11. Errors and Bias in the PPI

A. Introduction

11.1   A number of sources of error and bias have been discussed in the preceding chapters and will be discussed again in subsequent chapters. The purpose of this chapter is to briefly summarize such sources to provide a readily accessible overview. Both conceptual and practical issues will be covered. To be aware of the limitations of any PPI, it is necessary to consider what data are required, how they are to be collected, and how they are to be used to obtain overall summary measures of price changes. The production of PPIs is not a trivial task, and any program of improvement must match the estimated cost of a potential improvement in accuracy against the likely gain. In some

instances, one may have to take into account the user requirements necessary to meet specific needs or engender more faith in the index, in spite of the relatively limited gains in accuracy matched against their cost.

11.2   Figure 11.1 outlines the potential sources of error and bias in PPIs. The distinction between errors and bias is, however, first considered in Section B. In sampling, for example, the nature of the sample design (for example, the use of cutoff sampling—see Chapter 5) may *bias* the sample toward larger establishments whose average item price changes are below the average of all establishments. In contrast, an unrepresentative sample with

Figure 11.1. Outline of Sources of Error and Bias
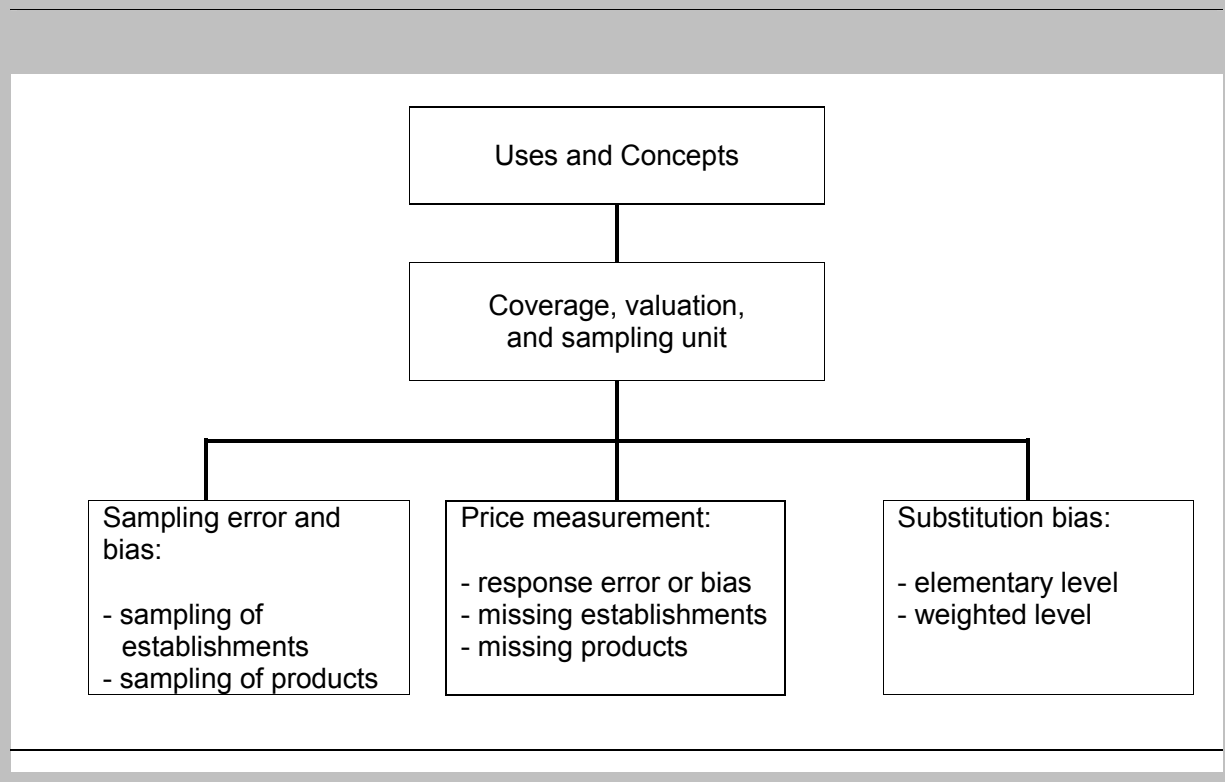
disproportionate larger establishments may be selected by chance and similarly include item prices that are, on average, below those of all establishments. This is *error* since it is equally likely that a sample might have been selected whose average price change was, on average, above those of all establishments

