

Guidance Note on the incorporation of alternative data sources into CPI compilation

Introduction and purpose

10.1 The environment in which statistical agencies operate is changing. New opportunities to access and interrogate large and detailed datasets from digital systems are becoming available, increasing the potential to provide new insights and possibilities for the compilation of consumer price indices (CPIs). At the same time, the statistical landscape is becoming more complex, and the expectations of decision makers are growing, including demands for greater timeliness and detail in an environment where non-official data and analytics are increasingly available. National Statistical Offices (NSOs) are therefore being challenged to deliver high-quality statistics in more efficient, cost-effective and innovative ways.

10.2 This note provides guidance on the incorporation of alternative data sources into CPI compilation and supplements Chapter 10 of the 2020 CPI Manual. While traditional survey-based price collection remains the foundation of most CPIs, the growing availability of large, high-frequency datasets from private and public digital systems offers significant potential to improve coverage, accuracy, and efficiency. This includes improved temporal coverage (for example, higher-frequency price observation), expanded spatial coverage (including the potential to better reflect regional variation), and enhanced representativity of items and outlets in some markets.

In addition to their direct use in index compilation, such data may also support other aspects of CPI production, including the development of expenditure weights, quality assurance and validation of existing methods, and improved understanding of market behaviour. The considerations discussed in this chapter generally apply whether alternative data sources are used directly in index compilation or to support these other stages of CPI production.

10.3 The use of alternative data sources is still evolving rapidly because both the data environment and the analytical methods required to process these sources are changing at pace. New data types— such as transaction data derived from scanner systems or other electronic records (including e-invoice data), as well as non-transaction data such as web-scraped prices—present challenges related to representativeness of the underlying data source (for example in terms of outlet or population coverage), data quality, classification, and integration with established CPI frameworks. At the same time, technological advances and international best practices continue to shape how these challenges are addressed. As a result, recommendations and implementation guidance require periodic review to remain aligned with emerging standards and practice.

10.4 Accordingly, this guidance note focuses on stable concepts, principles, and decision frameworks that are expected to remain relevant over time. This chapter does not replace existing guidance on sampling, quality adjustment, or index construction elsewhere in the Manual, but complements it where alternative data introduce new considerations. Detailed operational and methodological guidance is maintained in a dedicated [Handbook on utilising new data sources in the production of consumer price statistics¹](#) (referred to in this chapter

¹ <<handbook link>>

as “the Handbook”), which is updated periodically to reflect evolving data environments, methods and international practice.

Definitions and scope

10.5 For the purposes of this chapter, “alternative data sources” refers to data that differ from the traditional forms of data typically used by National Statistical Offices for CPI compilation, either by collection method (for example, web scraping or bulk data delivery) or by data type (for example, transaction records).

10.6 The classification of a data source as “alternative” is context-dependent: some NSOs may have long experience using certain sources (particularly administrative data) that may be considered novel by others.

10.7 Alternative data sources used in CPI compilation can be grouped into two broad categories according to whether they record actual transactions (transaction data) or posted prices without associated purchase information (non-transaction data). This distinction is important because it has implications for data preparation, weighting, index compilation and governance decisions discussed later in this chapter.

Transaction data

Transaction data record prices associated with actual purchases and typically include information on quantities or expenditures. Prices observed in transaction data reflect realised transactions, including promotions, discounts and other price adjustments at the point of sale, and may therefore differ from posted or advertised prices used in traditional price collection. When coverage and quality are sufficient, transaction data can support weighting and representativity assessments at detailed levels of aggregation. Transaction data may be accessed through a range of technical mechanisms, including bulk data delivery, secure file transfer or application programming interfaces (APIs).

Examples of transaction data include:

- **Scanner data:** Transaction-level data, or data aggregated over short time intervals (for example, daily or weekly) or to defined units such as retail outlets or product identifiers (for example, GTINs), obtained from retailers’ point-of-sale systems either directly or via third-party providers. These data typically include product identifiers, prices, quantities and turnover and can cover large shares of household retail expenditure for some product groups.
- **E-invoice data:** Electronic records generated for administrative or tax purposes that document transactions between sellers and buyers. Depending on national systems, e-invoice data may provide item-level price information with varying levels of product detail, quantities or values. Where sufficiently detailed and accessible for statistical use, these data may support CPI compilation and related weighting or validation activities.
- **Administrative transaction data:** Transaction-based data collected for regulatory or operational purposes, such as utility billing systems, regulated tariffs linked to consumption, or other administrative systems that record realised payments or charges experienced by households. While such data may follow administrative

concepts, they can provide consistent and timely transaction prices where alignment with CPI pricing concepts is satisfactory.

