

NBER WORKING PAPER SERIES

ERRORS OF MEASUREMENT
IN OUTPUT DEFLATORS

Frank R. Lichtenberg

Zvi Griliches

Working Paper No. 2000

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 1986

The research reported here is part of the NBER's research program in Productivity. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

Errors of Measurement in Output Deflators

ABSTRACT

In this paper we investigate the incidence of measurement errors in two independent estimates of long-term price change, within the framework of "multiple indicators" models of price measurement. We develop estimates of the measurement-error variances associated with both the Producer Price Index (PPI) and the Census Unit Value Relative (UVR). Our estimates provide support for the generally accepted view that the PPI is a far more reliable indicator of long-term price change: the estimated signal-to-noise ratios for the PPI and UVR are 2.72 and 0.53, respectively. Our estimates should be useful for both constructing an optimal indicator of price change, and for identifying econometric models including error-ridden price- or output-growth terms as regressors. Our analysis suggests that "scores" assigned to product deflators provide useful information about their reliability. By extending our model to explicitly incorporate product-quality change, we are able to assess the importance of the problem posed by quality change for price and productivity measurement. Less than half of quality change, which we estimate to occur at an average annual rate of 1.3 percent, appears to be adjusted for in the PPI. Consequently, estimates of productivity growth based on the PPI underestimate "quality-adjusted" productivity growth by an estimated 43 percent.

Frank R. Lichtenberg
Graduate School of Business
Columbia University
726 Uris Hall
New York, NY 10027

Zvi Griliches
Department of Economics
Harvard University
125 Littauer Center
Cambridge, MA 02138

Errors of Measurement in Output Deflators

Frank R. Lichtenberg and Zvi Griliches

Three of the most important indicators of a sector or country's economic performance are its rates of inflation, real output growth, and productivity growth. The accurate measurement of all three of these growth rates entails the development of reliable commodity price indices. Since much of the accounting data reported in surveys of producers is in current dollars only, such price indices are required to partition the growth in nominal output into its "desirable" and "undesirable" components -- real output growth and inflation, respectively.¹ Measured rates of growth of intermediate materials and producers' (investment) goods prices will determine not only the aggregate rate of total factor productivity (TFP) growth, but also the sectoral composition of this growth.²

Unfortunately, there are major obstacles to the accurate measurement of industrial price change. We believe that it is useful to classify errors of measuring price change into two main types: "short-run" and

¹In part due to the high inflation rates of the 1970's, the Financial Accounting Standards Board began requiring firms to report income statements and balance sheets both in current dollars and adjusted for the effects of inflation. Unfortunately, firms generally adjust or deflate financial statements using aggregate price indices such as the Consumer Price Index rather than industry-, firm-, or product-specific indices, so there is little information about price behavior to be gleaned from these "inflation-adjusted" statements.

²One of the central issues in the classic debate between Denison on the one hand and Jorgensen and Griliches on the other concerned the effect of mismeasurement of capital goods prices on aggregate TFP growth. For a discussion of the "transmission" of errors of price measurement via interindustry transactions and their effects on sectoral productivity growth estimates, see Griliches and Lichtenberg (1984).

"long-run." Many of the problems of measuring short-run (e.g., month-to-month) price change arise because rather than a unique, well-defined price, there typically exists a multiplicity of concepts of "the" price of a good, e.g., shipment versus order price, list price versus actual transaction price, gross price versus price net of transportation costs, and so forth. Moreover, the relationship between different concepts of price (e.g., shipment versus order) are likely to change over the course of the business cycle, as order backlogs and delivery lags fluctuate.

Stigler and Kindahl (1970) have shown that, despite the efforts of the Bureau of Labor Statistics' Division of Industrial Prices to secure reports on a uniform concept of price -- i.e., the actual shipment price, net of any discounts -- the official price indices are contaminated by reporters' failure to distinguish among these concepts.³

Errors in the measurement of price change in the short run are perhaps of greatest concern to business-cycle analysts and to economists interested in testing microeconomic theories of producer price determination.⁴ Because we are primarily interested in the measurement and analysis of long-run economic growth, this study is concerned with errors of measuring price change in the long run. In our view, the major source of such errors is un- or imperfectly-measured changes in product quality. Due to quality change, i.e., changes in the characteristics of

³Another problem with the Producer Price Index (PPI), noted by Ruggles (1977), is that, because the PPI is not based on a scientifically drawn sample of reporting firms, it is difficult or impossible to assess the magnitude of ordinary sampling error.

⁴See Garber (1985) for a discussion of the problem of discriminating empirically between alternative theories of short-run price determination in the presence of noisy price data.

goods whose prices we are trying to observe over time, intertemporal comparisons of directly observed prices of goods are of limited value. Quality change is likely to pose the most serious measurement problems in highly R&D-intensive industries.⁵ While the "hedonic" price index approach⁶ in principle provides a partial solution to the problem of treating quality change in price index construction, for various reasons this technique has not been widely adopted by the statistical agencies responsible for monitoring price change.

