Recent Productivity Puzzles in the Context of Zvi Griliches' Research*

Robert J. Gordon

Northwestern University
and
National Bureau of Economic Research

Paper presented to meetings of
the American Economic Association,
January 5, 2002

---

*This research has been supported by the National Science Foundation. I am grateful to Erik Schousboe for his work in developing our data file of productivity by industry from four government sources that is used in part in this paper. Time constraints in preparing this paper prevent me from including a list of references. These will be supplied subsequently when this paper is posted on my web site.

I. Introduction

Zvi Griliches was "Mr. Productivity." His amazing life's work treated every conceivable aspect of the production process, production function, output, and factor inputs, both broadly and deeply. Both by himself and with many co-authors, he brought modern econometrics to the study of agriculture, brought the hedonic price technique to the world, studied sources of measurement error in output and inputs, disentangled the joint contributions of education and ability, assessed the role of economies of scale, and led research teams examining the interplay of R&D and patents. His unique influence on his students, and their affection for him, was represented at his memorial service on January 10, 2000, by Ian Cockburn's "Tree of Zvi," tracing a trail of influence and citations to hundreds of people.

My connection with Zvi is not well known, since I was not his student, only briefly his colleague, and collaborated with him only on the Boskin commission report and several follow-up papers that explained our methods and responded to our critics. He interviewed me on the job market on a cold December day in Cambridge in 1966, and was primarily responsible for hiring me at Chicago for my first job. Then he promptly packed up and left to go to Harvard. He was my mentor, adviser, and friend after that, impatient at the 16 years that elapsed between the first and final draft of my book on durable goods price measurement, but one of those people who became a much better friend due to the ease of e-mail and our mutual enthusiasm for Broadway musical comedies.

This paper addresses two topics. The first concerns some contemporary perspectives on the hedonic price techniques that Zvi pioneered early in his career. My discussion includes both a reminder about perennial problems with the methodology and also some recent applications that raise new questions about conventional price indexes. The second topic is a revisitation of his Presidential Address, in which he suggested that some fraction of the post-1972 U. S. productivity slowdown had occurred because a growing fraction of output had become harder to measure. I revisit his exercise, creating a more fine-toothed comb to sort the difficulty of measurement across sectors, and ask how industry groups in the national accounts, disaggregated to a two-digit level, perform when sorted into four categories of measurement difficulty, both in the 1972-95 period of slow productivity growth and the 1995-2000 revival.

Hedonics: Longstanding Problems

It is ironic that Zvi's first application of the hedonic price technique was to automobiles, yet no official agency in any country has ever switched over to use of hedonic indexes to deflate automobiles in its CPI or durable goods deflator. Hedonics have been fruitfully incorporated into the national accounts for computer hardware, some other consumer electronic goods, residential housing, and apparel, but not automobiles. As Jack Triplett has written, automobiles may be "too complicated" to be suitable for hedonic measurement. Why?

One distinction is that for products with rapid declines in price like computer hardware, the disagreement among alternative hedonic specifications is small relative to the measured trend in prices, i.e., the signal-to-noise ratio is high. If alternative hedonic specifications conclude that computer prices are falling at 32, 35, and 38 percent per year, we surely can be confident that computer prices are falling rather than remaining fixed. For autos, however, the underlying price trend has typically been a slow rate of increase or decline, and so disagreement across alternative hedonic specifications is of greater consequence. In Zvi's original study of automobiles, for instance, the estimated change in price for the 1950-60 period ranges from -18.4 to 14.1 percent, depending on the time period used to measure the coefficients of the characteristics. This translates to a range of only -1.8 to 1.4 percent per year, a range of estimate which would be trivial for a study of computer prices but leaves an automobile study clouded with ambiguity.

Users of durable goods value them for their performance characteristics rather than their physical characteristics. The value of a locomotive depends on the ton capacity of freight cars it can pull at a specific speed, not on the horsepower of the diesel engine. If technical progress increases the efficiency of the driving mechanism that transmits locomotive power from engine to wheels, and if horsepower rather than pulling power is the explanatory variable in the hedonic regression, then the increasing ratio of pulling power to horsepower is an omitted variable that in this example causes an upward bias in the resultling price index.

