ESS guidelines on temporal disaggregation, benchmarking and reconciliation

2018 edition

MANUALS AND GUIDELINES

eurostat

ESS guidelines on temporal disaggregation, benchmarking and reconciliation

2018 edition

Authors: Dario Buono, Duncan Elliott, Gian Luigi Mazzi, Reiner Bikker, Markus Frölich, Riccardo Gatto, Barbara Guardalbascio, Stefan Hauf, Enrico Infante, Filippo Moauro, Erich Oltmanns, Jean Palate, Karel Safr, Peter Tibert Stoltze, Francesca Di Iorio

Manuscript completed in November 2018.

Luxembourg: Publications Office of the European Union, 2018

PDF ISBN 978-92-79-98007-7 doi: 10.2785/846595 KS-06-18-355-EN-N
PRINT ISBN 978-92-79-98006-0 doi:10.2785/79182 KS-06-18-355-EN-C

Foreword

Temporal disaggregation, benchmarking and reconciliation methods can be used to obtain temporal and spatial consistency in sets of time series. Such methods are widely used across the European Statistical System in the production of official statistics and the establishment of common guidelines for temporal disaggregation and related methods within the European Statistical System (ESS) is an essential step towards improved harmonization and comparability of official statistics, especially in macroeconomic indicators and labour market statistics.

These ESS guidelines address the need for harmonization expressed by many users from European and National Institutions. They are intended to support producers and users of official statistics in accordance with the European Statistics Code of Practice and complement other ESS guidelines by providing clear guidance on the use of temporal disaggregation, benchmarking and reconciliation methods and encouraging documentation and dissemination of best practice.

Acknowledgments

Eurostat and the European Commission express their gratitude and appreciation for the work carried out by the members of the ESS Task Force on Temporal Disaggregation for authoring these Guidelines: Dario Buono (Chair of the Task Force and co-editor, Eurostat), Duncan Elliott (Editor, Office for National Statistics, UK), Gian Luigi Mazzi (co-editor, Expert), Reiner Bikker (pen-holder for chapter 4, CBS, Statistics Netherlands), Markus Frölich (pen-holder for chapter 6, Statistics Austria), Riccardo Gatto (pen-holder for chapter 6, Eurostat), Barbara Guardalbascio (pen-holder for chapter 3, ISTAT, Italy), Enrico Infante (pen-holder for chapter 5, Eurostat), Filippo Moauro (pen-holder for chapter 3, ISTAT, Italy), Erich Oltmanns (reviewer, Destatis, Germany), Stefan Hauf (reviewer, Destatis, Germany), Jean Palate (reviewer, National Bank of Belgium), Karel Safr (reviewer, CZSO, Czech Statistical Office), Peter Tibert Stoltze (pen-holder for chapter 2, Statistics Denmark) and Francesca Di Iorio (reviewer, University of Naples Federico II, Italy).

Eurostat and the European Commission would like to thank also the chairs (Eduardo Barredo Capelot, Emanuele Baldacci, Martina Hahn, Ani Todorova, John Verrinder and Anne Clemenceau) of the DIME/ITDG Directors Group, the ESS Working Group Methodology, the ESS Working Group National Accounts, the ESS Working Group Labour Market Statistics, and the Seasonal Adjustment Expert Group and the Eurostat Methodology Advisory Committee (METAC) together with all members for their useful comments provided during the consultations and support.

The authors would also like to thank Eurostat colleagues for their continuous support and encouragement, in particular Matyas Tamas Meszaros (layout), Jacopo Grazzini (French translation), Sybille Luhmann and Britta Gauckler (German translation), Denis Leythienne (feedback) and Ruggeri Cannata Rosa (feedback).

The European Commission would also like to thank all the people from National Statistical Institutes, the ECB, National Central Banks, and Eurostat who have contributed with their comments to improve these Guidelines.

Table of contents

Introduction

Motivation for the guidelines

In official statistics there is an increasing demand for indicators at a higher frequency than those that have traditionally been observed. Direct measures of indicators at a high frequency can be very costly and difficult to achieve sometimes resulting in low quality results when the information set is not adequate. In such situations temporal disaggregation techniques can constitute a feasible alternative to the direct estimation of high frequency indicators. Additionally, even when high frequency indicators can be directly compiled, they are often not consistent over time with lower frequency versions. For example, annual surveys with larger samples may give more accurate estimates of the level of a variable compared to estimates from a small monthly survey that is designed to provide estimates of monthly change. Under the hypothesis that low frequency indicators are more reliable than high frequency ones, benchmarking techniques can be used to ensure the time consistency between high and low frequency indicators. Finally, directly or indirectly measured high frequency indicators may not necessarily meet required accounting and aggregation constraints. If that low frequency indicators meet accounting and aggregation constraints, reconciliation techniques can be used to restore them on high frequency indicators too.

Eurostat and the European Statistical System (ESS) developed these guidelines to help data producers derive high frequency data (e.g. quarterly or monthly) from low frequency data (e.g. annual) and to address related temporal and accounting constraints. Typical applications are known as temporal disaggregation, benchmarking, and reconciliation. The guidelines identify best practice to:

- achieve harmonization across national processes;
- enhance comparability between results;
- ensure consistency across domains and between aggregates and their components.

The establishment of common guidelines for temporal disaggregation within the European Statistical System (ESS) is an essential step towards better harmonization and comparability of official statistics, especially in macroeconomic indicators and labour market statistics. These guidelines address the need for harmonization expressed by users from European and National Institutions. This document presents both theoretical aspects and practical implementation issues in a user friendly and easy to read framework. They meet the requirement of principle 7 (Sound Methodology) of the European Statistics Code of Practice (CoP), and their implementation is consistent with principles 14 (Coherence and Comparability) and 15 (Accessibility and Clarity). The guidelines also provide transparency of temporal disaggregation, benchmarking and reconciliation practices by encouraging documentation and dissemination of practices. These guidelines are complementary to the ESS guidelines on seasonal adjustment (Eurostat, 2015 edition), the ESS guidelines on revision policy for PEEIs (Eurostat, 2013 edition), and the Eurostat and United Nations handbook on rapid estimates. They are also in line with the Handbook on quarterly national accounts (Eurostat, 2013 edition).

Scope of Guidelines

These guidelines are aimed at those involved in the production and analysis of infra-annual European statistics (compiled by Eurostat) and corresponding country specific official statistics compiled by National Statistical Institutes (NSIs). Topics covered in the guidelines and proposed recommendations should be of interest for public and private institutions working in compiling infra-annual statistics. They provide a consistent framework for temporal disaggregation, benchmarking and reconciliation, taking advantage of similarities in the process to define a common vocabulary to facilitate communication and comparison between practitioners. They have been conceived both for experts and beginners since they also present material on complex methods in an accessible way.

These guidelines cover important issues related to the choice of methods, to revisions and documentation. The guidelines do not cover the methods in detail, as a textbook would do, i.e. the guidelines describe neither the technical issues of applying the methods nor the actual calculations.

When using the guidelines for a specific set of time series, the first step should always be a thorough analysis of the data at hand, and when applying the different methods, careful evaluation of diagnostic measures should always be an integral part of the workflow. Applying the guidelines without paying due attention to the actual data should be avoided.

The guidelines are based on a set of principles, presented in section 1.5, which give some general rules to be followed when compiling infra-annual statistics using indirect techniques. The guidelines have 8 chapters devoted either to specific issues such as temporal disaggregation and benchmarking or to common issues such as revisions. The guidelines include a rich list of references for the different methods which should be consulted for more detailed technical information. For users of the guidelines who are not familiar with some of the terms used, relevant terminology on temporal disaggregation, benchmarking and reconciliation, as well as descriptions of the most relevant methods proposed in the literature and applied in statistical production are given in chapter 1. Each chapter is subdivided in sections or items describing a specific step of the process.

Each issue discussed in chapters 2 to 8 follows a standardized structure with three parts: a *description*, a list of *options* and a list of ranked *alternatives*. The description is free text presenting the problem. The options list, without pretending to be exhaustive, presents various possibilities to deal with the specific problem treated in the item. Out of these options, three ranked alternatives are highlighted:

- **(A)** Best alternative: should always be the target for producers.
- **(B)** Acceptable alternative: retained only if time or resource issues prevent alternative (A).
- **(C)** Alternative to be avoided: not a recommended option.

The objective of the guidelines is to help producers apply the best alternative whenever possible. It should then be a feasible target for producers. It should always be achievable with a reasonable effort, unless some production or institutional constraints prevent it.

The acceptable alternative (B) should be viewed as an intermediate step towards the achievement of alternative (A). It could also be considered the target for a limited number of cases when specific data issues, user requests, time or resource constraints prevent the achievement of the alternative (A).

The alternative to be avoided (C) includes some procedures that are not recommended.

Costs and benefit

The cost of compliance with these guidelines may be significant as temporal disaggregation, benchmarking and reconciliation require considerable resources. However, standard methods and tools are available to simplify these processes.

Being in line with these guidelines will ensure higher quality results, e.g. avoiding false signals and misinterpretation of the dynamics of the data.

1 Terms and methods

This chapter introduces important concepts for temporal disaggregation, benchmarking and reconciliation, providing definitions for terms used throughout the guide, including terms used to describe the problems that each method aims to address. Other considerations, including general principles that should be applied when using any of the methods discussed and some advice on software to use, are also given.

1.1 Terminology

The terminology in this section is applied throughout the guidelines. Some basic definitions of the types of data used in temporal disaggregation methods are provided followed by definitions of the main methods related to temporal disaggregation.

Flow

Definition: a *flow* is a measure of a phenomenon per time period.

Examples: Gross Domestic Product, gross fixed capital formation, monthly turnover measured as the sum of daily turnover within each month, crops production.

Stock

Definition: a *stock* is a measure of a phenomenon at a specific point of time. A stock is the result of cumulated flows.

Examples: population, inventories, capital stock at the end of a year, time averaged stock.

Time series

Definition: a *time series* is a set of time-ordered observations of a phenomenon taken at successive, periods or points of time. Time series are usually presented at a regular frequency such as monthly, quarterly or annually.

Examples: flows, stocks and indexes.

Index series

Definition: an *index series* measures the relative size of a variable in a time period relative to a base period. A reference period is usually scaled to 100 and may or may not be the same as the base period.

Examples: inflation, industrial production index, chain-linked indices of GDP.

Constraints

Definition: *constraints* describe relationships between data that should hold. Constraints are most often linear and binding. A common binding linear aggregation constraint is that annual sum of quarterly estimates is equal to given annual totals. The aggregation constraint becomes nonlinear when original quarterly data are subject to nonlinear transformations (for example, logarithmic). Nonbinding constraints occur when constraints are not met exactly but allow for some random error.

1.2 Problems to address

A variety of methods exist to solve the problems of temporal disaggregation, benchmarking and reconciliation. The aim of each method is to combine, in the best possible way, available time series (possibly of different frequencies) to make the series consistent according to a specified set of constraints.

