Volume Title: Models of Income Determination

Volume Author/Editor:

Volume Publisher: UMI

Volume ISBN: 0-870-14184-8

Volume URL: http://www.nber.org/books/unkn64-2

Publication Date: 1964

Chapter Title: Notes on the Measurement of Price and Quality Changes

Chapter Author: Zvi Griliches

Chapter URL: http://www.nber.org/chapters/c1823

Chapter pages in book: (p. 381 - 418)

Notes on the Measurement of Price and Quality Changes[1]

ZVI GRILICHES

UNIVERSITY OF CHICAGO

Introduction

AS ECONOMETRICIANS have accumulated experience in analyzing different bodies of data, their views of what are the major technical problems associated with a particular study have changed markedly. Early studies put most of the stress on "what is the relevant theory" for a particular piece of data, with "theory" providing such broad statements as "quantity purchased should be related to price and income." The next big step came when we started asking, "What is the relevant variable?" Is consumption related to measured income in the same period, to measured income in the previous period, or to some more elaborate but also rather intangible concept of expected or permanent income? Is the planned output of wheat related to current, past, or expected wheat prices? While earlier workers were not unaware of these problems, the general realization of the importance and fruitfulness of such a question owes much to the works of Friedman, Nerlove, and more recently, Muth.[2] It has now become standard operating procedure to inquire what kind of variable we "really" want when we write down price, income, or capital in one of our equations. More recently, we have started asking, "Is this a good or relevant measure of the variable?"; do the series that we actually have measure what we want or even measure well what they set out to measure?[3] As we begin to ask finer questions of the data, the quality of the available data becomes a major constraint on the work of the practicing econometrician.

The problems that I shall review in this paper arose during various

[1] This paper is based partly on the results of a larger research project on the econometrics of technological change supported by a grant from the National Science Foundation.

[2] See, for example, [5] [15] [13]. The numbers in brackets refer to the list of references at the end of the paper.

[3] See, for example, [14].

attempts to use available price data in two areas of special interest to me: the measurement of productivity and technological change and the study of investment behavior. On several occasions I found that what was being measured was not what I wanted, even though the two concepts had the same "name."[4] This, however, should not be interpreted as a complaint against the producers of these statistics. They have to provide all-purpose numbers and cannot guess in advance the particular combination or measure that I may want some day.

Economists use price series for two main purposes: (1) to deflate expenditures and receipts for the purpose of arriving at some conclusions about either changes in welfare (in the case of consumption expenditures and earning receipts) or changes in productivity (in the case of sales receipts, wage bills, and investment expenditures); and (2) to explain and predict changes in quantities used or purchased. In either case we are likely to have a broader concept of "price" in mind than just one of the particular numbers recorded during a transaction. Clearly, an item bought at the same "transaction" price but in one case paid a month in advance of delivery and in a second case paid three months after delivery did not "cost" the same amount. Since economists are likely to assume that it is *total* cost per unit that affects either behavior or welfare, they will usually try to convert these two different transactions into "equivalent" units, using some appropriate interest rate to achieve this transformation (and an argument always remains about the appropriate rate to use). It does not help to tell them that these are distinct transactions and cannot be compared perfectly, since this counsel leads either to despair or to the explicit introduction of an infinity of dimensions or qualities of a transaction and, hence, to despair again. To try and measure changes in welfare and productivity, or to explain the time pattern of investment or other economic series, we shall find it necessary again and again to put certain changes into the category of "price factors." Whether all these different attributes of commodities should be lumped into one index, and whether the official indexes should be broadened to include some of these, is a semantic and (perhaps) political question. The important point is that we want and need this information to evaluate and understand better the performance of our economy.

As the result of such consideration we shall usually want to define

[4] I have touched on some of these problems before in [7] [8] [9].

a commodity or transaction broadly enough to include all the characteristics and conditions of purchase and use which impose a cost or provide a benefit to the purchaser and adjust our measure whenever any *one* of these conditions changes. Thus, in a study of the demand for transportation services, not only the "price" of passage should be included but also the "time" of passage, since in almost all cases the time spent in transit is a cost rather than a "good" (except for cruise passengers). Also, since most econometric studies are implicitly "demand" studies (partly because our supply theories are much less well developed), we will be more interested in prices "paid" than in prices "charged." Unfortunately, most of the data (except wage data) is collected from sellers rather than buyers, and this brings in an additional source of bias. While much of what we want is very difficult to define, measure, collect, and compute, it is still important to keep in mind what it is that we "really" want as we go along and compromise with reality.

In the next several sections I shall describe the difficulties that arose when I tried to use some of the available price series for my purposes and shall suggest a few possible ways of arriving at more appropriate measures for the particular tasks.

Inappropriate Definitions or Measurement Procedures

Some available measures are just not very good price indexes, quality change and other problems aside. In the case of the United States Department of Agriculture's Prices Paid Index the bark is actually worse than the bite. Theoretically, it does not even desire to construct a conventional price index:

> . . . The method of pricing items for the Prices Paid Index seeks similarly to reflect changes in items bought by farmers, such as *grade*, *quality*, and *size of containers*, that is, to reflect accurately the *average* prices of things farmers actually buy under the economic conditions that exist at the time of purchase. For example, some items such as grease, corn meal, bread, oatmeal, and cornflakes are sold in different size containers. For these the attempt is made to estimate the average price for all such commodities bought, giving proper weight to the *changing* proportions bought in containers of different size.
>
> *Ideally*, to maintain conceptual similarity to the Index of Prices Received, the price paid, say for work shoes, should be the average price obtained by dividing the total sum spent by farmers for work shoes in a given period by the number of pairs bought. Thus the average price,

when multiplied by quantity bought, would equal total expenditures for shoes. However, it is impossible to obtain the data needed for such a computation, and the nearest approach appears to be to price "the kind of work shoes commonly bought by farmers." Thus, as in the case of the average price described above, any marked shift in kind or type of shoe would be reflected in the price reported.

Accordingly, in pricing most items emphasis is placed on the item "most commonly bought by farmers," or the "volume seller" . . .

Clearly, comparison of the cost of a certain make of automobile today with the cost of the same make 25 years ago is a legitimate comparison, irrespective of the fact that the two cars differ vastly as to quality and design. The car of today provides better, more dependable, and more comfortable transportation than its predecessors of 25 years ago. *But this is largely beside the point.*[5]

Two reasons are given by the USDA for desiring unit values rather than prices in its Prices Paid Index, and neither one of them is satisfactory. First, it is argued, since the Prices Received Index does not take into account quality change in the products farmers sell, the Prices Paid Index should not take into account quality change in the items farmers buy. The logical conclusion should be, however, to improve the Prices Received Index rather than ruin the Prices Paid Index. Since product quality change has probably been much less serious in agriculture than input quality change, the error introduced by ignoring it in the Prices Received Index is much smaller than the "compensating" error of ignoring quality change in the Prices Paid Index. The second reason given by the USDA for ignoring quality change is that most of it is "irrelevant" for the major purposes of the particular item. "A car is a car is a car." It provides transportation services, and it does not matter whether it is a Chevrolet or a Cadillac. And if farmers shift from Chevrolets to Cadillacs, the real price of transportation services to them has gone up. Obviously, whether certain quality changes are relevant or irrelevant is an empirical question, and it cannot be dismissed a priori.

Fortunately, the USDA does not practice all that it preaches. Where it is obviously wrong the USDA does not stick to unit values as its ideal. It does not price just "a tractor" but divides tractors into five size and type classes. Also, a substantial fraction of the recent revisions in the Prices Paid Index has been in the direction of making

[5] [23, pp. 32–33.] Italics supplied.

the definition of the commodity priced somewhat more specific. For example, "all soybean meal" has been changed to "41 per cent protein soybean meal," and so forth.[6] Nevertheless, many of the definitions are still quite vague, and the insistence on pricing items with all "the customarily bought" attachments leads to substantial bias in the USDA estimates of prices paid for more complicated pieces of machinery.[7]

Moreover, in some cases the adherence of the USDA to its definition leads to ridiculous and misleading results. For example, in October 1960 the USDA substituted Ramblers, Falcons, Corvairs, and Valiants for the previously priced Chevrolet, Ford, and Plymouth sixes (and the Buick Special for a standard Buick eight) without any adjustment or linking.[8] As a result of this the USDA index of prices paid for new automobiles fell from November 1959 to October 1960 by about *13 per cent*, while at the same time the Consumers Price Index (CPI) index of new automobile prices, which also introduced compacts into its list of automobiles priced in November 1960, fell only by about 2 per cent. The USDA explained: "This reduction reflects in part shifting consumer preference to the new compact autos."[9] In fact, however, this drop in the index is not the result of farmers shifting to compacts, on which we have very little evidence at the moment, but rather the consequence of the price-collecting agency's shift in its definition of the items priced without adjusting for it.

Pricing per Unit of Service Rendered

Ideally, a price index measures changes in the price of a well-defined commodity or service which is of interest to the investigator. Some indexes, however, are not even price indexes in this loose sense. For example, almost all of the construction "costs" indexes do not price a particular well-defined piece of construction (e.g., so many square feet, such and such materials, house or factory building); instead, they simply average building materials price indexes and construction wage series in the hope that the result will approximate movements

[6] See B. R. Stauber *et al.* [17].

[7] In a previous paper [7], I have estimated that the prices collected by the USDA for some of the more complicated farm machines, such as cornpickers and tractors, drifted upward between 1947–49 and 1958 by about 20 per cent relative to the same items in the Wholesale Price Index.

[8] [22, January 1960 and 1961].

