
- “It is hard to know when a hedonic function is good enough for CPI work” (page 143).
- “...Hedonic surfaces may change rapidly. The ability of BLS or any other agency to capture those changes in real time is, at best, doubtful. It is unclear whether usable estimates of hedonic surfaces can be routinely and rapidly computed for a wide variety of goods” (page 143).

Most of these points are relevant ones. They just say that doing hedonic research is not simple or mechanical or routine, as is of course true with any other economic research. There is nothing very surprising or controversial in them, though the last one (which is an empirical judgement) might be answered by an empirical review that the Panel did not undertake. And the list is neither very long (despite the report’s referring to a “long list” in more than one place), nor a particularly negative one, nor one that is particularly damaging to the concept of hedonic indexes. The Panel’s views, as summarised in the above quotations would probably be shared by researchers who have actually worked on hedonic indexes.

The Panel’s chairman, Charles Schultze, addressed perceptions of the Panel’s report in a paper written subsequently:

In recent months, a number of panel members have heard comments to the effect that the panel’s report takes a negative view about the potential of hedonic techniques—apparently because the report discusses some of the difficulties with hedonic techniques.

¹³ The Panel goes on to emphasize, appropriately, that consumer tastes differ, so that the quality adjustment that might be appropriate to one individual’s cost-of-living index may not be appropriate to another’s. See the discussion in section A.2.b.5 of the present chapter about differences in individuals’ price indexes for computers, which is a parallel point.

¹⁴ Curiously, the Panel does not acknowledge here that it is the index that matters, as Pakes (2003) points out, not just the hedonic function. Research suggests that the hedonic index is not very sensitive empirically to the functional form chosen (see Chapter VI). It is also odd that the Panel omits the simple point that standard econometric methods for choosing among functional forms exist and are widely used (Chapter VI); since the theory indicates that choice of hedonic functional form is entirely an empirical matter, as the Panel’s report itself notes, the Panel could profitably just have ended its discussion of functional forms with that observation, rather than bringing up the outmoded “no theory” contention. One suspects that the Panel’s report was somewhat inconsistently reflecting views of its various members.

¹⁵ For example, scanner data aggregated across stores.

Yet our report explicitly concluded: “Hedonics currently offers the most promising technique for explicitly adjusting observed prices to account for changing product quality.”

The issue is not whether hedonics is potentially of great usefulness. It is. Rather...[the BLS] could, as the as the panel suggests, channel its efforts principally into analyses, tests, and experiments aimed at exploring and resolving some of the methodological issues discussed in the panel’s report. *The results might well justify the modification of BLS item replacement procedures* and an expanded application of hedonics.... (Schultze, 2003, page 17, emphasis supplied).

The passage emphasised is easy to overlook. The Panel was impressed by the fact that the small number of BLS studies showed that using the hedonic quality adjustment method (explained in Chapter III, above) on forced replacements in the CPI sample made little change to the index (Schultze, 2002). This finding implies that the implicit quality adjustments that BLS had been making for forced replacements were surprisingly adequate, which in itself is an interesting empirical finding, that runs counter to the usual professional presumption that hedonic indexes will differ from conventional ones.¹⁶ But of course it applies only to the components that the BLS has studied and only to the treatment of forced substitutions in the CPI.

The Panel was also convinced (from BLS materials) that the item replacement rule used in the U.S. CPI assured that the quality changes observed for forced replacements were small, which is one reason why hedonic and conventional quality adjustments yielded similar indexes.¹⁷ Quality changes in the universe, however, were not necessarily small. Changing the BLS item replacement rule (for example, assuring that the replacement product was representative of the market, rather than similar to the item that disappeared) would likely introduce larger quality changes into the index, and for larger quality changes, it might not be true that conventional and hedonic quality adjustments were similar. The Panel, however, felt reluctant to recommend changes to the BLS item replacement policy because of its reservations about the stage of development of hedonic methods, at least in the BLS implementation of them. If hedonic indexes were more thoroughly developed, this might permit changing the product replacement rule. It might also justify some adjustment when the whole sample is rotated, what the Europeans call “many to many” replacements (Chapter III):

“...the application of hedonic adjustments in a different way and on a larger scale might produce more significant downward adjustments. The panel believes that...further research, testing, and evaluation of hedonic methodology and specific applications should precede expansion of its use, such as to sample rotation—something that the panel is not in principle opposed to—where the impact on index growth would likely be more significant (Schultze and Mackie, eds., 2002, page 140).

¹⁶ In a purely research mode, this finding might have suggested more hedonic studies to see if BLS conventional quality adjustments were equally adequate across a larger range of products. The Panel did recommend “audits” of CPI components, which goes in the same direction.

¹⁷ The item replacement rule the Panel cites is: Find the replacement product with the closest characteristics to the one that disappeared. However, this is not always the replacement rule, sometimes BLS resamples (makes a new probability selection) from the items currently for sale in the outlet. In this case, the characteristics of the replacement might not be similar to the item that disappeared. Moulton and Moses (1997) give as an example the replacement of a basketball with a tennis racket in the sporting goods index. It is doubtful that hedonic adjustments would be appropriate to such replacements, but it is also not clear that the example is typical of situations where resampling is employed when products exit the sample.

