Unit values and aggregation in scanner data – towards a best practice[1]

Jörgen Dalén, Consultant[2]

Abstract: The paper addresses some aspects of the treatment of scanner data by statistical agencies with special but not unique reference to the European HICP context. An operational definition of homogeneity as well as for the concept of consumption segments in aggregation are discussed and tentative recommendations are put forward. A distinction between dynamic and static approaches to scanner data is made and arguments in favour of a dynamic approach are given. Criteria for a successful scanner data methodology are proposed.

Introduction

The research literature on scanner data is increasing for each year and is now quite abundant. A recent paper by de Haan et al (2016) gives a detailed account of the current research frontier with an extensive reference list.

More and more countries are using or planning to soon use transactions (scanner) data for important segments of their CPIs and HICPs. But the practices adopted are quite divergent and often do not follow the leads in the research literature. In fact, the number of approaches used almost equals the number of countries. Perhaps it is even larger since some countries use different approaches for different product groups. Instead of convergence towards a best methodology that could serve as a benchmark, we are seeing divergence in applications.

In Dalén (2014) an account was given of the methods that were used at that time in Europe, mainly in the Netherlands, Sweden, Norway and Switzerland. Since then the new QU-GK method was presented in the Netherlands, for the first time in Chessa (2015). At the current meeting, several countries will also present their methods, already applied but mostly under development.

It is striking how different the various methodologies are. Why is it so? We will reflect on some of the more important choices that NSIs are faced with concerning the use of scanner data.

The most general aggregation setup of a CPI/HICP with scanner data can be expressed by the following steps:

1. The formation of homogeneous products within which prices are defined as unit values
2. The aggregation of homogeneous products to the lowest level in a scheme with fixed weights. This lowest level will here be referred to as consumption segments.

[1] The paper builds on reports on scanner data in 2014 to 2017 written within a Eurostat project organised by ICON Institute in Germany. The opinions expressed in the paper are the author's own and cannot be attributed to either Eurostat or ICON Institute.

[2] Email: jorgendalen@hotmail.com. Address: Solhagavägen 36, SE – 16347 Spånga.

3. The third step is the aggregation of consumption segments to higher level product groups and the All Products CPI.

In separate sections we will discuss steps 1 and 2 of this setup. (Step 3 is not affected by the use of scanner data so will not be discussed further.) Certain criteria regarding the best use of scanner data are also formulated.

On the basis of this discussion we put forward some recommendations.

Different forms of scanner data

All scanner data come in the form of sales values and quantities for a coded product, here called GTIN. But this is as much as can be said in general. In the following respects there are important differences:

- Values and quantities can be for a chain of outlets or for single physical outlets.
- Values and quantities can be for a week or a day. In principle, we could imagine also other demarcations in the time dimensions such a whole month or, in the other extreme, an hour or less.
- GTINs are simple identifiers in a number format. The extent and form in which additional information on attributes of the products is available varies from country to country, product to product and from one data provider to another.

The full picture of the kind of data countries are offered to work with is not known. In some countries data are for weeks and whole retail chains, whereas in other countries data are by single outlets and days and delivered by market research companies.

The content and structure of scanner data is one key factor that will influence the approaches that are possible and actually chosen by countries. However, it is still useful to imagine how a scanner data set should ideally be constructed.

Homogeneity

The unit value index should be used for a single homogeneous product. This is a basic axiom underlying the use of scanner data. Its motivation is that there needs to be a lowest level, where a single price is defined on the basis of actual transactions and the simplest and most natural definition is the quantity weighted average price. (At the very lowest level, the single transaction, it is not possible to establish a match between a reference and a current price.)

It is then imperative to develop a workable definition of homogeneity that takes the theoretical ideal as well as practical circumstances into account.

From the point of view of economic theory, a product is homogeneous if all product offers/ transactions within its specification are equivalent to the consumer. Otherwise expressed, the consumer is indifferent to which of the product offers she purchases. An implication of this is that the quantities are additive in an economic sense which in turn validates the use of the unit value index as a measure of price change. In scanner data we simply add up the quantities purchased per product thereby forcing the additivity property to the data. The interpretation is

then that each quantity unit is of the same utility to the consumer. If this is not the case in practice a unit value bias will occur.