[9] [22, October 1960, p. 3].

in the price of the product. As a result, little or no increase in productivity is allowed to appear in the construction industry, and the cumulation of such "deflated" building expenditures in various industries seriously underestimates the growth in the "true" quantity of capital invested in "structures."

Luckily, there is one area in which prices are collected for a well-defined unit of construction, and the resulting data show us how far off the results may in fact be in other areas, when the wrong kinds of measures are used because there is nothing better. The Bureau of Public Roads has collected price quotations of

> . . . prices actually paid, that is, the successful bids, ordinarily the low bid, at which federally aided road construction has been undertaken in the states. As an index of prices actually paid through the mechanism of competitive bidding, it can thus reflect all those changes in the components of price that may result from changing market conditions: discounts and premiums with respect to list prices of materials; *changes in labor productivity;* and changes in contractors' margins over costs.
>
> . . . [The resulting index constructed from these prices] employs fixed base period quantity weights, the latter referring to quantities required for the construction of one mile of road of "standard" quality in the base period, 1925–29.[10] The index consists of three components: common excavation, concrete paving and structures (bridges, underpasses, etc.), separate subindexes for which are published along with the overall index.[11] Component prices are expressed in terms of commonly employed physical units such as cubic yards of excavation and square yards of paving. Structures, an obviously heterogeneous item in contrast with the other components, are represented by the cost of three elements: reinforcing steel (per pound), structural steel (per pound) and structural concrete (per cubic yard). The price quotations refer to the unit prices charged by the low bidders to put the particular material in place or to perform the particular operation, that is, they include charges for materials, labor, overhead, and profit . . .[12]

This index and its components are available on an annual basis since 1924 and quarterly from 1931 on.

Table 1 records the postwar trend in this index and its components and compares it with some relevant alternatives. Over the whole

[10] Since this was written a revised version of this index using 1957–59 weights has been published by Stern [18]. The revision did not greatly affect the index.

[11] Bituminous paving has been added into the index in the above-mentioned revision.

[12] From Foss [4, pp. 375–376] (emphasis and footnotes supplied).

TABLE 1

BID-PRICES HIGHWAY CONSTRUCTION PRICE INDEX AND COMPONENTS, WITH COMPARISONS, 1957–61
(1947–49 = 100)

Title and Source	1957	1958	1959	1961[a] (first quarter)
1. Total "composite mile" (BPR)	118	116	114	110
2. Composite construction costs (Commerce)	137	138	141	144
3. Construction machinery and equipment prices (WPI)	160	166	172	178
4. Average hourly gross earnings: nonbuilding contract construction (BLS)	162	167	172	185
5. Common excavation (cu. yd.) (BPR)	102	101	98	97
6. Concrete pavement (sq. yd.) (BPR)	123	123	122	116
7. Power cranes, shovels, and draglines (WPI)	160	164	169	173
8. Mixers, pavers, and spreaders (WPI)	143	150	156	162
9. Concrete ingredients (WPI)	136	139	140	142
10. Structural steel (lbs.) (BPR)	150	133	127	117
11. Total "structures" (BPR)	127	120	117	113
12. Structural steel shapes (WPI)	192	195	200	200
13. Service buildings and other structures deflator (USDA)	125	131	135	n. a.

n. a. = not available.

SOURCE: Lines 1, 5, 6, 10, and 11: *Price Trends for Federal-Aid Highway Construction*, Bureau of Public Roads, second quarter 1961; lines 2 and 4: [27] and *Survey of Current Business*, Department of Commerce; lines 3, 7, 8, 9, and 12: [28, 1959, Bull. No. 1295, and June 1961]; line 13, unpublished Department of Agriculture figures.

[a] BPR data (lines 1, 5, 6, 10, and 11) refer to first quarter. All other data are as of February 1961.

period, the BPR index rose much less than any of the other construction cost indexes, and it has trended downward since 1957 while all the other indexes have kept on rising. Table 1 also compares the price paid per "cubic yard excavated" (line 5) with the price of excavating machinery (line 7) and the wage of construction labor (line 4). Since these and similar "input" prices have risen much more than the price of "output," either the quality of construction machinery or the quality of construction labor (or other aspects of productivity)

has risen very substantially over this same period.[13] This, by itself, throws some doubt on the "quality" of the construction machinery price indexes. Similar comparisons can be made for concrete pavement and concrete ingredients, for structural steel (put in place) and structural steel shapes, and for other components of this index. It is doubtful that these apparent productivity increases have been restricted to the road-building sector of the construction industry. It is my conjecture that this index is also closer to the "truth" for total construction than any of the other available alternatives. If this conjecture is true, we have seriously underestimated the growth in the structures components of our capital stock. Moreover, as Foss [4] has shown, we have also seriously underestimated both the flexibility and the variability of the prices paid for the output of the construction industry.

Among the few other series of prices per unit of closely specified service known to me are the USDA collected figures on "Average rates paid for hand harvesting 100 pounds of seed cotton," available by states since 1924. Table 2 presents a comparison of these series with the official farm wage indexes for two important cotton states, Mississippi and Texas, since 1947–49. The rates paid for picking cotton show a trend downward in the post-World War II period. The divergence between the cotton-picking rate and the average wage rates as reported by the USDA is particularly clear in the post-Korean War period. Also interesting is the substantial fall in the cotton-picking rate in Texas relative to that in Mississippi, reflecting the increasingly elastic supply of Mexican nationals to Texas agriculture. An examination of other changes that might have affected

[13] It can be shown, e.g., see Siegel [16], that the ratio of Laspeyres indexes of prices paid to prices received for an industry is equal to a Paasche total factor productivity index for this same industry. Kutscher and Waite [11] provide a breakdown of highway construction expenditures on various materials, on site wages, and equipment. By allocating "other" expenditures (13.5 per cent of the total) proportionally to the specified inputs, these figures can be used as weights for computing a highway construction *input* price index from Wholesale Price Index component price indexes for construction equipment and materials and BLS figures on the hourly earnings of construction workers. On a 1957 = 100 base, the resulting input price index stood at 66.2 in 1947–49 and 107.0 in 1961. At the same time and on the same 1957 base, the highway construction product price index stood at 84.7 in 1947–49 and 93.2 in 1961, implying jointly a total factor productivity index for this industry of 78.2, 100.0, 114.8, in 1947–49, 1957, and 1961, respectively, and an estimated 57 per cent rise in total factor productivity in the highway construction industry since 1947–49. Thus, without using capital stock figures an index of total factor productivity can be estimated for an industry which even the encyclopedic work of Kendrick [10] left uncovered.

these series differently in different states, such as the declining relative importance of cotton, the recent increases in mechanical harvesting, and changing yields per acre, lead me to believe that the cotton-picking-rate series probably represents rather well the marginal cost of additional low-skilled labor of constant quality to the

TABLE 2

COMPARISON OF COMPOSITE FARM WAGE RATE AND AVERAGE RATE FOR PICKING 100 POUNDS OF SEED COTTON, MISSISSIPPI AND TEXAS, 1954 AND 1961 (indexes, 1947–49 = 100)

	1954	1961[a]
Mississippi		
Cotton-picking rate	96	96
Average composite wage rate	107	117
Texas		
Cotton-picking rate	95	88
Average composite wage rate	119	139

SOURCE: [21, various issues] [24] [20].
[a] As of October–November 1961.

agriculture of these states. These series may reflect better than the average wage rates, which among other measurement problems are also affected by the changing skill mix of the agricultural labor force, the price of constant-quality labor. As in the case for the construction bid series, these series also exhibit a substantially larger annual variability than the alternative average farm wage series.

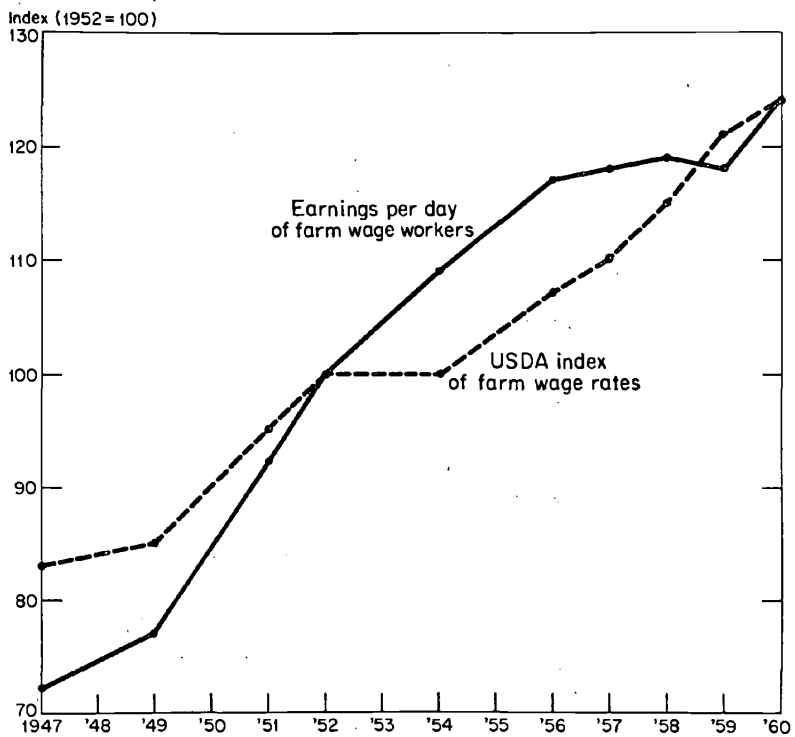
Collecting Data from the Other Side of the Market

Most of the price data used are employed to explain the behavior of buyers. Almost all of it, however, is collected from sellers. Both the experience of the Market Research Corporation of America and that of Michigan State University indicate that it is possible to collect price data using a "panel" of consumers. The USDA has used some of these panel data in its citrus fruit statistics; but, in general, little has been done to investigate whether in fact the prices as collected from different sides of the market differ, how they differ, and why.