Temporal disaggregation

Definition: *temporal disaggregation* is the conversion of a low frequency flow time series to a higher frequency time series.

Related terms: temporal distribution, benchmarking, interpolation, splining.

Example methods: regression-based methods (such as Chow-Lin).

Interpolation

Definition: *interpolation* denotes the generation of values of a time series for time points that have not been sampled within the interval of time of the original time series; for example, used for converting a low frequency stock time series to a higher frequency time series of stock data.

Related terms: splining, benchmarking, methods based on multivariate models.

Example methods: Cubic-spline interpolation, regression-based methods.

Benchmarking

Definition: *benchmarking* is adjusting a high frequency time series to have temporal consistency with a lower frequency version of the same variable usually measured from a different data source. This is also known as *binding benchmarking*.

Related terms: temporal disaggregation, temporal distribution, constraining.

Example methods: Denton methods, regression-based methods, growth rate preservation.

Reconciliation of time series

Definition: *reconciliation* is adjusting multiple time series for both contemporaneous and temporal consistency.

Related terms: balancing, benchmarking, constraining.

Example methods: multivariate Denton (and variants), two-step approaches.

Extrapolation

Definition: *extrapolation* is calculating values of a *time series* for time points that have not been sampled and are outside the interval of time of the original observed *time series*. In a temporal disaggregation problem, *extrapolation* relates to estimating the high frequency data for time periods where the low frequency data are not available.

Related terms: forecasting, nowcasting, backcasting, prediction, flash estimates.

Example methods: ARIMA models, exponential smoothing, regression-based methods, multivariate methods.

1.3 Methods for temporal disaggregation, benchmarking and reconciliation

This section provides a brief overview of the methods discussed in this guide and commonly used terms in official statistics that describe them.

Pro rata adjustment

Pro rata adjustment can refer to a benchmarking type problem where values of a high frequency time series are constrained to those of a low frequency series simply by multiplying each value of the high frequency series by a factor for the low frequency period that it belongs to. The factors are derived as the ratio of the low frequency observations to the high frequency series aggregated to the same frequency as the low frequency series.

Pro rata adjustment ensures that growth rates of the high frequency series within the low frequency period are not changed. However, the main disadvantage of this approach is a step problem in the adjusted high frequency series which arises when the adjustment factors change between low frequency periods. If the high frequency is a constant, then the low frequency series is temporally disaggregated by equally distributing each low frequency observation to the corresponding high frequency observations. For example, a quarterly flow time series is derived from an annual flow time series by dividing each annual time point by four and assigning that value to each quarter within that year.

Cubic-spline interpolation

Cubic-spline interpolation is a method of fitting a curve through a set of points from which it is then possible to derive a higher frequency series. It is a special case of a spline interpolation smoothing method that can be implemented with or without the use of a sub-annual indicator series. In the absence of an indicator series, the unknown trend can be conveniently described by a mathematical function of time.

Denton's variants

A method of benchmarking that has become known as the Denton method aims to minimize changes in the high frequency (indicator) series while meeting a set of benchmarking constraints. This uses constrained minimization of a quadratic form relative to the differences between disaggregated estimates and an indicator series. The penalty function can be specified as either arithmetic or proportional differences. See for example Denton (1971) and Dagum and Chollete (2006).

Stram and Wei

Stram and Wei (1986) propose a method based on the estimation of the autocovariance structure for the unobserved disaggregated series from the available autocovariances of an aggregate model. The method is applied in the absence of indicators.

Static Regression methods

Common static regression methods for temporal disaggregation include **Chow and Lin (1971), Fernández (1981)** and **Litterman (1983)**. These methods fit models with one or more high frequency indicator series regressed on the low frequency series to be disaggregated. The methods differ in the proposed models for the structure of the residuals. Chow and Lin extend the generalized least squares approach to temporal disaggregation, proposing a univariate regression of low frequency target data on high frequency indicators. The method also provides an optimal solution for extrapolation.

State-space methods

A univariate or multivariate modelling approach cast with two dynamic equations; the observation equation that describes how observations relate to the assumed unobserved components and the state or transition equation that describes how the unobserved components evolve over time. The state space form is general enough to represent all the most relevant model-based methods of temporal disaggregation. Its statistical treatment is based on Kalman filter and allows for nonlinear disaggregation of data transformed into logarithms.

Dagum and Cholette

Dagum and Cholette (2006) provide a unified regression framework for regressing high frequency indicators on available low frequency constraints, deterministic effects and autocorrelated errors. The model, also adaptable to a multiplicative form, nests common disaggregation methods such as Chow and Lin, Fernández and Litterman. It can be estimated using state-space methods and has a generalization to multivariate systems. Being a statistical model, this method yields information about its performance in the form of regression residuals.

Causey-Trager

Causey and Trager (1981) described an approach to benchmarking based on directly minimizing the changes to the proportional first differences of the indicator series, a principle called growth rate preservation. Because of this principle it could be considered an ideal benchmarking method. Its formulation however, involves a loss function which is nonlinear and even singular, and the model has no solution in a closed form. It must be solved by iterative algorithms, which make it impractical and therefore it is hardly used.

Dynamic regression methods

This approach for temporal disaggregation, developed by Proietti (2006), generalizes static regressions to autoregressive distributed lag (ADL) models. The order is limited to ADL(1,1) models, it nests static regressions, and it is treated by the state space methodology.

Multivariate approaches

Multiple low and high frequency indicators are taken simultaneously in a mixed-frequency multivariate model subject to temporal aggregation constraints. Several forms can be considered: vector autoregressions and error correction forms, structural time series and dynamic factor models, which can all be fitted with state space methods.

A comment on methods

When Denton methods are used, solutions and remarks proposed by the IMF Quarterly National Accounts manual (2014) should be considered, for correct initialization, extrapolation and preference for the proportional approach.

Mathematical methods such as the cubic spline are often chosen for their simplicity. However, most of them have a statistical method as a counterpart (see the first difference Denton's method well represented by Fernández). Therefore, the use of a statistical approach should be preferred for the advantages of computing and checking diagnostics including goodness of fit measures.

Within regression methods non-stationary residual models (for example, Fernández, Litterman, and ADL models in first differences etc.) are better suited for non-stationary and non-cointegrated series, whereas stationary models (Chow and Lin and ADL in levels) adapt better to stationary or co-integrated series. The use of a constant is suggested but it loses its importance for models well initialized. When more than one indicator is used, an accurate analysis of collinearity should be carried out. When no indicator is available these methods are still feasible considering simple regressions of deterministic effects (constant, linear trend, etc.).

1.4 Methods for variable selection and reduction

***Principal components (PCA)*:** based on summarizing the total variance of the indicators. With PCA, unities are used in the diagonal of the correlation matrix computationally implying that all the variance is common or shared. PCA summarizes information coming from the indicators only and does not account for their relationship with the target variable.

***Partial Least Squares (PLS)*:** maximizes the covariance between target variables and linear combinations of the indicators. PLS factors account for the co-movements between the target series and indicators, while PCA does not. Limitations of PLS include the fact that factors can only be computed at the lower frequency, and the information of target variables is used both for data reduction and temporal disaggregation.

***Clustering*:** a general algorithm for grouping large numbers of indicators into a few clusters.

Data reduction may lead to one or more composite indicators that meet both representativeness and degrees of freedom constraints. When both hard and soft indicators are to be used in temporal disaggregation it is recommended to keep at least two factors to avoid over-representing hard data.

1.5 Principles for temporal disaggregation, benchmarking and reconciliation

These guidelines provide recommendations on issues related to implementing methods of temporal disaggregation, benchmarking and reconciliation in the production of official statistics. However, there are some general principles that apply to all of these methods which should be considered:

- Validation: any method applied in the production of official statistics should have been generally accepted by peer-reviewed literature.
- Accessibility: methods should be available to all users and producers of official statistics (i.e. countries should in theory be able to implement them with sufficient effort and resource), with clear guidance on how to use them.
- Flexibility: while detailed modelling may be useful more generic methods should be flexible to deal with the variety of data and other complexities in regular production environments.
- Transparency: methods should be clear and methodological notes provided to explain what methods have been used in the production of official statistics.
- Quality: standard dimensions of quality should be considered in the choice of any method.
- Conformity: a method should not introduce artificial changes in the properties of the indicator. An important example of this is the so-called step problem.
- Time-symmetry: a method should yield the same results when performed on a set of benchmark and indicator series when time is reversed. If a method does not have this property, it can be shown that the timing of events (like the onset of a crisis) can be changed by the procedure.

1.6 Choice of software

Implementations of methods for temporal disaggregation, benchmarking and reconciliation exist in a wide range of free or proprietary statistical software. Some software is developed specifically for production of official statistics, for example JDemetra+.

For some methods there can be a lack of consistency between different software, especially in the case of multivariate methods. Hence, great care must be taken when it comes to the choice of a specific software for production of official statistics. It is important, for example, to consider: maintenance and support of the software; compatibility with these guidelines; cost and accessibility (possibly leading to a preference for open-source software); architecture and suitability for mass production; computational efficiency; and the extent of a user community in the production of official statistics.

The choice of software should allow a definition of an appropriate penalty function, according to the desired solution. In statistical methods the penalty function is also required that can be either a residual sum of squares, or a likelihood function more suitable for the well-known inferential properties.

The adoption in the software of the Kalman filter for statistical treatment is also recommended, especially when we have taken the logarithm of the data. As software is continually developing, these guidelines do not include specific recommendations on software to use. However, following the principles outlined above certain options will likely stand out as preferred. For example, thoroughly tested open-source software officially released by statistical institutions that is maintained and follows a clear release strategy such as JDemetra+.

2 General aspects

2.1 A general policy

DESCRIPTION

A general policy for the use of temporal disaggregation, benchmarking and reconciliation techniques in the production of infra-annual statistics should be based on the set of principles presented in chapter 1 of these guidelines. It should contain at least information related to: the statistical domains in which temporal disaggregation, benchmarking and reconciliation techniques should be used, the proposed methodologies, the quality framework of reference and the dissemination strategies proposed. The availability of such a policy will increase the transparency of the production process and trust in data producers.

OPTIONS

- A clear and comprehensive policy based on a set of agreed principles.
- A policy only partially based on a set of agreed principles.
- A policy not in line with a set of agreed principles.
- No general policy.

ALTERNATIVES

(A) Develop and disseminate a general policy on the production of infra-annual statistics using temporal disaggregation, benchmarking and reconciliation techniques based on a set of agreed principles, and detailing all important aspects related to the production, release, revision, quality assessment and dissemination of results.

(B) Develop and disseminate a policy on the production of infra-annual statistics based on temporal disaggregation, benchmarking and reconciliation techniques only partially in line with a set of agreed principles, covering at least partially important aspects such as production, release, revision, quality assessment and dissemination of results.

