This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: The Behavior of Industrial Prices

Volume Author/Editor: George J. Stigler and James K. Kindahl

Volume Publisher: NBER

Volume ISBN: 0-87014-216-X

Volume URL: http://www.nber.org/books/stig70-1

Publication Date: 1970

Chapter Title: Appendix A: Transcriptional Editing of Raw Data

Chapter Author: George J. Stigler, James K. Kindahl

Chapter URL: http://www.nber.org/chapters/c3319

Chapter pages in book: (p. 95 - 100)

Transcriptional Editing

of Raw Data

DATA COLLECTION

Firms and agencies made information available to us in one of three ways. Some gave permission for us to transcribe the relevant information directly from the records; others made available photo-reproduced copies of the records; still others preferred to send (at a date subsequent to the interview) specially prepared tabulations of the relevant data. The last procedure was the option chosen by most of the cooperating private firms; about two-thirds of the cooperating public agencies either allowed us to make reproductions of the relevant records or to copy information directly from the original records.

Where possible we copied the information from the records. We could then exercise a degree of quality control over transcription errors and look for relevant notations which might not be recorded by a firm's own transcriber. (The average degree of quality control actually exercised by others may have been higher or lower than ours, but others' transcriptions were undoubtedly more variable in quality.) However, the transcription of the data for us was only the final step in the processing of data by the firms: most reporters apparently took the information from summary records maintained for their own use, not from original invoices. (The number of invoices which come through the purchasing department of a large industrial firm is very large. The invoices themselves are typically used for no internal purposes other than the clerical operations of checking arrival of shipments and arranging payment.) Thus there were one or more steps of data transcription prior to enter-

ing the information on the form sent to us. Clerical errors inevitably appeared in some of the data sent to us or copied by us. The larger of these errors were, hopefully, caught by our checks for internal consistency and our review of large price changes and unusual timing of price changes.

Errors may also have arisen due to reports of nontransaction prices. In a few cases we encountered hostility to our study from subordinates assigned the task of cooperating with us. Usually it took the form of noncooperation, but on occasion it may have led to erroneous price quotations. Indolence, as well as hostility, led in some cases to the reporting of list prices (including in one case, copies of Bureau of Labor Statistics prices). We of course accepted list prices where they were reported to be paid, but not where they were given simply because the true records were inaccessible.

EDITING OF DATA

The price series reported to us were subjected to scrutiny from several points of view. We attempted to be sure that each price series represented quotations on a single commodity at different points in time; ideally, there would be no change in physical specifications of the commodity, its point of delivery, its approximate lot size, or in any of the other terms of sale or purchase. We also eliminated series whose information content was slim. In addition, certain prices were adjusted to take account of additional information in the report from the informant. These adjustments are discussed in detail below.

Minimal Coverage Requirements

Certain requirements were imposed on each series to eliminate those with very small amounts of information. First, no series was utilized if it had a gap of more than sixteen months between adjacent observations. Whenever a gap of such magnitude appeared, the total series was treated for computational purposes as if it were two independent series. Second, each series was required to extend over at least twenty-four months (i.e., the last quotation had to refer to a date at least twenty-four months after the earliest). Third, the number of months in which a price quotation was recorded had to be at least four. (These latter

two tests were applied to series *after* they had been broken into two or more series because of gaps of over sixteen months, if necessary.) A series which failed to meet these requirements of coverage was deleted from the study.

Requirements for Product Homogeneity

Clearly a price series is less valuable if there are changes in any of the specifications relevant to the price. Our editing procedure for this problem was as follows:

(1) Whenever a change in any of the specifications was discovered, the series was linked, if possible at that point. Linking could be accomplished only when information on the prices of both the old and new specifications were available on the same date. This information was usually not available.

(2) Whenever a change in specifications could not be linked in, the price series was marked as being subject to a specification change at that date. In further analysis, "the" series were treated as two different series, one ending with the last observation before the specification change, the other beginning with the first observations on or after the date of the change.

Often serious problems were faced in deciding whether there had been a specification change, and we called upon the respondent for additional information. In general, any of the following were treated as changes in specifications unless we were specifically assured by the informant that the change had no effect on price: any change in physical description; change in f.o.b. point; unquantified change in credit terms; change in services provided by seller as a condition of sale; sizable change in lot size or volume of purchase over a period of time. Usually our doubts could be cleared up by the price reporter. There remains, of course, the problem of unreported changes in specifications. Our main defense against unreported specification changes was to examine each series for large shifts in the level of prices, and to inquire specifically if there were changes in specifications at such times.