The objective of this paper is to investigate empirically the nature and extent of errors of measurement of relatively long-run changes in prices of manufactured goods. We analyze these errors within the framework of a "multiple indicators" (factor analysis) model of price change, estimated using (4-digit SIC) industry-level data on price change (and in one version, an indicator of product-quality change) between 1972 and 1977. In the next section, we describe the two (presumably independent) indicators of price change available to us, and formulate and present estimates of the simplest possible measurement model, in which both indicators are represented as unbiased indicators of true but unobserved price change. This model yields estimates of the measurement-error variance (hence the "signal-to-noise" ratio) corresponding to each indicator. These estimates are potentially useful for both (1) constructing a price index which has lower mean-squared-error than either

⁵Most R&D expenditure is product-, rather than process-, oriented, i.e., devoted to modifying existing products or developing new products.

⁶This approach is discussed in detail in several papers contained in Griliches (1971).

indicator alone, and (2) identifying econometric models which include error-ridden price- (or output-) growth variables on the right-hand side. We also briefly investigate the issue of heteroskedasticity with respect to errors of measurement.

Because, as we discuss below, the sample mean difference between the two indicators is significantly different from zero, the data are not consistent with the assumption maintained in Section II that both indicators are unbiased estimates of the true price change. In Section III we attempt to remedy this defect by specifying a more general measurement model, in which both indicators are postulated to be linear combinations of pure price change and of quality change. This model enables us, under certain assumptions, to estimate both the average rate of product-quality change in our sample, and the extent of bias attending conventional price and productivity estimates resulting from their treatment of quality change. A summary and conclusions are presented in Section IV.

II

In this section we formulate and present estimates of a simple measurement model of industrial price change. The data base for this investigation was assembled and provided by the Interagency Task Force for the Measurement of Real Output, a committee of representatives from five Federal agencies established to oversee the selection of price deflators used to construct the 1977 indexes of production.⁷ The Task

⁷ See the introduction of U.S. Bureau of the Census (1983) for a description of the task force's activities and procedures. We are grate-

Force attempted to obtain, for as many of the approximately 10,000 (7-digit) census products as possible, two independent measures of the change in the product's price between 1972 and 1977: the Producer Price Index (PPI), generated by the Bureau of Labor Statistics (BLS), and the Unit Value Relative (UVR), generated by the Bureau of the Census. Only about 2500 PPI's were available for use in the deflation of the 10,000 census products. Specifications for a number of the BLS products are far more refined than the census individual product categories, so that it is often necessary to average the PPI movements on the basis of the relative importance of each specified product to the overall 7-digit census product value to be deflated. Other producer prices apply to products which are broader than any individual 7-digit product and, consequently, have to be used for more than one of them. The census unit value is defined simply as the ratio of the value of shipments of a product to the number of units (quantity) of the product shipped; the UVR is the ratio of 1977 to 1972 unit value.

As noted by the task force, the PPI and the UVR each has both advantages and disadvantages with respect to the measurement of pure price change. Salient differences between the two indicators are summarized in Table 1. PPI's are price relatives relating to highly specified products of (it is hoped) unchanging quality. Census unit values are often an average of a mix of products of varying degrees of homogeneity. Consequently a change in unit values between two census years could be measuring not merely change in price, but also change in composition of

ful to Edward Robinson, Patrick Duck, and the Task Force for providing us with the data.

Table 1
Salient Differences Between PPI and UVR

<u>Attribute</u>	<u>PPI</u>	<u>UVR</u>
specificity of product	highly specified products	mix of products of varying degrees of homogeneity
type of price	list or book prices used if actual trans- actions prices un- available	reflects actual transac- tion prices at f.o.b. plant level
source of data	some data collected by BLS indirectly from other government agencies and from private trade asso- ciations	all data collected directly from producer

product mix and change in quality. On the other hand, UVRs, unlike PPIs, reflect actual transaction prices and are based on data collected directly from producers. In view of these different limitations, perhaps the comparative advantage of the UVR is in measuring short-run price changes, and that of the PPI, in measuring long-run changes.