(C) Adopt an incomplete policy or a policy not in line with agreed principles; lack of any policy.

2.2 Domain specific policies

DESCRIPTION

Temporal disaggregation, benchmarking and reconciliation techniques can be applied to several statistical domains with various characteristics. Statistical domains may also differ in terms of statistical constraints and legal requirements which need to be considered when applying temporal disaggregation, benchmarking and reconciliation techniques. For example, national accounts are characterized by different requirements and needs than labour market or short-term business statistics. To provide a clearer framework for the application of the guidelines, domain specific policies detailing all data specificities and peculiarities as well as specific user and legislative requirements should be described.

OPTIONS

- Develop domain specific policies for the compilation of temporal disaggregation, benchmarking and reconciliation techniques that are fully compliant with the general one.
- Derive domain specific policies for temporal disaggregation, benchmarking and reconciliation techniques that are only partially compliant with the general policy.
- Compile domain specific policies for temporal disaggregation, benchmarking and reconciliation techniques that are partially inconsistent with the general policy.
- Not developing any domain specific policy for temporal disaggregation, benchmarking and reconciliation techniques.

ALTERNATIVES

(A) Develop domain specific policies for temporal disaggregation, benchmarking and reconciliation techniques that are fully compliant with the general policy and accounting for all issues specific to each statistical domain.

(B) Develop domain specific policies for temporal disaggregation, benchmarking and reconciliation techniques that are only partially compliant with the general policy or covering only partially issues specific to each statistical domain.

(C) Develop domain specific policies for temporal disaggregation, benchmarking and reconciliation techniques that are inconsistent with the general policy or do not develop at all any domain specific policy.

2.3 Stability of policies

DESCRIPTION

Keeping general and domain specific policies for compiling temporal disaggregation, benchmarking and reconciliation techniques stable over time will increase users' confidence and will allow producers to work within a stable framework. On the other hand, it is not realistic never to revise the policies since the economic and statistical conditions as well as the relevance of a particular domain may change and consequently policies would need to be updated. Institutions should then find the right balance between keeping policies as stable as possible and ensuring that they are in line with an evolving world.

OPTIONS

- Keep general and specific policies for temporal disaggregation, benchmarking and reconciliation techniques as stable as possible over time and change them only after a thorough investigation, pre-announcing changes.
- Revise general and specific policies on the basis of a regular calendar, for example every five years, to ensure sufficient stability.
- Revise policies often, for example every year.
- Revise policies for temporal disaggregation, benchmarking and reconciliation techniques irregularly and without any pre-announcement.
- Never revise temporal disaggregation, benchmarking and reconciliation techniques policies.

ALTERNATIVES

(A) Keep the general and the domain specific policies for temporal disaggregation, benchmarking and reconciliation techniques as stable as possible over time and update them only when significant changes occur either in the economic system or in the statistical system; changes to the general and domain specific policies should be pre-announced.

(B) Update general and domain specific policies on a fixed schedule (for example every 5 years) to ensure sufficient stability; changes should be pre-announced.

(C) Update general and domain specific policies too frequently or on an irregular basis and without any pre-announcement; never update the general and the domain specific policies.

2.4 Quality assurance framework

DESCRIPTION

The quality assessment of temporal disaggregation, benchmarking and reconciliation techniques needs to consider all five dimensions of statistical output quality, as listed in the European Statistics Code of Practice and in the United Nations template for a Generic National Quality Assurance Framework:

- Relevance
- Accuracy and reliability
- Timeliness and punctuality
- Coherence and compatibility
- Accessibility and clarity

Measures identified for each dimension can be qualitative or quantitative. Qualitative measures will normally have a “Yes” or “No” value, while quantitative measures will normally be test statistics with the direct interpretation of “Pass” or “Fail”.

For example, relevance could be measurable qualitatively through consultation with users, while accuracy and reliability are generally measured quantitatively.

In principal, the use of temporal disaggregation, benchmarking and reconciliation techniques, can be treated within the standard quality framework without major adaptations. Minor improvements could be introduced to deal with specific modelling features of temporal disaggregation, benchmarking and reconciliation techniques.

OPTIONS

- Use the quality framework for official statistics suitably amended to incorporate specific features of temporal disaggregation, benchmarking and reconciliation techniques, particularly the timeliness and the accuracy dimensions.
- Use the default quality framework for official statistics.
- Use an ad hoc quality framework for temporal disaggregation, benchmarking and reconciliation techniques.
- Do not use any quality framework.

ALTERNATIVES

(A) Use the standard quality framework for official statistics suitably amended to account for all details of temporal disaggregation, benchmarking and reconciliation techniques.

(B) Use the default quality framework for official statistics.

(C) Use an ad hoc quality framework for temporal disaggregation, benchmarking and reconciliation techniques or not use any quality framework.

3 Temporal disaggregation

3.1 The design of the temporal disaggregation exercise

DESCRIPTION

When starting a temporal disaggregation exercise for the first time, involving several related low frequency variables (target variables) to be converted into higher frequency ones, several aspects need to be analysed. The first one is the quality of the low frequency indicators (at various disaggregation levels). The second is the availability and quality of high frequency indicators at various disaggregation levels. The last is legal requirements and/or the users' needs on the level of disaggregation. Statistical considerations of the quality of the output from the temporal disaggregation exercise at different disaggregation levels should also be considered. The choice of the level of disaggregation should be regularly assessed in the light of changes in legislation, new user requirements, and expert's appreciation of past experiences as well as the structural changes affecting the overall quality of statistical data due to major revisions or extraordinary events.

OPTIONS

- Choose the disaggregation level considering legislative requirements, users needs or both factors together with a detailed analysis of the quality and availability of information needed.
- Choose the disaggregation level considering legislative requirements, user needs or both without any further investigation.
- Identify the most suitable disaggregation level, compliant with legislative requirements, with an in-depth analysis of the quality of the high frequency estimates produced.
- Identify the disaggregation level without considering legal constrains and user requests.
- Do not perform any preliminary analysis to identify the most suitable level of disaggregation.

ALTERNATIVES

(A) Identify the most appropriate disaggregation level, compliant with legislative requirements and satisfying as much as possible user requirements, using an in-depth analysis considering the quality and availability of information and the quality of high frequency estimates.

(B) Base the identification of the disaggregation level only on legislative requirements and, possibly, also on user requirements.

(C) Absence of any preliminary analysis to identify the disaggregation level.

3.2 The choice of estimation strategy

DESCRIPTION

The choice of estimation strategies is strictly related to the identification of the best contemporaneous disaggregation level. When multiple series linked by aggregation constraints must be estimated, there are several possible alternatives. The first one is to estimate separately the aggregate variable and its individual components (direct method). In this case the fulfilment of contemporaneous aggregation constraints is not ensured. The second is the use of the direct method complemented by reconciliation techniques to ensure the fulfilment of contemporaneous constraints (see also chapter 5). The last possibility is to estimate individual components of an aggregate and derive the temporally disaggregated aggregate by aggregating the estimated components (indirect method).

There is neither theoretical reasoning nor empirical evidence favouring one approach, but decisions should be taken on case-by-case basis with clearly defined evaluation criteria. The direct method seems to be preferred when there are co-movements between aggregates and components and when the quality of components is not homogenous which can happen especially at a very disaggregated level. The indirect approach seems to be preferable when the quality of the components is high enough to ensure that the estimation of the aggregate will also be of high quality.

OPTIONS

- Use the direct approach at each disaggregation level.
- Use the direct approach complemented by reconciliation techniques at each disaggregation level.
- Use the indirect approach at each disaggregation level.
- Use the direct approach, complemented by reconciliation techniques, at the lowest disaggregation level and the indirect approach for the more aggregated series.

ALTERNATIVES

(A) For each level of disaggregation choose the most appropriate strategy by comparing the performance of the direct approach, complemented with reconciliation techniques, and the indirect approach.

(B) Either use the direct approach, complemented by reconciliation techniques, or the indirect one without any comparison of their relative performance but using expert judgment.

(C) Use an indirect approach, without checking the quality of the indirectly derived aggregate.

3.3 Choice of high frequency indicators

DESCRIPTION

An exhaustive set of high frequency candidate indicators (including hard and soft data when appropriate) should be considered. They should cover at least the same time span as the low frequency target indicator and be sufficiently timely. Furthermore, candidate high frequency indicators should closely approximate the expected short-term movement of the target variable and show a good correlation with the original target variable when converted to the low frequency. Furthermore, candidate high frequency indicators should be sufficiently regular, not too volatile and available non-seasonally adjusted and seasonally adjusted when needed. Similarities of definitions and production process between low frequency target variable and candidate indicators should also be considered. Using all candidate variables in the temporal disaggregation process is not generally recommended especially when static regression models are used. In such cases, using all candidate indicators can increase the risk of collinearity and reduce considerably the number of degrees of freedom of the regression model by compromising the overall quality of high frequency estimates of the target variable. Consequently, a variable selection step prior to the temporal disaggregation is needed. Both graphical and statistical methods can be used at this stage. The former can help in visually detecting similarities among indicators and target variables (at low frequency), while the latter can provide useful information in terms of cross-correlation between indicators and target variables (for example, leading-lagging relationships, presence of common variances). Statistical methods can also be used to study the residual behaviour from a low frequency static regression between target variable and indicators, helping to detect misspecification problems or missing information. Variable selection tools, such as general to specific (GETS) can be used at this stage to identify the best uncorrelated set of indicators.

OPTIONS

- Use both graphical and statistical methods, complemented when needed by variable selection techniques to identify the most appropriate set of indicators to be used in the temporal disaggregation exercise.
- Use only graphical inspection to select indicators.
- Rely on expert knowledge only to select variables.
- Do not perform any selection of the candidate variables.

ALTERNATIVES

(A) Select the most appropriate set of indicators for the temporal disaggregation exercise using graphical and statistical methods and variable selection techniques if needed, limiting the presence of collinearity among the selected indicators and fixing the number of selected indicators using the principle of parsimony.

(B) Only rely on graphical methods.

(C) Do not perform any selection of indicators or only select them based on expert knowledge not supported by statistical evidence.

3.4 Hard and soft data

DESCRIPTION

Hard indicators provide a quantitative measure of a given phenomenon available at high frequency. Hard data are often high-quality indicators, mainly official statistics, strongly correlated with the level of the target variable, but may only be available after a time lag with respect to the reference period so that their use can delay the temporal disaggregation exercise. Nevertheless, their high quality and their nature as official statistics make them ideal candidates to be used in the temporal disaggregation exercise. Soft indicators, generally derived from qualitative opinion surveys, offer a signal of the current and future tendencies of a phenomenon but not of their size. Signals are obtained directly from agents and they reflect their opinions about the present and near future. Their use requires a translation to a quantitative scale to preserve the levels of the temporally disaggregated target series. They are released much sooner than hard data and for this reason they are suitable for flash estimates. However, since they are based on opinions, they may be noisier than official (hard) statistics. Since soft data better describe the movement around the trend of a given variable, care has to be given when modelling them in a temporal disaggregation exercise. Hard and soft data are usually considered as complementary sources of information within a temporal disaggregation exercise.

