

CONTROL AND MEASUREMENT OF SPECIFICATION ERROR IN ESTABLISHMENT SURVEYS

Stanley R. Freedman, Energy Information Administration

KEYWORDS: Nonsampling Error, Federal Surveys, Operational Definitions

Specification error is often described as the difference between the quantity intended to be measured and the data collector's ability to obtain this quantity. It can result from poorly worded questionnaires and survey instructions or may reflect the difficulty of measuring abstract concepts. It most often occurs at the planning stage of the survey.

The purpose of this paper is to describe the sources, control, and measurement of specification error. Some examples are provided from establishment surveys in the Energy Information Administration.

SOURCES OF SPECIFICATION ERROR

There are three sources of specification error discussed in this paper:

- o inadequately specified uses and needs,
- o inadequately specified concepts,
- o inadequately specified data elements.

Inadequately Specified Uses and Needs

Behind every survey is some need for the data. It may be to report on economic conditions, support a legislative program, or allocate Federal funds. Whatever it is, the sponsor of a survey has a use for the data. When the uses and needs documented for a survey do not correspond to the actual uses and needs for the data, specification error occurs.

There are several causes for inadequately specified uses and needs. These include: (1) poorly stated uses and needs by the sponsor, (2) changing uses and needs over time, and (3) the population of inference not corresponding to the population surveyed.

Poorly stated uses and needs--The sponsor of a survey is responsible for specifying the uses of the data. This often requires the sponsor to conduct a special study or data needs assessment to identify data uses. If the uses are poorly defined and not specific, then it will be difficult to correctly specify what data are to be collected. This will result in specification error biasing the data from the outset.

The data collector is also responsible for specifying the uses of the data. Very often the data collector has experience in meeting a specific set of sponsor and user needs, and

knows what kind of data are needed to meet program requirements.

Finally potential users of the data must be consulted in articulating data needs. When a Federal agency sponsors a survey, a notice is published in the Federal Register asking for comments. Not only do potential respondents make comments, but potential users of the data often comment on whether the data will meet their needs. When the needs of other users do not coincide with those of the sponsor, even careful data specification may not satisfy all parties. While not an error in the traditional sense, this can be considered specification error because when one party uses data collected for the other's needs, it will not be properly specified.

Changing uses and needs--Data needs change over time; consequently the data needs must be reexamined on occasion. Even if the needs are clearly and unambiguously stated when the survey is undertaken, periodic review of data requirements is necessary. Changes in business and industry, changes in legislation, and changes in user requirements will affect what data need to be collected.

Population of interest differs from population surveyed--Specification error can occur when the survey respondents are not the same as the population for which the estimates are needed. This can occur when a survey is created for one sponsor and questions are added by another sponsor to save costs associated with creating an entirely new data collection. It can also occur when the population of interest is not obtainable because of frame deficiencies. In these cases the surrogate population is surveyed and estimates are produced. The surrogate population may not be able to answer the questions accurately or in the same way as the "real" population would have. This may not be an error in the strict sense of the word, but it would result in the estimated data measuring something different from what was intended by the survey sponsor.

Inadequately Specified Concepts

Once a need has been identified, it must be stated as a measurable concept. Specification error is the extent to which concepts defined for a survey do not reflect the primary uses and needs for the survey data. This may either be the result of using concepts that are poorly defined or of using existing concepts that do not fit the need.

Poorly defined concepts--Survey concepts must be unambiguously and carefully worded. Suppose an agency needs to know the amount of coal produced annually in the United States. It is critical to consider at the outset whether the types of coal produced--lignite, bituminous, and anthracite--need to be distinguished and whether production is defined as what is "dug out" of the ground or what has been cleaned and prepared for shipment.

Using an existing concept that does not really fit--A poorly specified data need is just as likely to cause specification error as a poorly defined concept. Consider the problem of determining energy consumption. Assume the sponsor or data user is interested in how much energy is used by a particular type of consumer, such as an industrial plant or commercial establishment, at the State level. The concept of interest is end-use consumption. This is most accurately measured by going to the end user. However, this would be very costly and time-consuming because of the large number of end users. Instead a surrogate measure, such as products supplied, may be used because there are fewer energy suppliers than consumers, and because a large number of end users would have to be sampled to get State-level estimates. Nevertheless, inaccuracies may result since supplied energy can be stored for later use or may be resold to other consumers. Thus, using the concept of "product supplied" in lieu of measuring end-use consumption may well introduce error into the estimates.

Inadequately Specified Data Elements

Data elements may be defined for inclusion on the questionnaire in such a way that they do not accurately reflect the survey intention. This is another source of specification error. Inadequate specification of data elements may result from:

(1) ambiguous definitions, (2) elements that do not fully reflect the survey concepts, (3) use of proxy data due to unavailability of primary data, and (4) poorly worded questions.