11.3   The discussion of bias and errors first requires consideration of the conceptual framework on which the PPI is to be based and the PPI's related use(s). This will govern a number of issues, including the decision as to the coverage or domain of the index and choice of formula. Errors and bias may arise if the coverage, valuation, and choice of the sampling unit fail to meet a conceptual need; this is discussed in Section C. Section D examines the sources of errors and bias in the sampling of transactions. The sampling of item prices for a PPI is undertaken in two stages: sampling of establishments and the subsequent sampling of items produced (or purchased) by those establishments. Bias may arise if establishments or items are selected with, on average, unusual price changes, possibly due to omissions in the sampling frame or a biased selection from the frame. Sampling error, as discussed previously and in Chapter 5, can arise even if the selection is random from an unbiased sampling frame and will increase as the sample size decreases and as the variance of prices increases. Sampling error arises simply because an estimated PPI is based on samples and not a complete enumeration of the populations involved. The errors and biases discussed in Section D are for the sample on initiation. Section E is concerned with what happens to sampling errors and bias in subsequent matched price comparisons.

11.4   Once the sample of establishments and their items has been selected, the sample will become increasingly out of date and unrepresentative as time progresses. The extent and nature of any such bias will vary from industry to industry. The effect of these dynamic changes in the universe of establishments and the items produced on the static, fixed sample are the subject of Section E. Sample rotation will act to refresh the sample of items, while rebasing may serve to initiate a new sample of items *and* establishments. Establishments will close, and items will no longer be produced on a temporary or permanent basis. Sample augmentation and replacement aid the sampling of establishments, although replacement occurs only

when an establishment is missing. Sample augmentation tries to bring into the sample a new major establishment. It is a more complicated process because the weighting structure of the industry or index has to be changed (Chapter 8). When item prices are missing, the sampling of items may become unrepresentative. Imputations can be used, but they do nothing to replace the sample. In fact, they lower the effective sample size, thereby increasing sampling error. Alternatively, comparable replacement items or replacements with appropriate quality adjustments may be introduced. As for new goods providing a substantively different service, the aforementioned difficulties of including new establishments extend to new goods, which are often neglected until rebasing. Even then, their inclusion is quite problematic (Chapter 8).

11.5   The discussion above has been concerned with how missing establishments and items may bias or increase the error in sampling. But the normal price collection procedure based on the matched-models method may have errors and bias as a result of the prices collected and recorded being different from those transacted. Such response errors and biases, along with those arising from the methods of treating temporarily and permanently missing items and goods, are outlined in Section F as errors and bias in price measurement. Section F is concerned with deficiencies in methods of replacing missing establishments and items so that the matched-models method can continue, while Section E is concerned with the effect of such missing establishments and items on the efficacy of the sampling procedure.

11.6   The final source of bias is substitution bias. Different formulas, as shown in Chapters 15 through 17, have different properties and replicate different effects depending on the weighting system used and the method of aggregation. At the higher level, where weights are used, substitution effects were shown to be included in superlative formulas but excluded in the traditional Laspeyres formula (Chapter 15). Similar considerations were discussed at the lower level. Whether it is desirable to include such effects depends on the concepts of the index adopted. A pure fixed-base period concept would exclude such effects, while an economic cost-of-living approach (Chapters 17 and 20) would include them. The concepts in Figure 11.1 can be used to address definitional issues such as coverage, valuation, and sampling, as well as

price measurement issues such as quality adjustment and the inclusion of new goods and establishments.

11.7 It is worthwhile to list the main sources of errors and bias:

(i) Inappropriate coverage and valuation (Section C);
(ii) Sampling error and bias, including
   a) Sample design on initiation (Section D), and
   b) Effect of missing items and establishments on sampling error (Section E);

(iii) Matched price measurement (Section F), including
   a) Response error/bias,
   b) Quality adjustment bias,
   c) New goods bias, and
   d) New establishments bias; and

(iv) Formula (substitution) bias (Section G), including
   a) Upper-level item and establishment substitution, and
   b) Lower-level item and establishment substitution.