- **Other electronic transaction records:** Including, where appropriate, data derived from payment systems, loyalty-scheme databases, digital platforms or other electronic transaction environments (including emerging data sources such as those from IoT enabled device data such as smart appliances, connected home devices and wearables), provided the data record actual purchases and meet CPI conceptual and quality requirements. The structure, coverage and accessibility of these data vary widely across countries and markets.

Non-transaction data

Non-transaction data provide information on posted or offered prices and product attributes but typically do not include quantities, expenditures or direct confirmation that a transaction has occurred. These data can expand price coverage and observation frequency but require careful design to avoid implicit weighting or representativity issues. Non-transaction data may be obtained through automated tools, APIs or direct provision by data owners, depending on the legal and institutional context.

Examples of non-transaction data include:

- **Web-scraped data:** Prices and product attributes collected automatically from online outlets using web-scraping tools or related automated mechanisms, including APIs. These data can capture online markets and support high-frequency price observation but usually lack information on purchase volumes or economic significance.
- **Online listings and advertised price data:** Prices and product characteristics derived from listing platforms or classified advertisements, where goods or services are offered but transactions are completed outside the platform (for example, car sales or property rentals). These data may be collected automatically or provided directly by the platform. The listed price may differ from the final transaction price, and while indicators of market activity (such as listing duration or removal) may be available, they do not constitute direct measures of transactions.
- **Administrative non-transaction data:** Prices or tariffs set or published by administrative bodies, regulators or public authorities, such as regulated fees, charges or official price schedules. These data record prices as set rather than prices paid in individual transactions and typically lack information on quantities or expenditures. Their suitability for CPI compilation depends on the extent to which the administered price aligns with the price faced by households.

10.8 Alternative data sources, whether transaction or non-transaction in nature, may be used to replace, supplement or validate traditionally collected CPI information. Their integration should be guided by the conceptual foundations of the CPI and assessed in terms of how well they support the index's objectives, coverage and interpretability.

Incorporating these sources is not solely a technical choice: it involves strategic decisions about representativeness, continuity, quality, and the coherence of the resulting indices with

existing CPI components. NSOs should therefore consider the implications of alternative data not only for data availability and efficiency, but also for alignment with CPI pricing concepts, the stability of long-term series, and the resources required to sustain access, processing and governance over time.

The broader CPI framework provides the basis for these assessments, ensuring that the use of alternative data strengthens the conceptual integrity and reliability of the CPI.

Guiding principles for the use of alternative data in CPI compilation

10.9 The guiding principles set out in this section reflect and adapt the core concepts and quality considerations described elsewhere in this Manual to the specific context of alternative data sources. They are not intended to introduce new CPI principles, but to support consistent decision-making when incorporating large, externally sourced and rapidly evolving datasets into CPI compilation.

10.10 Alternative data sources can affect multiple stages of CPI production simultaneously, including data acquisition, processing, index construction and dissemination. Considering these principles helps ensure that decisions about data sources and methods remain aligned with the CPI's conceptual objectives, governance framework and user expectations, while recognising the particular risks and opportunities associated with these data.

10.11 The use of alternative data sources in CPI compilation should therefore be guided by the following principles:

Conceptual consistency: The CPI should continue to measure price change for household final consumption expenditure in accordance with the concepts and scope defined in this Manual. Alternative data sources should be assessed for their suitability to represent consumer transactions and prices within this framework, and any conceptual differences should be clearly understood and managed.

Transparency and reproducibility: The main characteristics of alternative data sources, and the methods used to process them, should be documented and communicated in a way that allows users to understand their influence on the CPI. Sufficient documentation should be maintained to support reproducibility, audit and independent assessment, taking into account confidentiality and legal constraints.

Confidentiality and ethics: The acquisition and use of alternative data must comply with relevant legal, contractual and ethical requirements, including the protection of personal data and commercially sensitive information. NSOs should ensure that data use is consistent with their legal mandate and public trust obligations.

Data quality and suitability: Alternative data sources should be evaluated against standard quality dimensions used in official statistics, including relevance, accuracy, timeliness, coherence and comparability (These quality dimensions follow the quality framework used throughout the CPI Manual and are aligned with the IMF Data Quality Assessment Framework (DQAF), as described in **Chapter 13**). Quality assessment should consider not only the data themselves, but also the stability of access arrangements and the risk of structural change over time.

Continuity and comparability over time: Changes in data sources, processing methods or index compilation approaches should be managed to minimise breaks in series and unintended effects on measured price change. Where changes are necessary, their impact should be assessed, documented and communicated to users.