Almost every previous hedonic study has employed data on physical rather than performance data because the former are more readily available. But the study of locomotives is inherently easier than that of automobiles, because at least the traction power of locomotives can in principle be measured. It is far more difficult to measure the performance characteristics that automobile users care about. While fuel economy, acceleration, and braking performance can be measured, cornering, tightness of handling more generally, comfort, and smoothness of ride are more difficult or even impossible to measure. Over the history of the automobile industry, steady progress has been made in increasing performance relative to the quantity of physical characteristics. The 1955 Chevrolet Bel-Air two-door hardtop V-8 was a wonderful car, ahead of its time, and much valued on the automobile antique market. But today's owner of a Volkswagen Passat, chosen because it is the current darling of *Consumer Reports* for its handling and overall quality, would consider the 1955 Chevy to be an unacceptable bucket of bolts.

Early students of automobile hedonics, including Zvi and Jack Triplett, noticed that most of the fit came from a physical characteristic, weight, and this led to two problems. First, buyers of the highest priced cars seemed to be willing to pay an unreasonable premium for a small bit of extra weight, and what they really seemed to be buying was a subjective characteristic "prestige" rather than something that could be pinned down. In later work with Makoto Ohta, Zvi measured the prestige effect on the market for used cars and found that the prestige effect for a Cadillac persisted throughout the car's life on the used market while that for the Lincoln and Chrysler Imperial evaporated rapidly, although this difference between brands could reflect the poor repair records of the latter two brands in the set of years that made up the Ohta-Griliches sample.

The most serious complication introduced by using weight as the primary explanatory variable was in the treatment of fuel economy. Weight was associated with a larger, better-riding car, but it also was negatively correlated with fuel economy. The coefficient on weight could shift around a lot as energy prices changed and fuel economy became more or less important. As I showed in work both on automobiles and electric generating stations, the right way to measure improvements in fuel economy is to estimate a separate regression equation in which fuel economy is the dependent variable, and weight and other factors are the explanatory variables. Shifts in this relationship can then be valued at different fuel prices and used to adjust the standard hedonic price index. Put another way, because of the negative correlation between fuel economy and weight, the coefficient on weight is downward biased by a different amount each year, and the auxiliary regression that measures shifts in the relationship between fuel economy and weight can pick up changes in that bias and allow for its correction.

Compared to the complex automobile, where subjective user evaluations of subtle quality variations matter a lot for initial sales and resale value, electronic computers (first mainframes and later personal computers) were the perfect product for hedonic analysis. What mattered was primarily speed and memory, and in recent years for the PC such extra attributes as CD and other drives, and extra accessories like fancy speakers. Part of the appeal of the computer for hedonic analysis is that the underlying rate of price decline has been so rapid that there is a very high signal to noise ratio, and the quantifiable characteristics from the beginning were performance-based rather than physical. No one ever ran a hedonic regression in which the weight of the computer entered with a positive coefficient!

But the history of hedonic analysis of computers, which extends back to 1951, contains a major loose end. There are studies of mainframe computers going back to Gregory Chow in 1967 and culminating in the work of Rosanne Cole and Ellen Dulberger as adopted by the BEA into the national accounts in 1986 and extended in two papers by myself and by Jack Triplett in the Jorgenson-Landau volume of 1989. Work by Zvi, Ernie Berndt, and collaborators continued in this tradition and showed that the personal computer was an even better subject for hedonic analysis than the mainframe, due mainly to the proliferation of brands of PCs in contrast to the quasi-monopoly dominance of IBM in the mainframe market.

The loose end is that no one ever (to my knowledge) combined the mainframe, mini, and personal computers into a single sample at the time of overlap in the mid-1980s. Rates of price decline for mainframes were linked to rates of price decline for PCs, without anyone asking the question about the rate of price decline during the transition. I once made a crude calculation, based on the measure of MIPS contained in a 1972-era mainframe and a 1987-era PC using a 386 chip, that the annual rate of price decline over that 15-year interval would be increased from 21 percent in the published literature to more than 35 percent.