11.8 It is not possible to judge which sources are the most serious. In some countries and industries, the increasing differentiation of items and rate of technological change make it difficult to maintain a sizable, representative matched sample, and issues of quality adjustment and the use of chained or hedonic indices might be appropriate. In other countries, a limited coverage of economic sectors where the PPI is used might be the major concern. Inadequacies in the sampling frame of establishments might also be a concern.

11.9 There is no extensive literature on the nature and extent of errors and bias in PPI measurement, Berndt, Griliches, and Rosett (1993) being a notable exception. However, there is substantial literature on errors and bias in CPI measurement, and Diewert (1998a and 2002c) and Obst (2000) provide a review and extensive reference list. Much of this literature includes problem areas that apply to PPIs as well as CPIs.

B. Errors and Bias

11.10 In this section, a distinction is made between *error* and *bias*. The distinction is most appropriate to the discussion of sampling, although the same framework will be shown to apply to nonsampling errors and bias. Yet an error or bias can also be discussed in terms of how an existing measure corresponds to some true concept of a PPI and will vary depending on the concept advocated, which in turn will depend on the use(s) required of the measure. These issues are discussed in turn.

B.1   Sampling error and bias

11.11 Consider the collection of a random *sample* of prices whose overall population average (arithmetic mean) is μ.[1] The *estimator* is the method used for estimating μ from sample data. An appropriate estimator for μ is the mean of a *sample* drawn using a random design. An *estimate* is the value obtained using a specific sample and method of estimation, let us say $\bar{x}_1$, the sample mean. The population mean μ, for example, may be 20, but the arithmetic mean from a sample of a given size drawn in a specific way may be 19. This *error* may not be *bias*, it may simply be that by chance a random sample was drawn with, on average, below-average prices. If an infinite number of samples were drawn using sufficiently large samples, the average of the $\bar{x}_1$, $\bar{x}_2$, $\bar{x}_3$, ... sample means would in principle equal μ. The estimator is said to be *unbiased*; if it is not, it is called *biased*. The error caused by $\bar{x}_1$ being different from $\mu =$ 20 did not arise from any systematic under- or over-estimation in the way the sample was drawn and the average calculated. If an infinite number of such estimates were drawn and summarized, no error would be found, the estimator not being biased and the discrepancy being part of the usual expected sampling error.[2]

---

[1]The discussion is in terms of prices and not price *changes* for simplicity.

[2]This is *sampling error,* which can be estimated as the differences between upper and lower bounds of a given probability, more usually known as *confidence intervals.* Methods and principles for calculating such bounds are explained in Cochran (1963), Singh and Mangat (1996), and most introductory statistical texts. Moser and Kalton (1981)

11.12 It should be stressed that any one sample may give an inaccurate result, even though the method used to draw the sample and calculate the estimate is, *on average*, unbiased. Improvements in the design of the sample, increases in the sample size, and less variability in the prices (more detailed price specifications for the price basis) will lead to less error, and the extent of such improvements in terms of the sample's probable accuracy is measurable. Note that the accuracy of such estimates is measured in principle by confidence intervals, that is, probabilistic bounds in which μ is likely to fall. Smaller bounds at a given probability are considered to be more *precise* estimates. It is in the interest of statistical agencies to design their sample and use estimators in a way that leads to more precise estimates.

11.13 The calculation of such intervals requires a measure of the variance of a PPI in which all sources of sampling error are caught. However, the sampling of prices involves sampling of establishments and items, and probabilistic methods generally are not used at each stage. Judgmental and cutoff methods are often considered to be more feasible and less resource intensive. Estimates of the variance, however, require probabilistic sample designs at all stages. Yet it is feasible to develop partial (conditional) measures in which only a single source of variability is quantified (see Balk and Kerston, 1986, for a CPI example). Alternative methods for nonprobability samples are discussed in Särndal, Swensson, and Wretman (1992).

11.14 *Efficiency* gains (smaller sampling errors) may be achieved for a given sample size and population variance by using better *sample designs* (methods of selecting the sample) as outlined in Chapter 5. Yet it may be that the actual selection probabilities deviate from those specified in the sample design. Errors arising from such deviations are called *selection errors*.