Efficiency and proportionality: The use of alternative data sources should improve the efficiency, coverage or quality of CPI compilation where possible, while remaining proportionate to available resources and statistical needs. The adoption of new data sources or methods should be justified by clear benefits, relative to existing approaches.

10.12 These principles are consistent with the broader CPI framework set out in this Manual and with general quality considerations applied across official statistics. The **Handbook** provides practical guidance on how these principles can be implemented through operational workflows and methodological choices.

Opportunities and practical considerations

10.13 Alternative data sources present important opportunities but also specific limitations. These vary by national context, market structure, legal environment, access arrangements and technical capacity. This section summarises key opportunities and considerations, distinguishing between transaction data and non-transaction data while noting where issues are particularly relevant to specific data sources such as scanner, e-invoice, web-scraped or administrative data. These opportunities and considerations depend heavily on country context and will evolve rapidly; detailed guidance will be maintained in the **Handbook**.

Transaction data

- Opportunities

Transaction data, such as scanner data, e-invoice data and other electronic transaction records, can provide high levels of detail and coverage of household retail expenditure in many markets. Because these data record realised purchases, they typically include quantities or expenditures, enabling more representative weighting at the product level, and improved measurement of consumer purchasing behaviour. Prices observed in transaction data reflect actual transactions, including promotions and discounts, which may offer a closer representation of consumer experience than posted shelf prices used in traditional price collection.

Transaction data can support the monitoring of product entry and exit, enable the exclusion of products that are no longer sold, and facilitate more timely identification of new items and changes in consumer behaviour. Where the full universe of transactions is available, or where aggregation is performed at a fine level (for example by product identifier, outlet or short time interval), transaction data can support improved representativity and, in some cases, improve comparability across regions or countries.

The availability of high-frequency observations (for example daily or weekly data) may also allow the construction of new or supplementary indices or analytical outputs that are not feasible with traditional collection methods. In some product groups, transaction data can reduce—or in some cases replace—traditional field price collection, supporting efficiency gains and more consistent data capture over time.

- Considerations

The use of transaction data introduces specific challenges. Coverage may be uneven across sectors, outlet types or population groups, depending on market structure, retailer participation and data availability. Large-scale transaction datasets require substantial investment in data storage and processing, classification, quality assurance and system infrastructure, as well as robust procedures to manage product churn, relaunches and frequent changes in assortment.

Transaction data are often obtained from external providers, including private companies or administrative systems, introducing legal, confidentiality, cost and sustainability considerations. Changes to data definitions, access conditions or system designs at the provider can affect continuity and require careful governance and contingency planning. In some cases, fees are charged to access transaction data.

Where large volumes of transaction data are used dynamically, additional challenges may arise for index compilation and aggregation. These considerations may motivate the exploration of alternative index methods, including multilateral approaches, subject to governance decisions, operational feasibility and user expectations.

Non-transaction data

- Opportunities

Non-transaction data—such as web-scraped prices, online listings and advertised price data—can expand price coverage, particularly in online or rapidly evolving markets, and support high-frequency observation of price changes. These data may reduce the need for in-person price collection and can provide timely access to detailed product attributes across a wide range of goods and services. Automated data acquisition mechanisms, including web-scraping tools or APIs, may improve efficiency and consistency of collection.

In some contexts, non-transaction data provide a pragmatic entry point for introducing alternative data into CPI production, allowing NSOs to enhance coverage or automate aspects of price collection while maintaining established sampling and index-compilation frameworks.

- Considerations

Non-transaction data typically lack information on quantities, expenditures or confirmed purchases, limiting the ability to assess the economic significance of observed prices directly. As a result, there is an increased risk of implicit weighting, whereby products or outlets with a greater volume of observations exert disproportionate influence on measured price change if not carefully controlled.

Listed or advertised prices may differ from final transaction prices, particularly in markets characterised by negotiation, promotions or dynamic pricing strategies. Price information online may not reflect recent changes in prices because webpages may not be updated. Indicators of market activity (such as listing duration or removal) do not represent direct measures of consumption. Non-transaction data may also suffer from incomplete or inconsistent metadata, infrequent price updates, frequent product changes and coverage that does not reflect actual purchasing behaviour, especially where offline sales remain dominant.

Integration of non-transaction data therefore requires careful design, often through sampled approaches supported by external information or expert judgement. This helps to preserve representativity, interpretability and continuity with existing CPI components.

Legal and ethical considerations apply, particularly where data are obtained through web-scraping, including compliance with website terms of use and data protection legislation, as well as risks associated with dependence on platform policies or third-party intermediaries, which may change without notice. In some countries, laws prohibit NSOs from web scraping.