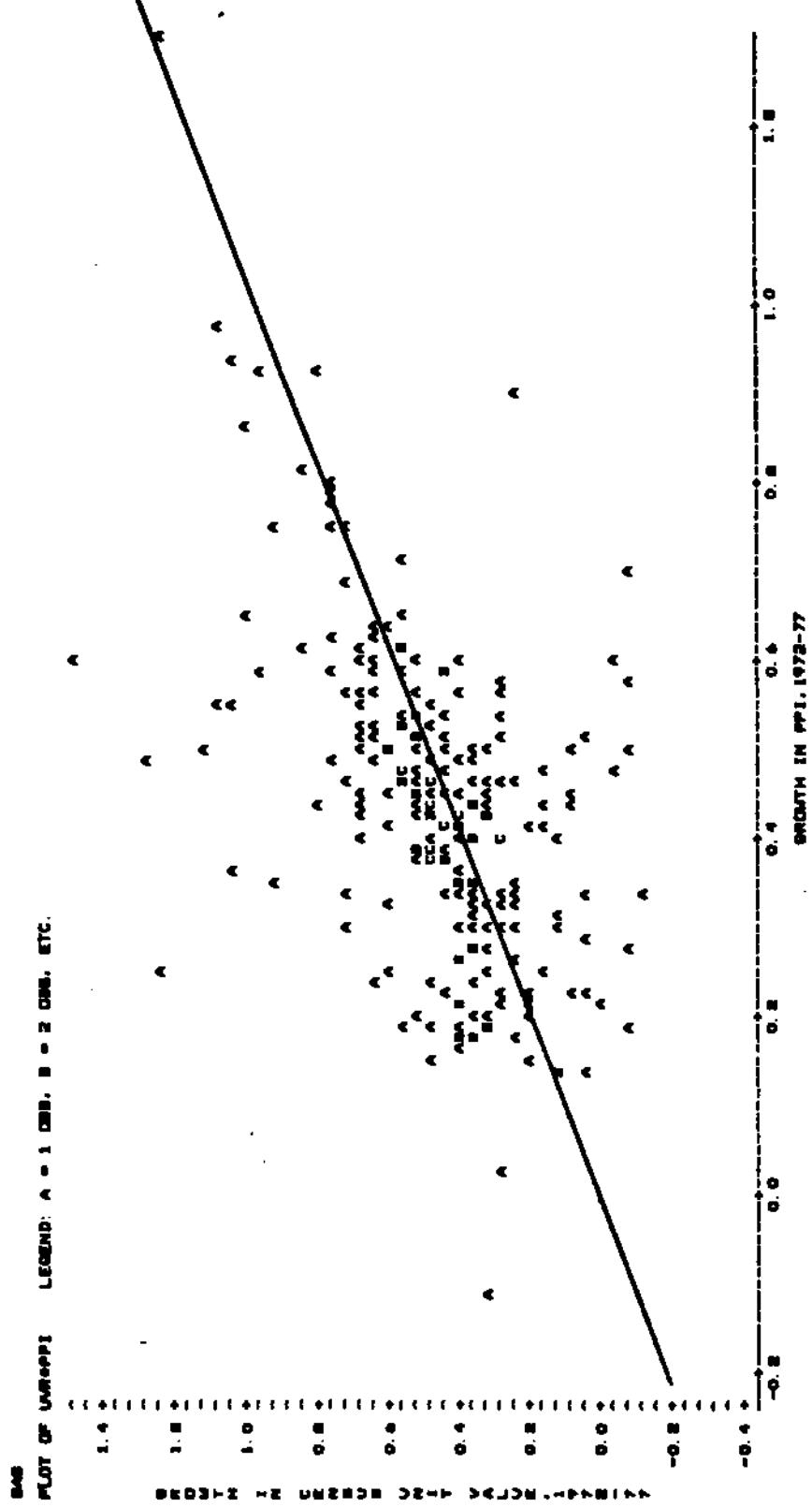
The task force provided us with a listing of PPI's and UVR's for all census products for which both measures were available and which did not pose any problem of disclosure of confidential company data.⁸ In order to be able to match these data to an indicator of quality change (required to estimate the model presented in Section III), we aggregated the task force's data to the 4-digit industry level, by computing weighted averages of (7-digit) product PPIs and UVRs, using value of product shipments as weights. Summary statistics for the resulting sample of 238 industries for which data were available are presented in Table 2. The average increase in prices between 1972 and 1977 (as measured by the difference in their logarithms) was somewhat higher according to the UVR than it was according to the PPI, 46.2 percent as opposed to 43.4 percent. Moreover, the mean difference between the PPI and the UVR, -2.8 percent, is significantly different from zero at almost the five percent level. The UVR exhibited substantially more variation across industries

⁸Products for which both indicators are available are probably not a random sample of all manufactured products, in terms of the extent of quality change and measurement error. It is possible that the problem of measurement error is less serious for products in our sample than for those not in our sample. If that is the case, our estimates of error variances and of the average rate of quality change should be interpreted as lower bounds.

Table 2
Summary Statistics: Logarithmic Changes in PPI and UVR, 1972-77,
Sample of 238 4-Digit Manufacturing Industries

	<u>PPI</u>	<u>UVR</u>
Mean	.434	.462
Standard Deviation	.178	.258
Minimum	-.112	-.118
Maximum	1.30	1.46

Figure 1



than the PPI: its standard deviation is about 45 percent greater. A scatter plot of UVR against PPI is shown in Figure 1.