11.15 While an unbiased estimator may give imprecise results, especially if small samples are used, a biased estimator may give quite precise results. Consider the sampling from only large establishments. Suppose such prices were, on average, less than μ, but assume these major establishments covered a substantial share of the revenue of the industry concerned, then the mean of the estimates from all such possible samples $\underline{m}$ may be quite close to μ, even if smaller establishments had different prices. However, the difference between m and μ would be of a systematic and generally predictable nature. On average, m would exceed μ, the bias[3] being $(\mu - m)$.

B.2 Nonsampling error and bias

11.16 The above framework for distinguishing between errors and biases is also pertinent to nonsampling error. If, for example, the prices of items are incorrectly recorded, a response error results. If such errors are *unsystematic*, then prices are overrecorded in some instances but, counterbalancing this, underrecorded in others. Overall, errors in one direction should cancel out those in the other, and the net error, on average, will be expected to be small. If, however, the establishments selected and kept in the sample are older and produce at higher (quality-adjusted) prices than their newer, high-technology equivalent establishments, then there is a systematic bias. The results are biased in the sense that if an infinite number of similar random samples of older establishments were taken from the population of establishments, the average or expected value of the results would differ from the true population average, and this difference would be the bias. The distinction is important. Increasing the sample size of a biased sample, of older establishments for example, when samples are rebased reduces the error but not the bias.

11.17 This distinction between errors and bias is for the purpose of *estimation*. When using the results from a sample to estimate a population parameter, both error and bias affect the accuracy of the results. Yet there is also a distinction in the statistical literature between types of errors according to their *source*: sampling versus nonsampling (response, nonresponse, processing, etc.) error. Although they are both described as errors, the distinction remains that if their magnitude cannot be estimated from the sample itself, they are biases, and some estimate of μ is required to measure them. If they can be estimated from the sample, they are errors.

provide a good account of the different types of errors and their distinction.

---

[3]Since μ is not known, estimates of sampling error are usually made; they are but one component of the variability of prices around μ.

B.3   Concepts of a true or good index

11.18   The discussion of errors and bias so far has been in terms of estimating μ as if it were the required measure. This has served the purpose of distinguishing between errors and bias. However, much of the *Manual* has been concerned with the choice of an appropriate index number formula. It is now necessary to consider bias in terms of the difference between the index number formula and methods used to calculate the PPI and some concept of a true index. In Chapter 17, *true* theoretical indices were defined from economic theory. The question is, if producers behave as optimizers and switch production toward products with relatively high price increases, which would be the appropriate formula to use? The result was a number of superlative index number formulas. They did not include the Laspeyres index or the commonly used Young index (Chapter 15), which give unduly low weights to products with relatively high price increases because no account is taken of substitution effects (see Chapter 17). For industries whose establishments behave this way, Laspeyres is biased downward. An understanding of bias thus requires a concept of a true index. According to economic theory, a true index makes assumptions about the nature of economic behavior of industries. These presuppositions dictate which formulas are appropriate and, given these constructs, determine if there is any bias.

11.19   A good index number formula can be defined by axiomatic criteria as outlined in Chapter 16. The Young and Carli indices, for example, were argued to be biased upward since they failed the time reversal test; the product of the indices between periods 0 and 1 and periods 1 and 0 *exceeded* unity.

11.20   In PPI number theory and practice there are quite different conceptual approaches. On the one hand, there is the revenue-maximizing concept defined in economic theory mentioned above. On the other hand, there is the fixed-basket approach.[4] An index based on the latter approach would not suffer, in the strictest sense of the concept, from the biases of substitution (formula) or new goods because the concept is one of measuring the prices

---

[4]A discussion of the debate is in Triplett (2001).

of a *fixed* basket of goods. However, it may be argued on the grounds of representativeness that the baskets should be updated and substitution effects incorporated.

C.   Use, Coverage, and Valuation

11.21   Errors and biases can arise from the inappropriate use of a PPI, regardless of the methodology used to compile it. Since price changes can vary considerably from product to product, the value of the price index will depend partly on which products or items are included in the index and how the item prices are determined (Chapter 15, Section B.1). In Chapter 2, different uses of the PPI were mentioned and aligned with different domains and valuation principles. Thus, the discussion of errors and biases starts with a need to decide whether the coverage and valuation practices are appropriate for the purposes required.