10.14 Many alternative data sources are substantially larger and more complex than traditional CPI inputs, often involving high-frequency observations, large numbers of products or transactions, and detailed metadata. While this scale underpins many of their potential benefits, it also introduces challenges related to data storage, processing, quality monitoring and system performance. These are discussed further in the High-level implementation framework section.

10.15 The inclusion of a data source in this chapter does not imply that it should be used in all countries or for all product groups. NSOs should undertake context-specific quality assessment before implementing any alternative data source in the CPI.

High-level implementation framework

10.16 The implementation of alternative data sources into CPI production is typically an incremental process. NSOs may choose to begin by applying new data sources in targeted areas of the basket to familiarise themselves with the data, processing requirements and governance implications, before extending their use to more complex markets or product groups with higher expenditure weights.

Introducing alternative data sources typically requires a period of parallel running, during which new processes operate alongside existing CPI production routines. While this ensures continuity, comparability and risk mitigation, it also necessitates additional resources during implementation, including staff time, analytical capacity and IT support.

10.17 The implementation of alternative data sources can be organised into several conceptual stages. This section outlines the main stages and considerations. This chapter contains the high-level implementation framework for introducing alternative data sources into CPIs. For more detailed technical procedures, options and examples, reference should be made to the [Handbook](#).

Data selection and acquisition

10.18 Selection of an alternative data source should begin with an evaluation of relevance to household consumption expenditure and coverage of the markets represented in the CPI. NSOs should assess whether the data capture a sufficiently large share of expenditure, whether the scope of products and outlets aligns with existing CPI coverage, and whether the information available supports reliable identification of products and consistent treatment of quality over time.

10.19 NSOs should also consider whether the data provider can deliver data regularly and sustainably, with sufficient stability in definitions, identifiers and coverage to support ongoing CPI production.

10.20 At the point of acquisition, NSOs are encouraged to consider potential uses of the data beyond the immediate calculation of consumer price indices, such as weighting, validation, or other statistical outputs (including Purchasing Power Parities, where relevant), as this may influence the variables required, the level of detail requested and the design of data access arrangements.

10.21 Acquisition must be supported by appropriate legal, contractual or collaborative arrangements, with attention to security, confidentiality, retention, and commercial sensitivities. Sustainability must be considered, as the data provision will need to be maintained over time with sufficient regularity and quality. Burden and incentives for the data providers should also be discussed, along with any measures that may be needed to support ongoing cooperation.

In some countries, legal or institutional frameworks increasingly support access to administrative and privately held data for official statistical purposes, alongside voluntary or cooperative arrangements with data providers. The scope and application of such frameworks vary across national contexts and data types and should be considered as part of sustainability and governance assessments.

10.22 Data selection should be informed by a structured assessment of these factors. In some cases, pilot work may be required to establish the feasibility and usefulness of the source before full implementation.

10.23 NSOs should consider resource implications, including the IT infrastructure required to process and securely store large and complex datasets, appropriate analytical tooling, and the availability of relevant staff skills. Contingency plans should also be established to manage potential interruptions to data supply. NSOs should also plan for additional temporary resource needs, as new alternative-data workflows must initially run in parallel with existing CPI production systems to ensure stable outputs and maintain continuity.

IT infrastructure and system design

10.24 The use of alternative data sources has significant implications for IT infrastructure and system design. Compared with traditional CPI inputs, alternative data are often larger in volume, higher in frequency and more complex in structure. These characteristics place greater demands on systems for secure data ingestion, storage, processing and analysis, and require approaches that can operate reliably within CPI production timeframes.

10.25 NSOs should ensure that the introduction of alternative data sources is supported by a coherent and forward-looking IT strategy. Experience from the use of scanner data has shown that system design choices made at an early stage can have long-lasting implications for methodological flexibility, production stability and resource requirements. IT systems should therefore be designed to be scalable, modular and adaptable, allowing new data sources, providers and methods to be incorporated over time without reliance on ad hoc or one-off solutions.

10.26 Given the scale of investment often required, NSOs are encouraged to test systems using pilot or test datasets and to draw on international experience where available. Systems and workflows should support automation, reproducibility and transparency, while remaining sufficiently flexible to accommodate evolving data characteristics and methodological

developments. The design of IT infrastructure should not constrain future method choices, particularly where alternative index approaches or enhanced quality assurance procedures may need to be introduced.

10.27 IT design should also support governance and risk-management requirements, including data security, access control, auditability, version control and documentation. Resource implications, including infrastructure capacity, analytical tooling and staff skills, should be considered as part of implementation planning. These considerations apply regardless of whether alternative data are obtained directly from data providers or via third-party intermediaries.

10.28 The successful use of alternative data sources may also require changes to organisational processes, skills and ways of working, particularly where automation and data-intensive methods are introduced.