Rather than putting additional resources into estimating more hedonic functions and using them to quality adjust forced replacements in CPI samples, in the Panel's view the same resources could yield more improvement in the CPI if spent in different ways. Though this was apparently the underlying motivation for the Panel's recommendations that the BLS pursue "a more cautious integration of hedonically adjusted price change estimates into the CPI" (Schultze and Mackie, eds., 2002, page 141), it is also fair to say that the eight recommendations in its chapter 4 do not explicitly make this point.¹⁸

If the foregoing interpretation of the Panel's thinking is correct, what is surprising is the Panel's characterisation of it. At one place, it says: "This recommendation [that the BLS slow its implementation of hedonic indexes] is based on theoretical grounds, not on empirical ones" (page 141). More telling, perhaps: "Our conservative view on integrating hedonic techniques [into the CPI] has more to do with concern for the perceived credibility of the current models" (page 141). Hulten's (2002) observation is relevant here: He said the Panel behaved as if old methods were good methods.

As with most economic measurement issues, analysis of the choice between hedonic and conventional methods requires balance. Refraining from adopting hedonic methods means relying on conventional methods.

Difficulties or problems in hedonic methods often have counterpart or parallel or corollary difficulties in conventional methods. For example, the Panel's report says that determining the relevant characteristics for hedonic analysis is difficult, which is clearly true; unless the characteristics are well chosen the hedonic index may be defective, as pointed out in Chapter V. But the conventional method also requires determining the characteristics because the characteristics must be built into the pricing specification. The specification determines when a match is successfully made and when some quality adjustment method must be brought into play because a match is not successful (see the discussion in Chapter II). Pakes (2003) points out perceptively that BLS (and other statistical agencies around the world) have been determining characteristics for years, because they are built into pricing specifications. Yet, BLS choice of characteristics has not generally been controversial and the Panel does not point to the difficulty of choosing the characteristics as a flaw in conventional methods. One can agree with the Panel's reservation that knowing the characteristics is a limitation on the validity of hedonic research without agreeing at all that this is an argument against hedonic methods and in favour of conventional methods.

Pakes (2004) goes further. He contends that most of the existing criticisms of hedonic indexes are either solved, in the sense that we know what to do to produce hedonic indexes that are not subject to the criticism, or they apply equally to hedonic indexes and to conventional matched-model ones. Moreover, in a number of the latter cases, the defect in the hedonic index is less severe than the corresponding defect in the matched-model index. For example, the hedonic index may not estimate adequately the loss to a computer buyer when the 100MHz computer the buyer prefers disappears (see the discussion in section A.2 of this chapter), but neither does the matched-model index.

Too often, the general discussion of hedonic and conventional methods has not considered adequately or put sufficient weight on cases where adopting hedonic methods can bring about corollary improvements in the index. Two examples illustrate.

¹⁸ Three of the eight recommendations focus on the difference between what the Panel called the "direct" and "indirect" methods. These recommendations concern essentially empirical matters, but they were unfortunately not backed by an empirical review. For example, the Panel instructed BLS not to put resources into the dummy variable method, but to explore the price index for characteristics method. Though I share some of the Panel's reservations about the dummy variable method, as pointed out in Chapter III of this Handbook, the review in Chapter III of hedonic indexes computed by different methods indicates that the method may not make as much difference empirically as the Panel apparently thought.

The Panel notes approvingly that hedonic functions can be used to perform statistical testing on the validity of BLS pricing specifications (first suggested in Triplett, 1971, and employed, rather sparingly to date, in BLS studies). The Panel also points out that the present BLS product replacement rules (which are effectively to find the next most similar product to the one that disappeared) create considerable potential for “outside the sample” quality change error (see in this Handbook Chapter II and the discussion of outside the sample error in Chapter IV). It judges that outside the sample error may be as significant for accuracy of the indexes as the usual concern for accuracy in pricing what is inside the sample, a judgement that is consistent with price index research. Yet, changing the item replacement rules and BLS sampling rules would require more reliance on hedonic methods to handle replacements that differ appreciably from the items in the sample previously.

Ball and Fenwick (2004) propose a novel use of hedonic methods to facilitate probability sampling. They use a computer hedonic function estimated in the Office for National Statistics in the U.K. to isolate the price-determining characteristics. Then, using scanner data, they array sales of computers by cross-classified characteristics. This gives a matrix from which they select with normal probability methods a particular specification for each pricing agent. The result is a probability sample that is much cheaper to implement than the probability sampling methods used by the BLS.

I would reverse the Panel’s recommendation to pursue “a more cautious integration of” hedonic methods. The potential error from outside the sample error, and from an inadequate replacement rule when old products exit the sample, is a defect in the CPI, and a reason for the BLS to accelerate its implementation of hedonic methods in order to facilitate ancillary changes elsewhere. Even if one agrees with the Panel’s reservations about the quality of some of the hedonic functions estimated in the past by BLS, surely it is much easier to correct these deficiencies than the Panel’s report suggests. Had the ONS followed the CNSTAT Panel’s recommendations to slow implementation of hedonic indexes, their innovative improvement to UK data would have been foreclosed because they would not have had a hedonic function available.

The Panel’s report does not make a case that hedonic methods have insurmountable flaws, and Schultze (2003) indicates that it did not intend to do so. It does make the supportable case that implementing hedonic indexes well is difficult, though not that it is impossible. It does validly criticise aspects of the BLS implementation, but (though the writing is insufficiently clear on the point) its criticisms are fundamentally of a particular, flawed, implementation of the technique, rather than of the technique itself. Finally, the report gave insufficient weight to whether some of the problems it raises with hedonic indexes have parallel problems in matched-model methods for handling quality change, and it also does not consider sufficiently or emphasise that abstaining from the use of hedonic methods in price indexes results, by default, in use of conventional methods, which have their own well-known deficiencies.