Hedonics and a New Problem with Conventional Price Indexes

In order to understand the reasons for differences between hedonic indexes and the conventional "matched-model" indexes that have long been used in official price indexes, several recent studies have attempted to construct both matched-model and hedonic indexes from the same data. In the case of personal computers, it is generally found that price indexes estimated from hedonic regressions generally decline more rapidly than matched model indexes. This has two complementary interpretations. First, the computer market is not continuously in equilibrium. New models are introduced with lower price-to-performance ratios, and new models that were introduced a few months or years earlier do not have their price marked down fully to reflect their now-inferior quality. Second, quality-corrected price decreases occur discontinuously when new chips or other sharp improvements in performance are introduced. Matched model indexes "link out" some of this price decline, since the matched model index by definition excludes comparisons of the previous model with the new model having a lower price-to-performance ratio.

An interesting parallel has recently surfaced in the history of apparel prices. Apparel prices have been suspected of a downward bias, because in earlier years the CPI tracked the decline in price as last season's clothing went on sale, but then linked out the jump back to full-priced clothing in the new season. The hedonic technique, by omitting any linking, is not subject to this weakness. The use of mail order catalogues can provide a range of clothing "models" with price and characteristics data, and preliminary results for womens' dresses from the Sears catalogue seem to indicate that the hedonic indexes rise substantially faster than matched model indexes created from the same data. In preliminary results, the Sears catalogue indexes for womens' dresses rise between 1914 and 1985 at a 1.1 percent annual rate for the matched model index and a 3.7 percent annual for the hedonic index. Both of these can be compared to a 2.2 percent increase for the closest comparable CPI.

Apparel is at the opposite end of the industrial spectrum from computer hardware. Quality change has been slow, negligible, or even negative, with the exception of the invention and diffusion of synthetic fabrics. Anyone taking a cursory glance at the womens' dress section of the Sears catalogue for, say, 1914 and 1993, will be struck by several contrasts. The impression that the catalog prices have increased far more than the 1993/1914 price ratio of 7.6 for the CPI apparel index can be quantified. Taking the median dresses (ranked from most to least expensive) sold by Sears in 1993 and the median sold in 1914, the 1993/1914 price ratio is 32.7. For the two most expensive the ratio is 27.4, while for the two least expensive the ratio is 59.5. It might seem easy to dismiss this discrepancy between the CPI increase and the median increase in catalog dress prices by arguing that quality has increased commensurately, wheras in fact an inspection of the photos in the catalogs suggests that, if anything, quality was higher in the earlier era, with higher quality fabrics (silk, cashmere) and more decorative elements (ruffles, braids, etc.).

Recent research of this type revives the view that at various points in the past the CPI for important products may have been downward biased, not just upward biased. A tendency of apparel manufacturers to disguise price changes at the point of model changes, which are then linked out in the CPI, may have been more important than previously recognized. Yet the challenge of carrying out hedonic studies in practice remains, in the face of small samples and quality characteristics that are difficult or impossible to quantify, including those ruffles and braids on 1914-vintage womens' dresses.

Hard-to-Measure Sectors in the Productivity Growth Slowdown and Revival

In his Presidential Address to the AEA in 1994, Zvi included an empirical investigation of the productivity performance of the economy subdivided into two components, a set of industries that he deemed "easy to measure" and the remainder that were "hard to measure". He showed that the share of the hard-to-measure sectors in total output had grown significantly, and that their productivity growth rate was low both in before and after the 1972 dividing line between rapid and slow U. S. productivity growth. Thus part of the productivity growth slowdown was due to a compositional change toward sectors with a greater downward measurement bias of output growth. This showed that measurement was, despite earlier doubts, a central ingredient in understanding the slowdown.

New data series suitable for testing Zvi's hypothesis have become available in recent years. The detailed two-digit data on value-added (Gross Product Originating or GPO) of the national accounts was many years behind at the time of Zvi's analysis but is now virtually up to date, with data for the year 2000 released two months ago. The analysis in this section begins by comparing the record of productivity growth in the new GPO data with the familiar BLS quarterly data that, because of its timeliness, dominates contemporary discussions of productivity accelerations and decelerations. The BLS data are published only for private business, private nonfarm business, durable manufacturing, and nondurable manufacturing.[1] The residual sector (nonfarm nonmanufacturing private business) is not published by the BLS, on the grounds that methodologies are inconsistent, but its productivity growth can be trivially calculated by applying value-added weights to the components. In contrast, the BEA GPO data allow for the calculation of productivity growth for almost 50 additional two-digit industries. The BEA has just published GPO data through 2000, and the agency deserves enthusiastic compliments for catching up in the past five years after more than a decade in which the publication of GPO data lagged years behind the remainder of the NIPA. The GPO data can be converted into labor's average product (ALP) by dividing through by hours worked in each industry.[1] At present real GPO data using the current improved methodology are published only back to 1977, but Table 1 treads where the BEA will not, by comparing the post-1977 data with the most recently published pre-1977 data.[1]