11.22   In general terms, a PPI can be described as an index designed to measure either the average change in the price of goods and services as they leave the place of production or as they enter the production process. Thus, producer price indices fall into two clear categories: input prices (at purchasers' prices) and output prices (at basic prices). In Chapter 15, a value-added deflator was described as a further PPI. This is used to deflate the value of a sector or economy, with outputs less the value of the intermediate inputs used to produce the output. First, some major uses are noted, and the domain or coverage of the index is considered. Second, the principles of valuation are reiterated.

C.1   Uses and coverage

11.23   The input PPI is a short-term indicator of inflation. It tracks potential inflation as price pressure builds up and goods and services enter the factory gate. Output PPIs or PPIs at different stages of production show how price pressure evolves up to the wholesaler and retailer. They are *indicators* of *producer* price inflation excluding the effect of price pressure from imports and including that which goes into exports. Separate import and export PPIs should form part of the family of PPIs. There may be a deficiency in the coverage of a PPI. If, for example, an output PPI is restricted to the industrial sector, this is a source of error when examining overall inflation if price changes

for other sectors that differ from the industrial sector.

11.24 The PPI indices may be biased when used for national accounts deflation. First, their coverage may be inadequate yet still be used by national accountants. For example, if only a manufacturing PPI is used to deflate industrial output, and price changes from the missing, quarrying, and construction sectors differ in the aggregate from those of manufacturing, there is a bias. The *undercoverage* bias is in the *use* of the index, not necessarily in its construction, although statistical agencies should be sensitive to the needs of users. Second, overcoverage bias means some elements are included in the survey that do not belong to the target population. The bias surfaces if their price changes differ on aggregation from the included ones. Third, the classification of activities for the PPI should be at an appropriately low level of disaggregation, and the system of classification should be the same as that required for the production accounts under the *1993 SNA*. Finally, the use of the Laspeyres PPI formula as a deflator induces a bias, since a Paasche formula is theoretically appropriate (see Chapter 18) for the measurement of changes in output at constant prices. The extent of the bias will also increase as the weights become more out of date.

11.25 Highly aggregated PPIs are used for the macroeconomic analysis of inflation. Certain industries or products with volatile price changes may be excluded. Such indices may be excluded because they introduce substantial sampling error into the aggregate indices, and their exclusion helps with the identification of any underlying trend.

11.26 The preceding discussion has considered the coverage or domain of the index in terms of the activities included. However, such issues also may extend to the geographic scope. The exclusion of establishments in rural areas, for example, may lead to bias if their price changes differ from those in urban areas. Such issues are considered in Sections D and E under *sampling*.

C.2 Valuation

11.27 The valuation of an output PPI is to value output at basic prices with any VAT or similar deductible tax, invoiced to the purchaser, excluded. Such tax revenues go to the government and

should be excluded because they are not part of the establishment's receipts. Transport charges and trade margins invoiced separately by the producer should also be excluded. An input PPI should value intermediate inputs with nondeductible taxes included, since they are part of the actual costs paid by the establishment. For input PPIs, changes in the tax procedures—due to a switch to import duties on intermediate inputs, for example—can lead to bias. In such instances, ex-tax or ex-duty indices might be produced. In any event, it is necessary to ensure that establishments treat indirect taxes in a consistently appropriate way, especially when such tax rates fluctuate.

D. Sampling Error and Bias on Initiation

11.28 In Chapter 5, appropriate approaches to sample design were outlined. The starting point for potential bias in sample design is an inadequate sampling frame. It is one of the most pernicious sources of error because the inadequacies of a sampling frame are not immediately apparent to users. Yet a sampling frame biased to particular sizes of establishments or industrial sectors will yield a biased sample irrespective of the probity of the sample selection. Since sampling is generally in two stages—the sampling of establishments and the items within establishments—a sampling frame is required for establishments and for items within establishments. The latter relies on the establishment producing data on the revenues, quantities, and prices (or revenue per unit of output) for the items produced. Any bias here, perhaps because some components produced are priced and recorded at the head office, may lead to bias. It should be kept in mind that even when purposive sampling is used, there is an implicit frame from which the respondent selects items. It should be clear to the respondent what the frame should be.

11.29 The selection of the sample of establishments from the sampling frame should be random or, failing that, purposive. In the latter case, the aim should be to include major items whose price changes are likely to represent overall price changes. Chapter 5 provided a fairly detailed account of the principles and practice of sample selection and the biases that may ensue. The distinction has already been drawn between *bias* and sampling *error,* and the possibility has been raised that unbiased selection will be accompanied by es-

timates with substantial error, due to high variability in the price (change) data and relatively low sample sizes.