Quantity discounts are a particularly important and pervasive example of specification changes. Both the size of the individual order and the volume of annual purchases may affect the price for each buyer. Quantity discount schedules usually deal with a single aspect of the problem.

Thus the steel industry charges extra for orders of less than 20,000 pounds of a given type of steel but amalgamations of orders are often allowed, and amalgamations of shipments to several plants of the buyer may be allowed. The procedures used to deal with quantity discounts were:

(1) As part of general instructions, informants were asked to identify (but may not always have done so) any changes in volume or lot size which they considered to have an effect on price. These were generally marked as specification changes.

(2) Movements from one known or published quantity discount bracket to another either were marked as specification changes, if there was a once-and-for-all shift, or only observations within a bracket were kept in any one series. No attempt was made to compare prices in different quantity discount brackets.

(3) Volume changes beyond the stated quantity discount schedule were not recorded as a specification change or eliminated from consideration unless the change were of the order of magnitude of twofold or more. Such large changes called for a consultation with the informant.

We seldom got price reports for very small quantity purchases, for two evident reasons: purchases were less likely to be continuous; and the records are not as well kept. Therefore our price series underrepresent prices to buyers of small quantities.

Internal Consistency of the Data:
Special Editing by Frequency of Data

Although we sought to obtain data at monthly intervals, this was not always possible. Some firms' records were not sufficiently detailed and other firms bought on a purchase contract specifying a price (or formula to be used for computing a price) to be used for some period of time. All data were classified by a frequency code, being classified as (1) monthly series, (2) averages of prices paid over a period (annual, semi-annual, or quarterly), (3) observations on prices in purchases at regular discrete intervals (again, annually, semi-annually, monthly); (4) as contract prices prevailing over a period, or (5) as irregular data. This coding was essential for correct processing and interpretation of the data.

Contract data, if not subject to escalation provisions, report prices which are firm over the specified period. Such contracts are recorded with the specified price treated as the price in each month covered by the contract. In the case of escalated contracts, the price is computed (by us) according to the formula, and the resulting figure recorded for each month of the contract. (Since all escalations required that a specified formula be applied to a published price, this can be done unambiguously.)

All data had to be combined eventually into a monthly index. In the absence of additional information, series which were not on a monthly basis after the above operations could not provide direct information about the behavior of price in the market for the months in which no prices were reported. Nevertheless, in numerous and important cases the series implied information on the missing months which was superior to the use of interpolation by other reporters' prices, which is the basic alternative method of using broken series. These cases are as follows:

(1) A particular series has one or more months with no data. The prices on either side of the gap are identical.

In this case, it is preferable to assume that the price for the missing months was the same as it was in the months before and after the gap than to believe that it rose and then fell, or fell and rose again, to exactly the level it was originally. We have accordingly filled in all such gaps. The slight bias toward price rigidity introduced by this procedure would appear to be more than counterbalanced by the loss of information entailed by not doing it.

(2) A series of prices, each averaged over a quarter or longer period, has two or more consecutive *averages* identical. In such a case, we assume that the same price prevailed in each month covered by the averages, and that the price was equal to the average. Again, it is vastly more probable that the price was constant than that two consecutive averages of fluctuating prices came to the same average.

(3) A single company submits data on an irregular basis for two or more closely related items—e.g., two sizes of brass rod. The dates given do not always coincide; where they do, it appears that one series is a multiple of the other, or differs by a constant amount. In such cases, the several series are combined employing the constant relationship to

make one series with fewer gaps. (Of course, the two series cannot be available only for identical dates or one series would not add to the information in the other.)

In many cases a firm submitted information on closely related items— items which fell under the same product classification used for purposes of constructing index numbers, for example, six sizes of window glass. Such series were combined into one prior to construction of index numbers. This gives any one company the same weight as any other company in an unweighted average. The only problem in the mechanics of combining arose when the series to be combined had observations at different dates but there was evidence that one or more series were not linear transformations of another. Here we had to choose between two alternatives in combining the series: Do two price series for closely related commodities from a particular firm behave more like each other than they behave like the average of all series from all reporters for the commodity in question? We chose to combine series on the basis of other series from the same reporter. For most of the multiple series, no problems arose, fortunately, for they were observations on the same time period for the different series.