10.29 Detailed technical architectures and tooling considerations are discussed in the **Handbook**.

Data preparation and classification

10.30 Before detailed processing begins, NSOs should determine how alternative data sources will be used within CPI compilation, including whether a static (sampled) or dynamic (full-universe) approach is intended. This choice is a design decision and is not necessarily determined by the type of data source itself.

Either a static or dynamic approach may be applied to transaction data (such as scanner or e-invoice data) or non-transaction data (such as web-scraped or company listings data), depending on how products are selected and used within the CPI framework. This decision has important implications for data preparation, product identification, and the treatment of product entry, exit and change over time. Some NSOs follow a phased approach, using a static approach initially and moving to a dynamic approach over time.

10.31 Under a static approach, the key design choice concerns how the subset of products or varieties is selected from the available data. For transaction data, such as scanner and e-invoice data, selection may draw on detailed expenditure or quantity information. For non-transaction data, such as web-scraped data, product selection often relies on expert judgement, price-collection rules and external information, in a manner broadly consistent with traditional CPI practice. Once selected, products are tracked using fixed item definitions for a defined period, mirroring traditional CPI compilation. Data preparation therefore focuses on preserving stable product specifications, consistent classification and continuity of price comparisons over time.

10.32 Under a dynamic approach, the full universe of products observed in the data source may be used, allowing products to enter and leave the index as market conditions change. This approach places greater emphasis on robust product identification, grouping and matching rules, and on procedures to manage high product turnover and changes in assortment.

When transaction data are used dynamically, products can be weighted to reflect their economic importance using quantity or turnover information available within the data. For

non-transaction data, where such information is typically unavailable, dynamic use is more challenging and requires particular care to avoid implicit weighting and resulting bias.

10.33 Alternative data must be processed and standardised to ensure consistency across time and coherence with other CPI data. Decisions will be required as to how to identify a product, and how to group observations, for example by time, outlet and product.

10.34 Products should be mapped to the CPI classification at an appropriate level of detail, and the information available should support reliable tracking of products or product groups over time. Where product identifiers are unstable or subject to relaunching (where the same product is re-released under a different unique identification number and needs to be retrospectively linked to avoid potential biases), additional linking or grouping strategies may be required to preserve continuity and avoid unintended bias.

Depending on the characteristics, quality and structure of the underlying dataset, different classification and relaunch linking techniques may be appropriate. These can range from simple rule-based approaches that use available product attributes, to more advanced methods such as supervised or unsupervised machine-learning algorithms for product grouping or matching. Detailed operational guidance and examples of these techniques are provided in [the Handbook](#).

10.35 NSOs should also establish clear rules for the treatment of observed prices and transactions. Alternative data sources often record transaction prices, including discounts, promotions and refunds, rather than list or shelf prices. Decisions are therefore required on how such features are reflected in CPI price measures, ensuring consistency with CPI pricing concepts. In some cases, this treatment may differ from that applied to traditionally collected prices, reflecting historical limitations in the information available to identify and account for promotional take-up.

10.36 Similarly, alternative data may reveal frequent changes in product characteristics, such as package size, formulation or bundled offers. NSOs should consider whether the available information is sufficient to identify and manage quality-related changes over time, and how such changes are handled within the chosen data preparation and index compilation approach.

10.37 Transactions or observations affected by known errors, returns or cancellations, or structural changes may be removed or addressed through quality controls, while ensuring that genuine price variation, such as temporary promotions, is retained.

10.38 For some administrative data, conceptual alignment with CPI pricing concepts (for example, the timing of price reference) should be considered.

Quality assurance and monitoring across the production process

10.39 Quality assurance and monitoring are integral to the use of alternative data sources in CPI compilation and should be applied throughout the production process. Given the scale, complexity and external origin of many alternative data sources, quality issues may arise at multiple stages and may not be detectable through a single set of checks. Effective quality assurance therefore requires a combination of input-level checks, process monitoring and assessment of outputs.

10.40 At the data ingestion and processing stages, NSOs should establish procedures to monitor the coverage, structure and stability of alternative data sources over time. This may include checks on provider participation, product assortment, classification consistency, volumes or transaction counts, and the prevalence and pattern of promotions or other pricing features. Regular review of outliers, missing data and unusual patterns can help identify data errors, delivery issues or structural changes that may affect index results.

10.41 Quality assurance should also consider risks associated with reliance on external data providers, including changes to data definitions, identifiers, delivery mechanisms or access arrangements. Clear documentation should be maintained on the characteristics of the data source and any changes that could influence measured price movements or comparability over time.