E.   Sampling Error and Bias: The Dynamic Universe

11.30   Chapters 7 and 8 also considered sampling issues. Under the matched-models method, prices will be missing in a period if the item is temporarily or permanently out of production. If overall imputations are used to replace the missing prices, the sample size is being effectively reduced and the sampling error increased. In a comparison between prices in period 0 and period *t,* imputation procedures (Chapter 7) ignore the prices in period 0 of items whose prices are missing in period *t*. If such old prices of items no longer produced differ from other prices in period 0, there is a bias due to their exclusion. Similarly, new items produced after period 0, and thus not part of the matched sample, are ignored; if their prices in period *t* differ, on average, from the prices of matched items in period *t*, there is a bias. Sampling error and bias, therefore, may arise due to the exclusion of prices introduced after initiation and dropped when they go missing. This is over and above any errors and bias in the sample design on initiation. Its concern is ensuring that the sample is representative of the dynamic universe.

11.31   As the sample of establishments and items deteriorates, the need for rebasing the index—to update the weights and sample of establishments and items or the rotation of sample items— becomes increasingly desirable. However, these are costly and irregular procedures, and, for some industries, more immediate steps are required. Rebasing and sample rotation are used to improve the sampling of establishments and items. Strategies for dealing with missing establishments and missing prices also have an effect on the sampling of establishments and items. Such strategies involve introducing *replacement* establishments and items that replenish the sample in a more limited way than rebasing and sample rotation. Quality adjustments to prices are required if the replacement establishment or item differs from the missing ones, although this is the concern of price measurement bias in Section F. New establishments and goods may also need to be incorporated into the sample to avoid sampling bias. There is a need in such instances to *augment* the sample. Such augmentation

may require a change to the weighting system and, as discussed in Chapter 8, should be undertaken only when the incorporation of major new establishments or goods is considered necessary. Thus, bias in sampling due to differences between the dynamic universe and the static one on initiation may, to some extent, be militated by sample replacement and augmentation (Chapter 8).

11.32   Circumstances may arise in which there is a serious sample deterioration due to missing items as differentiated items rapidly turn over. In such cases, hedonic indices or chaining based on resampling the universe each month was advised in Chapter 7, Section G.

F.   Price Measurement: Response Error and Bias, Quality Change, and New Goods

F.1   Response error and bias

11.33   Errors may happen if the reporting or recording of prices is inaccurate. If the errors occur in a systematic manner, there will be bias. The item descriptions that define the price basis should be as tightly specified as is reasonably possible, so that the prices of like items are compared with like. Allowing newer models to be automatically considered comparable in quality introduces an upward bias if quality is improving. Similar considerations apply to improvements in the service quality that accompanies an item. The period to which the prices relate should be clearly indicated, especially where prices vary over the month in question and some average price is required (Chapter 6). Errors in valuation can be reduced by clear statements of the basis of valuation and discussions with respondents if the valuation principles of their accounting systems differ from the valuation required. This is of particular importance when there are changes in tax rates or systems. Diagnostic checks for extremely unusual price changes should be part of an automated quality assurance system, and extreme values should be checked with the respondent and not automatically deleted. Price collectors should visit establishments on initiation and then periodically as part of a quality assurance auditing program (see Chapter 12).

F.2   Quality change bias

11.34   Bias can arise, as discussed in Section E, because newly introduced items do not form part of the matched sample, and their (quality-adjusted) prices may differ from those in the matched sample. This *sampling* bias from items of improved quality and new goods was the subject of Section E. It was also noted that statistical agencies may deplete the sample by using imputation or use replacements to replenish the sample. The concern here is with the validity of such approaches for price measurement, not their effects on sampling bias.