Similarly, most earnings data come from employers rather than employees. In a few areas where we have some data, it appears that these two groups may see or at least report the same transaction in

quite different terms.[14] Chart 1 plots the official USDA index of farm wage rates (based on employers' and other "informed" persons' reports on the "prevailing" wage in their community) and an index of

CHART 1



average earnings per day of farm work collected from the household side by the Current Population and Labor Force sample for the USDA. The two series have moved rather differently over time, with the employee-reported series according perhaps somewhat better

[14] In 1945, for example, farm wage workers reported receiving on the average $44 as noncash wages (perquisites) during the year. There were 3.2 million such workers, resulting in an estimate of $141 million total noncash wages *received* by hired farm workers. In the same year, the USDA estimated on the basis of farm operator reports that noncash wages *paid* amounted to $347 million, or more than twice as much as was "received." Similarly, in the same year cash wages paid to farm workers were estimated at $1,358 million from a sample of farm workers and at $1,839 from a sample of farm operators. This type of difference has persisted. In 1959, farm wage workers reported receiving a total of $1,876 million from cash farm wages, while the USDA estimated that farm operators paid out $2,523 million in cash farm wages during this same year (see [19] [25] [26]). All of this, of course, *could* be explained away by the assumption that employers are a much better and more reliable source of data than employees, but it still leaves me wondering what actually is going on here.

with other information and my own impressions of what has been happening in this market.

In general, I would suggest that where possible the data be collected from the particular side of the market whose behavior is to be explained and in terms that are most directly relevant to the behavior phenomena that are to be studied. This, of course, is unlikely to lead to all-purpose numbers.

An Approach to the Measurement of Quality Change

While there is not much one can do about truly new products, and while some quality and style changes can be adequately taken care of by "linking," most of the quality changes are gradual and are not priced separately. In some areas, last year's models are not available any more on the market, and, hence, it is impossible (even if it were desirable) to use linking procedures to adjust for these changes. I have recently resurrected an old suggestion for dealing with this type of problem and have presented computations indicating that the suggested method may be both feasible and powerful.[15] In essence, it consists of viewing a commodity as a bundle of qualities, each one of which contributes (positively or negatively) to the utility or productivity derived from the commodity in question, with many or most of these dimensions or qualities quantifiable. Moreover, since at any point of time it may be possible to observe different "quality" combinations selling at different prices, one may be able to estimate (impute) the price (value) of these dimensions at the margin. One way of doing this is through cross-section price specifications regressions. Armed with these estimated "prices" of particular qualities, which may not remain constant over time, we can adjust the price of the total bundle for changes in the level of different qualities, either by adjusting the bundle at current quality prices to base period quality levels, or by valuing the change in the various dimensions by base period dimension prices. In principle, we are doing nothing more, except in a more complicated fashion, than discovering that prices of different size rugs can be compared, since apparently there is a fairly constant relationship between the size of a rug measured in square yards and its price, holding other variables constant. Armed with the computed price per square yard, price indexes can be computed, even though the previously priced 6 x 9 rugs have been replaced in the market by 8 x 9's and 6 x 8's.

[15] See Griliches [9] and Adelman and Griliches [1], and the literature cited there.

Tables 3 through 8 illustrate the possibility of and the problems associated with making such quality-change adjustments for United States passenger cars in the postwar period and the probable magnitude of the resulting adjustments. Table 3 illustrates the relationship between list prices of automobiles and several specification variables in selected years. The regressions for 1952 and 1961 had not been previously presented or computed and represent, in a sense, a successful test of the model. Because of multicollinearity of the various specifications the estimated quality prices, or weights, are not very stable from year to year. I have, therefore, preferred to use combined two-year regressions to estimate the weights to be used in the

TABLE 3

SINGLE-YEAR CROSS-SECTION REGRESSIONS, RELATING THE LOGARITHM OF NEW U.S. PASSENGER CAR PRICES TO VARIOUS SPECIFICATIONS, 1950–61

Model Year	*N*	*Constant*	*Coefficients of*				R^2
			H	*W*	*L*	*V*	
1950	72	1.2709	.158 (.048)	.048 (.029)	.832 (.115)	−.024 (.014)	.892
1952	51	1.7174	.097 (.042)	.105 (.030)	.578 (.127)	−.020 (.015)	.927
1957	95	2.7370	.051 (.013)	.059 (.017)	.171 (.057)	−0.11[a] (.010)	.967
1961	99	2.2530	.026 (.011)	.132 (.017)	.309 (.080)	−.011[b] (.012)	.940

NOTE: Dependent variable is the logarithm (to the base 10) of "list" (advertised delivered) price. To convert the results to natural logarithms multiply all the coefficients by 2.3. The resulting coefficient, if multiplied by 100, would measure the *percentage* impact on price of a *unit* change in a particular specification or "quality," holding the other specifications constant. Figures in parentheses are standard errors.

H = advertised brake horsepower in 100's
W = shipping weight in hundreds of pounds
L = over-all length, in hundreds of inches
V = 1 if the car has a V-8 engine; = 0 if it has a six-cylinder or smaller engine
T = 1 if car is a hardtop; = 0 if not
A = 1 if automatic transmission is "standard" equipment (i.e., is included in the price); = 0 if not
P = 1 if power steering is standard; = 0 if not
B = 1 if power brakes are standard; = 0 if not
C = 1 if the car is designated as a "compact"; = 0 if not

[a] Plus significant coefficients for *T*, *A*, *P*, and *B*.
[b] Plus *T*, *P*, and *C*.

TABLE 4

U.S. PASSENGER CARS: REGRESSIONS OF LOGARITHM OF PRICE ON SELECTED SPECIFICATIONS, TWO-YEAR CROSS-SECTIONS, 1947–48 THROUGH 1960–61

Model Years	N	Constant	H	W	L	V	D	R^2
1947–48	110	2.3854	.184 (.067)	.182 (.028)	.014 (.047)	−.047 (.016)	.0558 (.0119)	.761
1948–49	100	2.4813	.182 (.073)	.159 (.037)	.032 (.048)	−.038 (.019)	.0540 (.0138)	.736
1949–50	114	1.3494	.167 (.047)	.029 (.029)	.831 (.110)	−.014 (.014)	−.0217 (.0087)	.851
1950–51	127	1.3522	.136 (.036)	.026 (.021)	.839 (.082)	−.002 (.010)	−.0092 (.0070)	.893
1951–52	106	1.5283	.102 (.034)	.061 (.021)	.709 (.087)	−.001 (.010)	.0608 (.0075)	.920
1952–53	105	1.8785	.105 (.031)	.108 (.024)	.493 (.092)	−.027 (.012)	−.0012 (.0089)	.904
1953–54	119	2.1737	.114 (.028)	.087 (.029)	.379 (.090)	−.027 (.012)	−.0227 (.0096)	.855
1954–55	121	2.4713	.105 (.026)	.004 (.026)	.358 (.067)	−.013[a] (.010)	−.0403 (.0087)	.904
1955–56	143	2.4615	.040 (.024)	.105 (.024)	.232 (.067)	−.019[b] (.013)	.0085 (.0077)	.924
1956–57	186	2.5868	.041 (.012)	.092 (.017)	.195 (.048)	−.016[b] (.009)	.0116 (.0046)	.947
1957–58	202	2.8814	.017 (.011)	.118 (.016)	.030 (.056)	.002[b] (.011)	.0116 (.0052)	.929
1958–59	191	3.0820	.027 (.011)	.124 (.015)	−.076 (.057)	−.011[b] (.014)	.0024 (.0061)	.915
1959–60	165	3.0610	.050 (.008)	.092 (.012)	−.025 (.047)	−.026[c] (.010)	−.0010 (.0005)	.943
1960–61	177	2.7311	.033 (.008)	.107 (.014)	.120 (.058)	−.013[d] (.009)	−.0157 (.0049)	.929

NOTE: See Note to Table 3. In addition:

M = 1 if car has aluminum engine; = 0 if not
D = 1 in the second of the two years; = 0 in the first

[a] Plus coefficients for A, P, and B.
[b] Plus coefficients for T, A, P, and B.
[c] Plus coefficients for T, A, P, and C.
[d] Plus coefficients for T, P, C, and M.

TABLE 5

ESTIMATED QUALITY WEIGHTS, OR "PRICES": PERCENTAGE CHANGE IN PRICE OF CARS AS RESULT OF UNIT CHANGE IN SELECTED "QUALITIES," 1937–61

	Percentage Change in Price per		
	10-Unit Change in Horsepower	100-Pound Change in Weight	One-Inch Change in Length[a]
1937–39[b]	7.1	3.0	0.15
1947	5.1	3.1	0.69
1950	3.6	1.1	1.92
1952	2.1	3.0	1.34
1957	1.2	1.4	0.39
1959	1.2	2.4	−0.16
1960	1.2	1.4	0.15
1961	0.6	3.0	0.71

[a] Wheelbase length, 1937–39; over-all length thereafter.
[b] From Court [2, p. 111].

adjustment for specification changes between these two years. Table 4 presents the combined two-year regressions for 1947–61. Again, the results for 1947–54 and 1960–61 represent a successful extension of the model into previously unexplored periods.[16] Table 5 summarizes the trend in the estimated prices of the various automobile qualities or specifications. It does make clear that substantial changes have occurred in these prices over time.