10.42 Given these risks, NSOs should establish contingency plans to manage potential interruptions to data supply or material deterioration in data quality. Such plans may include identifying alternative collection modes, temporary fallback methods, or transitional arrangements to support continuity of CPI production. The feasibility and implications of such measures should be considered in advance, rather than at the point of disruption.

10.43 In addition to monitoring input data and processing outcomes, NSOs should assess the behaviour of indices produced using alternative data sources as part of the quality assurance process. This may include monitoring index movements, volatility and contributions, examining coherence with related CPI components, and comparing results with indices derived from other data sources where appropriate. Such analysis can help identify issues that may not be apparent at the microdata level alone.

10.44 Quality assessment may lead to decisions to adapt, restrict or suspend the use of a particular data source, for example limiting its application to specific product groups or using it primarily for validation purposes. NSOs should establish internal criteria and escalation procedures for remedial action and ensure that these are reflected in standard operating procedures.

10.45 Effective quality assurance and contingency planning support both methodological soundness and user confidence. The outcomes of quality monitoring should inform decisions at subsequent stages of CPI production, including index compilation, integration with other data sources and user communication.

Index compilation and integration

10.46 When integrating alternative data sources into CPI compilation, NSOs should consider how these data will be incorporated into the existing index structure and how continuity with previously published series will be maintained. The introduction of new data sources or compilation approaches should be managed in line with established CPI practices, typically through testing, parallel runs and appropriate linking strategies (see the [Handbook](#)), to ensure that changes do not introduce unintended breaks or distortions in the CPI.

10.47 The role of alternative data within the overall CPI framework should be clearly defined. This includes decisions on whether the data are used to replace or supplement traditionally collected prices, how coverage overlaps with existing sources are managed, and how hybrid

approaches combining multiple data sources are implemented without duplication or double counting, using explicit market-share weights where possible.

When combining multiple data sources, NSOs should ensure that the relative importance of each source reflects its economic significance rather than its volume. Alternative data sources may contain substantially more observations than traditional collections, which can lead to implicit weighting if not carefully managed. NSOs should therefore consider approaches such as aligning weights with expenditure shares or outlet importance, maintaining consistency with existing sampling structures or applying explicit aggregation rules that prevent over-representation of high-volume data sources. The choice of approach should ensure the resulting index remains representative of household consumption and consistent with CPI concepts.

10.48 Index compilation choices are closely linked to earlier decisions on data preparation and use, including whether a static or dynamic approach has been adopted (see 10.XX). This choice reflects differences in how products are selected, tracked and weighted over time, rather than the type of data source itself. Under a static approach, alternative data may be used to support the selection and pricing of a fixed sample of products, maintaining continuity with traditional CPI compilation while benefiting from improved coverage or efficiency. Under a dynamic approach, index compilation must accommodate frequent product entry and exit and changing expenditure patterns, placing greater demands on data completeness, weighting information and methodological robustness. Both transaction and non-transaction data sources may be used under either approach, provided the associated design choices are consistent with CPI concepts and governance requirements.

10.49 The availability of expenditure or quantity information is a key consideration at this stage. Transaction data such as scanner and e-invoice data often provide detailed information that can be used to support explicit weighting below the elementary aggregate level.

In contrast, non-transaction data sources—such as most web-scraped data or advertised price listings—typically provide prices and product attributes without information on purchase volumes. When such data are used, compilers need to take particular care to avoid implicit weighting effects, whereby products with a larger number of observed prices exert disproportionate influence on measured price change, irrespective of their relative importance in household consumption.

In these cases, non-transaction data such as web-scraped data are commonly integrated through a static or sampled design, supported by external information or expert judgement, in line with traditional CPI compilation practice. Dynamic use may be appropriate where suitable weighting information can be obtained; however, in practice external sources or well-justified proxies are generally not available to develop weights and NSOs tend to follow the static or sampled approach.

10.50 Where alternative data sources are integrated with existing CPI components, NSOs should assess the implications for index behaviour, weighting structures and aggregation, and ensure that the resulting indices remain conceptually consistent and interpretable by users. Particular care is required when incorporating large-scale alternative data sources to ensure that the volume of observations does not translate into disproportionate influence on measured price change.

Documentation, change management and user communication

10.51 Clear documentation and transparent communication are essential to maintaining user confidence in CPI outputs, particularly when alternative data sources or new compilation approaches are introduced. Given the scale and complexity of many alternative data sources, NSOs should ensure that appropriate documentation, metadata and communication practices are embedded throughout implementation.

10.52 NSOs should document the origin and scope of alternative data sources, key processing steps, index compilation approaches and any assumptions or limitations affecting interpretation. Metadata should describe the role of alternative data within the CPI, including coverage, weighting information, and the relationship to traditionally collected data. Where changes occur over time—such as modifications to data sources, processing rules or compilation methods—the nature and timing of these changes should be clearly recorded.