11.35   In Chapter 7, a host of explicit and implicit quality adjustment methods were outlined. From a practical perspective, the quality change problem involves trying to measure price changes for a product that exhibited a quality change. The old item is no longer produced, but a replacement one or alternative is there. If the effect of quality on price is, on average, either improving or deteriorating, then a bias will result if the prices are compared as if they were comparable when they are not. An explicit quality adjustment may be made to the price of either of the items to make them comparable. A number of methods for such explicit adjustments were outlined in Chapter 7, including expert judgment, quantity adjustment, option and production costs, and hedonic price adjustments. If the adjustment is inappropriate, there will be an error, and, if the adjustments are inappropriate in a systematic direction, there will be a bias. For example, using quantity adjustments to price very small lots of output, for which customers pay more per unit for their convenience, would yield a biased estimate of the price adjustment due to quality change (Chapter 7, Section E.2).

11.36   There are also implicit approaches to quality adjustment. These include the overlap approach overall and targeted mean imputation; class mean imputation; comparable substitution, spliced to show no price change; and the carryforward approach. Imputations are widely used, whereby the price changes of missing items are assumed to be the same as those of the overall sample or some targeted group of items. Yet such approaches increase error through the drop in sample size and may lead to bias if the items being dropped are at stages in their life cycle where their pricing differs from that of other items. Such bias is usually taken

to overestimate price changes (Chapter 7, Section D).

11.37   The choice of appropriate quality adjustment procedure was argued in Chapter 7 to vary among industries to meet their particular features. There are some products, such as consumer durables, materials, and high-technology electronic products, in which the quality change is believed to be significant. If such products have a significant weight in the index, overall bias may arise if such changes are ignored or the effects of quality change on price is mismeasured. Whichever of the methods are used, an assumption is being made about the extent to which any price change taking place is due to quality; bias will ensue if the assumption is not valid.

F.3   New-goods bias

11.38   Over time, new goods (and services) will appear. These may be quite different from what is currently produced. An index that does not adequately allow for the effect on prices of new goods may be biased. Introducing new goods into an index is problematic. First, there will be no data on weights. Second, there is no base-period price to compare the new price with. Even if the new good is linked into the index, there is no (reservation) price in the period preceding its introduction to compare with its price on introduction. Including the new good on rebasing will miss the price changes in the product's initial period of introduction, and it is in such periods that the unusual price changes are expected if the new good delivers something better for a given or lower price. Similar considerations apply to new establishments (Section G.4). New-goods and new-establishment bias is assumed to overstate price changes, on average.

F.4   Temporarily missing bias

11.39   The availability of some items fluctuates with the seasons, such as fruits and vegetables. A number of methods are available to impute such prices during their missing periods. Bias has been shown to arise if inappropriate imputation approaches are used. Indeed, if seasonal items constitute a large proportion of revenue, it is difficult to give meaning to month-on-month indices, although comparisons between a month and its counterpart in the next year will generally be meaningful (see Chapter 22).

G. Substitution Bias

11.40   Given the domain of an index and the valuation principles, the value of the revenue accruing to the establishment can be compared over two periods, let us say, 0 and 1. It is shown in Chapter 15 that the change in such values between periods 0 and 1 can be broken down into two components: the overall price and overall quantity change. An index number formula is required to provide an overall, summary measure of the price change. In practice, this may be undertaken in two stages. At the higher level, a weighted average of price changes (or change in the weighted average of prices) is compiled with information on revenues (quantities) serving as weights. At the lower level, the summary index number formulas do not use revenue or quantity weight, and use only price information to measure the elementary aggregate indices of average price changes (or changes in average prices). It is recognized that in many cases, only weighted calculations are undertaken. Five approaches were used in Chapters 15 through 17 to consider an appropriate formula at the higher level, a similar analysis being undertaken for lower-level elementary aggregate indices in Chapter 20.

G.1   Upper-level substitution bias

11.41   Different formulas for aggregation have different properties. At the upper-weighted level, substantial research from the axiomatic, stochastic, Divisia, fixed-base, and economic approach has led to an understanding of the bias implicit in particular formulas. Chapters 15 through 17 discuss such bias in some detail. The Laspeyres formula is generally considered to be used for PPI construction for the practical reason of not requiring any current-period quantity information. It is also recognized that the appropriate deflator that generates estimates of output at constant prices is a Paasche one (Chapter 18). Thus, if estimates of a series of output at constant prices is required, the use of Laspeyres deflator will result in bias. In practice, for a price comparison between periods 0 and t, period 0 revenue weights are not available, and a Young index is used, which weights period 0 to t price changes by an earlier period b revenue shares. Chapter 15 finds this index to be biased. Superlative index number formulas, in particular the Fisher and Törnqvist indices, have good axiomatic properties and can also be justified using the fixed-base, stochastic, and economic approaches.