This conceptual background to homogeneity can serve as a theoretical reference point. But it cannot be directly applied to arrive at an operational definition. First of all, not all consumer have the same preferences. One consumer may prefer a certain colour, one prefers a particular outlet before another, one likes to shop on Sundays etc. So all aspects of a product offer (combination of product, variety, outlet and time) may potentially disqualify an assertion of homogeneity.

The approach to homogeneity in statistical practice therefore needs to apply some kind of convention with the theoretical ideal in mind. SNA (2008, 14.139) speaks of "sufficiently homogeneous" to justify the use of a price index and then (14.144) takes as an example different types of screws which are certainly not equivalent for a user and argues:

"It is obviously impracticable to introduce a degree of disaggregation that would identify each of these types of screw separately and the thought of identifying screws separately from nails and other metal construction materials is already implausible. The problem of non-homogeneity is thus inevitable but may be reduced by considering the level of detail available"

With this thinking in mind the challenge in price indexes is to establish what should count as *sufficiently homogeneous* in order to justify the use of unit values. This needs to be a pragmatic approach to determine what is on balance the best solution. But it will always be less than perfect in an ideal sense.

Product homogeneity and the relaunch problem

A product defined by a GTINs can in practice be considered as homogeneous, since GTINs are assigned to only one article. But different GTINs may sometimes be identical, almost identical or perfect or near perfect substitutes to each other. A GTIN may disappear but another GTIN appear in the next period that is virtually identical to the previous one and equivalent to all or most consumers. The GTINs may differ in very minor respects that do not matter to most consumers such as package or a small change in volume. In order to maintain comparability over time it is in such cases necessary to use a product characterisation that is above the GTIN code and allows for a price comparison.

This problem is strongly accentuated by so-called **relaunches**. The relaunching phenomenon occurs when a GTIN code is discontinued but more or less the same product enters again with a new code.

Chessa (2013) describes this phenomenon as follows:
"This term is used in situations where articles are replaced by items that mainly differ from their predecessors by external appearance of their packaging rather than content or ingredients. The follow-up items are assigned a new EAN and are usually introduced at a higher price. Price increases with respect to the preceding items are therefore missed when calculating price changes per EAN."

In this paper Chessa dealt with drugstore data, especially for the product groups toiletries (toothpaste, shampoo etc.) and beauty articles (lipstick, deodorants, eau de toilette etc.), both of which belong to COICOP 12.1.3. Such articles are also sold in supermarkets in most European countries. He finds that for these types of products (and to a lesser extent other

products as well) it is quite common that an outgoing GTIN/EAN code is relaunched in a new type of package with a new code but exactly the same characteristics. Typically, the new code comes with a price increase despite the fact that all the visible characteristics are the same.

In later papers, e.g. Chessa (2016) it is made clear that the relaunch phenomenon in a wider sense, is even more pronounced for clothing articles. Here "relaunches" are often not exactly identical to their predecessors but rather close substitutes. The pattern of a reduced price in the last month(s) of the outgoing product and a relatively high price in the first month of the "relaunched" product creates the same kind of problem as with "identical" relaunches.

Relaunches mean that the simplest way to use scanner data – using GTIN as identifier –does not always work. For certain products, GTINs need to be grouped into higher level entities so that a continuous series of sufficiently homogeneous products is obtained. For this purpose, additional characteristics of the products are needed such as, in the case of clothing, brand, material, etc.

Best methods are clearly data dependent. For example, it will be necessary to look at verbal article descriptions which are not standardized. If the number of products is large then, to be able to do the data editing work, a sampling approach may be necessary. If so a rigorous probability sampling approach should be sought that make unbiased inferences to the ideal universal parameter possible and at the same time reduces the resources needed. Sho

The relaunch problem is not limited to scanner data. It occurs in general where a product life tends to end with one price level (typically lower) and the new substituting product enters at another price level (typically higher). Sometimes the two products are not of the exact same quality but still the price difference is much larger than the quality difference. One could here speak of "near relaunches". In such situations, it is imperative to enable direct price comparisons in order to avoid large biases. With scanner data, the best approach is to define homogeneity so that relaunched products are "internalized" in a sufficiently homogeneous product.

Outlet homogeneity - service level

Not only the homogeneity of the physical product but also the service level, which is determined by the outlet selling the product, matter when determining homogeneity.

The consumer values not only the physical product but also the shopping experience. This includes for example long opening hours including weekends, nearness to shop from workplace or home, a large assortment (but also ease of finding what you are looking for), access to service staff in the shop etc. This makes the application of unit values across shops with different service levels problematic and will create some kind of unit value bias, large or small.