Table 1 shows the BLS and BEO GPO record of productivity growth for three periods, 1948-72, 1972-95, and 1995-2000. For the private nonfarm business economy, the growth rate of productivity is

---

1. The BLS also publishes a series for the nonfinancial corporate sector, which does not mesh with industry definitions in the BEA GPO data and thus is neglected here.

2. The BEA publishes hours data only at the one-digit level. The exercise carried out here requires the assumption that hours per employee in each two-digit industry within a one-digit industry are identical to the one-digit industry average.

3. The "old" BEA data for 1948-77 are expressed in fixed 1982 dollars and are linked to the contemporary data for 1977-2000 calculated as chain-linked quantity indexes. Minor adjustments are made for changes in industry definition in 1977 and 1987. Complete details of the compilation of the data are provided elsewhere.

somewhat more rapid in the BLS than in the BEA data. Both data sets should be comparable, in that both exclude the farming, real estate, and private household sectors. Otherwise there are two interesting discrepancies. First, in the BEA data the "residual" (nonfarm nonmanufacturing nonhousing) sector registers productivity growth slightly faster than in manufacturing between 1948 and 1972, whereas that sector grows significantly slower in the BEA data. The largest single discrepancy is the much slower growth in nondurable manufacturing in the BEA than the BLS data for 1995-2000. As a result the BLS shows a productivity revival after 1995 in that sector (right-hand column) whereas the BEA registers a significant slowdown.

The Griliches "Hard to Measure" Hypothesis

In his Presidential Address, Zvi divided U. S. industry into two groups corresponding to his assessment as to whether output was relatively easy or hard to measure. He grouped all of agriculture, mining, manufacturing, transportation, communication, and public utilities into an easier-to-measure category and the remaining industries (construction, trade, FIRE, and services) into a harder-to-measure category. He showed that the harder to measure groups had grown as a share of the total, suggesting that part of the post-1972 productivity slowdown could have been due to a growing share of output growth that remained unmeasured. With the BEA GPO data much more current than it was when Zvi wrote, we can extend his distinction to all of the 57 two-digit industries for which the BEA publishes output data. We can examine how difficulty of measurement relates to the magnitude of the post-1972 slowdown and also in the post-1995 productivity growth revival.

Industries are classified not just as easy to measure (group "A") or hard to measure (group "D"), but in two intermediate categories "B" and "C". Some industries with easily obtainable measures of physical volume are in group "A", such as agriculture, mining, parts of nondurable manufacturing producing relatively crude materials, most of transportation, and utilities. Group "B" includes most of durable goods manufacturing (where there is a good chance that price indexes miss some quality change), as well as trucking, telephone, hotels, and miscellaneous repair services. Group "C" includes construction, trade, insurance, and miscellaneous services. Group "D" includes most of the really hard to measure components of finance and services, especially business, health, legal, and educational services. The full list is displayed as Table 3. Our finer-grained categories reflect the fact that within the finance, insurance, and service sectors various industries may differ in their intrinsic difficulty of measurement. For instance, personal services (e.g., barber shops and beauty parlors) is rated in the "A" category; the product changes little over the decades, there are few if any unmeasured product attributes, and there has never been a suggestion that the underlying Consumer Price Indexes for haircuts or beauty treatments contain a bias.

The main conclusion of this section of the paper is reached in Table 2, where the top section aggregates the 57 industries using 1996 weights and the middle section using 1948 weights, and the bottom section records the difference made by the two weighting schemes. The top section of the table indicates that productivity growth was positively correlated with "easy to measure" in both the 1948-72 and 1972-95 period, but that in 1995-2000 the easiest-to-measure group faltered and had a slower rate of productivity growth than the other three groups. As for the productivity growth slowdown, it was greatest in groups A and C and smallest in groups B and D.