10.53 The introduction of alternative data sources or new methods represents a form of change to CPI production and should be managed in line with established CPI change-management practices. This typically includes pilot studies, testing and, where appropriate, parallel runs to assess impacts before full implementation. Decisions on implementation timing, linking to existing series and continuity should be guided by the need to preserve interpretability and minimise disruption to users.

10.54 User communication should accompany material changes affecting CPI outputs. NSOs are encouraged to explain the purpose and rationale for changes, the product groups affected, and the expected implications for index behaviour, such as changes in level, volatility or revision policy. Where relevant, communication should clarify whether historical series are revised or whether continuity is maintained through linking or other techniques.

10.55 Effective documentation and communication support accountability, transparency and trust in CPI statistics. The level of detail published will depend on national circumstances, but users should be able to understand the main implications of the use of alternative data for CPI compilation and interpretation.

10.56 The first period for which the new data sources and methods are used to compile the CPI should be announced well in advance and should include detailed metadata for the media and other key data users. This will ensure that the methodological changes implemented to the CPI are well understood. Further information relating to the CPI dissemination can be found in Chapter 14 of this Manual.

Index number methods with alternative data sources

10.57 The use of alternative data sources introduces specific challenges for index number methods that are not fully addressed by traditional CPI compilation approaches. These challenges arise from features such as high product turnover, frequent price changes, and, in some cases, the availability of detailed expenditure or quantity information. As a result, method choice becomes closely intertwined with data characteristics and governance considerations.

10.58 This section provides conceptual guidance on index number methods in the context of alternative data sources. It does not replace the general treatment of index numbers provided in Chapter 8 of this Manual, nor does it prescribe specific formulas or parameter

choices. Detailed methodological guidance and implementation options are documented in the **Handbook**.

Bilateral methods and the static (sampled) approach

10.59 Traditionally, the CPI is produced using bilateral index methods, where prices from two periods (for example, the current and reference month) are compared using a fixed set of representative goods and services. Even when alternative data sources are available, NSOs may choose to adopt a static or sampled approach by selecting a subset of products from the full dataset, defining item specifications that remain fixed for some period, and calculating price indexes using standard bilateral formulas (for example, Jevons price index see Chapter 8).

10.60 This approach can support continuity with existing CPI methods while benefiting from improved efficiencies in data collection, sample selection, product coverage and, where available, the derivation of expenditure weights. It maintains close conceptual alignment with established CPI design and avoids the need for revisions to previously published series. NSOs with limited prior experience in using alternative data sources may apply the static approach as a transitional strategy to build institutional capability and methodological confidence in working with larger and more complex datasets.

10.61 However, a static approach does not fully exploit the richness of alternative data sources and may be less responsive to rapid changes in product assortments and consumption patterns. Where product turnover is high or where new products play an important role in consumer expenditure, static sampling may introduce bias if the selected product set becomes unrepresentative over time.

Bilateral methods and the dynamic (full dataset) approach

10.62 When alternative data sources are used to represent a dynamic universe of products traditional bilateral index methods face several challenges. High rates of product entry and exit, frequent promotions, and rapidly changing market shares can undermine assumptions of stable item samples and fixed expenditure patterns that underpin conventional bilateral comparisons.

10.63 One way of managing these conditions while retaining a bilateral framework is to apply methods over shorter intervals and link the resulting indices through chaining. Compared with long-span bilateral comparisons to a fixed base period, chaining allows products to enter and exit the comparison more flexibly and can improve continuity in the presence of changing assortments. However, the performance of chained bilateral approaches depends critically on how chaining is implemented, including the choice of link periods, the treatment of expenditure or quantity weights, and the handling of product turnover.

10.64 Early empirical studies using scanner and transaction data have shown that frequent chaining of weighted bilateral indices can lead to systematic chain drift, particularly in environments characterised by intensive promotional pricing and asymmetric quantity responses. Ivancic (2007), using market research scanner data on goods sold in supermarkets, found a downward drift in chained Fisher price indices (see also Ivancic and others 2009, 2011). Drift in chained matched-model superlative price indices has been documented for durable goods as well. Here, the drift is likely due to seasonal fluctuations in prices and quantities. De Haan and Krsinich (2014), using scanner data, found a downward

drift in chained Törnqvist price indices for consumer electronics goods sold. Silver and Heravi (2005) presented evidence of downward bias in chained Fisher indices using scanner data on televisions. These findings highlighted that, even when chaining helps manage product turnover, the accumulation of small linking effects can distort measured price change over time.