Since the CPI has priced only the "low-priced three" cars in its index (until 1961), I shall present and describe the construction of comparable quality indexes for these cars. The quality changes evaluated are changes in horsepower, weight, and length. A time series of these specifications for the low-priced three is presented in Table 6. None of these changes has apparently been adjusted for or "linked out" in the construction of the CPI.[17] Using weights derived from Tables 3 and 4, it is possible to estimate for each pair of years the change in price that is due to changes in specifications. The exact magnitude of these estimates depends on the particular set of weights used. Table 7 presents such annual estimates based on two sets of weights:

[16] The coefficient of D (the year dummy variable) in these regressions, if multiplied by 2.3×100, becomes itself an estimate of a price index. It is an estimate of the percentage change in the average *unweighted* list price of cars in the sample between the two years *holding the various specification variables constant.*

[17] See, for example, the discussion of linking in Larsgaard and Mack [12].

TABLE 6

SPECIFICATIONS AND LIST PRICES OF AN AVERAGE[a] "LOW-PRICED THREE" CAR, 1939–61

	Horsepower	Weight (Pounds)	Over-all Length	Price[b]	N[c]
		SIX-CYLINDER ENGINES			
1939	80	2784	189.7	720	6
1947	92	3166	197.4	1238	6
1948	92	3134	197.6	1424	6
1948	91	3152	197.8	1363	4
1949	94	3034	195.5	1522	4
1950	95	3075	195.9	1508	4
1950	95	3081	195.6	1518	6
1951	103	3303	199.3	1779	6
1952	97	3106	194.6	1739	6
1953	103	3115	194.3	1751	6
1953	104	3138	194.5	1768	7
1954	111	3147	195.5	1781	7
1955	120	3129	198.7	1839	7
1956	135	3172	199.7	1938	7
		V-8 ENGINES			
1955	163	3185	198.7	1939	7
1956	176	3246	199.7	2039	7
1957	184	3354	203.6	2240	7
1958	210	3440	206.6	2390	7
1959	202	3525	209.6	2533	7
1960	190	3615	211.5	2537	7
1961	187	3566	209.6	2542	7

[a] Average for three Chevrolet, three Ford, and two (the two lower-priced series) Plymouth models since 1953. The 1939 sample consists of two Chevrolets, two Plymouths, and two eight-cylinder Fords. The 1947–48 samples are the same as the 1939 one except that the Fords are "sixes." The eight-cylinder Fords in 1939 were included to raise the sample size to approximately the same levels as in the subsequent years. Since these eights (not V-8's) had a lower list price than comparable sixes in 1939, their inclusion will, if anything, bias the quality indexes downward. The 1948–50 samples consist of one Chevrolet, two Fords, and one Plymouth. The 1950–53 samples are back to two Chevrolets, two Fords, and two Plymouths.

[b] Arithmetic average.

[c] Number of models in average.

1. Fixed 1950 quality prices. This is comparable to the rest of the CPI, which is based on 1950 weights throughout.
2. Adjacent years weights, in which a specification change from, say, 1952 to 1953 is valued at specification prices derived from a price specification regression using data for both years jointly.

TABLE 7

QUALITY INDEXES FOR THE LOW-PRICED THREE CARS, 1947–60
(six-cylinder engines to 1956, V-8's thereafter)

	Estimated Percentage Change in Price Due to Changed Specifications	
	1950 Weights[a]	Adjacent-Year Weights[b]
1947–48	0.0	−1.6
1948–49	−4.6	−3.3
1949–50	1.6	1.2
1950–51	12.4	10.8
1951–52	−13.4	−11.7
1952–53	1.7	1.5
1953–54	4.5	3.0
1954–55	9.3	5.7
1955–56	8.1	2.9
1956–57	12.4	4.8
1957–58	16.9	3.4
1958–59	4.3	1.4
1959–60	0.6	0.3
1960–61	−5.3	−2.0
1939–47[c]	23.3	
1947–52[c]	−4.0	−5.8
1952–60[c]	57.8	25.2

[a] For all comparisons, 1950 weights used. For example, the 1937–50 figure is arrived at by multiplying the *change* in the average specifications given in Table 6, by the 1950 weights given in Table 4 and adding them together.

[b] Weights from Table 4. I.e., the 1954–55 comparison uses average 1954–55 weights, and so on.

[c] Derived for the second column by adding 100 to each of the relevant observations, multiplying, and subtracting 100. For the first column, where a fixed system of weights was used, derived by adding up the appropriate percentage changes. This is done because the underlying equation says, e.g., that a 100-pound increase in weight leads to a 2.5 per cent increase in price; and 200 pounds, to a 5 per cent increase in price rather than the 5.1 per cent that one would get by multiplying the two changes. The resulting index is lower than it would be if it were computed from annual links.

This last, and to my mind preferable, method allows the weights to shift smoothly over time, employing essentially a two-year moving average of current-specification prices to weight changes in these specifications. In Table 8 these estimated quality changes are cumulated, expressed as an index to the 1947–49 base, and used to deflate the new-automobile prices component index of the CPI.

The resulting indexes, together with the undeflated CPI, are graphed in Chart 2. While the price index deflated by the 1950 weighted quality index drops substantially more than the index based

TABLE 8

PRICE AND QUALITY INDEXES FOR THE LOW-PRICED THREE CARS, 1939–61
(1947–49 = 100)

Model Year	Quality Indexes[a]		New-Automobile Component of CPI[b]	CPI Deflated by Quality Indexes	
	G_{50}	G_A		CPI / G_{50}	CPI / G_A
1939	78.5		57.1	72.7	
1947	101.5	102.2	88.8	87.5	86.9
1948	101.5	100.6	95.3	93.9	94.7
1949	97.0	97.3	108.8	112.2	111.8
1950	98.6	98.4	109.9	111.5	111.7
1951	110.8	109.0	113.1	102.1	103.8
1952	97.6	96.3	124.9	128.0	129.7
1953	99.2	97.7	126.5	127.5	129.5
1954	103.7	100.6	129.7	125.1	128.9
1955	112.8	112.5	127.5	113.0	113.3
1956	120.8	115.8	126.4	104.6	109.2
1957	133.0	121.3	132.8	99.8	109.5
1958	149.7	125.5	138.4	92.5	110.3
1959	153.9	127.2	144.2	93.7	113.4
1960	154.5	127.6	144.3	93.4	113.1
1961	149.3	125.0	139.1	93.2	111.3

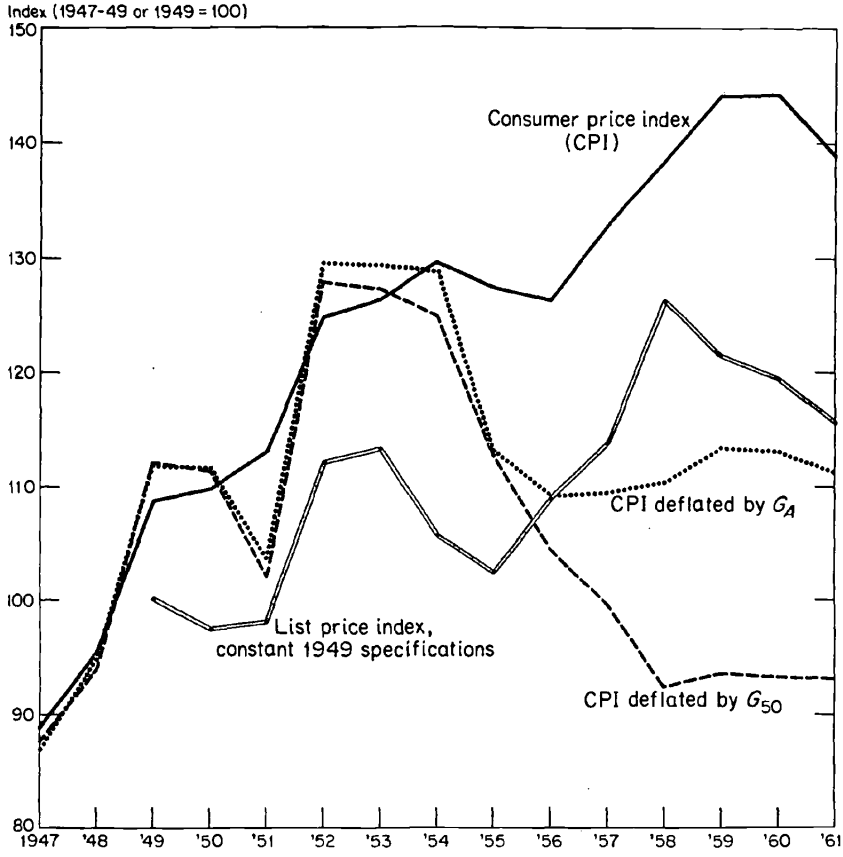
CPI = Consumer Price Index.

[a] Computed from Table 7. G_{50} is quality index based on fixed 1950 weights. The numbers in the first column of Table 7 were treated as index number points of an index (1950 = 100) added together (linearly) and translated to a 1947–49 = 100 base. G_A is a quality index based on adjacent-year weights, linked together (by multiplying through, assuming 1948 = 100) and translated to a 1947–49 = 100 base.

[b] From [29] and various CPI releases. For 1939 and 1947–53, as of March of the same year; for 1954, as of January 1954; for subsequent years, as of November of the preceding year.

on changing adjacent-year quality weights, even the latter index is still in the neighborhood of its 1949–50 values in 1961, indicating little or no rise in the "real" price of new automobiles. The quality-adjusted indexes are also somewhat more variable than the CPI, dropping sharply from 1950 to 1951 and rising much more abruptly from 1951 to 1952. Apparently, the 1951 pre-Korean-design models represented the manufacturers' response to the running out of the postwar demand backlog by 1949–50 and were the first substantial model changes in the postwar period. But then came Korea, price ceilings, and steel shortages, and by 1952 the manufacturers retrenched sharply, without dropping list prices, by cutting down on the power and size of their automobiles. The horsepower and size

CHART 2



"race," which was about to begin in 1951, was interrupted by the increased demand and price ceilings generated by the Korean conflict and did not start again in earnest until after 1954. Whether the above is either an accurate or a useful interpretation of history is a question for industry specialists. I want only to suggest here that quality-adjusted price indexes may in fact be helpful and do better in explaining to us what happened than indexes that do not take such changes into account.