An important paper by Ivancic and Fox (2013) throws light on the topic of unit value aggregation across stores**.** The paper uses a so-called Bertrand framework to define homogeneity in the seller dimension as follows:

"*The same item sold by different sellers is viewed as homogenous if the price of the item is found to be consistently the same across sellers in the long term.*"

It is important to stress here *in the long term* since this is what is correlated to the service level. So the point is not that the price is the same on each occasion – it is that overall and over the long term the sellers have the same price level. This may be taken as a criterion for when unit values across stores will not give rise to a unit value bias.

Ivancic and Fox find after an empirical analysis that sellers within several Australian chains of stores are homogeneous by this criterion but that there is not always homogeneity across different chains.

The analysis of homogeneity of European retail chains remains to be done. But a look (in 2014) at the websites of some major chains provides some first ideas. *Carrefour* (a French chain) operates stores in many European countries and divides their stores into four types with clearly different service levels – hypers, supers, convenience and cash & carry. *Tesco* presents 14 types of stores in its UK operations. *Aldi* and *Lidl* in Germany and some other countries do not present any classification of their outlets. In Sweden *Coop* and *ICA* present 4-7 outlet categories that are related to certain levels of service as well as size of the outlet. A tentative idea that should be confirmed by further explorations is that each chain's own outlet types can normally be taken as homogeneous with respect to service levels. In large countries, a geographic subdivision may also be warranted. If this idea is accepted it would facilitate index computations from scanner data of chains which are not separable by outlet, which seems to be currently the case for Dutch data.

Special care needs to be taken to separate out internet sales in a chain. Several chains operate an online business in parallel to their physical outlets. Online sales clearly represent a different service level and to the extent that price levels are different they should be kept separate from physical outlets in unit value aggregation.

If scanner data are supplied by the retail chains it may not distinguish between different outlets or outlet types. It may include outlets in different areas of the country, which could be more problematic in a large than a small country. If the data are supplied by a market research company or the like (Nielsen, GfK etc.) it would more likely have single outlets with their addresses included. In the latter case, it opens up for defining homogeneous outlets across different chains but in a limited geographical area. With respect to consumers' opportunities to substitute, this may be considered as a close to ideal unit of aggregation. For example, neighbourhood stores (say, where cars are not needed for shopping) within medium-sized cities in a not too large country would perhaps be a close to ideal unit for UVI aggregations for an also homogeneous product.

Sampling procedures may also be used in the outlet dimension. For the foreseeable future, we have to accept that the whole outlet universe could not be covered by scanner data. Some kind of sampling procedure is then needed to ensure outlet representativity also in a scenario which includes scanner data for a large part of the outlet universe. Depending on the general calculation scheme in a country this may call for various solutions including sampling among scanner data outlets.

4.4 The time dimension

Time is an issue for homogeneity only to the extent that it is related to different service levels. If all points in time during a certain time period are equivalent to the consumer and there are no price level differences between weekdays or hours of the day, then the whole time period (month or week) can be considered as homogeneous for the purpose of price aggregation.

Today it is quite rare for goods sold in physical outlets to have different prices for different days or hours.[3] (For services this happens more often but they are not so far covered by scanner data.) If such practices become more common decomposition of time periods into homogeneous subperiods would have to be considered. Since the cost for late opening hours may be higher it is not far-fetched to imagine that such market practices could evolve.

In the time dimension the length of the period for which a unit value should be applied has also been discussed. Taking a unit value over a whole week or even a month would only be a problem if service levels differ within, which as said is rarely the case. Also, if unit values would only be used per day it would mean that expenditure weights per day would be needed, or that days would have equal weights regardless of purchase intensity which is not ideal. So it appears that a unit value over a period of up to a month is a reasonable approach.

4.5 Attempts at recommendations

An operational definition of homogeneity has to address the product, the seller/geography as well as the time dimensions.

In the product dimension, we need to use a product definition that sometimes combines several GTIN codes into one product.

In the seller dimension, we have a choice between unit values in a single outlet or in a homogeneous chain of outlets. In addition geographical demarcations may be needed, at least in large countries with different price levels in different regions.