OPTIONS

- Use a set of candidate indicators containing both hard and soft data.
- Use only hard data as a set of candidate indicators.
- Use only soft data as a set of candidate indicators.
- Use soft data as a tool for extrapolating hard data for the latest period before using them as candidate indicators.

ALTERNATIVES

(A) Use both hard and soft data when constructing a set of candidate indicators.

(B) Use soft data only for extrapolating the latest period of hard indicators before their inclusion in the set of candidate indicators.

(C) Use a set of candidate indicators only composed of soft data when hard data is also available.

3.5 Data reduction

DESCRIPTION

When a rich information set, containing several candidate variables, is available, an alternative to the selection of indicators is represented by data reduction techniques. Such methods aim at reducing the information set by compiling a small number of orthogonal linear combinations of the original set of indicators (factors) keeping the most relevant information contained in the original set of candidate indicators. The orthogonality of factors guarantees the absence of any collinearity in the temporal disaggregation exercise. The limited number of factors also ensures that the low frequency regression will have a sufficient number of degrees of freedom. Well known data reduction methods are: principal components analysis (PCA) and partial least squares (PLS). They differ mainly in terms of the optimization rule adopted. When a set of candidate indicators contain different categories of data such as hard and soft data, it is recommended to construct separate factors for each data category to avoid the dominating effect of one category on the others which could happen if all kinds of categories go simultaneously in the same factors. The optimal number of selected factors can be decided according to empirical considerations or to statistical criteria. Single factor models are used when dealing with a homogenous set of indicators while two factors are usually preferred whenever hard and soft data coexist in the same indicators set. More factors can also be selected to synthetize more complex data structures adequately.

OPTIONS

- Consider the simultaneous presence of different data categories (hard and soft) when applying empirical or theoretical methods to select the ideal number of factors.
- Always rely on a single factor model extracted either by PCA or PLS.
- Define the number of factors according to statistical and empirical criteria.
- Do not use any specific algorithm to select the number of factors.

ALTERNATIVES

(A) Use all available criteria to identify the right number of factors, taking into account the presence of different data categories in the original set of indicators.

(B) Always rely on a single factor model.

(C) Do not use any specific criteria to select the ideal number of factors.

3.6 Choice of temporal disaggregation methods

DESCRIPTION

A temporal disaggregation exercise, particularly with many target variables and indicators, involves several challenges. These include, among others, the characteristics of the target variables (sparse observations, values close to zero...); the availability of indicators, the reliability of both target variables and indicators, etc. Although some temporal disaggregation methods perform reasonably well for almost all possible cases, there is no single method systematically outperforming the others across all possible situations. For example, a simple pro rata method can work well when the target variable is sparse with values close to zero and irregularly distributed. Similarly, in the absence of indicators, the use of ARIMA decomposition-based methods such as Stram and Wei may be appropriate. Finally, it must be noted that regression-based temporal disaggregation methods usually provide good results in most cases.

OPTIONS

- Rely on a wide range of methods chosen for their ability to deal with specific situations (preferably implemented within the same software).
- Chose only a very limited number of methods providing reasonable results in various cases and possibly implemented in the same software.
- Rely on methods implemented in a variety of software.
- No preselection of methods but decisions are taken on a case-by-case basis.

ALTERNATIVES

(A) Predefine a wide list of methodologically sound, well-tested and stable methods implemented in the same software, providing answers to all possible empirical situations.

(B) Predefine a restricted list of methodologically sound, stable and well-tested methods, implemented in the same software, working reasonably well in various cases.

(C) Do not provide any preselection of methods and decide case by case; rely on unstable or not adequately tested methods.

3.7 Temporal disaggregation without indicators

DESCRIPTION

Temporal disaggregation can be performed even when no high frequency indicators are available. In this case it is possible to consider regression methods using deterministic variables such as a constant or linear trend as an indicator. Alternatively, the use of time series methods, which do not require any specification of indicators, such as Stram and Wei, can be investigated. Also, the use of simple mathematical methods such as the pro rata one is not excluded. Apart from the methods selected, the estimated high frequency pattern will not be very informative due to the impossibility of properly estimating seasonal and cyclical movements. If the low frequency target variable, for which high frequency indicators are not available, is not a relevant component of an aggregate (representing a small share of the aggregate), its high frequency data can be computed without the risk of compromising the quality of the corresponding high frequency aggregate. By contrast if this target variable is a relevant component of an aggregate, or is an aggregate itself, the possibility of not providing any high frequency result should be seriously considered.

OPTIONS

- Not estimating high frequency data in absence of indicators.
- Performing temporal disaggregation by using regression methods with deterministic indicators.
- Using Stram and Wei or other ARIMA based temporal disaggregation methods.
- Using Cholette-Dagum with or without deterministic trend.
- Using unobserved components methods.
- Using mathematical methods such as pro rata.

ALTERNATIVES

(A) Carefully evaluate the risks and benefits of computing high frequency estimates of a low frequency variable without any indicator, taking also into account legislative requirements and users' needs. If high frequency data need to be provided and the associated risks are not too high, compare a wide range of methods and select the one providing the most informative results. Otherwise, do not provide any high frequency information or limit as a much as possible its release.

(B) Always provide high frequency estimates even if no indicators are available using the best performing method selected within a small scale comparative exercise.

(C) Always use pro rata adjustment in the absence of indicators; never compute high frequency data in the absence of indicators even when requested by legislation.

3.8 Temporal disaggregation with indicators

DESCRIPTION

When one or more indicators are available, ideal candidate methods to be used for temporal disaggregation are those belonging to a family of regression models. The core of this family is represented by the Chow-Lin method and its variants proposed by Fernández and Litterman that differ essentially in terms of the structure of the residuals. They can also be considered as special cases of the Cholette-Dagum model and they can also be used with an unobserved component framework with a state-space representation. Furthermore, non-linear models expressed either in logarithms or logarithmic differences have been proposed in the literature, as well as dynamic models with or without error correction terms. Besides the Chow-Lin method and variations of this approach, some regression models are characterized by increased computational complexity and are not always necessary. The use of mathematical methods allowing for the inclusion of indicators such as Denton do not seem to represent a reasonable alternative because, unlike regression methods, they do not use all available information in optimal ways. Furthermore, they automatically allow extrapolation only by means of a naïve approach as suggested by the IMF handbook of quarterly national accounts. In contrast, all regression models can extrapolate automatically.

OPTIONS

- Perform the temporal disaggregation using always the original Chow and Lin specification.
- Select one among the Chow-Lin specifications and variants and use it in the whole temporal disaggregation exercise.
- Select the most appropriate specification and variant of the Chow-Lin method to be used for each single target variable to be temporally disaggregated.
- Perform the temporal disaggregation exercise by using non-linear or dynamic methods.
- Use non-regression-based methods, allowing for the use of indicators.

ALTERNATIVES

(A) For each temporal disaggregation exercise, select the most appropriate regression-based method using a large set of descriptive and parametric criteria including graphical analysis. Selection should also account for the computational complexity and usability of methods in regular production.

(B) Select the best method for the whole set of low frequency indicators to be temporally disaggregated, or subsets of them, within a small scale comparative exercise involving only Chow-Lin, Fernández or Litterman specifications.

(C) Use of non-regression-based methods.

3.9 Extrapolation with temporal disaggregation

DESCRIPTION

In a temporal disaggregation exercise, it may be important for publication purposes to have higher frequency estimates that extend beyond the available lower frequency time series. For temporal disaggregation with or without high frequency indicator series we define this more properly as a problem of extrapolation.

Many temporal disaggregation methods, especially regression-based methods, deal with extrapolation as part of the method. Ideally the choice of a temporal disaggregation method should provide the best extrapolated values. However, if consistency of methods within a larger exercise is important for processing considerations, it may be preferable to make some compromise, by choosing a method that performs best for most series.

It is possible to test the performance of extrapolation from alternative temporal disaggregation methods by analysing revisions to extrapolated values in a similar way that forecasts can be evaluated using out-of-sample forecast tests. In this case extrapolated high frequency values are the forecasts and final temporally disaggregated values using complete low frequency series are the values against which these forecasts should be compared.

For some methods of temporal disaggregation, such as cubic splines or pro rata, it may be preferable to use alternative explicit forecasting methods such as ARIMA models to forecast the lower frequency series before temporally disaggregating.

OPTIONS

- Test the performance of a wide range of temporal disaggregation methods using appropriate out-of-sample forecast tests and analysis of revisions to temporally disaggregated series choosing the method that minimizes out-of-sample forecast error and hence revisions.
- Use extrapolated values from the chosen temporal disaggregation method without having evaluated the revisions performance.
- Perform no extrapolations of the high frequency series for which there is no corresponding low frequency data.
- Perform extrapolation that is inconsistent with the historical structure of the series.

ALTERNATIVES

(A) Extrapolation is determined by the method of temporal disaggregation with minimizing the size of revisions of extrapolated values used as one of the empirical tests when deciding on the choice of method.

(B) Extrapolation is determined by the method of temporal disaggregation, but the choice of method does not include empirical tests on revisions to extrapolated values or a restricted choice of methods is tested.

(C) No extrapolation, or extrapolation by naïve or other methods inconsistent with the historical structure of the series.

4 Benchmarking

4.1 Choice of benchmarking method

DESCRIPTION

Benchmarking is a specific case of temporal distribution, where the high frequency indicator series and the low frequency benchmark series describe the same phenomenon. For instance, we might have quarterly and annual series of production of mining companies from different data sources and the quarterly sums might not equal the annual figures. Other aspects of the benchmarking exercise with respect to temporal distribution are that both low and high frequency series must be measured in the same units. In the benchmarking exercise only one high frequency indicator, which represents the preliminary estimates of the phenomenon, is present.

A benchmarking procedure restores temporal consistency between the high and low frequency series, by adjusting the high frequency series while preserving as much as possible the short-term information contained in the high frequency series. Exactly what this high frequency information is, and how it is preserved, is where benchmarking methods differ. The most commonly applied principle in benchmarking methods is *movement preservation*. This means that the benchmarking method aims to preserve either the proportional or additive first differences of the high frequency series, or a mix of these.

OPTIONS

- Use the pro rata method.
- Use any one of the benchmarking methods based on movement preservation (Denton, Cholette-Dagum and Chow-Lin).
- Use the Causey-Trager GRP approach.

ALTERNATIVES

(A) Choose the best specification for a specific time series using the Cholette-Dagum model and make full use of its diagnostics.

(B) Use the Denton method, provided that an analysis of the benchmark to indicator ratios is used for diagnostics and care is taken to avoid the tail effects; or use a method from the Chow-Lin family.

