Quality Assessment for Price Indexes:
Statistics Canada's Performance Measure Grading Scheme

By:

Kate Burnett-Isaacs and Jean-François Côté[1]

Prepared for the 32nd Meeting of the Voorburg Group on Service Statistics

New Delhi, India

October, 2017

[1] Kate Burnett-Isaacs: Producer Prices Division, Statistics Canada, Ottawa, Canada, K1A 0T6 (email: kate.burnett-isaacs@canada.ca); Jean-François Côté: Producer Prices Division, Statistics Canada, Ottawa, Canada, K1A 0T6 (email: jean-françois.coté6@canada.ca); The authors would like to thank Jean-Sébastien Provençal, Kathleen MacEachern, Steve Matthews, Wesley Yung, Vasuki Tharmarajan and Steve Morrison for their contribution to the concepts of this paper.

Quality Assessment for Price Indexes: Statistics Canada's Performance Measure Grading Scheme

The Producer Prices Division (PPD) at Statistics Canada has developed a Performance Measure Grading Scheme to evaluate each PPD index on key performance indicators to promote sound methodological practices and convey overall quality and reliability of published index numbers. This grading scheme was developed to improve the program assessment tools for the divisional Performance Measurement Strategy and aligns with recommendations of the agency wide Quality Assurance Review Committee. Its components were drawn from the OECD Generic Statistical Business Process Model and Statistics Canada's six dimensions of quality. Assessing the quality of an index is multi-faceted because of the complexities of index numbers and calculations and the different components of index compilation. An index number is comprised of price relatives, weights and a variety of forms of treatments to these data. The quality of an index must be assessed on the individual parts as well as the whole. The grading scheme intends to capture this and provide a measure of quality for the entire index, as well as its individual components, starting from a qualitative conceptual assessment, to a quantitative processing perspective. PPD produces over 25 indexes that cover a wide scope of the business sector including goods production and manufacturing, construction, financial, transportation and professional services. These industries each have their own sources of data and standards of price measurement. The diversity of PPD's index coverage brings with it a complexity when developing a standard method to assess data quality. This paper will discuss the complexities of measuring data quality for indexes, explain the development of the grading scheme and choice of performance measures and highlight the challenges of making a standard measure of quality for the broad range of producer price indexes published by Statistics Canada.

Keywords: Data quality, price indexes, quality assessment, standardization.

1. Introduction

The Producer Prices Division (PPD) at Statistics Canada publishes 25 indexes and is continually expanding its portfolio. The program includes Service Producer Price Indexes along with indexes covering the goods production sector and the construction sector. These price indexes have a wide range of uses, including deflation of the macroeconomic accounts and by external users for contract escalations. It is important that the price indexes made by PPD adhere to the latest standards and practices so that they meet users' needs and reflect the current market conditions.

Statistics Canada is committed to transparency of concepts, methods and practices used to estimate its statistics and how well these statistics are fit for user needs. Statistics Canada has a department-wide Policy on Informing Users of Data Quality and Methodology to ensure users have the necessary information to determine the accuracy and overall quality of the statistics.

Assessing the quality of an index is multi-faceted because of the complexities of index numbers and calculations and the different components of index compilation. Due to these complexities, standard means of measuring quality, such as response rates and coefficients of variance are not sufficient to capture the overall quality of an index. In this light, PPD developed the Performance Measure Grading

Scheme with the purpose of evaluating each index on key performance measures in order to promote sound methodological practices and overall quality and reliability of published index numbers.

This paper describes how the Performance Measure Grading Scheme is structured (section 2), the intuition behind specific grading criteria and how the grading is conducted (section 3), and finally the major challenges in designing a grading scheme for price indexes and specifically for the large variety of prices indexes produced by PPD (section 4). This paper provides an overview of the Performance Measure Grading Scheme; additional details can be found in appendix 1.

2. Structure of the Performance Measure Grading Scheme

The structure of the Performance Measure Grading Scheme highlights the main areas of quality measurement and is based on Statistics Canada's six dimensions of quality. Below is a brief description of each of the six dimensions[2]:

> **Relevance**: The relevance of statistical information reflects the degree to which it meets the real needs of users. It is concerned with whether the available information sheds light on the issues of most importance to users. The assessment of relevance needs to take into account the varying needs of users.
>
> **Accuracy**: The accuracy of statistical information is the degree to which the information correctly describes the phenomena it was designed to measure. It is usually characterized in terms of error in statistical estimates and is traditionally decomposed into bias (systematic error) and variance (random error) components. It may also be described in terms of the major sources of error that potentially cause inaccuracy (e.g., coverage, sampling, non-response, response).
>
> **Timeliness**: The timeliness of statistical information refers to the delay between the reference point (or the end of the reference period) to which the information pertains, and the date on which the information becomes available. It is typically involved in a trade-off against *accuracy*. The *timeliness* of information will influence its *relevance*.
>
> **Accessibility**: The accessibility of statistical information refers to the ease with which it can be obtained by users. This includes the ease with which the existence of information can be ascertained, as well as the suitability of the form or medium through which the information can be accessed. The cost of the information may also be an aspect of *accessibility* for some users.
>
> **Interpretability**: The interpretability of statistical information reflects the availability of the supplementary information and metadata necessary to interpret and utilize it appropriately. This information normally covers the underlying concepts, variables and classifications used, the methodology of collection, and indicators of the accuracy of the statistical information. This Policy aims to ensure the interpretability of our information.
>
> **Coherence**: The coherence of statistical information reflects the degree to which it can be successfully brought together with other statistical information within a broad analytic

[2] Statistics Canada (2000). Policy on Informing Users of Data Quality and Methodology. http://www.statcan.gc.ca/eng/about/policy/info-user

framework and over time. The use of standard concepts, classifications and target populations promotes coherence, as does the use of common methodology across surveys. *Coherence* does not necessarily imply full numerical consistency.

The number of performance measures in each dimension/section varies, with the largest number in Accuracy, followed by Relevance. The basis for the performance measures themselves are founded upon the guidelines in the UNECE/Eurostat/OECD Generic Statistical Business Process Model (GSBPM) for producing official statistics. The Scheme focuses mainly on the Design Phase of the GSBPM.[3] This phase covers development and design activities needed to define the statistical outputs, concepts, methodologies, and operational processes. This phase usually occurs when developing a brand new index, for first publication of the index, or when improvements are made, for example during basket updates or survey re-designs.