We begin our econometric analysis of errors in the PPI and UVR by specifying the following extremely simple measurement model:⁹

$$\begin{aligned} \text{PPI} &= P^* + e_1 \\ \text{UVR} &= P^* + e_2 \end{aligned} \tag{1}$$

where P^* denotes true but unobserved industry price change and e_1 and e_2 are classical measurement errors assumed to have the following properties:

$$E(e_i) = E(e_i \cdot e_j) = E(e_i \cdot P^*) = 0$$

$$i, j = 1, 2, i \neq j$$

This model may be regarded as a factor analysis model, in which P^* is a "common factor" and e_1 and e_2 are "specific factors." The model is exactly identified: there are three unknown parameters to be estimated -- the variances of P^* , e_1 , and e_2 (denoted by $\sigma_{\text{P}^*}^2$, σ_1^2 , and σ_2^2 , respectively) -- and three sample moments (the variances and covariance of PPI and UVR) which can provide (maximum likelihood) estimates of them.

In this model, PPI and UVR are both postulated to be unbiased indicators of P^* . Subtracting the UVR equation from the PPI equation and applying the expectation operator, we obtain

$$E(\text{PPI} - \text{UVR}) = E(e_1) - E(e_2) = 0.$$

⁹Industry subscripts are suppressed to simplify notation.

This model implies that the expected value of the difference between the PPI and the UVR is zero, whereas we noted above that the sample mean difference is significantly different from zero. This inconsistency can be eliminated by suitably generalizing the model (increasing the number of parameters), but a larger model will be underidentified in the absence of additional indicators. In the next section we develop an expanded model which includes an indicator of product-quality change; that model provides an "explanation" of the nonzero mean difference between the two measures of price change. We believe, however, that despite the inconsistency noted, the above model provides a useful starting point for our analysis, and we pursue it in the remainder of this section.

Equations (1) constitute a purely measurement model of price change; they do not embody any theory of long-run industry price determination. Although our empirical analysis is not predicated on any specific theory of price determination, we thought that it might be of interest, before presenting our estimates, to briefly discuss our view of the process of long-run industry price formation. We postulate that the long-run change in an industry's output price is determined by exogenous changes in its total factor productivity and input prices according to the formula:

$$P^* = PI^* - TFP^*$$

where PI^* denotes the change in (a Divisia index of) input prices, and TFP^* denotes the change in total factor productivity.¹⁰ This model of

¹⁰ This is the dual to the perhaps more familiar formula equating TFP^* to the change in output quantity minus the change in (a Divisia index of) input quantity.

price formation is based on the assumptions of constant returns to scale technology and competitive (cost-minimizing) producer behavior. Industries might be thought of as drawing at random from a bivariate distribution of PI* and TFP* (which may or may not be independent). Industries "lucky enough" to have extensive "technological opportunities" will tend to draw high values of TFP*; those "unlucky enough" to rely on inputs which are becoming increasingly relatively scarce (e.g., oil in the 1970s) will tend to draw high values of PI*. We can think of reasons why PI* and TFP* might be either negatively or positively correlated across industries. Theories of "induced innovation" or "organizational slack," for example, may imply that industries faced with higher input price increases may devote more resources to (factor-saving) research and development, and hence (subsequently) experience higher productivity growth.¹¹ Another hypothesis is that an industry with high exogenous productivity growth will tend to exhibit either above- or below-average (depending on the elasticity of demand for the industry's product) growth in demand for factors of production, which may influence the prices of inputs specialized to the industry.

We now present estimates of the parameters underlying the measurement model formulated above. That model implies that the "population" variances and covariance of PPI and UVR are related to σ_x^2 , σ_1^2 , and σ_2^2 as follows:

$$\text{var(PPI)} = \sigma_x^2 + \sigma_1^2$$

$$\text{var(UVR)} = \sigma_x^2 + \sigma_2^2$$

¹¹See Lichtenberg (1986) for evidence that industries experiencing the largest energy cost increases during the 1970s maintained the highest growth of R&D expenditures, conditional on output growth.

$$\text{cov}(PPI, UVR) = \sigma_{\epsilon}^2$$

The sample moments are maximum likelihood estimates of the corresponding population moments. The following is the computed sample moment matrix:

	PPI	UVR
PPI	.0316	.0231
UVR		.0668

The parameter estimates implied by these sample moments are:

$$\begin{aligned}\hat{\sigma}_{\epsilon}^2 &= .0231 \\ \hat{\sigma}_1^2 &= .0085 \\ \hat{\sigma}_2^2 &= .0437\end{aligned}$$

The UVR measurement-error variance is over five times as large as the PPI error variance. From another perspective, the fraction of observed interindustry variance which is due to "noise" (measurement error) is 27 percent for the PPI and 65 percent for the UVR. The evidently greater reliability of the PPI appears to offer some justification for the decision by the Interagency Task Force to give "more consideration to BLS price indexes than to census unit values" when comparing their merits as indicators of (long-term) price change.