Before leaving this subject, I would like to mention an alternative way of computing quality-adjusted price indexes. What I have done above is to take each change in specifications and value it by estimated "prices," which either do not change or change slowly and smoothly over time. The result is a chain-link quality index with changing weights, which is eventually used to deflate a price index.

Alternatively, we could have asked the question directly: What is the price of a particular fixed bundle of qualities, say, those of 1949 models, in each of the subsequent years, allowing both the general price of the commodity and the implicit "prices" of the various specifications to change? The answer to this question can be had directly from the estimated annual regressions by inserting the appropriate fixed specification levels and reading off the predicted (interpolated) price for these base-period quantities at current prices. This type of calculation is described elsewhere and will not be reproduced here, but I have also graphed on Chart 2 an index of list prices (1949 = 100) that uses "constant" average 1949 specifications for all cars (not just the "low-priced three") and is based on the annual list-price regressions subsequently.[18] This index tells a similar story, though it rises more in the 1954–58 period. It is biased upward, however, since it does not take into account either the growth in discounting or the increased number of attachments, such as directional signals and electric windshield wipers, included in the list price. But it does represent the most desirable way of computing such an index if annual cross sections of actual transaction prices were available.

A Critique of a Criticism

. . . I believe these challenges are conceptually wrong; they rest on the assumption that intangible quality improvements can be brought into the sphere of quantitative measurement. In the end, they would make it impossible to construct measures of output and price changes that are useful to the study of economic growth.

. . . economic welfare as a measurable idea must be restricted to telling us if we are better off only by our having more goods. Any broader idea of welfare which would take account of the character of the goods available, or the satisfaction they give, may be a perfectly valid subject for speculative appraisal, but it is not measurable.

. . . our units of measurement are fixed transactions because they are the only measurable units.

. . . the advance of medical practice cannot be allowed for. If this makes it possible for a patient to be cured in half the hospital time, it would be just nonsense to say that production has remained constant while prices have fallen by 50 per cent; we must record that production has declined

[18] For details of these computations see Fisher, Griliches, and Kaysen [3].

while prices have remained constant. Of course, the patient is better off in a very real sense because he needs less hospital care. However, there is no way to measure his needs in a production index; what must be measured are the goods he buys in response to those needs.[19]

Stated in so bold a fashion these criticisms almost answer themselves. As far as I can see, Gilbert's objections to quality change measurement can be summarized as follows: (1) quality cannot be measured; (2) only "goods" and "fixed" transactions can or should be measured; (3) changes that are measured should be measured by cost and not by value to purchaser; and (4) welfare measures should be based on "goods" only. I shall try to discuss these point by point, though they are all interrelated.

The first assertion is either a tautology, achieved by adding the adjective "intangible" (unmeasurable?) to the concept of quality change, or a statement about the empirical impossibility of getting "interesting," "useful," or "good" measures of quality change. Philosophically speaking, it is hard to conceive of any phenomenon that is *in principle* unmeasurable. The simple act of "naming" or "defining" a phenomenon has the seeds of measurability in it. Moreover, history is littered with the remains of impossibility assertions. Quality changes are being measured. Perhaps they are being measured wrongly or, more likely, not well enough, and the proposed measures may thus deserve criticism on this score, but to assert that what is being done cannot be done is not particularly helpful.

The second point rests on the assumption that the relevant notions of "goods" and "fixed transactions" can be defined in the abstract, without recourse to "subjective" concepts of productivity, utility, or welfare. But the economist or statistician does not have to restrict himself to the particular form in which the data come to him. For example, a farmer may be buying "all-purpose" fertilizer in 100-pound sacks and paying a price that is quoted "per sack." Both we and the more intelligent farmer know that what he is interested in is not total poundage but the "plant nutrient content" of it. From this it is only one step, and a feasible one at that, to translate the "fixed per sack transaction" into prices paid per plant nutrient unit and perhaps to complicate it further by distinguishing different plant nutrients (nitrogen, phosphoric acid, and potash) and assigning different units to each one of these.[20] It may be necessary to make some

[19] Gilbert [6].

[20] I have an example of this in [7].

imputations, but this is no reason for shying away from such a procedure—the national accounts are full of imputations: the contribution of home ownership is imputed, the value of home consumption of farm products is imputed, and the various interpolations used in constructing series from scraps of data are also "imputations".

The choice of a transaction unit is not obvious. If I have appendicitis, my family tries to contract for a "cure." Neither the number of the doctor's visits, nor the amount of drugs used, nor the length of my stay in the hospital are under my control. They are decided by the doctor (jointly with my family) on the basis of the quantity needed to "cure" me. The cost of a *successful* appendectomy consists of the direct monetary cost of room and board, the cost of the various drugs required in conjunction with the appendectomy, physicians' and nurses' services, the indirect monetary costs of income foregone while incapacitated by the illness, and the more subjective costs of pain and of the utility or disutility of getting away from the family for a while.[21] Both the direct and indirect monetary costs are "measurable"; and if there are improvements in medical efficiency, I see no reason why they should not show up as a reduction in the "real price per unit of service" (appendicitis cure) to consumers. After all, technical advances in other areas do show up as price reductions to consumers. Why should not the same be true of medical services? Or do we want to argue that there have not been any advances in the productivity of the "health-maintaining" industry?

A similar type of "what *is* the transaction" example brings us directly to the third point: what should quality change be measured by—"cost" or "value"? The dosage of the new birth control pills (Enovid) has been recently cut in half, reducing thereby the price of this contraceptive method by half. This came about as the result of additional research which showed that half of the previously recommended dose is really enough to achieve the desired result. What is being bought here is a method of contraception. It comes in packages of twenty pink pills to be used during each menstrual cycle. The same method of contraception is now available at half the price of a year ago. True, each of the pills is somewhat smaller today, but what does that matter if the user is assured that they are equally effective for her purposes? After all, the doctor prescribed "twenty pills a month" before and is prescribing "twenty pills a month" now

[21] The cost of *unsuccessful* appendectomies is somewhat harder to estimate, but it could probably also be quantified. On this, see Weisbrod [31].

for exactly the same purpose. Why is this not a decline in price? True, this will lead to some trouble in our measure of this industry's output. It is producing now an equal or larger number of smaller pills than before. How we should treat this change depends on our definition of "productivity." I would choose a measure that showed no decline in output, since in this way output would be defined in units comparable to the "market" for it, and such a definition would show a substantial increase in the productivity (the satisfaction of a given set of wants with a smaller use of resources) of this industry. In fact, this is a rare actual example of the "pure-knowledge" no-increase-in-costs type of technological advance which crowds our textbooks. Since it falls so well within the usual definition of productivity, it should be measured as such.

By now it should be clear that "goods" do not mean much independently of a welfare or utility calculus. Nor does it make sense to restrict welfare and "economic growth" measures to goods-only-based output measures. An increase in aspirin consumption may indicate that we have more headaches, but this is not an obvious measure of "*well*-fare."

A large part of the disagreement, however, is purely semantic. There is no real dispute that quality changes affect both behavior and welfare. If they could be measured, they *might* help us toward better explanations of consumer and producer behavior and better measures of economic growth. The argument really narrows down to whether they are measurable, how good the proposed measures are (imprecise measures may be better than none), and finally, and most controversially, whether these aspects should be incorporated into the various official price and output measures. The last problem *is* a semantic one. As long as all aspects of economic change are adequately measured and reported, the question of which of them should be combined into one over-all, all-purpose measure is not a "scientific" one. It is probably possible to *define* aggregate price and output indexes so that they would not include these types of changes. Personally, I doubt whether the resulting indexes would be very interesting or useful. Be that as it may, the thing that I am trying to get across is that these type of changes are important, that if we have good measures of them they would prove useful, that some measures may be feasible, and, therefore, that we should try to measure them as well as we can. Whether the resulting measures should be incorporated into any one particular index is a secondary matter. The

important thing is to measure these changes, since they may represent the essence of economic progress.

References

[1] ADELMAN, IRMA, and GRILICHES, ZVI, "On An Index of Quality Change," *Journal of the American Statistical Association*, September 1961.

[2] COURT, A. T., "Hedonic Price Indexes with Automotive Examples," *The Dynamics of Automobile Demand*, New York, General Motors Corp., 1939.

[3] FISHER, F. M., GRILICHES, ZVI, and KAYSEN, C., "The Costs of Automobile Model Changes Since 1949," *Journal of Political Economy*, October 1962 (abstract in *American Economic Review*, May 1962).

[4] FOSS, MURRAY F., "How Rigid Are Construction Costs During Recessions?," *Journal of Business*, July 1961.

[5] FRIEDMAN, MILTON, *A Theory of the Consumption Function*, Princeton for NBER, 1957.

[6] GILBERT, MILTON, "The Problem of Quality Changes and Index Numbers," *Monthly Labor Review*, September 1961, reprinted from *Economic Development and Cultural Change*, April 1961.

[7] GRILICHES, ZVI, "Measuring Inputs in Agriculture: A Critical Survey," *Journal of Farm Economics*, December 1960.

[8] ———, "Capital Stock in Investment Functions: Some Problems of Concept and Measurement," in Don Patinkin (ed.), *Measurement in Economics* (Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld), Stanford, California, 1963.