More specifically, three sub-processes of the Design Phase - Design Variable process, Design Frame and Sample Methodology process and Design Statistical Processing Methodology - form the basis for the performance measures.[4] The Design Variable sub-process, as named, defines the statistical variables to be collected. In the case of price indexes, these variables could be the prices themselves or any other variables used in index estimation, such as revenues, quantities and quality adjustment factors. The Design Frame and Sample Methodology sub-process identifies the population of interest, sampling frame, and sampling methodology. The Design Statistical Processing Methodology sub-process examines the processes and methodologies of coding, editing, imputing, estimating, integrating, validating and finalizing data.

The Process Phase of the GSBPM is also touched upon in the Performance Measure Grading Scheme through criteria that looks at imputation methodology, weight calculations and methods to calculate aggregates. There is only one specific performance measure for the Collect phase of the GSBPM in the Performance Measure Grading Scheme (the collection response rate), however the source and method of collection (survey versus administrative) does play a role in determining the wording and grading of many performance measures. This topic will be expanded on in the Challenges section of this paper.

The GSBPM is a model applicable to all statistical products, but how does it apply to price indexes? From start to finish, there are six stages to creating price index statistics:

1. Sample and variable design
2. Collection
3. Microdata processing
4. Index estimation
5. Validation
6. Dissemination

[3] See appendix 2 for an outline of the GSBPM.
[4] http://www1.unece.org/stat/platform/display/GSBPM/GSBPM+v5.0

The first four stages of index compilation are assessed in the Performance Measure Grading Scheme. Other initiatives by the Producer Prices Division address stages 5 and 6.[5] Table 1 highlights the key activities of stages 1 through 4 and outlines how the GSBPM phases relate to each stage.

Table 1: Index calculation stages and their related GSBPM phase

Index Compilation Stage	Description of Stage	GSBPM Phase
Sample and variable design	At this stage, the definitions of the prices and other variables, such as weights and classifications, are identified. The population, sample size, frame and methodology are determined.	Design Variable Process and Design Frame and Sample Methodology Process phases
Collection	The activity of collection, whether through survey, internet collection or administrative sources	Collect Phase
Microdata processing	The activities used to quality adjust, impute and create derived variables	Process Phase
Index estimation	The method of aggregating price and weight data to result in an index number	Process Phase

The six dimensions of quality – relevance, accuracy, timeliness, accessibility, interpretability and coherence - transcend the stages of index compilation and the GSBPM phases. Relevance plays an important role in the sampling frame and methodology and variable definitions. Accuracy is a crucial part of all index compilation stages of microdata processing, index estimation and even components of collection, sampling and dissemination. Timeliness is an important factor of the collection and dissemination phase and accessibility, interpretability and coherence are important factors of quality in the variable design and dissemination phases. It is for these reasons that both the quality dimensions and GSBPM phases were used to form the underlying direction of the Performance Measure Grading Scheme.

3. Grading

Grading specific performance measures was a difficult task because there is no standard method to do so. International guidelines outline preferred methods for price definitions, basket update frequencies and index formulas to use, but quantifying this preference is new ground.

The task of assigning a grade scale to performance measures was based on Statistics Canada's data accuracy rating from the Policy on Informing Users of Data Quality and Methodology.[6] The policy states

[5] A Data Validation Checklist has been created to assess the quality of the Validation Stage. A Dissemination Checklist has been created to assess the quality of the quality of the Dissemination Stage.

[6] Statistics Canada (2000). Policy on Informing Users of Data Quality and Methodology. http://www.statcan.gc.ca/eng/about/policy/info-user

that accuracy ratings should be based on expert judgment or analysis. The goal when creating grading criteria is to limit subjectivity in the evaluation of each measure. In many cases, international standards and guidelines fulfilled the role of expert judgment. In particular, the Methodological Guide for Developing Producer Price Indices for Services was consulted. [7] In other cases, index and industry experts within Statistics Canada were consulted to provide a hierarchy of which methods or practices deserve a higher grade. The accuracy of the data for a given measure or the level of confidence with which that data may be used determined the specific grade a certain method, definition, procedure or rate received.

3.1 Grading each performance measure

The evaluation of each performance measure is based on the risk level that the index is not representing reality to users. This method of grading was adopted from Statistic Sweden's Index Valuation Model.[8] The risk levels are defined as:

1 = High risk; the index likely has a clear bias, index review required

2 = Medium risk; index should be representative in the long term but may be misleading in some periods; index review required

3 = Ok, approved (not prioritised for immediate review)

4 = Low risk. High confidence in the representativeness of this index; not a review priority

5 = Very low risk. Not a review priority.

The risk level was chosen as the grading criteria to highlight the relevance and accuracy of the index measures and to foster an approach of improvement in methods and practices and not disciplining areas for poor performance.

3.2 Overall grade

Each dimension of quality will have a section grade, calculated by taking the average of the points achieved for each performance measure within the section. An average is chosen as the means to determine the section grade in order to not penalize those indexes for which a performance measure does not apply. For example, an index that uses an administrative data source will not have a collection response rate.

An overall grade for the index is derived by taking the weighted average of the section grades. The weights used in the Performance Measure Grading Scheme are highlighted in table 2:

[7] Eurostat-OECD (2014). Eurostat-OECD Methodological Guide for Developing Producer Price Indices for Services. 2nd ed.

[8] Draper, R., & Schoultz, C. – Statistics Sweden (2016). *A model for index auditing – the Swedish experience*. Presentation, Nordic SPPI Seminar, 19-20 May 2016, Copenhagen.

Table 2: Weighting of the Performance Measure Grading Scheme

Section	Section Name	Number of Measures	Weight
1	Relevance	6	35
2	Accuracy	15	40
3	Timeliness	2	10
4	Accessibility	1	5
5	Interpretability	1	5
6	Coherence	3	5

There is no standard method to determining the weight of each dimension of quality. The main basis for determining the weight of each section was the number of performance measures in the section and its importance for providing a measure of how well the index represents reality.

The number of measures was largely based on the GBSPM suggested criteria and how important each section is to evaluating the risk that the index does not represent reality. Though all six dimensions of quality contribute to the index, values representing reality, such as Accuracy and Relevance, play a larger role.