Knowledge of the error variances (or signal-to-noise ratios) of the PPI and UVR is potentially useful in at least two respects. First, these estimates can be used to construct an optimal (in the sense of minimum variance, conditional on available information) index of price change. DeLeeuw and McKelvey (1983) have shown that when two indicators, each hypothesized to be a linear function of a single unobservable, are

available, an appropriately weighted average of the two indicators has a lower mean squared error than the more reliable indicator alone. In the case of our model, their formula for the index \hat{P} which minimizes the mean squared error reduces to

$$\hat{P} = \lambda \cdot PPI + (1-\lambda) \cdot UVR,$$

where $\lambda = \sigma_2^2 / (\sigma_1^2 + \sigma_2^2)$.

Substituting the above estimates of the measurement-error variances yields the value $\hat{\lambda} = .837$.

The second respect in which these estimates may be useful is in identifying otherwise unidentified econometric models which include error-ridden price- (or output-) growth variables on the right-hand side. Suppose, for example, that we are interested in estimating the output-elasticity of labor demand, and that we propose to do this by running the regression of the change in labor input on the change in deflated (by the PPI) output, using cross-sectional industry level data. Assuming that the change in nominal output is measured without error, the measurement error in the PPI will be transmitted, with an opposite sign, to the measured change in real output. The probability limit of the least-squares coefficient of measured output change is $\beta(1-\gamma)$, where β is the true output elasticity and γ is the ratio of measurement-error variance (σ_1^2) to the total variance of measured real output. In the absence of extraneous information about γ (or an instrument for real output growth), β is not identified; only a lower bound is provided by the least-squares

coefficient. The estimates presented above, however, would enable us to compute an estimate of γ , and hence identify β .¹²

It is implicitly assumed in the above analysis that the error variances of both the PPI and the UVR are constant across industries. The Task Force, however, clearly believes that the "quality" or reliability of both indicators varies considerably across products, and in fact assigns "scores" to both indicators for every product to reflect its assessment of data quality. Here is a brief description of the Task Force's procedure for determining the score of a given product's PPI:

Price data from a producer is direct and considered good. More indirect sources, such as trade journals, are given worse scores. If the reported data [sic] is the transaction price, it is considered good. Other types of pricing, such as list price, are not as good, and therefore, are given worse scores. The two factors were used together to calculate a single score for each product priced... A score of "0" was assigned if the BLS index had no transaction problems and was directly priced by that agency."

Table 3A shows the procedure used by the Task Force for assigning scores, which range from zero to six, to PPI observations on the basis of the source and type of price.

The assignment of quality scores to the UVR was based on somewhat different considerations. The logic underlying this procedure is described below:

Census unit values are based on product description[s] that often contain items which may vary in both quality and by price line. Thus, changes in unit values may result [from] changes in product mix and quality, as well as changes in pure price.

¹² Estimates of γ can also enable identification of parameters in more complicated (i.e., multiple regression) models. See Griliches (1984), p. 21 ff. for details.

Table 3A. Quality Scores for Producer Price Indexes of the Bureau of Labor Statistics

Source	Trans- action	List with dis- count	List Order	Other, un- known	Not avail- able
Producer.....	0	1	2	3	4
Buyer.....	0	(NA)	(NA)	(NA)	4
Government agency.	0	1	2	3	4
Importer	1	3	4	5	6
Trade journal	0	2	3	4	5
Other, unknown....	1	3	4	5	6
Not available.....	6	6	6	6	6

Table 3B. Quality Scores for Census Unit Values Relatives

1977 concentration (Percent of 1977 shipments with unit values between 0.8 and 1.2 multiplied by the mean unit value)	Difference (points) between 1977 and 1972 concentrations			
	5.0 or less	5.1 to 10.0	10.1 to 15.0	15.1 or more
90-100.....	1	2	4	6
80-89.9.....	2	4	5	6
70-79.9.....	4	5	6	6
0-69.9.....	6	6	6	6
Not available.....	6	6	6	6

Note: Unit values derived from Current Industrial Reports (CIR series) were not analyzed. They were assumed to rate a score of "1", but many were reviewed by the committee. Price relatives other than PPI, Census, and CIR unit value relatives were adopted only on recommendation of the committee.

Source: U.S. Bureau of the Census (1983), pp. 141-142.