[9] ———, "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change," *The Price Statistics of the Federal Government*, New York, NBER, 1961.

[10] KENDRICK, JOHN W., *Productivity Trends in the United States*, Princeton for NBER, 1961.

[11] KUTSCHER, R. E., and WAITE, C. E., "Labor Requirements for Highway Construction," *Monthly Labor Review*, August 1961.

[12] LARSGAARD, OLGA A., and MACK, LOUISE J., "Compact Cars in the Consumer Price Index," *Monthly Labor Review*, May 1961.

[13] MUTH, JOHN F., "Rational Expectations and the Theory of Price Movements," *Econometrica*, July 1961.

[14] PRICE STATISTICS REVIEW COMMITTEE, *The Price Statistics of the Federal Government*, New York, NBER, 1961.

[15] NERLOVE, MARC, *Distributed Lags and Demand Analysis for Agricultural and Other Commodities*, Department of Agriculture, Ag. Handbook No. 141, June 1958.

[16] SIEGEL, IRVING H., "Concepts and Measurement of Production and Productivity," Bureau of Labor Statistics, March 1952 (mimeographed).
[17] STAUBER, B. R., *et al.*, "The January 1959 Revision of the Price Indexes," *Agricultural Economics Research*, April–July 1959.
[18] STERN, E. L., "A New Base for the Highway Construction Bid Price Index Compiled by the Bureau of Public Roads," *Public Roads*, October 1961.
[19] U.S. DEPARTMENT OF AGRICULTURE, BUREAU OF AGRICULTURAL ECONOMICS, *Employment and Wages of the Hired Farm Working Force in 1945*, June 1946.
[20] ———, *Farm Wage Rates by States, Revised, 1910–1948*, January 1951.
[21] ———, *Farm Labor*, various issues.
[22] ———, *Agricultural Prices*, various issues.
[23] ———, *Major Statistical Series of the U.S. Department of Agriculture*, Ag. Handbook No. 118, Vol. 1: *Agricultural Prices and Parity*, December 1957.
[24] ———, *Statistics on Cotton and Related Data*, SB 99, June 1951.
[25] ———, *The Farm Income Situation*, July 1961.
[26] ———, *The Hired Farm Working Force of 1959*, Ag. Information Bulletin 238, April 1961.
[27] U.S. DEPARTMENT OF COMMERCE, *Business Statistics*, 1961 edition.
[28] U.S. DEPARTMENT OF LABOR, BUREAU OF LABOR STATISTICS, *Wholesale Prices and Price Indexes*, various issues.
[29] ———, *Consumer Prices in the U.S., 1953–58*, Bull. No. 1256, 1959.
[30] ———, *Consumer Price Index*, various releases.
[31] WEISBROD, BURTON A., "The Valuation of Human Capital," *Journal of Political Economy*, October 1961.

COMMENT

GEORGE JASZI, Department of Commerce

Zvi Griliches touches upon several problems of widely different character.

I agree with his discussion of the well-advertised defects of the Department of Agriculture indexes, which proclaim it a virtue to count as price change, changes in unit values that are due to shifts among qualities. I share his uneasiness about the construction cost indexes, which, as we know, tend to measure prices of inputs rather than outputs. But I do not think that his discussion contributes much to diminishing our ignorance as to the quantitative importance of

the resulting bias. Griliches seems to imply that it is quite significant. I note that R. A. Gordon, in the December 1961 *American Economic Review*, has written in a manner that implies the opposite.

The problem of whether we should collect price information from buyers or from sellers is also an important one. But I am not quite sure whether I understand and agree with the way in which Griliches sorts out the issues that are involved. Does he imply that there are two valid prices in every transaction—a seller's price and a buyer's price—and that we should turn to sellers or to buyers for our information depending on whether we are interested in supply theory or in demand theory? I always thought that there was only one price to one transaction, and that the answer to the question of whether we should canvass sellers or buyers depended on who was more likely to give correct information. If this is the case, a great deal of detailed investigation is required if this question is to be dealt with fruitfully.

I share with Griliches part of his disagreement with Milton Gilbert. I think Gilbert is wrong in saying that our disabilities in measuring quality change do not lead to defects in our price and volume index numbers. I think they lead to serious defects; it would be very useful if quality change could be measured.

However, I do agree with Gilbert rather than with Griliches as to what can and what cannot be achieved in this field.

I shall spend the rest of my time in trying to explain why I am rather pessimistic about the progress that can be made. To do so I shall compare the approach to the measurement of quality change recently put forward in the United States by Griliches and by Richard Stone in England with the more conventional approach that preceded their contributions. I shall start by outlining the conventional method and explaining its limitations.

The Conventional Method

From the examination of extant price and volume indexes, it would appear at first that a bewildering variety of techniques is used to cope with the appearance of products of altered quality and of new products. On closer examination, however, this impression is strongly modified. The bulk of the procedures used reduce to one fundamental method, or at least to close variants of it, which I shall call the "conventional" method.

The essence of the conventional method is to translate quality into quantity by reference to market prices. If a new variety of a good is

introduced, one physical unit of the new good is not simply equated to one physical unit of the old good. Instead, one unit of the new good is regarded as equivalent to one unit of the old good times the ratio of the price of the new good to that of the old good in an overlap period. If such an overlap period does not actually exist, an estimate is made of what the relative cost of producing the two goods would have been had both been produced in a common period.

Note that my discussion proceeds in terms of the construction of real-volume measures. It could equally well be conducted in terms of the construction of price indexes: All the problems that arise are identical.

This procedure for introducing new goods is, it may be noted, quite in harmony with that adopted for volume measurement in the absence of quality change. More expensive grades of goods are given a greater weight in such measurement than less expensive varieties —in proportion to their relative market prices.

The fundamental shortcoming of this method is, of course, that quality change will show up as volume change only if, and to the extent that, it is reflected in the relative prices of the new and old variety. If, for instance, one unit of the new good does not cost more than one unit of the old good, it will be considered to represent the same physical volume even though it renders better services. Indeed, if the new good costs less than the old good, it will be considered to represent less real volume even though it is superior. These, I should add, are general propositions which would have to be modified to apply to some of the variants of the conventional method.

It has long been felt that in the presence of quality improvement this treatment in common sense understates the increase in the volume of production. One way of expressing this is to say that the additional services rendered by the new good must be more than its additional price; otherwise, it could not be introduced. A formulation which suggests more graphically the importance of what the conventional method misses has recently been put forward by Edward F. Denison. On the basis of certain simplifying assumptions, which I need not state here, it can be said that a comparison of today's output with that of an earlier period by means of the conventional method in fact measures the physical volume of the goods known in the earlier period that could have been produced today, using the resources and the improved technology, etc., that are employed today. In other words, what our measure misses is that we do not in

fact produce the outmoded products of the earlier year, but a superior collection instead. Introspection shows that this is a good deal to miss.

Although some have upheld the results obtained by the conventional method, this is not the generally accepted position. The limitations of that method have been widely admitted—even by those who see no good prospect of replacing it; and alternative techniques have been put forward with the claim that they embody more effective ways of dealing with quality change.

The Suggested Alternative

The alternative method that has received widest notice is that put forward by Stone and Griliches. It aspires to get at quality directly by deriving from a study of the past consumer response to old goods a measure of the consumer evaluation of the qualities embodied in the new good. The physical quantities represented by the new goods are then taken to be in proportion to these evaluations. More specifically, the new good is regarded as the result of the combination of two or more qualities that are associated with identifiable physical features—such as coffee bean and chicory content in the case of coffee—which have been embodied in the old products also, but not in the same proportions as in the new product. The consumers' separate evaluation of these several qualities is determined on the basis of his past behavior to the old products embodying these qualities; and the new product is valued by attaching to the quantity indicators of the several quality features which it embodies, the consumer evaluations that have been derived from this analysis. The possibility of such an analysis hinges on the existence in the past of goods embodying the characteristics in different proportions. It can then be carried through on the basis of multiple regression techniques in complex cases, and simpler ones in others.

To take a very simple case, assume that in period 1 consumers paid $2.00 for one pound of pure bean coffee and that they paid $1.50 for one pound of coffee containing half a pound of an admixture of chicory. In period 2 a new, intermediate brand containing one-quarter of a pound of chicory replaces the two old brands. According to the proposed method it will be valued at $1.75. This is an estimate, based on an analysis of consumer evaluations in period 1, of what consumers would have paid for the intermediate brand in period 1 had it been available in that period. As a result of the procedure, one

pound of the new coffee will be considered as equivalent to seven-eighths of a pound of the pure bean coffee, and seven-sixths pounds of the mixed coffee which it has replaced.

What are the limitations and advantages of this new method as compared with the conventional one?

It should be noted, in the first place, that it is obviously incapable —as incapable as the conventional method—of taking into account any quality change that introduces something that is really new. For it hinges on the assumption that the new differs from the old only by embodying in new proportions qualities that have existed before. This has been pointed out before.

But, secondly—and to my knowledge this point has not been made previously—the new method would in principle have to produce exactly the same results as the conventional method. This is so because the relative costs of the quality characteristics that are utilized in implementing the conventional method must be identical to the relative consumer evaluations on which the new method is based. In the coffee example just given, for instance, the conventional method of estimating the cost of the new, intermediate brand of coffee in a common period would produce a result identical to that of the new method, namely, $1.75.

The reason that this must be so is simple. The observed market price relations are compatible only with cost conditions in which the cost of one pound of pure bean coffee is $2.00 and that of one pound of chicory is $1.00. If these were not the respective costs, the prices could not be the observed market prices. But if the relative costs are as stated, the conventional method would calculate the price of the new intermediate brand at $1.75 also—three-fourths of a pound of pure bean coffee at $2.00 a pound plus one-fourth of a pound of chicory at $1.00 a pound.