The final grade for an index provides an idea of the overall quality. This grade can be used to provide a total risk level of data quality to users. However, where it comes to identifying areas of improvement to reduce the level of risk that the index value is not representing reality, the section grades should be examined.

4. Challenges

Measuring the quality of price indexes is a challenge not seen in other areas of statistics. Aspects such as no true population of prices or the fact that a price index itself is a compilation of many different statistics and procedures makes standard measures insufficient for price indexes. These challenges are the reason for the Performance Measure Grading Scheme in the first place; however PPD faced its own set of challenges when trying to create one assessment model for 25 very different indexes. The challenges specifically are creating measures that apply to both survey and alternative data source statistics, common measures for a wide variety of industries being measured and adapting generic measures of response rates and variance to price indexes.

4.1 Survey versus alternative data sources

One of the key challenges in creating a grading scheme applicable to 25 indexes is creating performance measures that can be applied to indexes using survey data, those using alternative data sources, and those using a combination of both. At Statistics Canada, the term alternative data source includes prices collected via the internet as well as administrative sources such as private companies. The term encompasses any data not collected by a survey. Table 3 lists the PPD indexes and their sources of data. These indexes fall into three broad categories: survey data only, a combination of survey and alternative sources and pure alternative source.

Table 3: List of PPD indexes and their data sources

Price Index Name	Data Source
Accounting Services	Survey
Architectural & Engineering Services	Survey
Commercial Rents	Survey
Export Import Price Index	Survey
Machinery and Equipment Rental Services	Survey
Retail Services	Survey
For-hire motor carrier services	Survey
Wholesale Services	Survey
New Housing Price Index	Survey
Consulting Services	Survey
Courier and Messenger Services	Survey and alternative
Industrial Product Price Index	Survey and alternative
Raw Materials Price Index	Survey and alternative
Informatics and Professional Services	Survey and alternative
Apartment building construction price index	Survey and alternative
Non-residential building construction price index	Survey and alternative
Traveller Accommodation	Alternative
Computer and Peripherals Price Index	Alternative
Electric Power Selling Price Index	Alternative
Farm Input Price Index	Alternative
Machinery and Equipment Price Index	Alternative
New Lending Services	Alternative
Construction Union Wage Rate Index	Alternative
Commercial Software Price Index	Alternative
Passenger Air Services Price Index	Alternative

There are certain measures that do not apply to alternative data sources such as sampling method or collection response rate, simply due to the nature of alternative data sources. Alternative data sources rarely entail sampling and there is no 'respondent' in the traditional sense of the word. Other measures, though originally designed for survey-based indexes can still apply to alternative data sources. For example, industry coverage, price definitions and weight updates.

A few key measures had to be modified from their traditional definitions in order to accommodate alternative data sources. These in particular include the Estimation Response Rate. One of the challenges of working with alternative data is that you are limited by the source. This impacts the price definition, the frequency of weight updates and the type of quality adjustment that can be done. This aspect not only poses a challenge when creating grading criteria, but it can limit the ability of an index to improve in its performance measure if it is restricted to what comes with the data source.

4.2 Variety of sector coverage

The Producer Prices Division (PPD) of Statistics Canada publishes 25 indexes that cover 5 broad economic sectors: goods, distributive trades, transportation services, financial and professional services and the construction sectors. These broad categories are different in how prices are set, how prices

should be measured, and the frequency of entries and exits in each industry. Even within each of these areas, more specific industries differ greatly. The challenge of creating a generic scheme is to determine standard measures for price definition, weight update frequency and quality adjustment that can still capture the intricacies of each industry.

4.2.1 Price Definition

In standard price index compilation, transaction prices are characterized as the ultimate price definition to use.[9] However, over the last 10 years and the increased effort by statistical agencies to measure the service industries, it has become clear that transaction prices do not necessarily apply to services and they cannot always be properly captured. Measuring price change for services is difficult because services output in and of itself can be hard to identify and even harder to measure. For example, a key component to index calculation is that quality is kept constant over time. In an industry like Architectural and Engineering services, the service of designing a bridge is measured. However, the company providing this service prices their contracts on the specifics of the bridge – and this bridge will never be built again. Therefore, it is not possible to repetitively price the same transaction price for the same service. Instead, a model pricing method might be used. With the growing need to measure the service sector of the economy, international guidelines on price measurement for the services have been developed through the work of the Voorburg Group on Services Statistics and endorsed by international bodies in the latest edition of the Eurostat-OECD Methodological Guide for Developing Producer Price Indices for Services.[10] These guidelines detail the best pricing methods for a wide variety of services and the Performance Measure Grading Scheme used these guidelines to determine the best pricing methods for each industry. The price definition for each index produced by Statistics Canada is measured against these guidelines. These definitions are summarized in appendix 4. An index is given a top grade (or very low risk rating) if it uses the optimal pricing method as determined by international standards. The standards also outline second best options which, if used in PPD, receive a lower grade.

4.2.2 Weight update frequency

Another area of challenge when dealing with a wide variety of industries is the frequency of sample updates. When measuring price indexes, one must balance the relevance of the basket with a measure of pure inflation. In many cases, more frequent sample or weight updates are preferred because they make a basket more relevant and so more reflective of reality to users. If we exclude the issues of the practical problems of sample updates (such as frame availability, weight availability and resources to implement the update), some industries do not change as fast as others, and therefore have no reason for frequent updates.

4.2.3 Quality Adjustment

Quality adjustment methods can differ depending on the product, service or industry being measured. Price indexes measure pure inflation and so quality adjustments need to take place so index movements do not reflect price changes attributed to quality change. In most cases in PPD indexes, prices are

[9] IMF/ILO/OECD/UNEC/World Bank (2004). Producer Price Manual Theory and Practice. Washington D.C: IMF.
[10] Eurostat-OECD (2014). Eurostat-OECD Methodological Guide for Developing Producer Price Indices for Services. 2nd ed.

collected for defined products of the same quality over time. If this can be achieved, then no quality adjustment needs to take place.

However, products and services change over time, and quality adjustment methods need to be used to handle these changes. There are a variety of methods of quality adjustment including imputation, direct comparison, explicit adjustment and hedonics.