(C) Use the pro rata method or the Causey-Trager growth rate preservation approach or any other method that introduces artificial changes in the high frequency series.

4.2 Dealing with large discrepancies

DESCRIPTION

In benchmarking the assumption is made that the differences between the indicator and the benchmarks are small. In this case, all methods based on movement preservation perform well and will yield very similar results. It is important, therefore, that this assumption is always checked.

It is not easy to define what constitutes a large discrepancy. A practical threshold value for most series will be somewhere between 5 and 10%. However, it really depends on the characteristics of the time series and how much effort enhancing the quality of a time series is worth.

Large discrepancies are a sign that something is wrong with either the indicator or the benchmark series and always warrant an investigation. If the discrepancies are systematic and large, you are in the realm of temporal disaggregation and an approach from the previous chapter should be used. If the discrepancy is incidental, a survey may have produced an outlier, or a mistake may have been made in the compilation of either series. The cause of an incidental discrepancy is unlikely to fit the underlying error model of any benchmarking method. Therefore, large incidental discrepancies should be manually corrected.

OPTIONS

- Use a systematic (preferably automated) setup to check for large discrepancies, using predefined threshold values for the differences. Investigate and correct large incidental discrepancies and do not use benchmarking methods in the presence of large systematic discrepancies.
- Do not perform benchmarking in the presence of large discrepancies.
- Perform benchmarking regardless of the presence of large discrepancies.

ALTERNATIVES

(A) Use of a systematic (preferably automated) setup to check for large discrepancies with predefined threshold values for the differences. Investigate and correct large incidental discrepancies and do not use benchmarking methods in the presence of large systematic discrepancies.

(B) Do not perform benchmarking in the presence of large discrepancies.

(C) Perform benchmarking regardless of the presence of large discrepancies.

4.3 Combining benchmarking and seasonal adjustment

DESCRIPTION

If you have a high frequency series that requires both benchmarking and seasonal adjustment you have the option of benchmarking first and then seasonally adjusting it, possibly with further benchmarking applied to obtain temporal alignment, as discussed in the ESS guidelines on seasonal adjustment (2015), or seasonally adjusting the time series first and then benchmarking it.

When performing seasonal adjustment and/or calendar adjustment, small differences with the original annual total of the indicator may arise. It is possible to use benchmarking techniques to reconcile these differences. The ESS guidelines on seasonal adjustment (2015) recommend not to benchmark, unless there is an external requirement for temporal alignment, as there is in both national accounts and labour statistics. The main reason for not wanting to benchmark seasonally adjusted series is that there is no methodological requirement to do so and it may harm the quality of the seasonal adjustment.

OPTIONS

- Always first perform seasonal adjustment and then do benchmarking.
- Perform benchmarking first and then do seasonal adjustment. In this case one ends up with small differences in the annual alignment. At this point, one has the option to leave these differences as they are or perform benchmarking again to restore alignment.

ALTERNATIVES

(A) Perform benchmarking on the unadjusted data and then seasonally adjust the benchmarked data following the ESS guidelines on seasonal adjustment.

(B) Seasonally adjust the data and then benchmark to an appropriately calendar adjusted low frequency series using a method that does not introduce seasonal or calendar effects into the final series only when required by legislation or other reasonable user requirements.

(C) Regardless of legislative or user requirements, always seasonally adjust the data and then benchmark to a low frequency series and/or do so in a way that causes residual seasonal and or calendar effects in the final series.

4.4 Extrapolation and benchmarking

DESCRIPTION

When the low frequency benchmark is not yet available at the current end of the series, the high frequency observations in a benchmarking exercise are extrapolated as described in section 3.9 on extrapolation with temporal disaggregation.

An additional consideration concerns the publication of the final high frequency period in a low frequency period (for example, the last quarter of a year). In this situation, if the final high frequency data is available before the low frequency one, then extrapolation can be used as discussed in section 3.9. However, if the low frequency series is sufficiently timely, or if there is strong evidence that the gain in reliability from waiting, based on historical revisions performance, outweighs the possible reduction in timeliness, then it may be preferable to wait and use the low frequency period for benchmarking rather than extrapolating.

OPTIONS

- Always wait for the low frequency observation before estimating the last point of the high frequency data.
- Estimate the last high frequency observation using extrapolation determined by the benchmarking methods as discussed in section 3.9 and then revise the data when the low frequency observation is available to benchmark to.
- Estimate the last high frequency observation using extrapolation but do not revise when the benchmark period becomes available.
- Estimate the last high frequency observation by adjusting only that value to give consistency with the low frequency benchmark but not revising other high frequency extrapolated values using the chosen benchmarking method.

ALTERNATIVES

(A) Decide between extrapolation of the last high frequency point or waiting for the low frequency observation to perform benchmarking by assessing the quality attributes of timeliness of the low frequency data and the reliability of extrapolations based on revisions between past extrapolations of the high frequency series and the final benchmarked series.

(B) Choose to extrapolate or wait for the low frequency data when publishing the last high frequency period, where the choice is made without any analysis of the quality dimensions of timeliness and reliability.

(C) Estimate the last high frequency observation by adjusting only that value to give consistency with the low frequency benchmark but without revising the other high frequency extrapolated values by using the chosen benchmarking method. Or estimate the last high frequency observation using extrapolation but do not revise extrapolated values when the benchmark period becomes available.

5 Reconciliation and multivariate temporal disaggregation

5.1 Choice of multivariate temporal disaggregation method

DESCRIPTION

Standard temporal disaggregation methods consider one target series at a time and do not consider relationships between other series. Therefore, a set of separately temporally disaggregated or benchmarked series may not form a consistent picture where contemporaneous or accounting constraints are met. For example, temporally disaggregated quarterly estimates of GDP from the production side may differ from the temporally disaggregated quarterly estimates of GDP from the expenditure side, even though the annual data are consistent. Reconciliation techniques are defined as the statistical processes that aim to temporally disaggregate a system of time series without losing contemporaneous consistency.

Multivariate temporal disaggregation techniques can be divided into two categories: simultaneous methods, and two-steps methods.

Possible approaches include the use of state-space models, the multivariate extension of the Chow-Lin model, the multivariate random walk model and its extensions, dynamic factor models and some semiparametric approaches such as spline methods (Mazzi and Proietti, 2017).

When the dimension of the system is large, it may become difficult to apply simultaneous techniques using standard algorithms. In this case a version of the two-step reconciliation procedure applying a temporal disaggregation technique in the first step could be used (Quenneville and Rancourt, 2005, Di Fonzo and Marini, 2011). Such methods solve the temporal constraint in the first step, by applying univariate temporal disaggregation techniques on each series of the system, and the contemporaneous constraint in the second step, by applying a constrained optimization technique to each of the low frequency periods, without altering the dynamic movements of the series.

OPTIONS

- Multivariate Chow-Lin method and its variants.
- Multivariate Denton method.
- Two-step procedure.
- State-space models.
- Dynamic factors models.
- Semiparametric models or some other relevant observation-driven methods.
- Combinations of some of the models mentioned above.

ALTERNATIVES

(A) State-space models for multivariate temporal disaggregation.

(B) Any other kind of simultaneous methods, for example a version of the multivariate Chow-Lin method or a version of the two-step procedure using a temporal disaggregation technique in the first step.

(C) Any other methods that can be shown to cause unreasonable distortions to the movements of the original series or that do not meet the temporal and contemporaneous constraints.

5.2 Choice of reconciliation method

DESCRIPTION

As for temporal disaggregation techniques, standard benchmarking methods also consider one target series at a time. Reconciliation techniques are statistical or mathematical processes that aim to restore consistency in a system of time series with regards to both contemporaneous and temporal constraints.

Reconciliation is typically needed when the total and its components are estimated independently (direct approach), while in the case where only the components are estimated, the contemporaneous constraint is solved by the aggregation of the components (indirect approach).

Reconciliation techniques may be divided into two categories: simultaneous methods, and two-step methods.

Amongst the simultaneous methods, the most popular is the multivariate Denton method (or the multivariate Cholette method), a multivariate extension of the univariate Denton method which includes all the high frequency series in the system where the minimization problem is extended to include the constraints. An extension of multivariate Denton method for benchmarking is proposed for large data sets by implementing an optimization solver as in Dutch national accounts (Bikker et al., 2013). Di Fonzo and Marini (2015) propose to use a multivariate version of the GRP method. When the sample sizes are small, but a large number of series are to be benchmarked and reconciled according to a complex hierarchical scheme, a state-space model might be used, as discussed in Tiller and Pfeffermann (2011). This method is especially useful for situations when time series data are combined with cross-sectional data and when the incomplete information is to be treated by small area estimation. See also US Bureau of Labor Statistics (BLS) methodology for producing monthly employment and unemployment estimates where benchmarking is implemented together with seasonal adjustment using state-space models.

When the dimension of the system is large, it may become difficult to apply simultaneous techniques (as the multivariate Denton approach) using standard algorithms. In this case a two-step reconciliation procedure could be used (Quenneville and Rancourt, 2005, Di Fonzo and Marini, 2011). Such methods solve the temporal constraint in the first step, by applying univariate benchmarking techniques on each series of the system, and the contemporaneous constraint in the second step, by applying a constrained optimization technique to each of the low frequency periods, without altering the dynamic movements of the series. Such methods might approximate the results of the multivariate Denton method.

Some possible alternative approaches include dynamic factor models and some semiparametric approaches such as spline methods (Mazzi and Proietti, 2017).

OPTIONS

- Multivariate Denton method and its variants.
- Two-step procedure.
- State-space models.
- Semiparametric models or some other relevant observation-driven methods.
- Combinations

ALTERNATIVES

(A) Apply an appropriate multivariate method such as a state-space model or multivariate proportional Denton method for benchmarking. When the dimension of the system is too large to be solved efficiently in a single step, a two-step procedure may be used.

(B) Always apply a two-step procedure.

(C) Any other methods that can be shown to cause unreasonable distortions to the movements of the original series or that do not meet temporal and contemporaneous constraints.

6 Specific Issues

6.1 Outliers identification and treatment

DESCRIPTION

The presence of outliers can reduce the overall quality of temporal disaggregation, benchmarking and reconciliation exercises. It is important to understand the reasons for apparent outliers before deciding how to deal with them. If outliers appear in the indicator series only as a single extreme value (additive outliers) and appear to be an error rather than explained by statistical or economic reasons, they should be corrected before running temporal disaggregation, benchmarking or reconciliation procedures. If they appear in both the high frequency indicator and low frequency target variable and are statistically or economically explicable, they should be either modelled during the estimation process or removed before its start and reintroduced at the end. Finally, if more than one structural outlier appears in a few indicators reflecting changes in the production process which are not reflected in the low frequency target variable, the possibility of removing such high frequency indicators from the set of candidate variables should be evaluated.