This is the main point of my remarks: I think it is a mistaken idea that we can get at a better measure of quality by a study of consumer evaluations of the various features inherent in a good than by a comparison of costs. If we are confined to market data, we can measure these evaluations only to the extent that they are reflected in costs. This is all the market will reveal. Departures of relative evaluations from relative costs are prevented by adjustments in the quantities traded, and cannot show up in the observable value dimensions attached to the goods.

Future Work

If these conclusions are correct, two lines of investigation are indicated. First, we should explore the extent to which the results of the new method can be approximated by a careful, detailed application of the conventional method and the pros and cons of the two methods in statistical practice. Is it possible, on the one hand, that an application of the conventional method would be simpler and easier to evaluate step by step, as it were, than that of the multiple regression techniques usually employed in conjunction with the new method? Or, on the other hand, is it possible that the new method is a more effective way of implementing the old one in statistical practice?

Second, some consideration should be given to yet another approach, which is distinct from the two I have examined. The essence of this approach is that in the measurement of quality change it cuts loose entirely from valuations that can be derived from past cost relations or consumers' behavior. Instead, extraneous standards that seem reasonable and useful for particular purposes are introduced. For instance, calorie content may be taken as the basis of volume measurement in certain types of analysis of the volume of food consumption. The durability of automobile tires may be taken as the basic quantity dimension in some studies of transportation. It seems to me that this approach holds more promise than the two outlined before in some specific cases of applied economic analysis.

However, its limitations are severe, as further consideration of the two examples just given will show. In most studies of food consumption, it will be necessary to take account also of features other than the calorie dimension—for instance, the content of carbohydrates, fats, and proteins. Similarly, durability is not the only relevant dimension of a tire; safety and comfort in riding are also important. The difficulties involved in selecting the relevant quality characteristics, in finding good quantity indicators for them, and in assigning appropriate weights to these indicators tend to become unmanageable in most cases even of specific *ad hoc* analysis. Certainly, there is no chance at all that the method can be generalized to supersede the conventional method in the construction of over-all, all-purpose indexes of real volume.

EDWARD F. DENISON, Brookings Institution

The paper by Zvi Griliches is constructive and ingenious in criticizing available price series and suggesting alternative procedures. But Griliches does not distinguish the various issues sufficiently, and I fear this may contribute to confusion in an already confused field. Terminological differences lead him to erroneous criticism of an excellent article by Milton Gilbert.

I agree with George Jaszi's description of the conventional method of dealing with product changes and his statement that the procedures used by Griliches in his automobile calculations can only be construed as alternative statistical procedures for application of the conventional method. The method postulates that quantities of two products that sell for the same amount in a base period when both are in quantity production, or would do so if both *were* in quantity production, are the same quantity of product. (In competitive equilibrium this is equivalent to saying that quantities of two products that do, or would, use the same quantity of resources are the same amount of product, and this test is sometimes used where both products are not simultaneously produced.) The auto calculations by Griliches rely exclusively on relative prices set by sellers at points in time and consequently must conform to this standard. They give no consideration to the number of buyers who think it worthwhile to incur the additional cost of any particular new feature, let alone of any "consumer surplus" obtained by those who consider the feature more than worth its cost.

If one product partially or wholly displaces another over time it is likely that the resources devoted to its production better satisfy the wants of its buyers than would the same quantity of resources devoted to the displaced product.[1] If both products remain in production (like train and plane travel between Washington and New York) a welfare gain clearly arises because the consumer can buy the one he prefers, whereas he previously had only one choice. The fact that many travelers have shifted from train to plane indicates they prefer this alternative.[2] If the older product disappears the case is less

[1] This probability does not exist where there is a cycle in fashion, or other "change of tastes." In that case, consumers of each period prefer the products offered in that period, and there is no presumption that product change implies an increase in welfare.

[2] The fact that a shift takes place gradually rather than abruptly is sometimes construed as evidence of a change in tastes. On this interpretation, even displacement is not a valid indication of welfare gain. Aside from cycles in fashion, gradualness seems to me more often to reflect the time needed for information about new products to circulate and for adjustment in buying habits to occur.

definitive, since both before and after its introduction the consumer has only one choice, and some consumers would prefer the product that is not available. However, a total welfare gain is reasonably surmised on the grounds that production of the older product would have continued if many consumers preferred it to the new product, given relative prices of the two products corresponding to their relative production costs. Complete displacement thus indicates most consumers prefer the new product to the old. Displacement is perhaps the only acceptable indication of superiority, and the amount of displacement the only criterion by which to judge the number of consumers who prefer the new product. But it provides no measure of the value of the superiority and, hence, no useful criterion for price index construction. In any case, Griliches does not rely upon this test.

Griliches may say, with justice, that many price indexes do not in fact correspond well to the criteria of the conventional method. His calculations present a legitimate challenge to the CPI automobile index. Like almost everyone who has commented upon the construction cost indexes (and his comparisons here are of much interest) he stresses that they do not correspond at all to any reasonable concept of a price index. Many other price series are surely poor. If compilers of price indexes were clearly instructed to accept equivalent cost as a criterion the correspondence might improve. New techniques such as Griliches proposes could prove helpful.

However, with one exception, these techniques are not related to his criticism of Milton Gilbert. Gilbert's article stresses clearly and repeatedly, with examples, that when he says adjustment for quality change is not possible or desirable, he refers only to adjustment over and beyond that which results from the conventional method. In national income circles, at least, it is customary to use the term "quality change" in this restricted sense. If a price series fails to distinguish adequately between a more and a less costly article by simply pricing "cars" or "pills," this is considered a simple error in the price index rather than a failure to "adjust for quality." Gilbert obviously favors the correction of such errors. Once this is understood, and Jaszi's finding is accepted that the Stone-Griliches approach is irrelevant to measurement of the kind of quality change not caught by the conventional method, the most telling point that Griliches makes against Gilbert disappears. Gilbert is not asserting "that what is being done cannot be done"; what Gilbert says is not being done in fact is not being done.

Griliches' fertilizer example parallels Jaszi's coffee example; we need only substitute nitrogen, phosphoric acid, and potash for pure bean coffee and chicory. Gilbert certainly does not argue for counting bags of fertilizer. There is no issue here.

The appendicitis case that Griliches introduces under the category of the choice of the transaction unit is the classic example of quality change as the term is used by national income estimators. In indicating an adjustment can be made, and is called for, Griliches is truly in disagreement with Gilbert, although there is no dispute that the consumer may be better off with the new techniques.[3] This is the only part of his paper in which Griliches can claim he has a method for the kind of adjustment that Gilbert says cannot be made. But does he? Griliches cannot logically rest his case for quality improvement on the grounds that consumers get well more quickly and cheaply than before unless the whole concept of consumer choice underlying income and price measurement is changed. The evidence of superiority of new techniques seems to me rather to lie in the fact that consumption patterns have shifted. If consumers ignore the scientific evidence and shift, as they often do, to nostrums that are more costly and less effective, is there quality improvement or deterioration? If, as I suspect, Griliches' answer is improvement, his criteria for measuring its amount are not applicable; they will probably show deterioration. If the answer is deterioration, is not this a judgment based on expert rather than consumer opinion? And if his recourse is indeed to expert opinion, to how many products would he apply this standard?

Griliches says "the choice of transaction unit is not obvious," but then proceeds to specify one. I see no unique reason to agree with him that the patient really wants to buy a cure for appendicitis. Until the diagnostician told him he had appendicitis he had no such desire; when the patient entered the doctor's office he simply wanted a cure for what ailed him. But this is not satisfactory either; there is such a thing as preventive medicine, and the patient would be still better pleased if he had not become sick. We might more reasonably say the consumer's real desire is good health and long life. But then we are in worse trouble; the consumer may stubbornly

[3] See the quotation Griliches gives from Gilbert: "Any broader [than is provided by the conventional method] idea of welfare which would take account of the character of the goods available, or the satisfaction they give, may be a perfectly valid subject for speculative appraisal, but it is not measurable."

insist on smoking, or attending football games in the rain at the risk of incurring pneumonia, because these provide him even more satisfaction than a better chance of long life. If good health and long life are specified as the "transaction unit," is increased smoking or sports attendance in the rain, which increases his medical expenses, to be construed as a price increase for the "commodity" good health and long life? Switching the criterion for the "commodity" to be priced from what the consumer actually buys (hospital care, surgeon's time, drugs, etc.) to what he "really" wants is a dangerous and inconclusive game for the statistician to play.

I agree, on the evidence of consumption shifts to new or altered products, that output deflated by the conventional method is "biased" downward over time if viewed as an index of the satisfactions output can provide and, also, that it is unfortunate such satisfactions cannot be measured directly. But I see no reasonably objective way to measure or eliminate the bias.

The Enovid case, in which the consumer buys pills half as large as formerly because she now knows this dosage is adequate for her needs, raises still a different issue; even if we were to agree that there has been a productivity increase somewhere, has it occurred within the drug industry, or the medical care industry, or, as seems to me most reasonable, within the consumer sector, which is outside the market economy and not reflected in the national accounts or in price indexes?[4]

I am not sure why Griliches introduces his example of more headaches increasing aspirin consumption, but I trust he agrees that it has nothing to do with quality change or with price, income, or output measurement; what has changed is consumer needs or tastes.