More and more recently, national statistics offices have been using hedonics in their price indexes. Hedonics refers to the method whereby a price is determined by a hedonic regression. This method is the best approach to use when the products being priced are very heterogeneous in quality or when quality change is very rapid. However, if products are homogeneous or consistent in quality over long periods of time, hedonics is not necessary. Other explicit methods such as adjusting for difference in option costs when the option costs are collected or directly observed, or even direct comparison when quality changes are virtually non-existent. Therefore, the optimal measure of quality adjustment highly depends on the price structure of the market being measured.

4.3 Response Rates

Response rates are a measure of responding units compared to in-scope and unresolved units. Response rates measure whether enough data with the right sample representation are being collected to provide reliable results. A response rate is a term generally used in reference to collection from a survey, but more recently Statistics Canada has extended it to include all data obtained either directly from respondents or from administrative (or alternative) data.[11] Yet it is still unclear what exactly a 'response' from an alternative source is. We want to find a measure that includes alternative data sources because if response rates were to solely refer to survey based data collection, then this measure would apply to only 40 percent of PPD's indexes.

The response rate used in the Performance Measure Grading Scheme has been defined as an Estimation Response Rate in order to assess the data available for use in the index calculation. The benefits of using such a rate is that it adheres to Statistics Canada's standard reporting practices of providing response rates at the estimation stage, it is in line with similar measures in other Statistics Canada business surveys and lastly, it incorporates both survey and administrative data because the goal is to assess the data used at the estimation stage, not just collected from a survey.

4.3.1 Estimation Response Rate Methodology

The Estimation Response Rates for PPD are calculated at the index estimation stage and are a weighted estimate. The weights are the economic weights used for the calculation of the index they relate to. Economic weights is a general term and can refer to industry, firm or product revenue or expenditure and may incorporate the sampling design weight if sampling has taken place. Therefore, the rate can be interpreted as the revenue proportion for which prices have been collected and are subsequently used for the estimation of the index.

It is important that weights are used in the estimation of the response rates. Supply and demand factors drive price levels in the economy. Under monopolistic competition, similar, yet differentiated products are produced and prices firms charge are influenced by rival firms. Larger firms in the market tend to

[11] http://www.statcan.gc.ca/pub/12-539-x/2009001/response-reponse-eng.htm

have larger influence on the price taken by rival firms. Therefore, the presence or absence of a large firm in an industry can greatly impact the overall accuracy of the price movement measured in a price index. Since the Estimation Response Rate is a measure of accuracy, it is important that weights are included in the calculation.

Respondents can be divided in two, mutually exclusive, categories – firms and products. The dichotomy is important to establish since our definition of response is based on what the primary unit, or elementary aggregate (EA), is. Consequently, if the EA is a firm then we define response as the provision of at least one valid price by the primary unit. For example, if a particular survey is collecting prices for 8 different products for a selected firm and the firm has only provided 1 of those prices, it is considered a respondent under our definition. In the case where an EA is a product (as in the case of manufacturing output prices), if the price for a product is missing than the primary unit is considered a non-respondent, regardless of where the information is sourced. Since Statistics Canada's SPPIs are all industry-level indices, the EA is currently a firm.

Now let us consider the calculation of the response rate. In the index calculation, weights are applied at every level of aggregation starting from the elementary aggregate, which is also the basis for our response definition. In essence, we sum the weights of EAs (firms or products) that have been imputed by the Estimation System[12] upon calculation of the index. We then divide this number by the total revenue (total weight) included in the sample used for index calculation.

Suppose that index calculation is performed using N elementary aggregates. Now let w_k be the weight of primary unit *k* and W be the total weight, and define the indicator function $I(k)$ as

$$(1)\quad I(k) = \begin{cases} 1, & if EA\ k\ is\ imputed \\ 0, & otherwise \end{cases}$$

Then the response rate r is calculated as

$$(2)\quad r = 1 - \frac{\sum_{k=1}^{N} w_k \cdot I(k)}{W}$$

Note that if $\sum_{k=1}^{N} w_k = W$ then our formula can be expressed as

$$(3)\quad r = 1 - \frac{\sum_{k=1}^{N} w_k \cdot I(k)}{\sum_{k=1}^{N} w_k}$$

The case where $\sum_{k=1}^{N} w_k \neq W$ happens when external sources are used to obtain more accurate weights at higher levels of aggregation.

The drawbacks of this method are that it includes imputation done outside of the Estimation System (microdata cleaning) as a response. It is also still unclear how to interpret such a rate from a pure alternative data source. If no imputation has occurred, common with alternative data sources, then should the rate be 100% or non-applicable?

[12] The Estimation System refers to the computer system used in PPD to calculate price indexes.

4.4 Measuring Variance for Producer Price Indexes

Estimates of the magnitude of statistical errors are an important component of the assessment of any statistical product. Wherever possible, Statistics Canada publishes measures of the sampling variance to express the accuracy of data. For the most part, such measures have not been available for producer price indexes but work is underway to produce accurate measures of variance. The quality grading scheme proposes a variance measure where possible.

4.4.1 Statistical Survey Error

The error of a statistical estimate can be decomposed into two parts, namely the bias and the variance.

$$E(\hat{Y} - Y)^2 = E\left(\hat{Y} - \bar{Y}\right)^2 + E(\bar{Y} - Y)^2$$

Where $\hat{Y}$ represents the estimate, $\bar{Y}$ represents the average over all possible outcomes of the survey, and Y represents the true value of the parameter. In this expression, the term on the left hand side is the total error, while on the right hand side of the equation the first term is the variance and the second term is the square of the bias. In producer prices indexes, the concept of the true population value is not clearly defined since the population of all products and prices that we aim to measure in all producers is somewhat abstract, and impossible to measure. For this reason, we generally select estimators that are known to be unbiased and use only the variance to represent the total error of the survey. In practice, this assumption is validated by monitoring different aspects of the survey process like, for example, the coverage of the target population and the response rate.

To further consider the variance, many specific steps in a survey process can contribute to the variance of the final estimate. The step that is most easily represented in a statistical framework is sampling, and we are able to estimate the sampling variance for any probabilistic sample design. Variance for non-probabilistic designs like cut-off or judgemental sampling are more difficult to evaluate, as there is no explicit random mechanism that leads to the variability in the estimates. Consequently, the variance for each index is considered on a case-by-case basis. For some price indices, it is possible to estimate variance using classical variance estimation methods, while for other indices, such methods are not applicable.