Ambiguous definitions--Ambiguous definitions may result in respondents reporting different data than was intended by the survey designers. For example, in a survey of crude oil production, it would be important to carefully define the term "crude oil." Otherwise, respondents would be left guessing whether to include lease condensate, a natural gas liquid recovered from gas-well gas, in their crude oil production figures. Because

lease condensate is generally blended with crude oil for refining, some producers might automatically include it in reported volumes of crude oil production. Others might not include it in the reported volumes, or might report it separately. Thus, if crude oil were not clearly defined on the data collection instrument, respondents would likely use varying definitions in reporting production figures. Precise specification then is the key to achieving consistent responses that measure the intended concept accurately.

Elements not reflecting survey concepts--All research entails describing or analyzing certain theoretical concepts. In establishment surveys it might be the money flow among Federally chartered banks, the supply of petroleum products, or the behavior of producer prices in the economy. Before data can be collected and analyzed, these concepts must be reduced to specific, empirical indicators. The data collector must specify observations that may be taken as indicators of the attributes of a given concept. An operational definition must be created that will measure that concept. The process is complicated in establishment surveys because economic statistics are usually byproducts of other business or government activities and have to be collected as part of that process. Thus, data collectors often lack control over what is collected, how it is defined, and how closely the definition conforms to the concept being measured.

Moreover, when several variables are used to create a composite measure, such as a producer price index, the analyst has created a measure of an abstract concept that does not exist in any real economic sense. Error can then result not only from error in the individual variables, but it can be compounded when these statistics are combined.

Proxy data requested due to unavailable primary data--Even where concepts are clearly defined, respondents may be unable to supply the requested data because the data are not available. Another energy-related example involves the disaggregation of natural gas supplied by end-use sector. Generally, utilities keep track of gas supplied by rate class--industrial, commercial, and residential. However, these classes are determined not by the actual activity of the energy consumer, but by the flow rate or amount of energy consumed. This is also how utility rates are determined. Thus, master-metered apartment buildings may get

billed at the commercial rate rather than at the residential rate. As a result, the utility may be unable to provide accurate information broken down by end-use sector even when the sectors are clearly defined. Moreover, because of the great differences in rate classes in different States, inconsistencies between States could lead to errors that would be hard to detect and quantify.

Questionnaire wording, definitions, classification, or instructions--Once an operational definition has been specified, a survey instrument is constructed, questions are formulated, terms are defined, and instructions for completing the questionnaire are written. Ambiguous questions, questions without unique answers, and unclear instructions all cause response errors. Misclassification may occur when respondents are asked to report familiar data in ways that are unfamiliar to them or in inconsistent ways. For example, companies reporting on imported petroleum products are asked to classify commodities one way for the U.S. Customs Service and another way for the Department of Energy. Both schemes have legitimate conceptual foundations, but the disparity in definitions causes difficulty both to the respondents and to the data collectors.

Respondent classification is another major source of specification error, particularly when multi-functional conglomerates are assigned SIC codes, or when parent/subsidiary relationships have to be untangled. Moreover, the risk of double counting increases when data are aggregated from several surveys in which the rules for classification are unclear or inconsistent.

CONTROL OF SPECIFICATION ERROR

Control of specification error relies on the tenets of good questionnaire design as well as some of the techniques used in its measurement. These control mechanisms include: (1) requirements reviews, (2) industry consultation, (3) expert review panels, (4) cognitive studies, and (5) pretests.

Requirements Reviews

A requirements review determines what data in a subject matter area are needed. Potential data users and analysts are contacted to find out if new data are required and how these data would be used. Data that are currently being collected are evaluated to determine if they meet the users' analytical needs. If they do not, this may suggest that the

wrong data are being collected. This can frequently be remedied by changing some of the definitions used in the survey in lieu of collecting new data. The steps involved in conducting a requirements review are:

- (1) assembling available background information on the phenomenon to be measured,
- (2) developing a description of the phenomenon,
- (3) researching and formalizing the evidence from which to infer information requirements,
- (4) generating a matrix of data requirements with relationships mapped to the need for the information,
- (5) developing a rationale for selecting the required data,
- (6) developing the "justified" data requirements by applying the rationale to the data requirements matrix, and
- (7) identifying new data elements or changes in existing elements that need to be implemented.

Industry Consultations

Whenever a new collection or changes to an existing one are proposed, the sponsoring agency should discuss the proposed collection with those who will be supplying the data. This can be done through discussions with trade associations and industry representatives as well as directly with potential respondents. Operational definitions can be discussed, recordkeeping practices reviewed, and data collection methodology explained. Allowing potential respondents to provide input into the data specification process helps insure that the survey elements will be properly specified.