The only available measure of homogeneity of the product category and comparability in quality and product mix between censuses is the concentration of the product's unit values reported by individual establishments around the mean unit value based on the published quantity and shipment aggregates. If a high percentage of the total value of shipments of the product is associated with unit values that are close to the mean, and if this percentage did not change significantly between censuses, the unit value relative was assumed to be a reasonably good measure of price change.

The decision rule adopted by the Task Force for assigning scores to UVR observations is indicated by Table 3B.

We sought to determine whether the PPIs and UVRs of products (or industries) receiving good scores tended to contain less measurement error than those of products receiving poor scores. To test this hypothesis, we performed the following procedure. First, we computed for each 4-digit industry the weighted average of the PPI scores of the 7-digit products sampled in that industry, using the product's share in the industry's "covered" shipments as weights. Next, the industries were classified into two groups: those with "low" (essentially, below the median) weighted PPI scores, and those with "high" scores. Finally, the sample covariance matrix, and the implied model parameter estimates, were computed separately for each group. This procedure was performed a second time, partitioning the sample on the basis of UVR score instead of the PPI score.

The parameter estimates produced by this procedure are displayed in Table 4. The estimated values of both σ_1^2 and its ratio to $\text{var}(\text{PPI})$ are much lower for the subset of industries with low ("good") PPI scores than they are for the high-PPI-score subset. Similarly, the subset of industries with high UVR scores has considerably larger values of $\hat{\sigma}_2^2$ and its ratio to $\text{var}(\text{UVR})$ than the subset with low UVR scores. These estimates

Table 4

**Estimates of PPI and UVR Error Variances,
By Value of Industries' Weighted PPI or UVR Scores**

<u>Quantile (mean score)</u>	<u>$\hat{\sigma}_1^2$</u>	<u>$\hat{\sigma}_1^2/\text{var(PPI)}$</u>	<u>$\hat{\sigma}_2^2$</u>	<u>$\hat{\sigma}_2^2/\text{var(UVR)}$</u>
A. Industries classified by weighted PPI score				
Low PPI score (mean score = 0.45)	.0010	.035	.0474	.631
High PPI score (mean score = 2.33)	.0139	.428	.0392	.667
B. Industries classified by weighted UVR score				
Low UVR score (mean score = 3.28)	.0030	.085	.0317	.493
High UVR score (mean score = 5.98)	.0119	.492	.0560	.807

are quite consistent with the proposition that the scores assigned to the two measures of price change provide useful information about their reliability.

III

In the model analyzed in the previous section it was assumed that the PPI and UVR measurement errors were uncorrelated both with one another and with P^* . Recall that in our introductory discussion of problems in measuring price changes we distinguished (primarily long-run) errors associated with changing product quality from (largely short-run) sampling, reporting, and miscellaneous errors. It is probably reasonable to assume that the errors of the second type present in the PPI and the UVR are distributed independently of one another and of everything else. Both indicators, however, mis-measure quality change -- although we hypothesize that the extent of mismeasurement differs. Since the same unobservable product-quality change contributes to errors in both indicators, the covariance between the errors should be allowed to differ from zero. We therefore proceed to analyze a model in which both the PPI and UVR measurement errors are specified as linear combinations of two components, one of which (interpreted as "quality change") is common to both indicators.

We begin by defining the growth rate of product quality, Z^* , as the difference between the growth rate of the "effective quantity" of output Q^* and the growth rate of number of units sold, Q :

$$Z^* = Q^* - Q. \quad (2)$$

Defining the growth in the price of effective output as $P^* \equiv VQ - Q^*$, where VQ denotes the growth in nominal output, and the growth in the price per unit sold as $P \equiv VQ - Q$ yields the identity

$$P = P^* + Z^*. \quad (3)$$

The change in price per unit sold equals the sum of the change in price per unit of effective output and the change in quality. The symbol P^* , which we used earlier to denote "true" price change, is now interpreted more precisely as "true, quality-adjusted" price change.

Because the Census unit value is defined simply as the ratio of value of shipments to number of units, the UVR is, in effect, an indicator of P , not of P^* :

$$UVR = P + e_2, = P^* + Z^* + e_2. \quad (4)$$

BLS' practice of "linking in" new products and discontinuing old products suggests that the PPI "adjusts for" some, but not all, product quality change. As Ruggles observes, "BLS adjusts the price observations only in those cases where the change in the specification of the product is directly associated with a change in price reflecting the cost change... where there are design changes which affect the performance of the product but are unrelated to the cost of production, there is no recognition of this fact other than the publication of the new specification."¹³ We therefore postulate the following measurement model for the PPI:

$$PPI = P^* + a_1 Z^* + e_1. \quad (5)$$

¹³Op. cit., p. IV-1.