10.65 In practice, different configurations of bilateral methods—under static or dynamic use of alternative data—entail trade-offs. Shorter chaining intervals may reduce sensitivity to product churn but can increase the risk of drift through repeated linking. Approaches that rely on fixed or infrequently updated weights may lag behind rapidly changing consumption expenditure patterns, while more frequent weight updating, particularly when combined with chaining, can amplify drift effects. The interaction between data characteristics, processing choices and index design therefore plays a critical role in determining index behaviour.

10.66 As a result of these findings, NSOs who may have access to expenditure or quantity information may nevertheless choose to compile indices using an unweighted chained bilateral approach. This can arise where the primary concern is the potential for chain drift when frequently chaining weighted bilateral indices in highly dynamic markets with intensive promotional activity. In such cases, expenditure information can be used to support representativeness through the design of the product universe used for index compilation, rather than through explicit weighting within the index formula. For example, expenditure data may be used to apply dump filters to exclude rarely purchased or transient products and expenditure or turnover filters to restrict the product set to items accounting for a meaningful share of consumption, before compiling an unweighted chained bilateral index (Eurostat, 2017).

10.67 This strategy reflects a trade-off. Using expenditure information for product selection can improve representativeness and stability, while the unweighted chained bilateral compilation may reduce some risks associated with chaining weighted bilateral indices. However, the resulting index remains sensitive to the choice of filters, update frequency and product inclusion rules, and these design choices should be governed, tested and monitored over time.

Multilateral index methods: concepts and motivation

10.68 Multilateral index methods compare prices across multiple periods simultaneously rather than one pair of periods at a time. They have traditionally been used in spatial price comparisons and have received increased attention in the context of alternative data sources, particularly in response to empirical evidence that fixed bilateral and chained bilateral approaches can perform poorly when applied to highly dynamic product universes. Experiences with transaction data, such as scanner data, —especially in markets characterised by frequent product turnover and promotional pricing—provided a key motivation for the development and adaptation of multilateral methods for CPI purposes.

10.69 Multilateral methods can offer improved handling of product entry and exit, reduced drift when assortments or expenditure patterns change over time, and closer alignment with cost-of-living-based index concepts. These properties make them a promising option for alternative data sources, particularly when the full dynamic universe of products is used.

10.70 However, multilateral methods raise specific challenges for CPI production that are less prominent in traditional CPI settings. A key issue is the revision problem: because each new period is evaluated relative to all periods within the comparison window, historical index values may change as new data become available. Since the CPI prioritises stability and guards against revisions, this issue must be carefully managed.

10.71 Extension methods, such as splicing, address this concern by using a rolling window of recent periods and extending the index to ensure that previously published values are not revised. While extension methods limit revisions, they can reintroduce some undesirable index number properties, including the potential for drift, and therefore involve trade-offs.

10.72 From a CPI perspective, the implementation of multilateral methods involves some strategic design decisions. These include the initial choice of multilateral formula and its underlying assumptions, the length of the comparison window, the approach used to splice or extend the index, and the treatment of weights and product turnover within the window. Each of these choices involves trade-offs between conceptual accuracy, responsiveness to market change, stability of published series and operational complexity.

10.73 NSOs may wish to consider multilateral methods particularly in situations where the full dynamic universe of products is used, product turnover is high, market structures are changing rapidly, or traditional bilateral approaches produce instability or drift. The availability of quantity or expenditure information can support weighting within multilateral methods but is not necessarily a prerequisite, particularly where such information can be obtained from external sources or where well-justified approximations can be derived.

10.74 The choice of index method should reflect trade-offs between conceptual accuracy, stability of published series and user expectations. Multilateral methods and their implementation options continue to evolve, and NSOs should adopt them cautiously, with appropriate testing, governance and communication with users. The conceptual basis of multilateral methods is discussed more fully in the Consumer Price Index Theory publication. Detailed methodological guidance and implementation examples are provided in the **Handbook**.

Relationship to the Handbook and future development

10.75 This chapter sets out the conceptual and governance framework for the use of alternative data sources in CPI compilation. It is complemented by the **Handbook** that provides detailed operational guidance, including:

- data preparation workflows for scanner, web-scraped and administrative data
- alternative index calculation methods and their properties
- practical examples, including country experiences and case studies
- tools and templates for documentation and quality assessment

10.76 The **Handbook** is expected to be updated more frequently than the Manual, reflecting evolving data practices and research. Future editions of the Manual may expand coverage as experience with alternative data sources accumulates.

References

[other references in this redraft have been retained from the old chapter and shouldn't need to be updated].

Handbook on utilising new data sources in the production of consumer price statistics.
Maintained by the UN Task Team on Scanner Data.