Variance is often used to assess accuracy in a statistical survey as it provides an indication of how similar estimates from different samples are expected to be. The total error of an estimator (including both bias and variance) can be decomposed into errors arising directly from sampling (instead of conducting a census of the entire population) and those from non-sampling sources (such as response errors, non-response errors, among others).

We estimate only the sampling error in our variance estimation. Non-sampling errors can be difficult to estimate, and often need to be addressed individually for each component rather than developing procedures that reflect all possible sources of error. Note that some non-sampling errors are accounted in other components of the grading scheme.

4.4.2 Estimation of Sampling Variance

In the case of producer price indexes, we consider three situations:

i) The prices are collected through a probabilistic design such as Poisson sampling

ii) The prices are collected through a non-probabilistic design

iii) The prices are collected from a comprehensive administrative source (i.e. on a census basis and there is no sampling error) as in the Passenger Air and the New Lending indexes.

In case i), established statistical techniques can be used to estimate the sampling variance of survey estimates when a probabilistic design is used. In cases where the form of the estimator is more complex, as is the case for price indexes, re-sampling methods such as the bootstrap approach or linearization techniques can be used. These approaches can be applied to estimate the variance of the index for any two specific periods (e.g. for two consecutive periods or for the current period relative to the base period)

In case ii), the sampling error cannot be directly estimated. However, if the mechanism used to select the non-probabilistic sample can be well approximated by a probabilistic design, it may be possible to estimate an approximate sampling variance. More research is required in order to identify variance estimation approaches that would apply in the context non-probabilistic samples.

In case iii) there is no random sample selected and thus the sampling variance is 0.

In cases estimating the variance from a probabilistic design two main approaches can be used. Bootstrap variance estimates can be used, where a pseudo-population is created based on the sampled units, and repeated sampling is applied to this population using the actual sampling design to generate a set of plausible estimates. The variability between the plausible estimates is then used to estimate the variance of the estimator. Alternatively, a linearization technique can be applied where a linear approximation to the estimator is derived, and the variance of this approximation can be estimated and attributed to the original estimator.

More details and examples of the application of these methods to price indexes can be found in Beaumont and Patak (2012) for the bootstrap and Patak and Rais (2007) for the linearization technique.

4.4.3 Special Considerations and Limitations

There are a number of characteristics of price indices that present challenges in applying either of these methods.

Small sample sizes – since the variance is calculated at the sampling stratum, it is generally required that we have reasonably large sample sizes within each stratum in order to reliably estimate the variability between units in the stratum. Variance estimates based on small sample size should be used and interpreted with caution.

Variability of price movements between products - . For price indices where variance estimation approaches can be applied, the sampling variance measures the variability introduced by the random selection of a sample of firms and will not capture the variability between products within a firm i.e. the variability related to the fact that prices for only a subset of products or services are collected. In practice, we use strategies and processes to insure that we select a subset of products or services that is going to provide a complete and representative picture of the firm in terms of price movement.

Non-sampling errors - In cases where, for example, our samples have a low response rate or our survey frame have a low coverage rate of the target population, estimates of the sampling variance will not be good measures of the total error. Those examples of non-sampling errors could contribute to the bias component of the total error in a non-negligible way.

4.4.4 Interpretation of accuracy measures

In most Statistics Canada surveys, the coefficient of variance is used as a measure of accuracy. The reason why we use the coefficient of variation instead of the standard error or variance is to give us a relative measure of dispersion (variability) in our sample. It tells us the size of the dispersion relative to the population mean, which in our case is just the sample mean. The coefficient of variation for a given classification level, call it k, is

$$(4)\quad CV_k = \frac{\sqrt{\widehat{\sigma^2_{b,k}}}}{\overline{x_{full,k}}}$$

The interpretation of a coefficient of variation is fairly straightforward in that we are technically only capturing the variance due to our sample selection. A high coefficient of variation would indicate that our sample is unlikely to be representative of the population and should therefore be used with caution. In contrast, a low coefficient of variation indicates that our sample is a good representation of the population.

5. Conclusion

The Performance Measure Grading Scheme provides a comprehensive framework to evaluate all components of a price index program. Statistics Canada has a ten-year planning window for investments in the Continuity and Quality Maintenance of its base statistical program. These indicators can be used to plan investments in the redesign, basket updates and resampling activities required to ensure the quality of data is fit for the intended purpose. In some programs, like the mission critical Industrial Product Price Index (IPPI) which provides deflators of manufacturing output, the risk tolerance is lower than for a deflator of an industry that has a small weight in overall GDP.

New measures of variance are being developed and will be reviewed by a technical committee of methodologists at Statistics Canada. As Producer Prices Division finalizes its measures and grading scheme and applies it to the full suite of prices, the grading scheme may evolve.

In addition, Producer Prices Division needs to determine the indicators that will be most relevant to data users in their assessment of fitness for use for contract escalation and other purposes. A subset of these indicators will be included in the metadata that is available with every data release. Departmental standards recommend publication of response rates and coefficients of variation but, as noted, these measures are not always applicable to every statistical program, particularly as we make increased use of alternative data sources. Alternative indicators that help users understand the strengths and limitations of our data will be introduced accordingly.

Finally, Producer Prices Division is currently investigating fitness for use indicators for each sub-index of a published series. These indicators will be included with every index number released from our public dissemination database, CANSIM. Such indicators will be based on a subset of indicators of accuracy.

References

OECD/Eurostat (2014), *Eurostat-OECD Methodological Guide for Developing Producer Price Indices for Services: Second Edition*, OECD Publishing.
http://dx.doi.org/10.1787/9789264220676-en

GSBPM v 5.0 http://www1.unece.org/stat/platform/display/GSBPM/GSBPM+v5.0

Statistics Canada (2015 last modified). Statistics Canada Quality Guidelines – Response and nonresponse. http://www.statcan.gc.ca/pub/12-539-x/2009001/response-reponse-eng.htm

IMF/ILO/OECD/UNEC/World Bank (2004). Producer Price Manual Theory and Practice. Washington D.C: IMF.

Statistics Canada (2000). Policy on Informing Users of Data Quality and Methodology. http://www.statcan.gc.ca/eng/about/policy/info-user

Eurostat-OECD (2014). Eurostat-OECD Methodological Guide for Developing Producer Price Indices. 2nd ed.