Indeed, Laspeyres can be shown to suffer from *substitution* bias if particular patterns of economic behavior are assumed. For example, producers may seek to maximize revenue from a given technology, and inputs may shift production to items with above-average relative price increases. The Laspeyres formula, in holding quantities constant in the base period, does not incorporate such effects in its weighting, giving unduly low weights to items with above-average price increases. Therefore, it suffers from a downward bias. It can be similarly argued that the fixed, current-period weighted Paasche index suffers from an upward bias, while the Fisher index is a symmetric mean of the two, falling within these bounds. Calculating the Fisher index retrospectively on a trailing basis will give insights into upper-level substitution bias.

11.42   The extent of the bias depends on the extent of the substitution effect. The Laspeyres index is appropriate if there is no substitution. However, the economic model assumes that the technology of production is the same for the two periods being compared. If, for example, the factory changes its technology to produce the same item at a lower cost, the assumptions that dictate the nature and extent of the bias break down.

G.2   Lower-level substitution bias

11.43   In some countries or industries, elementary aggregate indices at the lower level of aggregation are constructed that use only price information. The prices are aggregated over what should be the same item. In practice, however, item specifications may be quite loose and the price variation between items being aggregated quite substantial.

11.44   The axiomatic (test), stochastic, and economic approaches can also be applied to the choice of formula on this lower level (Chapter 20). The Carli index, as an arithmetic mean of price changes, performed badly on axiomatic grounds and is not recommended. The Dutot index, as a ratio of arithmetic means, was shown to be influenced by the units of measurements used for price changes and is not advised when items do not meet tight quality specifications. The Jevons index, as the geometric mean of price changes (and equivalently, the ratio of geometric means of prices), performed well when tested by the axiomatic approach but incorporates a substitution effect that goes the opposite way to that predicted by the aforementioned economic model. It has an implicit

unitary elasticity, which requires revenues to remain constant over the periods compared. For a consumer price index, the economic model is one of consumers substituting *away* from items with above-average price increases so more of the relatively cheaper items are purchased. Constant revenue shares is an appropriate assumption in these circumstances. However, producer theory requires producers to substitute *toward* items with above-average price increases, and assumptions of equal revenues are not tenable. Chapter 20 details a number of formulas with quite different properties. However, it concludes that since the axiomatic, stochastic, fixed-base, and economic approaches, as noted in Section G.1, find superlative index numbers to be superior (Chapters 15 to 17), a more appropriate course of action is to attempt to use such formulas at the lower level, rather than replicate their effects using only price data, a task to which they are unsuited. Respondents should be asked to provide revenue or quantity data as well as price data. Failing that, an appropriate index number formulas are advocated depending on the expected nature of the substitution bias.

G.3   Unit-value bias

11.45   Even if quantity or revenue data were available at a detailed item level, there is still potential for bias due to the formula used to define prices. If an establishment produces thousands of an item each day, the price may not be fixed. Minor variations in the nature of what is produced may affect the price if it is estimated as the total revenue divided by the quantity produced. If production moves to higher-priced items, then average prices will increase simply because of a change in the mix of what is produced; there will be an upward bias.

G.4   New-establishment (substitution) bias

11.46   The need to include new establishments in the sample has already been referred to in Section E under *sampling bias*. Products produced by new establishments may not only have different (usually lower) prices, arguing for their inclusion in the sample, but gain increasing acceptance as purchasers substitute goods from new establishments for goods from old establishments. Their exclusion may overstate price changes. When an establishment in the sample closes, an opportunity exists to replace it with a new establishment, thus militating against sampling bias as discussed in Section E. However, the quality of not only the item being replaced, but also the level of service, geographical convenience, and any other factors surrounding the terms of sale, must be considered in any price comparison to ensure that the pricing is for a consistently defined price basis.

11.47   The sections above are merely an overview of the sources of error and bias and are intended to be neither exhaustive nor detailed accounts. The detail is to be found in the individual chapters concerned. The multiplicity of such sources argues for statistical agencies undertaking audits of their strengths and weaknesses and formulating strategies to counter such errors and bias in a cost-effective manner.