OPTIONS

- Remove all outliers before running temporal disaggregation, benchmarking or reconciliation.
- Ignore the presence of outliers.
- Model within the estimation process all the outliers in the indicators and low frequency target variables.
- Remove all outliers present in both indicators and low frequency target variables before starting the estimation process and reintroduce them at the end.
- Always remove additive outliers before starting the estimation process without reintroducing them at the end.
- Discard high frequency indicators with transitory or permanent level shifts caused by changes in the production process not reflected in the low frequency target variable.

ALTERNATIVES

- **(A)** Carefully analyse the typology and the nature of the outliers and undertake the most appropriate correcting measures. Modelling outliers in both the indicators and low frequency target variables within the estimation process or removing them before starting the estimation process and reintroducing them at the end.
- **(B)** Remove all outliers regardless of their typology and nature.
- **(C)** Ignore the presence of outliers.

6.2 Treatment of short series

DESCRIPTION

Most popular temporal disaggregation and benchmarking methods are based on a statistical regression model for the unknown values to be estimated, e.g. Cholette and Dagum (1994) or Chow and Lin (1971) and their variants. When only short time series are available (for example less than ten years), these methods may not be appropriate since the involved parameters can be poorly estimated, which would negatively affect the quality of the temporally disaggregated or benchmarked series.

Appropriate methods for temporal disaggregation and benchmarking where only short series are available will depend to some extent on the nature of the data (e.g. volatility). The tables below provide some guidance on methods to consider.

Temporal Disaggregation

Number of low frequency time points	**Potentially appropriate methods**
10 or more	Regression based methods
5 to 10	Regression based methods if there is a clear and stable relationship (Fernández may be more appropriate if there is instability in estimating the correlation coefficient) otherwise spline or pro rata
2 to 3	Spline methods or pro rata
1	Pro rata

Benchmarking

Number of low frequency time points	**Potentially appropriate methods**
5 or more	Cholette-Dagum
3 to 5	Denton (or Cholette-Dagum with fixed AR model e.g. rho=0.7)
Less than 3	Pro rata

OPTIONS

- Use of regression-based methods without considering the length of the series and diagnostics associated with the method.
- Use of mathematical based methods (e.g. spline or pro rata).
- Only use pro rata method.
- Choose a method for temporal disaggregation or benchmarking based on the recommended methods for the length of the series with an assessment of the methods using appropriate diagnostics.

ALTERNATIVES

(A) Choose a method for temporal disaggregation or benchmarking based on the recommended methods for the length of the series with an assessment of alternative methods using appropriate diagnostics where possible.

(B) Choose a method for temporal disaggregation or benchmarking only considering the length of the series.

(C) Application of complex statistical methods where the series are too short.

6.3 Dealing with chain-linked series

DESCRIPTION

If temporal disaggregation or benchmarking of chain-linked measures is required, there is a question of whether to do this before or after chain-linking. Chain-linking techniques are discussed in more detail with further references in the Handbook on Quarterly National Accounts (Eurostat, 2013 edition). An important consideration in the decision is whether the time series in the temporal disaggregation or benchmarking exercise are consistent. Chain-linked series can be considered consistent time series and therefore temporal disaggregation and benchmarking methods can be applied directly.

In statistical processes, aggregating chain-linked Laspeyres-type volume measures for example, involves 'unchaining' the chain-linked measures to provide series described as being in previous year's prices (PYPs), aggregating PYPs up the aggregation structure and then re-chaining. In this situation it may be important to consider the stage at which temporal disaggregation or benchmarking is applied and should be considered alongside the discussion in section 3.2 of these guidelines. An 'unchained' series in PYPs is not strictly speaking a consistent time series and should not be directly temporally disaggregated or benchmarked. The method of chain-linking can distort temporal constraints, for example the quarterly overlap method, and therefore it is important to restore the temporal consistency in these cases by benchmarking after chain-linking.

If a direct approach to temporal disaggregation or benchmarking is taken for a set of sectoral or geographic series other reconciliation techniques should not be performed to restore sectoral or geographic consistency as the chain-linked series are not additive.

OPTIONS

- Temporal disaggregation or benchmarking is performed on chain-linked series. If after consulting section 3.2 of these guidelines sectoral or geographical aggregation of the temporally disaggregated or benchmarked series is required, these are 'unchained' to provide previous year price measures that can be aggregated and then re-chained. Depending on the method of chain-linking, benchmarking may be required to provide temporal consistency in the final temporally disaggregated or benchmarked aggregated chain-linked high frequency series.
- Do not consider issues of direct or indirect methods of temporal disaggregation and benchmarking and only ever perform temporal disaggregation or benchmarking directly on individual chain-linked series.
- Temporal disaggregation or benchmark methods applied to 'unchained' previous year's price measures.
- Directly temporally disaggregate or benchmark series and use other reconciliation techniques to restore sectoral or geographic consistencies.

ALTERNATIVES

(A) Temporal disaggregation or benchmarking is performed on chain-linked series. If after consulting section 3.2 of these guidelines sectoral or geographical aggregation of the temporally disaggregated or benchmarked series is required, these are ‘unchained’ to provide previous year price measures that can be aggregated and then re-chained. Depending on the method of chain-linking (for example, the quarterly overlap method), benchmarking may be required to provide temporal consistency in the final temporally disaggregated or benchmarked aggregated chain-linked high frequency series.

(B) Do not consider issues of direct or indirect methods of temporal disaggregation and benchmarking and only ever perform temporal disaggregation or benchmarking directly on individual chain-linked series, ignoring issues of sectoral or geographic consistency.

(C) Temporal disaggregation or benchmarking is performed on ‘unchained’ data. Or directly temporally disaggregate or benchmark series and use other reconciliation techniques to restore sectoral or geographic consistencies.

6.4 Calendarization

DESCRIPTION

Time series data do not always coincide with calendar periods. Examples include fiscal years starting in March or April or retail survey data which may be collected on a four or five-week period instead of calendar month basis.

Calendarization is the process of transforming the values of a flow time series observed over varying time intervals into values that cover given calendar intervals such as month, quarter or year. The process involves two steps. The first is the temporal disaggregation of the observed values into a high-frequency (often daily) series. The second step reaggregates the resulting high-frequency (daily) values into the desired calendar reference periods.

In most cases, we have little information on the high-frequency series. The temporal disaggregation process could then be based on a method of disaggregation without indicator as discussed in chapter 3. The disaggregation might also involve a high-frequency indicator varying in function for instance of seasonality, of trading day or of other calendar effects. In such cases, the first step of calendarization will consist in benchmarking that indicator to the reported values, using for instance a modified Denton procedure.

Compared to temporal disaggregation or benchmarking, calendarization – especially in the case of daily disaggregation – implies solving large systems of equations, which may raise tricky practical issues. Approaches based on spline interpolation or on appropriate state-space representations can provide efficient solutions, contrary to standard matrix computations, which are often unfeasible. The state-space framework has the additional advantage of allowing an easy derivation of the errors that the calendarization process generates on the final aggregated results.

OPTIONS

- Assignment procedure, where the data relating to non-calendar periods is allocated to a specific calendar-period according to arbitrary rules.
- Fractional or pro rata method, allocating data to calendar periods as weighted sums of overlapping non-calendar period values.
- Temporal disaggregation without indicator or equivalently benchmarking using a constant indicator.
- Benchmarking using a priori information on the seasonality or on the trading day/calendar effects.

The two last options should be implemented using algorithms based on spline interpolation or on state-space forms.

ALTERNATIVES

(A) Benchmarking using a priori information or, failing that, temporal disaggregation without indicator. When an estimation of the standard errors generated by the process is desirable, state-space forms should be the preferred approach. Otherwise, fast methods based on spline interpolation are equivalently advisable.

(B) Fractional or pro rata method.

(C) Arbitrary assignment procedure.

7 Revisions

7.1 Routine revisions

DESCRIPTION

Routine revisions usually characterize the regular production process of high and low frequency statistics. They are usually due to some factors such as the incomplete information set used to produce preliminary estimates of official statistics, the replacement of some sources with more reliable ones, the update of parameters of seasonal adjustment and estimation techniques etc. As for seasonal adjustment in this context it is of great importance to define a clear policy for the re-estimation and re-identification of temporal disaggregation and benchmarking models to avoid too many revisions and to increase the transparency or procedures. Re-estimating and re-identifying models too frequently might lead to more precise results but also to more revisions; on the other hand, re-estimating and re-identifying the models quite rarely could contribute to the stability of data over time but might compromise the accuracy of the estimates. Therefore, a compromise between these two extremes has to be identified.

OPTIONS

- Re-estimate parameters each time the method is used and re-identify the models once a year.
- Re-estimate parameters and re-identify the models each time the method is used.
- Re-identify and re-estimate models and parameters once a year.
- Re-estimate and re-identify parameters and models on an irregular base.
- Never re-estimate and re-identify parameters and models.

ALTERNATIVES

(A) When past data are revised for less than two years and/or new observations are available, re-estimate the parameters and the models once a year or whenever the seasonal adjustment models are revised. In case of data revised for more than two years and/or unexpected events, evaluate the case for re-estimating parameters during the year.

(B) Re-estimate the parameters each time the method is used and re-identify the models once a year.

(C) Any other solution.

7.2 Period of routine revisions

DESCRIPTION

Usually, revisions characterizing official statistics tend progressively to converge to zero so that, after a certain number of periods, data are not revised further and considered as definitive unless major revisions occur. This process is led by the low frequency data when both low and high frequency estimates exist for the same variable such as in the case of national accounts and labour market statistics. Ideally high frequency data should follow the revision behaviour of the low frequency ones not being anymore subject to revisions when low frequency data are definitive. Unfortunately, due to the use of temporal disaggregation, benchmarking and reconciliation techniques, this is not necessarily the case. To avoid the phenomena of a continuous revision process for high frequency data even when the corresponding low frequency figures are not subject to revision, some restrictions to statistical methods used on high frequency indicators may be required.

In some statistical domains the period of revisions is set by existing agreements that provide a compromise between constraints of the data production process and requirements for a common revision policy. For these cases the number of periods that are subject to revision is fixed and could differ at different times.

OPTIONS

- Estimate the benchmarking, temporal disaggregation and reconciliation models over the full time-span of low frequency data but revise only high frequency data corresponding to revised low frequency data and freeze the previous data.
- Estimate the benchmarking, temporal disaggregation and reconciliation models over a significant subsample of low frequency data such as the last 10-15 years (or similar span for quarterly low frequency data) and freeze high frequency data corresponding to non-revised low frequency data.
- Do not use any restriction letting high frequency data to be revised theoretically over the full time-span.

ALTERNATIVES

(A) Identify the most suitable time span for estimating the benchmarking, temporal disaggregation and reconciliation models but freeze high frequency data consistently to the non-revised low frequency ones checking regularly to avoid the presence of any break to this procedure.

(B) Always estimate the benchmarking, temporal disaggregation and reconciliation models on the whole time-span and freeze high frequency data consistently with low frequency data.