U.S. Bureau of Labor Statistics (2016). *Variance estimates for price changes in the Producer Price Index.* http://www.bls.gov/ppi/ppivariance.htm

Draper, R., & Schoultz, C. – Statistics Sweden (2016). *A model for index auditing – the Swedish experience*. Presentation, Nordic SPPI Seminar, 19-20 May 2016, Copenhagen.

Beaumont, J.-F. and Patak, Z. (2012), On the Generalized Bootstrap for Sample Surveys with Special Attention to Poisson Sampling. International Statistical Review, 80: 127–148. doi:10.1111/j.1751-5823.2011.00166.x

Patak, Z. and Rais, S. (2007), Survey Methodology for New Business Services Price Indices. American Statistical Association. https://ww2.amstat.org

Appendix 1: Performance Measure Grading Scheme

Section 1: Relevance

#	Performance Measure	Grading Criteria	Grading Guidance	Risk Level	Grade
1.	**Macroeconomic Accounts relevance**	How is the index used by the Macroeconomic Accounts?	Used directly in deflation or other volume estimates	5	
			Consistently used in consultation	4	
			Infrequently used in consultation	2	
			Not used	1	
2.	**User Relevance**	How many users are there?	Used by SNA, other Stats Can divisions and/or external clients	5	
			Used by SNA only	3	
			No known users	1	
4.	**Price representation**	How does the price definition[13] predominantly used in index calculation meet international standards?	Meet ideal international standards[14]	5	
			Meet acceptable international standards	3	
			Does not meet international standards	1	
5.	**Proportion of Price Definition Used**	What proportion of the price definition (from question above) is used in sample?	Greater than 80%	5	
			Between 50% and 80%	3	
			Less than 50%.	1	
6.	**Weight Type**	What type(s) of weight is used at the lowest level of weights?	Expenditure or revenue	5	
			Other	1	
7.	**Additional Weights**	Do you use a different source of weights at a higher level of aggregation than the lowest level of weights?	YES if additional source is available	5	
			NO because additional source is not available	4	
			Addition source available but not used.	1	
Grade for Relevance					

[13]See Appendix 3: Glossary of Key Terms

[14] See Appendix 4: International Standards for Price Definitions

Section 2: Accuracy

#	Performance Measure	Grading Criteria	Grading system	Risk Level	Grade
1.	**Industry Coverage**	What (weighted) percentage of the target population does the sample cover?	> 90%	5	
			80-89%	4	
			70-79%	3	
			60-79%	2	
			<60%	1	
2.	**Collection Response Rate**	If using survey data, provide the rate calculated as what was received from collection compared to what was sent out for collection.	> 80%	5	
			60-79%	4	
			40-59%	3	
			<39%	2	
3.	**Estimation Response Rate**	The data is used in estimation versus what is expected (annual average).	> 80% 60-79% 40-59% <39%	5 4 3 2	
4.	**Measure of variance**	TBD	TBD		
5.	**Product Substitution**	When does product substitution (replacement, or model update) occur?	Occurs during a production period	5	
			Occurs during a sample or basket update	3	
			Does not occur	1	
6.	**Quality Adjustment for Product Substitution (replacement)**	If product substitution occurs which method of quality adjustment is used? (mark all that apply - take performance measure average when marking)	Comparable replacement, Hedonics, or Differences in production and option costs	5	
			Subject Matter Expert Judgement	4	
			Parental Imputation	3	
			Carry Forward, Overlap Method, Linked-to-show-no-price change or does not occur	2	
7.	**Quality Adjustment for Product Specification Changes**	If changes occur to the product specifications, but the product is not substituted, which method of quality adjustment is used? (mark all that apply - take performance measure average when marking)	Hedonics or Differences in production and option costs	5	
			Subject Matter Expert Judgement	4	
			Parental imputation	3	
			None or other	2	

8.	**Revision Magnitude (if applicable)**	Within a year, what is the average magnitude of the revision of the total index (calculated by first time month is published vs. last time months is published and then take average of 12 months).	<1%	5	
			1-1.9%	4	
			2-2.9%	3	
			>3%	2	
9.	**Source of Revision(if applicable)**	What is the source of the majority of revisions?	Late data	5	
			Incorrect data	1	
Grade for Accuracy					

Section 3: Timeliness

#	Performance Measure	Grading Criteria	Grading system	Risk Level	Grade
1.	**Time Lag**	What is the time lag between the end of reference period and release date? (If monthly and quarterly indexes are calculated, use quarterly)	Within 2 periods after reference period	5	
			Between two and three periods after reference period	4	
			Between three and six periods after reference period	3	
			Greater than six periods after reference period	2	
2.	**Basket Updates**	How often is the basket of goods or services reviewed?	Basket updated every 5 years or updated when deemed necessary by industry standards	5	
			Basket reviewed every 5 years and basket update deemed necessary but still in progress	3	
			Basket not reviewed every 5 years	1	
10.	**Sample/Weight Update Frequency**	If survey data, how frequently is the sample of respondents updated? If admin data, how frequently are the weights updated?	< 2 years	5	
			2 -5 years	3	
			>5 years	1	
Grade for Timeliness					

Section 4: Accessibility

#	Performance Measure	Grading Criteria	Grading system	Risk Level	Grade
1.	**Suppressions**	What percentage (in terms of relative weight) of the CANSIM published indexes is suppressed?	No suppressions	5	
			<10%	4	
			10-30%	3	
			>30%	2	
Grade for Accessibility					