Expert Review Panels

Sometimes it is useful to convene a panel of experts in the subject matter area of the survey to review the specification of data. The panel is usually assigned a specific task--such as a review of definitions of petroleum products or unemployment. The panel's recommendations help ensure that questionnaires and instructions meet the stated objectives of the study and measure what they purport to measure.

Cognitive Studies

Cognitive studies, which are discussed in more detail in the following section on measurement of specification error, can be used both to measure specification error and to control it. In the process of measuring an error, the causes for that error are often uncovered. Steps can then be taken to control the problem by revising the definitions, changing the wording of the questionnaire, or modifying the instructions.

Questionnaire Pretests

Pretesting questionnaires is another activity essential for both measuring and controlling specification error. Identifying and resolving problems with the survey instrument before it is used in a full scale data collection reduces specification error in the final study.

MEASUREMENT OF SPECIFICATION ERROR

Specification error can be measured either directly or indirectly. Direct measurement of the error involves comparing the data value against some benchmark known to be true and accurate. The benchmark need not be the same as the data value, but the difference between the two should be a known constant. A method of direct measurement is a records check study.

Indirect measurement techniques identify discrepancies or possible errors in the data. These techniques establish the existence of an error, often providing a qualitative description of it. An indirect measure can be quantified, but in the absence of a benchmark or "true" value against which to measure its magnitude and direction, the measure is only indirect. Indirect measures include cognitive studies, questionnaire pretests, and comparisons to independent estimates.

Records Check Study

Specification error can be measured directly by checking survey responses against administrative records. This can involve auditing a company's books or matching survey responses against tax records or licensing information. Administrative records are not always available, however, because of privacy restrictions. When reviewing administrative records it is important to determine whether definitions used in recordkeeping are the same as those used in the survey instrument. It is also important to determine whether there is an inherent bias in the recordkeeping because respondents over- or underreport for business or economic reasons.

Cognitive Studies

A cognitive study, or validation study, is an indirect approach to measuring specification error. It entails examining each stage of the data collection process from beginning to end including a review of data requirements, questionnaire and survey frame construction, data processing and editing procedures, nonresponse followup, and data aggregation and publication of results. Examination of each of these stages can detect errors caused by improper operational

definitions. Generally, a site visit to selected respondents is the most useful way for identifying error associated with poor questionnaire design or disparate recordkeeping practices. Actually "walking through" the industrial or commercial process with the respondent is helpful. Seeing at what points the data are collected, how they are measured, and how they are used by the respondent indicate whether the intended concepts are being accurately measured. In many respects this process is similar to a pretest or pilot study, except that it is conducted after a survey is underway.

The disadvantage of cognitive studies is that they are very costly and labor intensive. Moreover, because the review concentrates on a very few respondents, it may be difficult to know whether the identified problems are widespread. This makes it difficult to quantify the magnitude of the errors discovered, even if it is possible to quantify the magnitude for that subset of the respondents.

Questionnaire Pretests

Before a questionnaire is used it should be pretested and the results analyzed in the same way the actual data will be collected and analyzed. Many problems involving unclear definitions, question wording, and instructions will become apparent at this point.

Comparisons to Independent Estimates

A less costly technique for measuring specification error involves comparisons of data series. The data series in question is compared with similar, independent estimates. When the two estimates match, both are usually presumed accurate. When the two estimates differ systematically, it is an indication that one of estimates is biased. Sometimes the "true" value is considered bounded by the two estimates. If there is an indication of bias, one or more of the following procedures is instituted: (1) matching individual respondent records from the two data series, (2) contacting respondents, and (3) contacting the survey managers and data processing specialists to try to determine the source of the bias.

For example, the Energy Information Administration (EIA) recently compared its coal production data with similar coal production data series from other sources. In comparing EIA production data with information from the Mine Safety and Health Administration (MSHA), the MSHA data were found to be systematically lower than the

comparable EIA data. The discrepancy ranged from 4.7 percent in 1978 to 2.6 percent in 1982. The comparisons were then disaggregated by coal type, mine type, and selected States to determine the possible causes for the discrepancies. It turned out that definitional differences in clean versus raw coal accounted for some of the discrepancies in the production figures.

CONCLUSION

Specification error is one of several sources of nonsampling error in established surveys. Other sources

of error include coverage, response, nonresponse, and processing errors. In practice it is often very difficult to classify an error in the data exclusively to one of these categories. The techniques for measuring and controlling these errors also overlap. The critical point is that applying the measurement and control techniques described in this paper will help identify and correct all source of nonsampling errors in established surveys.