(C) Revise the whole high frequency time output where there are no revisions to the low frequency series and no significant improvement in the quality of the output.

7.3 Major revisions

DESCRIPTION

Major revisions rarely occur and are due to important changes in the statistical production system such as: adoption of new legislation, methodological changes, adoption of new definitions and/or nomenclatures etc. Furthermore, they can also be associated to a periodic review of the sources used in compiling statistics, in an update of the base year for fixed year base indices etc. In both cases major revisions are normally scheduled in advance, pre-announced and well documented also with comparative investigation of methods and models. The publication of major revisions is considered a good opportunity to review the temporal disaggregation, benchmarking and reconciliation techniques used to concentrate revisions in a single publication.

OPTIONS

- Carry out an in-depth review of temporal disaggregation benchmarking and reconciliation techniques whenever major revisions take place and provide detailed documentation about the changes planned in these fields.
- Perform a detailed review of temporal disaggregation benchmarking and reconciliation techniques together with major revisions providing only a summary documentation on the changes implemented.
- Only review temporal disaggregation, benchmarking and reconciliation techniques in a marginal way without any in-depth investigation.
- Review temporal disaggregation, benchmarking and reconciliation techniques independently from other major revisions.
- Never conduct a general review of temporal disaggregation, benchmarking and reconciliation techniques.

ALTERNATIVES

(A) Always schedule a general review of temporal disaggregation, benchmarking and reconciliation methods together with other major revisions. Define the content and the depth of the review also in relation to other foreseen changes. Provide a detailed documentation on the changes performed.

(B) Conduct a review of temporal disaggregation, benchmarking and reconciliation techniques together with some major revisions already foreseen without an in-depth investigation of all methodological aspects. Provide basic information to the users.

(C) Do not review temporal disaggregation, benchmarking and reconciliation techniques in the case of major revisions.

8 Presentation

8.1 Quality assessment of temporal disaggregation

DESCRIPTION

A standard check for quality of temporal disaggregation is provided by graphical and/or correlation analysis comparing disaggregated series with indicators. Sign analysis, distance measures and related tests are further evaluation methods.

Recursive exercises on different (increasing) subsamples provide a good check for stability of the disaggregation.

When statistical models are adopted the evaluation concerns standard residual diagnostics, goodness of fit statistics and tests on reliability and significance of parameters. The analysis on the stability extends in this case also to model parameters.

OPTIONS

- Limit evaluation to graphical and correlation analysis.
- Extend the analysis to sub-periods to investigate the stability of the temporal disaggregation.
- Extend the analysis to standard regression diagnostics when statistical methods are adopted.
- Do not perform any quality assessment.

ALTERNATIVES

(A) Check of goodness of fit and stability of temporal disaggregation. Use standard diagnostic checks when statistical methods are adopted.

(B) Limit evaluation to few distance measures.

(C) Absence of evaluation.

8.2 Accuracy of benchmarking and reconciliation

DESCRIPTION

Benchmarking and reconciliation procedures are generally applied at the end of the data production process. For the quality of the data it is critical to obtain a good quality benchmarked target series, where the revisions have limited effects. The benchmarking procedure cannot be considered a tool for the improvement in quality of initial poor data and the overall quality of benchmarked series will depend on both the quality of initial data and the quality of the benchmarking procedure. A measure of the quality of the benchmarking procedure should then not account for the original quality of the input data. The quality of the benchmarking procedure can be measured by the distance between the original series and the benchmarked series in terms of RMSE or in terms of growth rate preservation.

Graphical comparison between the preliminary high frequency series and the benchmarked one can provide useful information. Furthermore, it can be interesting in addition to the usually applied sign concordance rate also to investigate the difference of growth rate and the associated mean and variance. Obviously, the quality assessment of the benchmarking exercise should not ignore the size of the discrepancy to be distributed since it affects the final quality of the results of the exercise.

OPTIONS

- Carry out a graphical comparison of preliminary and benchmarking series.
- Complement a graphical comparison with a basic analysis of the growth rate to check the movement preservation principle.
- Carry out an in-depth comparison of preliminary and benchmarked series based on graphical and descriptive statistics. Also including a detailed investigation of the growth rate difference.
- Do not perform any quality assessment.

ALTERNATIVES

(A) Detailed quality assessment of the benchmarking exercise based on a graphical comparison of the preliminary and benchmarked series complemented by a detailed comparison of the growth rate using the sign concordance rate as well as investigation of the growth rate differences.

(B) Quality assessment based on a simple graphical comparison of preliminary and benchmarked series.

(C) Do not perform any quality assessment of the benchmarking exercises.

8.3 Metadata

DESCRIPTION

Metadata is data providing additional information about the main supplied data. When temporally disaggregated data (including benchmarked/reconciled data) are supplied it is important to complement it with information regarding their nature (i.e. temporally disaggregated), the used inputs, the used statistical methods, associated quality measures, links to relevant and publicly available policies and release calendar, reference to international standards, reference contacts. Thanks to this information the user will be able to use the data consciously, fully understanding its informational content and avoid misuse. Information provided should put the users in a position to replicate the procedure. Metadata should not just cover the methodological choices but also the reasons behind them.

Complement temporally disaggregated data (including benchmarked/reconciled data) with complete metadata including nature of the supplied data, inputs, statistical methods, associated quality measures, links to relevant and publicly available policies and release calendar, reference to international standards, reference contacts. A section of the metadata shall explain methodological choices and their reasons. A bibliography will also help the users.

OPTIONS

- Provide full metadata including all relevant information.
- Provide basic metadata.
- Provide incomplete basic metadata.
- Do not provide any metadata.

Alternatives

(A) Provide full metadata covering all relevant information.

(B) Provide basic metadata on the nature of the supplied data, inputs, statistical methods and reference contacts.

(C) Do not provide metadata or providing incomplete metadata.

References

Barcellan, R. (2002), ECOTRIM: a program for temporal disaggregation of time series, Workshop on Quarterly National Accounts, Eurostat, *Theme 2 Economy and finance:* pp.79-95.

Barcellan, R. and D. Buono (2002), Temporal Disaggregation Techniques – ECOTRIM Interface (Version 1.01), User Manual, Eurostat.

Bérubé, J. and S. Fortier (2009), PROC TSRAKING: An in-house SAS procedure for balancing time series. In *ASA Proceedings of the Business and Economic Section*. American Statistical Association.

Bikker, R., Daalmans, J. and Mushkudiani, N. (2010). A multivariate Denton method for benchmarking large data sets. Statistics Netherlands, Discussion Paper.

Bikker, R., Daalmans J. and Mushkudiani, N. (2013), Benchmarking large accounting frameworks: a generalized multivariate model, *Economic Systems Research*, 25:4, 390-408, DOI: 10.1080/09535314.2013.801010. Available at: http://dx.doi.org/10.1080/09535314.2013.801010

Bisio, Laura & Moauro, Filippo (2017). "Temporal disaggregation by dynamic regressions: recent developments in Italian quarterly national accounts," *Statistical Neerlandica*, Volume 72, issue 4 (Special Issue), 471-494.

Boot, J.C.G., W. Feibes, and J.H.C. Lisman (1967), Further Methods of Derivation of Quarterly Figures from Annual Data, *Applied Statistics*, Vol. 16 (1), pp. 65–75.

Bozik, J.E. and Otto, M.C. (1988) *Benchmarking: Evaluating methods that preserve month-to-month changes.* U.S. Census Bureau – Statistical Research Division, EE-88/07. http://www.census.gov/srd/papers/pdf/rr88-07.pdf

Brent G. (2018) Maximum likelihood estimation framework for table-balancing adjustments, *Statistical Neerlandica*, Volume 72, issue 4 (Special Issue), 520-532.

Causey, B., and Trager, M.L. (1981) *Derivation of Solution to the Benchmarking Problem: Trend Revision.* Unpublished research notes, U.S. Census Bureau, Washington D.C. Available as an appendix in Bozik and Otto (1988)

Chen B. Di Fonzo T., Howells T., Marini M. (2018) The statistical reconciliation of time series of accounts between two benchmark revisions, *Statistical Neerlandica*, Volume 72, issue 4 (Special Issue), 533-552.

Chen B. Di Fonzo T., Mushkudiani N. (2018) Benchmarking, Temporal Disaggregation, and Reconciliation of Systems of Time Series, preface to the Special Issue, *Statistical Neerlandica*, Volume 72, issue 4 (Special Issue), 402-405.

Cholette, P.A. and E.B. Dagum (1994), Benchmarking Time Series with Autocorrelated Survey Errors, *International Statistical Review*, Vol. 62(3), 365–377.

Chow, G.C. and Lin. A.-L. (1971), Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. *The Review of Economics and Statistics*, 53(4):372–375.

Cubadda G., B. Guardabascio (2012), A medium N approach to macroeconomic forecasting, *Economic Modelling, 29, 1099–1105.*

Cubadda G, B. Guardabascio (2012), On the use of Partial Least Squares for forecasting large sets of cointegrated time series, in Di Ciaccio et al. (Eds.), *Advanced Statistical Methods for the Analysis of Large Data-Sets*, Studies in Theoretical and Applied Statistics, Springer-Verlag, 171-180.

Daalmans J. (2018) On the sequential benchmarking of subannual series to annual totals, *Statistical Neerlandica*, Volume 72, issue 4 (Special Issue), 406-420.

Dagum, E.B. and P.A. Cholette (2006), *Benchmarking, Temporal Disaggregation, and Reconciliation Methods for Time Series*, Springer edition.

Denton, F. (1971), Adjustment of Monthly or Quarterly Series to Annual Totals: An Approach based on Quadratic Minimization, *Journal of the American Statistical Association*, Vol. 66, pp. 99–102.

Durbin, J. and Quenneville, B. (1997). Benchmarking by State Space Models. *International Statistical Review*, 65, 23-48.

ESS (2015) Guidelines on Seasonal Adjustment. https://ec.europa.eu/eurostat/documents/3859598/6830795/KS-GQ-15-001-EN-N.pdf/d8f1e5f5-251b-4a69-93e3-079031b74bd3

ESS (2013) Handbook on Quarterly National Accounts https://ec.europa.eu/eurostat/documents/3859598/5936013/KS-GQ-13-004-EN.PDF/3544793c-0bde-4381-a7ad-a5cfe5d8c8d0

ESS (2013) Guidelines on Revision Policy for PEEIs. https://ec.europa.eu/eurostat/documents/3859598/5935517/KS-RA-13-016-EN.PDF/42d365e5-8a65-42f4-bc0b-aacb02c93cf7

Eurostat and United Nations, Handbook on Rapid Estimates (2017) https://ec.europa.eu/eurostat/documents/3859598/8555708/KS-GQ-17-008-EN-N.pdf/7f40c70d-0a44-4459-b5b3-72894e13ca6d

IMF (2014), *Quarterly National Accounts Manual - Concepts, Data Sources, and Compilation*, ch. 6 Benchmarking and Reconciliation. https://www.imf.org/external/pubs/ft/qna/2000/textbook/ch6.pdf

Di Fonzo T. and Marini M. (2011), Simultaneous and two-step reconciliation of systems of time series: methodological and practical issues, *Journal of the Royal Statistical Society*: Series C (Applied Statistics), 60, 143-164.