We expect the value of α_1 to lie between zero and one. Substituting for Z^* in (5) using (3), and rearranging,

$$PPI = [(1-\alpha_1)P^* + \alpha_1 P] + e_1. \quad (6)$$

so the PPI may be interpreted as an error-ridden indicator of a weighted average of P^* and P , with weights determined by α_1 . We assume that $E(e_i)$ = $E(e_i \cdot e_j)$ = $E(e_i \cdot P^*)$ = $E(e_i \cdot Z^*)$ = 0 for $i, j = 1, 2, i \neq j$. In order to obtain the results derived below, we are also forced to assume that $E(P^* \cdot Z^*) = 0$, i.e., that quality change and quality-adjusted price change are uncorrelated. We have some reservations about this assumption, since both Z^* and P^* might be postulated to be functions of industry R&D activity, for example.

The system of equations (4) and (5) contains two more parameters -- α_1 and $\text{var}(Z^*)$ -- than the system (1) and is no longer identified. But by adding an indicator for Z^* we can obtain estimates of both α_1 and of the mean change in product quality over the sample period. As noted earlier, in constructing industry output deflators BLS both introduces, or "links in," new products, and drops old products from the price index. We hypothesize that the relative extent to which the statistical agency introduces and drops products from price indexes over an extended period is a reflection of the incidence of quality change. Data compiled by Ruggles (1977) allow us to compute (unfortunately only at the 2-digit level) the fraction of products which were ever used in constructing an industry's price index during the period January 1967 - January 1975 that were not used at the beginning of the period. We define this ratio of the number of products introduced during the period to the total number

of products ever included during the period as NEW, and postulate the indicator equation

$$\text{NEW} = \alpha_2 Z^* + e_3. \quad (7)$$

The value of NEW ranges from zero for the tobacco, furniture, and printing and publishing industries, to 43 percent in electrical equipment and supplies and 59 percent in rubber and plastic products.

Because the NEW variable plays a key role in identifying an important parameter of our model, we would like to provide some evidence that NEW is a reasonable indicator of product-quality change. We believe that the major cause of such quality change is product-oriented R&D expenditures undertaken by industry. We therefore postulate a "quality-change production function" of the form

$$Z^* = \beta RD + u \quad (8)$$

where RD is a measure of R&D activity. Since Z^* is unobserved, equation (8) cannot be estimated, even if we have data on R&D. By substituting the equation (8) for Z^* into the indicator equation (7), however, we obtain a reduced-form equation which can be estimated from available data:

$$\text{NEW} = \alpha_2(\beta RD + u) + e_3 = \Pi RD + v$$

where $\Pi \equiv \alpha_2\beta$ and $v \equiv (\alpha_2u + e_3)$. The hypotheses that (at least some part of) R&D expenditures cause product-quality change and that NEW is an indicator (or an "effect") of quality change suggest that NEW should be an increasing function of RD. We tested this implication of our theory, using the data on industry R&D expenditures developed by Scherer

(1984).¹⁴ From Scherer's data it is possible to distinguish three mutually exclusive categories of R&D related to a given industry: product-oriented R&D performed within the industry (OWNPROD); product-oriented R&D "relevant" to the industry performed by its suppliers of capital and materials (SUPPROD); and process-oriented R&D performed within the industry (PROCESS). We hypothesized that OWNPROD, and perhaps also SUPPROD, would be significantly positively related to NEW (via their effects on Z*), but that this would not necessarily be the case with PROCESS. The estimated regression of NEW on the three components of industry R&D (each divided by industry sales) confirmed our suppositions (standard errors are reported in parentheses):

$$\text{NEW} = .349 \text{ OWNPROD} + .878 \text{ SUPPROD} - .064 \text{ PROCESS} \quad \bar{R}^2 = .176$$

$$(.063) \qquad (.341) \qquad (.216)$$

Consistent with our expectations, we cannot reject the hypothesis that the coefficient on PROCESS equals zero, whereas both of the other coefficients are significantly different from zero at about the one percent level. The finding that the coefficient on SUPPROD is much larger than the coefficient on OWNPROD is rather surprising, but we cannot reject the hypothesis of equality of the two coefficients at conventional significance levels.

The above analysis lends support to the view that NEW is a valid indicator of (or instrument for) Z*. We now show how this instrument can be used to identify an important parameter of our measurement model, α_1 .

¹⁴For further discussion and analysis of these data, see Griliches and Lichtenberg (1984).

The indicator equations (4), (5), and (7) imply that the population covariances $\text{cov}(\text{PPI}, \text{NEW})$ and $\text{cov}(\text{UVR}, \text{NEW})$ may be expressed as follows:

$$\text{cov}(\text{PPI}, \text{NEW}) = \alpha_1 \alpha_2 \text{var}(z^*)$$

$$\text{cov}(\text{UVR}, \text{NEW}) = \alpha_2 \text{var}(z^*)$$

Hence the ratio of these covariances is equal to α_1 . The sample covariances of PPI and UVR with NEW are, respectively, .0021 and .0037, yielding an estimate of α_1 of .568. As expected, the ratio is less than unity, suggesting that roughly half as much product-quality change is erroneously reflected in the PPI as in the UVR.