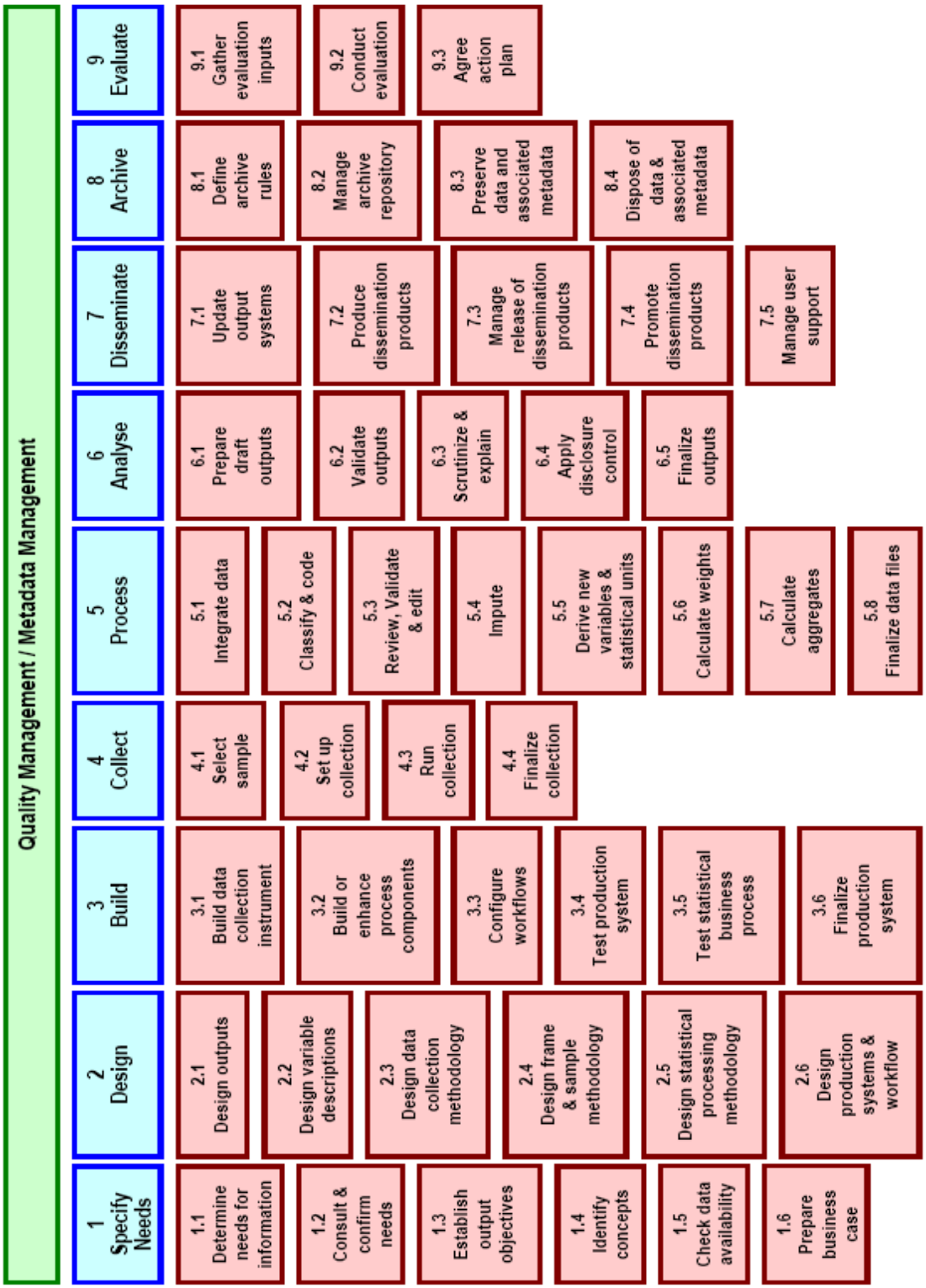
Section 5: Interpretability

#	Performance Measure	Grading Criteria	Grading system	Grade Scale	Grade
1.	**Notifications to users**	Is the IMDB page up to date with the latest methodological practices?	Yes, up to date	5	
			No, out of date	1	
Grade for Interpretability					

Section 6: Coherence

#	Performance Measure	Grading Criteria	Grading system	Risk Level	Grade
1.	**Use of standard classifications**	Does the index use a Statistics Canada standard classification? Examples of standard classifications include NAICS, NAPCS and HS Codes.	Full classification follows a standard or approved variant classification	5	
			Partial classification follows standards	3	
			No standard classification	1	
2.	**Version of classification**	What version (year) is the current classification from?	Most recent version (i.e. NAICS 2012)	5	
			Previous version	3	
			Older than previous version	1	
3.	**Coherence across baskets**	If classifications have changed between baskets, are concordance tables available to help users connect the old classification to the new one?	Yes	5	
			No	1	
Grade for Coherence					

Appendix 2: Generic Statistical Business Process Model[15]



[15] Image from GSBPM v 5.0 http://www1.unece.org/stat/platform/display/GSBPM/GSBPM+v5.0

Appendix 3: Glossary of key terms

1. **Average Revision Magnitude:** Assess how much an index value has changed due to late responses or corrections to the data between the first time it is published and the last time it is revised. It is calculated by taking the percentage change between the first time an index value for a given period is published versus the last time that same index value is published, then taking the average of this magnitude over a year.

   For example, for the first time January is calculated, the index value 101.2. Six months later, January can no longer be revised, and its index value is 101.4. The revision magnitude for January is 0.2%. Repeat this process until you have a revision magnitude for each month. Take the absolute average of these 12 revision magnitudes to get the Average Revision Magnitude.

2. **Basket of Goods or Services:** A fixed set of products, services or firms that are representative of market or industry production conditions that allows index calculation to measure pure inflation.

3. **Post Revision Response Rate:** The response rate after the revision period. It is calculated as the number of respondents divided by the total number of sampling units within a year.

4. **Price Definition:** Pricing Definitions, also known as Price Methods, refer to procedures put in place by index compilers to make price data suitable for use in index compilation. The guidelines for the following price definitions come from the Producer Price Index Manual, the *Methodological Guide for Developing Producer Price Indices for Services*, and the Export and Import Price Index Manual.
   a) **Component pricing method**
      - The component pricing method is particularly relevant when only information on total revenues and number of sales are typically available, but where the products sold are relatively heterogeneous – invalidating the unit value approach. The approach specifies a representative product and estimates its price on the basis of the prices of identifiable components that determine the overall price.
      - Example: Purchasing a computer and pricing its multiple components which could include a CD-ROM, Floppy Disk Drive and USB ports.

   b) **Contract pricing**
      - Contract pricing refers to the use of prices in long term contracts for the repeated delivery of the same (or a very similar) service. This is a special case of using real transaction prices.
      - Example: A cleaning company's contract to clean an office building each day over a period of time.

   c) **Direct use of prices of repeated services**
      - Direct use of prices of repeated services is the use of real transaction prices or, less preferably, list prices, of the same service product in successive survey periods.