Zvi's hypothesis was that there was a shift in the composition of economic activity toward the poorly measured sectors. This is confirmed clearly in comparing the top and middle sections of Table 2, where the weight of group A declined from a 1948 weight of 37.3 percent to a 1996 weight of 22.9 percent, reflecting primarily the decline in importance of both agriculture and nondurable manufacturing. The big increase in share was in group D, the hardest to measure category, clearly supporting the basic message of Zvi's hypothesis. The overall effect of the change in weighting is shown by the difference in the magnitude of the post-1972 slowdown with the 1948 and 1996 weights for the total economy. Clearly, the shift in weights made the slowdown greater, supporting Zvi's hypothesis, but only by about 14 percent of the slowdown (-1.06 percent with 1996 weights versus -0.92 percent with 1948 weights). This leaves 86 percent of the slowdown to be explained by within-group or within-industry phenomena. The biggest event occurring within groups was the shrinking share of agriculture in group A; since agriculture had a minimal post-1972 slowdown compared to other industries within group A, its sharply declining weight from 1948 to 1996 places a greater weight within group A on industries that had larger slowdowns. The opposite inside-group shift occurred within group D, the hardest-to-measure industries.

The post-1995 productivity growth revival seems impervious to changes in weighting and is registered at +1.31 percent with the 1948 weights and +1.29 percent with the 1996 weights. However, looking more closely at the four groups, there is a big difference within group A, because agriculture (with a shrinking share) had very rapid post-1995 productivity growth yet some of the nondurable manufacturing industries in group A performed dismally after 1995. The reverse occurred within group D, presumably because of the outstanding performance of financial institutions, where there was an improvement of measurement and a capturing of significant computer-driven productivity change in the last decade or two of the postwar era.

Conclusion

This paper treats two themes dear to the heart of Zvi Griliches. We find that hedonic price regressions are extremely problematic for automobiles, the topic of Zvi's seminal early work that "launched" the hedonic technique into the profession. This difficulty occurs because automobiles are a consumer good with important aspects of subjectivity in measuring performance, because of difficult-to-measure changes in performance relative to physical characteristics, and because automobiles exhibit only modest "true" changes in price, implying that hedonic studies yield a low ratio of "signal to noise." Hedonic regression studies for mainframe and personal computers are different. As producer goods, there is less room for consumer subjectivity; the quality characteristics in regressions studies have always been focussed on performance rather than physical characteristics, and because the list of relevant characteristics is short.

In recent years new evidence has emerged, both from studies of personal computers and of apparel, that price changes tend to occur together with model changes. This means that the conventional method of collecting matched model indexes tends to understate the rate of price decline for computers and to understate the rate of price increase for apparel. Preliminary results for apparel suggest that that the CPI and the consumption deflator may have understated the inflation rate for apparel over a long period. Preliminary evidence on the rents of apartments and prices of houses also suggest the possibility of a downward CPI bias. Any attempt to extrapolate backwards for several centuries the Boskin commission rate of upward CPI bias, by 1.1 or 1.5 percent, as attempted by Nordhaus in his famous paper on light, inevitably leads to the implausible result that households two centuries ago were unable to afford more than a sack of potatoes per week, with nothing left over for apparel or shelter. A well-researched finding that official price indexes were biased downward at some point in the past would be welcome in trying to resolve the fundamental puzzle posed by what I have called the "Hulten-Bruegel paradox."

Zvi's Presidential Address emphasized the growing importance in GDP of sectors where output is hard to measure. Our revisitation of his work is based on more recent data which is much more finely tuned across 57 two-digit industries, and these are newly divided into four rather than two difficulty-of-measurement categories. We find little evidence that the post-1972 productivity growth slowdown was caused by a shift of output into the hard-to-measure sectors. In fact, more was going on within the sectors than across them. Further, the transition of the American economy into a post-1995 productivity growth revival is completely unrelated to difficulty of output measurement. Indeed, the hardest to measure categories had larger productivity growth revivals than the easiest to measure categories, which is perhaps the most important refutation advanced here of Zvi's basic hypothesis.