The estimate of α_1 enables us to compute an estimate of the average change in product quality in our sample, and of the extent to which the PPI and UVR tend to overstate true price change (and productivity indices based on the PPI understate true productivity change) due to their failure to correctly account for quality change. To calculate the sample mean change in product quality during the period 1972-77, subtract (4) from (5), and denote this difference by DIFF:

$$\text{DIFF} \equiv \text{PPI} - \text{UVR} = (\alpha - 1)Z^* + e_1 - e_2$$

Hence, the mean difference between the two price indices is (denoting means by bar superscripts)

$$\overline{\text{DIFF}} = (\alpha_1 - 1)\bar{Z}^*.$$

$$\begin{aligned}\bar{Z}^* &= \overline{\text{DIFF}} / (\alpha_1 - 1) \\ &= -0.0282 / (.568 - 1) = 0.0654.\end{aligned}$$

On average, product quality increased about 6.5 percent, or 1.3 percent per annum, during this period. This implies that the sample mean values of UVR and PPI overstated true price growth P^* by 16.5 percent and 10.9 percent, respectively. Estimates of total factor productivity based on these output price indices are, of course, biased downward, since "effective" output grew faster than measured output. Assuming for simplicity that errors in output deflators are the only source of errors in TFP calculations, TFP calculated using the PPI, $\overline{\text{TFP}}_{\text{PPI}}$ is related to true TFP, TFP^* , by

$$\overline{\text{TFP}}_{\text{PPI}} = \text{TFP}^* - \alpha_1 Z^* - e_1$$

Given our sample mean values of \bar{Z}^* and $\overline{\text{TFP}}_{\text{PPI}} (= .0492)$, this implies that $\overline{\text{TFP}}_{\text{PPI}}$ underestimates true productivity change by about 43 percent.¹⁵

¹⁵The productivity data upon which this calculation is based are described in Griliches and Lichtenberg (1984).

IV

In this paper we have investigated the incidence of measurement errors in two independent estimates of long-term price change, within the framework of "multiple indicators" models of price measurement. We have developed estimates of the measurement-error variances associated with both the Producer Price Index and the Census Unit Value Relative. Our estimates, based on a possibly nonrepresentative sample of products from 238 industries, provide support for the generally accepted view that the PPI is a far more reliable indicator of long-term price change: the estimated signal-to-noise ratios for the PPI and UVR are 2.72 and 0.53, respectively. Our estimates should be useful for both constructing an optimal indicator of price change, and for identifying econometric models including error-ridden price- or output-growth terms as regressors. Our analysis suggests that "scores" assigned to product deflators provide useful information about their reliability. By extending our model to explicitly incorporate product-quality change, we were able to assess the importance of the problem posed by quality change for price and productivity measurement. Less than half of quality change, which we estimated to occur in our sample at an average annual rate of 1.3 percent, appears to be adjusted for in the PPI. Consequently, estimates of productivity growth based on the PPI underestimate "quality-adjusted" productivity growth by an estimated 43 percent.

REFERENCES

- deLeeuw, Frank, and Michael J. McKelvey (1983), "A 'true' time series and its indicators," Journal of the American Statistical Association, Vol. 78, No. 381, March, pp. 37-46.
- Garber, Steven (1985), "The reserve-labor hypothesis, short-run pricing theories and the employment-output relationship," unpublished manuscript, Carnegie-Mellon University, June 5.
- Griliches, Zvi (ed.) (1971), Price indexes and quality change: studies in new methods of measurement (Cambridge: Harvard University Press).
- Griliches, Zvi (1984), "Data problems in econometrics," NBER Technical Paper No. 39, July.
- Griliches, Zvi, and Frank R. Lichtenberg (1984), "Interindustry technology flows and productivity growth: a reexamination," Review of Economics and Statistics, Vol. LXVI, No. 2, May, pp. 324-329.
- Lichtenberg, Frank R. (1986), "Energy prices and induced innovation," Research Policy, forthcoming.
- Ruggles, Richard (1977), The wholesale price index: review and evaluation, Executive Office of the President, Council on Wage and Price Stability Council Report, June (Washington, D.C.: U.S. Government Printing Office).
- Scherer, F. M. (1984), "Using linked patent and R&D data to measure interindustry technology flows," in Zvi Griliches (ed.), R&D, patents, and productivity (Chicago: University of Chicago Press).
- Stigler, George J., and James K. Kindahl (1970), The behavior of industrial prices (New York: National Bureau of Economic Research).
- U.S. Bureau of the Census (1983), 1977 census of manufactures, indexes of production, subject series MC77-SR-14, June (Washington, D.C.: U.S. Government Printing Office).