Adjustments will be needed to account for any changes that occur when the observed product is replaced or if its quality changes.

- Example: A telecommunication company's observed monthly package which may include different services such as a fixed amount of talk time and texting at a fixed fee.

d) **Hedonics**

- In this method, data on the market prices and characteristics of various models of a product are collected. A regression is carried out to investigate which characteristics are the determinants of the price differences between the models.
- Example: A regression model is developed to measure prices of real estate properties. Properties which have desirable characteristics such as larger square feet or a nice location will have higher prices.

e) **Margin pricing method**

- Margin prices are defined here as those prices that are not directly observable but where the value of the service can be measured as the difference between the observed acquisition and selling price of a given product.
- Example: A shipment of clothing being bought at one price and being sold at a different price. The pricing method would measure the acquisition price minus the selling price.

f) **Model pricing method**

- The model pricing method is typically applied in cases where the service provided is unique. The approach specifies a standardised product, that is sufficiently representative of the type of service provided, and respondents are asked to provide a price quote for this standardised product.
- Example: The estimation of prices of engineering services such as building inspections or project management.

g) **Percentage fee method**

- The percentage fee method calculates the value of the service as the product of the percentage fee and value of the product to which the fee relates.
- Example: A real estate agent selling a property to a client and collecting a commission fee.

h) **Time based methods**

- Time based methods reflect cases where a service is specified in terms of the time spent providing a particular service and not necessarily in terms of the actual service provided.
- Example: The hourly charge out rate for IT support staff.

i) Transaction Prices

- Transaction prices are actual prices paid to or received from producers for goods or services. These prices include all discounts or rebates given. Transaction prices are similar in concept to direct use of prices of repeated services but used in the measurement of goods prices.

- Example: The price of a manufactured chair sold to a wholesaler.

j) **Unit value method**

- The unit value method constructs a price index based on observed revenue and quantity data.
- Example: The cost of local phone calls, for instance the total revenue from the local calls divided by the total volume (minutes) of local phone calls.

5. **Product Substitution:** A product substitution occurs when an item becomes obsolete and a new product replaces the old one. If model pricing applies, the models are updated to reflect changes. Product substitution is also synonymous with product replacement, where a new product takes the place of an old product even though they may not be perfect substitutes for each other.

6. **Sampling Units:** The elements considered for selection when sampling. This could be a product, service, establishment or contractor.

7. **Suppression:** Data presented on CANSIM is not published in order to protect confidentiality of respondents or due to poor quality of underlying data.

8. **Target Population:** The full population that an index is trying to cover. This might be an industry, a market or a classification structure like NAICS or NAPCS.

9. **Timely manner:** In this context, the timely manner in which methodological changes are disseminated to the public refers to the period in which these changes have been implemented in production.

10. **Type of Weight:** The category of values used to weight price or index movements in the index calculation. Weight types include expenditure, quantities and company revenue.

Appendix 4: International standards of price definitions for PPD indexes.

Price Index	Ideal	Acceptable
Accounting Services	**Model Pricing** Model prices easily track changes of prices for the same product from quarter to quarter.	**Time based methods**
Architectural & Engineering Services	**Model Pricing** Model pricing is preferred since it can capture productivity changes over time unlike charge out rates.	**Time based methods**
Commercial Rent Services	**Direct Use of prices of repeated Services** The actual transaction (rental) price is preferred to the list price.	**List Price**
Courier and Messenger Services	**Contract Pricing** Contract pricing is ideal because it measures the price of real services.	**Model Pricing** **Unit Value Method**
Export Import	**Transaction Prices**	**Unit Value Method**
Industrial Product	**Transaction Prices**	
Machinery and Equipment Rental	**Direct Use of prices of repeated Services** **Model Pricing – For Unique Services** For most goods using the actual rental price is ideal while model pricing should be used for unique services.	
Raw Materials	**Transaction Prices**	
Retail Services	**Margin Price Per unit** Margin Price per unit is the commonly accepted methodology. This is done in order to measure change in the price of the distributive trade service not the price of the product.	
Traveller Accommodation	**Direct Use of prices of repeated Services** **Unit Value Method**	

	The above methods are what are commonly used by other statistical agencies	
For-hire motor carrier services	**Direct Use of prices of repeated Services** **Contract Pricing** **Model Pricing** The above methods are considered the most appropriate methods to measure prices by Eurostat	
Wholesale services	**Margin Price Per unit** Margin Price per unit is the commonly accepted methodology. This is done in order to measure change in the price of the distributive trade service not the price of the product.	
Computer and Peripherals	**Hedonics** The Hedonic method defines and measures variables that determine computer prices and can adapt to technological change.	**Component Pricing** **Re-sampling**
Electric Power Selling	**Margin Prices** are used for Net Recording Margin prices are best for electricity as they reflect quality changes in the price as well. Gross recording methods should deflate output by an appropriate PPI to get basic prices.	
Farm Input	**Unit Value Method** Eurostat uses the Unit Value Method to measure prices of agricultural products this however may not be the only ideal method.	
Informatics and Professional Services	**Time based Methods** Most countries use time based Methods, such as charge out rates, to measure prices. Other Methods are acceptable as well.	**Model Pricing** **Direct Use of prices of repeated Services**
Machinery and Equipment	**Transaction Prices**	
New Lending Services	**Direct Use of prices of repeated Services** **Unit Value**	

	Percentage Fee Method The above methods are considered the most appropriate methods to measure prices by Eurostat	
Construction Union Wage Rate Index	**Union Wage Rate** The Methodology is similar to recording prices of staffing agencies who set wages.	
New Housing	**Model Pricing** **Direct use of prices of repeated services**	**List price**
Apartment building construction	**Direct use of prices of repeated services** **Model pricing**	**Hedonics**
Non-residential building construction	**Direct use of prices of repeated services** **Model pricing**	**Hedonics**
Commercial Software	**Direct use of prices of repeated services** **Unit Value Method** **Component Prices** **Model Prices** The above methods are considered the most appropriate methods to measure prices by Eurostat.	
Passenger Air Services	**Direct use of prices of repeated services** **Unit Value Method**	
Consulting Services	**Time Based Methods** **Model Pricing** Both methods are used internationally by different statistics agencies.	