ERNEST W GROVE, Department of Agriculture

Zvi Griliches is to be complimented on his paper. The pleasure in reading it was threefold. First, it was a pleasure to find an econometric model-builder who recognizes that the elaborate superstructure on many of his models rests on statistical data foundations of shifting sand—and not solid rock as most such builders have assumed. Sec-

[4] The concept of productivity increase within the consumer sector is not well developed, and I am not sure whether the Enovid case ought to be considered a productivity increase at all. If a family learns it is happier and healthier eating pieces of cake only half as large as formerly, this would not seem to be a productivity increase, but if the mother learns from a new cookbook to waste fewer ingredients in baking it would. The Enovid case might be likened to either of these situations.

ond, it was a pleasure to find someone with both the energy and the courage required to attempt a measure of changes in quality, and to publish his results for the critical wolves to snap at. And third, it was a pleasure to see the Department of Agriculture's index of prices paid by farmers attacked so explicitly and straightforwardly—and with such thorough justification.

That the Prices Paid Index is not a well-designed measure of price changes has long been recognized. The inappropriate procedures used, however, have not been open to any substantial criticism within the department, largely because they have had evident and authoritative backing from the top. Persons directly engaged in this work are not to blame, for such decisions have been made on a higher level.

There has been serious misguidance of the price index work in the Department of Agriculture, but there is some hope that officials now in charge will do better.

REPLY by Zvi Griliches

> That was a way of putting it—not very satisfactory . . . . Leaving one still with the intolerable wrestle with words and meanings.[1]

Jaszi calls that which is rarely applied "the conventional method." Denison says that I do not provide a solution for the quality problem, since I do not deal explicitly with the case of "new" goods. I had assumed that *quality change* is a term restricted to changes which are occurring along some dimension on which we have had at least one previous observation. Gilbert is defended for desiring to measure only goods, even when they are just a reflection of "bads." And so it goes.

"The essence of the conventional method is to translate quality into quantity by reference to market prices. . . . If . . . an overlap period does not actually exist, an *estimate is made* of what the relative cost of producing the two goods would have been . . . in a common period."[2] This is a general exhortation against sin and not an operational prescription for daily virtuous living. I can find nowhere in the "conventional" literature an actual operational *method* for estimating these changes. The method advocated in my paper—and the credit for it should really go to A. T. Court for his original

[1] T. S. Eliot, *East Coker*.
[2] Jaszi's comment. Italics mine.

path-breaking paper—provides an operational procedure which I believe to be applicable to a wide variety of (though not necessarily all) cases of quality change. Whether or not it would give the same answer as the so-called conventional method, I cannot tell. The "conventional" method has not been spelled out operationally. How would Jaszi *know* that it is the relative chicory and coffee bean content that determines the price of the mixture? Perhaps it is the eggshell and the label on the can that are the essence of the matter. Unless he performs an analysis similar to mine and *tests* it in some acceptable sense, he would not know the facts that he assumes. Whether we would come out with the same number at the end would depend on *how* and *when* the estimates were to be made. My impression is, however, that the "conventional" method would tend to underestimate the importance of quality change by valuing it too late and too little. The tendency is to make this estimate only when one is forced to by the disappearance of the original goods, and to make it at the prices of that time. This introduces a Paasche element into an otherwise Laspeyeres-type index and usually leads to an underestimate of the contribution of quality change. Similarly, the reliance on *cost* rather than *value* (which may have to be imputed) for making the adjustment will also tend to result in underestimation of it in disequilibrium periods. But technical change usually occurs in disequilibrium periods; in fact, it creates them.

Consider the previously discussed case of automobiles: The hedonic price indexes method would divide an automobile into five (or more) not necessarily additive components or qualities, each with its own prices: horsepower, weight, length, "V-eightness," and the "rest" (measured by the constant term). The following table sets up the data for a 1950–61 index number computation. As can be seen from it, if a quality-change adjustment were to be made in 1961, it would matter substantially whether it were made in 1950 or 1961 prices. Of course, a series of chain links would reduce the problem, but would not eliminate it entirely. The basic problems would still remain: how to measure the change, when to start measuring it, how often to do so, and how to weight the new variety of a commodity in the total.[3] Since the "conventional" *method* has no operational *procedure* for dealing with the type of changes that I described, it usually ignores them. In this sense, I am trying to

[3] Neither my paper nor the comments deal adequately with the problem of weighting the new commodity. But to open this up here and now would take us too far afield.

TABLE A

HEDONIC LIST PRICE INDEX COMPUTATION FOR LOW-PRICED THREE, 1950–61

Quality or Dimension	*Price*[a] 1950	*Price*[a] 1961	*Quantity per Car*[b] 1950	*Quantity per Car*[b] 1961
Horsepower (in 100's of units)	0.158	0.026	0.950	1.870
Weight (1000's of pounds)	0.048	0.132	3.081	3.566
Length (100's of inches)	0.832	0.309	1.956	2.096
Cylinders (0 if six, 1 if V-8)[c]	−0.024	−0.011	0.000	1.000
"Other" (constant term)	1.271	2.253	1.000	1.000

Components for Index Computations[d]	Number	Indexes for 1961 (1950 = 100)[e]	
1. $\Sigma P_{1950}Q_{1950}$	3.1963	5. P_L	123.7
2. $\Sigma P_{1961}Q_{1961}$	3.4090	6. P_p	89.5
3. $\Sigma P_{1961}Q_{1950}$	3.2888	7. Q_L	182.6
4. $\Sigma P_{1950}Q_{1961}$	3.4574	8. Q_p	131.9
		9. $V = PQ$	163.2

[a] Coefficients of the 1950 and 1961 cross-section regressions, from Table 3, above.

[b] From Table 6, above.

[c] Since the Consumer Price Index switched to pricing only V-8's in 1956, I am following suit. One could, however, substitute the actual fraction of cars sold with V-8 engines here.

[d] The logarithm (to the base 10) of the predicted price for a given combination of quality prices and quantities per car.

[e] Antilogarithm of (3) − (1) for (5), (2) − (4) for (6), (4) − (1) for (7), (2) − (3) for (8), and (2) − (1) for (9). The result is a set of constant-quality price indexes (Laspeyeres and Paasche), a similar set of quality-level indexes per car, and a total-value index.

provide a tool which would allow conventional practice to approximate its own goals closer.

The difference between the Denison-Gilbert-Jaszi position and my own seems, however, to be more than a disagreement about the way of handling particular "index number problems." If I interpret them correctly, they believe in the possibility and desirability of constructing a value-free set of price and output indexes, independent of a welfare framework or of production or utility function considerations. I do not believe that this is feasible, but even if it were I would not be interested in it. I am interested in these indexes only to the extent that they measure changes in aggregate economic welfare or illuminate other aspects of economic behavior. Thus, while the particular regression method can perhaps be encompassed within the conventional framework, I view it as only a first step toward

the construction of constant utility or productivity level price indexes. To deal with really "new" commodities, I suggest in the last part of my paper that the transaction unit be redefined to be a broader concept of a service flow, corresponding more closely to the units that enter the appropriate utility or production functions, and compute (price) the cost of achieving the same level of service (often using different commodities) in different periods. This is a direct generalization of the preceding concept of a "car" as a bundle of qualities to the case of a "cure" as a bundle of services.[4] I see no more operational difficulty in specifying a base period number of appendectomies, tonsillectomies, and nervous breakdowns than the base period number of all the various pills purchased. The suggested alternative would seem to be much more informative.

An additional question raised by these comments is: To what industry should certain productivity increases be attributed? An improvement in the quality of a particular input, its nominal price remaining constant, will usually be ignored in the computation of conventional input price indexes. Thus, this improvement in quality will show up as a rise in the productivity of the input-using rather than the input-producing industry. For example, it is now possible to produce higher strength (concentration or "proof") fertilizers at a lower price per nutrient unit. This has resulted in a substantial decline in the "real" price of fertilizer to farmers and a large shift by them toward higher strength mixtures. Since the new mixtures are linked in without any direct comparison, this decline does not appear in the official fertilizer price indexes. Agriculture is thus credited with a productivity increase which is perhaps best described and understood as a technical advance in the fertilizer industry. Such a forward shifting of productivity increases, while misleading, would not matter much if we were only interested in the growth of aggregate productivity, since a consolidation of industries would eliminate this difficulty. However, in one of the largest input-using sectors, the consumption sector, we do not measure productivity at all. If we did, we might be indifferent between calling this a decline in a constant-quality price index or a rise in the productivity index. Either way, we would know what has happened. But since we are not measuring the one, we should at least try to measure the other.

[4] Such an approach would consider a car as only one of the inputs in the production of transportation services and would include changes in gasoline consumption per mile in its definition of quality change.

In addition, I believe that it is more illuminating and useful to attribute productivity to the "originating" rather than to the "using" industry. This, of course, argues for quality-corrected price indexes even in the consumption sector.

Brief remarks on some of the specific comments: (1) I find that if the highway construction price index were computed in the same way as other construction *cost* indexes, it would have risen by 47 per cent more since 1947–49. This, to me at least, implies that the bias in the other construction cost indexes is quite substantial and provides an order-of-magnitude estimate for it. (2) Besides problems of measurement and the empirical possibility that different people see the same thing differently, the evaluation of a transaction may differ if tie-ins are involved. For example: Is the price paid and received the same for all individuals when trading stamps are "given away" as part of the purchase? Additional examples of this sort occur when some of the perquisites or fringe benefits given by the employer are really inputs in the production process and have no direct utility of their own to the employee—such as nurses' uniforms. (3) "We can measure [quality changes] . . . only to the extent they are reflected in costs." Counter example: consumer evaluation of various past quality changes could be measured in the used car market—where there is no direct relationship of prices to costs. (4) The aspirin example was introduced to show that "goods only" indexes are of limited use or interest.