TABLE 1

Output per Hour,
BLS Quarterly vs. BEA Gross Product Originating,
1948-2000, Selected Intervals, Annual Percentage Growth Rates

	1948-72	1972-95	1995-00	Slowdown (1972-95 vs. 1948-72)	Recovery (1995-2000 vs. 1972-95)
BLS					
Private NonFarm	2.82	1.41	2.46	-1.40	1.04
Manufacturing#	2.70	2.51	4.88	-0.19	2.36
Durables#	2.55	3.01	6.61	0.45	3.60
NonDurable#	2.87	1.93	2.85	-0.94	0.92
Residual#	2.83	1.00	1.74	-1.83	0.74
BEA					
Private NonFarm Business	2.45	1.19	2.13	-1.26	0.94
Manufacturing	3.16	2.91	4.31	-0.25	1.40
Durable	2.81	3.16	6.57	0.35	3.41
NonDurable	3.63	2.58	1.30	-1.05	-1.28
Residual	2.02	0.54	1.50	-1.48	0.96
BLS minus BEA					
Private NonFarm Business	0.37	0.22	0.33	-0.14	0.10
Manufacturing	-0.46	-0.40	0.57	0.06	0.96
Durable	-0.26	-0.15	0.04	0.10	0.19
NonDurable	-0.76	-0.65	1.55	0.11	2.20
Residual	0.81	0.46	0.24	-0.35	-0.22

BEA data exclude the farming, real estate, and private household sectors
#- Data for these categories begin in 1949, not 1948

TABLE 2

Output per Hour, Measurement Groups, Alternate 1948 and 1996 Weights, Selected Intervals, Annual Growth Rates

	1948-72	1972-95	1995-00	Slowdown (1972-95 vs. 1948-72)	Recovery (1995-00 vs. 1972-95)	Weight
1996 Weights						
Group A	4.14	2.35	1.41	-1.79	-0.94	22.85
Group B	3.32	2.83	4.60	-0.49	1.77	25.83
Group C	2.11	0.80	3.18	-1.31	2.37	27.19
Group D	0.61	-0.08	1.61	-0.69	1.69	24.13
Total	2.52	1.47	2.76	-1.06	1.29	100.00
1948 Weights						
Group A	3.97	2.90	3.09	-1.06	0.19	37.33
Group B	3.05	2.77	4.43	-0.28	1.66	22.60
Group C	1.98	0.80	3.17	-1.18	2.37	33.92
Group D	0.65	-0.39	0.66	-1.04	1.04	6.16
Total	2.88	1.96	3.27	-0.92	1.31	100.00
1996 Weights minus 1948 Weights						
Group A	0.17	-0.55	-1.68	-0.72	-1.13	-14.48
Group B	0.27	0.06	0.17	-0.21	0.11	3.24
Group C	0.13	0.00	0.00	-0.13	0.00	-6.73
Group D	-0.04	0.31	0.95	0.35	0.64	17.97
Total	-0.36	-0.49	-0.51	-0.13	-0.02	0.00

TABLE 3

Identification of Subindustries by Measurement Group

Group A (Easiest)	Group B	Group C	Group D (Hardest)
Farms	Furniture and Fixtures	Construction	Radio and TV
Agriculture services etc.	Primary metal industries	Wholesale Trade	Depository institutions
Metal mining	Fabricated metal industries	Retail Trade	Nondepository institutions
Coal mining	Machinery (except elect)	Insurance carriers	Security and commodity brokers
Oil and gas extraction	Electric Equip..	Insurance agents, brokers, service	Holding, other investment offices
Nonmetallic minerals, except fuels	Motor vehicles and equipment	Auto repair, services, parking	Business services
Lumber and wood products	Other transportation equipment	Motion Pictures	Health services
Stone, clay, and glass products	Instruments and related products	Amusement and recreation services	Legal services
Food and kindred products	Miscellaneous manufacturing industries		Educational services
Tobacco manufactures	Apparel and textile products		Social services, Membership org..
Textile mill products	Printing and publishing		Other services
Paper and allied products	Rubber and miscellaneous plastic products		
Chemicals and allied products	Leather and leather products		
Petroleum and coal products	Trucking and Warehousing		
Railroad transportation	Transportation services		
Local, interurban passenger transit	Telephone and telegraph		
Water transportation	Hotels and other lodging places		
Transportation by Air	Miscellaneous repair services		
Pipelines (not natural gas)			
Electric, gas, sanitary services			