Di Fonzo, T. and M. Marini (2012a), Benchmarking Time Series according to a Growth Rates Preservation Principle, *Journal of Economic and Social Measurement*, Vol. 37(3), pp. 225–252.

Di Fonzo, T. and M. Marini (2012b), On the Extrapolation with the Denton Proportional Benchmarking Method, *IMF Working Paper Series*, WP/12/169.

Di Fonzo, T. and M. Marini (2015), Reconciliation of systems of time series according to a growth rates preservation principle, *Statistical Methods & Applications*, Vol.24 (4), pp. 651-669.

Durbin, J. and B. Quenneville (1997), Benchmarking by State Space Models, *International Statistical Review*, Vol. 65(1), pp. 23–48.

Elezović S., Xie Y. (2018) Reconciliation of seasonally adjusted data with applications to the Swedish quarterly national accounts, *Statistical Neerlandica*, Volume 72, issue 4 (Special Issue), 590-602.

Fernández, R. B. (1981), A Methodological Note on the Estimation of Time Series, *The Review of Economics and Statistics*, Vol. 63(3), pp. 471–476.

Frale C, Marcellino M, Mazzi G., Proietti T. 2010. Survey data as coincident or leading indicators, *Journal of Forecasting*, 29: 109–131.

Frale C, Marcellino M, Mazzi G.L., Proietti T. 2011. Euromind: a monthly indicator of the euro area economic conditions. *Journal of the Royal Statistical Society*, A174(2): 439–470.

Girardi, A., Guardabascio, B., and Ventura, M. (2016), Factor Augmented Bridge Models (FABM) and Soft Indicators to forecast Italian Industrial Production, *Journal of Forecasting*, 10.1002/for.2393.

Grassi S., Proietti T., Frale C., Marcellino M. and Mazzi G. 2014. EuroMInd-C: a Disaggregate Monthly Indicator of Economic Activity for the Euro Area and member countries. *International Journal of Forecasting*, Volume 31, Issue 3, July-September 2015, Pages 712-738.

Guerrero V. M., Corona F. (2018) Retropolating some relevant series of Mexico's System of National Accounts at constant prices: The case of Mexico City's GDP, *Statistical Neerlandica*, Volume 72, issue 4 (Special Issue), 495-519.

Harvey, A. C., Pierce R. G. 1984. Estimating missing observations in economic time series. *Journal of the American statistical Association*, 79: 125-131.

Harvey, A. C. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Harvey A. and C.H. Chung. 2000. Estimating the underlying change in unemployment in the UK. *Journal of the Royal Statistical Society*, A, 163: 303-328.

Latendresse, E., Djona, M., and Fortier, S. (2007). Benchmarking sub-annual series to annual totals – from concepts to SAS procedure and SAS K12089 Chapter: 10 page: 253 date: February 14, 2012 Restoring Accounting Constraints in Time Series 253 enterprise guide custom task. In Proceedings of the SAS Global Forum 2007 Conference. SAS Institute Inc.

Latendresse, E., Djona, M. and Fortier, S. (2007). Benchmarking sub-annual series to annual totals – from concepts to SAS procedure and SAS K12089 Chapter: 10 page: 253 date: February 14, 2012 *Restoring Accounting Constraints in Time Series* 253 enterprise guide custom task. In *Proceedings of the SAS Global Forum 2007 Conference*. SAS Institute Inc.

Litterman, R. B. (1983), A Random walk, Markov model for the Distribution of Time Series, *Journal of Business and Economic Statistics*, Vol. 1(2), pp. 169–173.

Marini M. (2016), Nowcasting Annual National Accounts with Quarterly Indicators: An Assessment of Widely Used Benchmarking Methods, *IMF Working Paper* 16/71.

Mazzi, G.L. and Proietti, T. (2015). Multivariate Temporal Disaggregation. *Handbook on Rapid Estimates*, Ch. 8, Eurostat (Nov 13, 2015).

Mazzi G.L. Proietti T. (2017) “Multivariate Temporal Disaggregation” in *Handbook on Rapid Estimates*, Eurostat and United Nations, Manuals and Guidelines, 2017 edition

Mazzi G.L., R. Ruggeri Cannata (2017), " Rapid Estimates: Different Products for Different Purposes”, *Handbook on Rapid Estimates*, Ch. 2, Eurostat.

McElroy, T. (2018) Seasonal adjustment subject to accounting constraints, *Statistical Neerlandica*, Volume 72, issue 4 (Special Issue), 574-589.

Moauro F. and G. Savio. 2005. Temporal disaggregation using multivariate structural time series models. *Econometrics Journal*, 8: 214-234.

Moauro F. (2014) Monthly Employment Indicators of the Euro Area and Larger Member States: Real-Time Analysis of Indirect Estimates. *Journal of forecasting*, 33, 5: 339-49.

Mushkudiani N., Daalmans J., Bikker R. (2018) Solving large-data consistency problems at Statistics Netherlands using macro-integration techniques, *Statistical Neerlandica*, Volume 72, issue 4 (Special Issue), 553-573.

National Bank of Belgium (2016): JDemetra+, https://github.com/jdemetra/jdemetra-app.

Pfeffermann, D. and Tiller, R. (2011). State-Space Modeling with Correlated Measurements with Applications to Small Area Estimation Under Benchmarking Constraints. Southampton Statistical Science Research Institute, University of Southampton. Methodology Working Paper M03/11.

Proietti T. (2005) Temporal disaggregation by state space methods: dynamic regression methods revisited, in: “Rapporto finale della Commissione di studio sul trattamento dei dati ai fini dell’analisi congiunturale”, ISTAT, Ottobre 2005, pp.85-115, with description of Ox programmes, http://www3.istat.it/istat/eventi/2005/contieconomici3112005/rapportofinale.pdf.

Proietti T. (2006) Temporal disaggregation by state space methods: dynamic regression methods revisited, Econometrics Journal, 9, 357–372

Proietti T. (2006) On the Estimation of Nonlinearly Aggregated Mixed Models. *Journal of Computational and Graphical Statistics*, 15,1: 18-38.

Proietti T. and F. Moauro. (2006) Dynamic factor analysis with non-linear temporal aggregation constraints. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 55,2: 1467-9876.

Proietti T. and Moauro F. (2008). Temporal Disaggregation and the Adjustment of Quarterly National Accounts for Seasonal and Calendar Effects. *Journal of Official Statistics*, vol. 24, pp. 115–132

Proietti, T. (2011). Estimation of Common Factors under Cross-Sectional and Temporal Aggregation Constraints, *International Statistical Review*, 79, 455–76.

Proietti, T. (2011). Multivariate Temporal Disaggregation with Cross-Sectional constraints, *Journal of Applied Statistics*, 38, 1455–1466

Proietti, T., Marczak, M. and Mazzi, G. (2016). EuroMInd-D: A density estimate of monthly gross domestic product for the Euro Area. *Journal of Applied Econometrics*, doi: 10.1002/jae.2556.

Quenneville, B. and S. Fortier (2012), Restoring Accounting Constraints in Time Series – Methods and Software for a Statistical Agency, in *Economic Time Series Modeling and Seasonality*, ed. T.S. McElroy, Chapman and Hall, pp. 231–253.

Quenneville, B. and E. Rancourt (2005), Simple Methods to Restore the Additivity of a System of Time Serie, presented at the workshop *Frontiers in Benchmarking Techniques and their Application to Official Statistics*, Eurostat, Luxembourg, April 2005.

Quillis M. (2018) Temporal disaggregation of economic time series: The view from the trenches, *Statistical Neerlandica*, Volume 72, issue 4 (Special Issue), 447-470.

Santos Silva J.M.C. and F. N. Cardoso. 2001. The Chow-Lin method using dynamic models. *Economic Modelling*, 18 : 269-280.

Sanz, R. 1981. Metodos de desagregacion temporal de series economicas. Banco Espana. Servicio Estudios Economicos, 22.

Sax C. and Steiner R. (2013) Temporal Disaggregation of Time Series, http://journal.r-project.org/archive/2013-2/sax-steiner.pdf.

Stram, D. O. and Wei, W. W. (1986), Temporal aggregation in the ARIMA process. *Journal of Time Series Analysis*, 7: 279-292. doi:10.1111/j.1467-9892.1986.tb00495.x

Temursho U. (2018) Entropy-based benchmarking methods, *Statistical Neerlandica*, Volume 72, issue 4 (Special Issue), 421-446.

United Nations template for a Generic National Quality Assurance Framework (2012) https://unstats.un.org/unsd/dnss/docs-nqaf/2012-13-NQAF-E.pdf

Getting in touch with the EU

In person
All over the European Union there are hundreds of Europe Direct Information Centres. You can find the address of the centre nearest you at: https://europa.eu/contact

On the phone or by e-mail
Europe Direct is a service that answers your questions about the European Union. You can contact this service
– by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
– at the following standard number: +32 22999696 or
– by electronic mail via: https://europa.eu/contact

Finding information about the EU

Online
Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu

EU Publications
You can download or order free and priced EU publications from EU Bookshop at: https://bookshop.europa.eu. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/contact)

EU law and related documents
For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex at: https://eur-lex.europa.eu

Open data from the EU
The EU Open Data Portal (https://data.europa.eu/euodp/en/data) provides access to datasets from the EU. Data can be downloaded and reused for free, both for commercial and non-commercial purposes.

KS-06-18-355-EN-N

ESS guidelines on temporal disaggregation, benchmarking and reconciliation

In official statistics there is an increasing demand for indicators at a higher frequency than have traditionally been observed. Eurostat and the European Statistical System (ESS) developed these guidelines to help data producers derive high frequency data (e.g. quarterly or monthly) from low frequency data (e.g. annual) and to address related temporal and accounting constraints. These guidelines are aimed at those involved in the production and analysis of infra-annual European statistics (compiled by Eurostat) and corresponding country specific official statistics compiled by National Statistical Institutes (NSIs). They have been conceived both for experts and beginners. In order to rank the different methods, each issue is discussed following a structure with three parts: a description (free text presenting the problem), a list of options (various possibilities to deal with the specific issue) and a list of ranked alternatives (A,B,C). Namely:

(A) Best alternative: should always be the target for producers.

(B) Acceptable alternative: retained only if time or resource issues prevent alternative (A).

(C) Alternative to be avoided: not a recommended option.

Being in line with these guidelines will ensure higher quality results, e.g. avoiding false signals and misinterpretation of the dynamics of the data.

For more information
https://ec.europa.eu/eurostat/

Publications Office

978-92-79-98007-7