It should be noted that there is something of a trade-off situation in the choice between unit value in a single vs. many outlets. Looking at homogeneity alone it would seem that we should always choose the smallest unit possible. But there are at least two important advantages of moving to a larger aggregation unit. One is the automatic weighting of the outlets according to their actual sales instead of imperfect outlet weights (or none). Another is the smaller amount of missing data caused by zero sales in either time period for new or disappearing products or outlets. Missing prices result in the use of replacements and imputations, which are very difficult to include in a coherent methodology and therefore lead to non-transparent and hard-to-assess procedures. So if heterogeneity is modest between outlets it could well be better to combine them into one aggregation unit.

The time dimension is at least for supermarket products not so critical in today's markets. Points in time can in practice be taken as comparable with each other. Unit values over weeks or even a whole month are thus not problematic. Scanner data are often supplied on a weekly basis and if so some compromise is needed for making them represent a month.

The above discussion boils down to a couple of recommendations. First, the most obvious one:

Unit values are to be applied for homogeneous groups of product offers/GTINs

[3] But offers of 5% discount for everything for pensioners on Wednesdays has been seen in the author's home area.

Regarding a recommendation on homogeneity, consumer valuation is key. If consumers value two product-offers with the same use equally and are prepared to pay the same price for each of them, then the two product-offers belong to the same homogeneous group and the simple unit value should be applied across them. However, the re-launch issue, and the zero sales of a very narrowly defined product-offers necessitate a somewhat wider definition of homogeneity.

Striking the best balance between Scylla and Charybdis will in practice be a question of judgement. When exercising this judgement, it has to be recognised that failure to deal with re-launches will lead to an underestimating bias that can be very severe, since the price increases at re-launches are often big, often in the order of 20-50% and sometimes even more. Unit value (mix) biases do not have a known sign (could be over- as well as underestimating) and when the underlying quality mix is moderate the aggregate error does not need to be large. Therefore, the best recommendation would be to first ascertain that all re-launches are handled and thereafter put the threshold so that mix biases can be considered of minor influence. The details of these judgements will eventually have to be subject to empirical research.

Recommendations for homogeneity that take these factors into account are:

Product-offers (GTINs) shall be considered homogeneous if they have the same use and most consumers are judged to consider them of equal value (choose between them on the basis of price only). Furthermore, GTINs with life-cycles that frequently end with price reductions must be combined into larger homogeneous groups with a long duration also where small differences in quality exist. (This recommendation gives a lower limit for homogeneity.)

Product-offers (GTINs), which have different price levels over a longer time period, shall not be considered homogeneous. (This recommendation gives an upper limit for homogeneity.)

The proposal apply to a GTIN in an outlet in a certain time period. They give ample room for judgement according to local and special circumstances.

Consumption segments

A consumption segment consists of a normally large number of homogeneous products which are all intended for broadly the same use. Between such products, calculation procedures have to be used that take account of quality differences but within each homogeneous product simple unit values are to be used for aggregating all transactions.

The concept of consumption segment has played an important role in HICP methodology. In the traditional setting its role is the following, according to the new version of the HICP Manual:

*A **consumption segment (CS)** is a partition of the product universe below the harmonised ECOICOP 5-digit sub-class level. … The following points provide a general guidance as to what is intended:*

- *CS are fixed at least throughout an index link (normally a year). This distinguishes them from product-offers which change frequently, often several times a year.*
- *Product-offers within a CS may be considered by consumers to be either equivalent or equivalent by applying appropriate quality adjustments in the sense of being substitutable subject to a monetary compensation to reflect differences in quality.*
- *Replacements should be made within a CS and not between CS*"

When traditional sampling practices are replaced by scanner data, the CS concept could play another important role: to determine **the level of fixity in the aggregation system**. Since CS are relatively stable entities, they will by design be fixed within an index link and can only be changed with the new link in December. Each included CS will have an expenditure weight according to the Laspeyres-type (Lowe) definition.

But within CS, the dynamics of product offers and outlets need to be accounted for. With scanner data, new methods to do this become available.

What would be the most appropriate level in the aggregation tree to place the CS? The main criterion should be that products within a CS have the same use and could therefore be substituted for each other, although not necessarily having the same quality/utility to the consumer.

Certain products, although maybe belonging to the same low level COICOP group cannot be said to belong to the same CS:

- Within Consumer Electronics a TV and a stereo system have different uses and few specifications in common
- Within Men's clothing shirts and trousers have different uses and few specifications in common
- Within Non-motorised small tools hammers and screwdrivers have different uses

At the other end of the spectrum it should be clear that certain product offers do belong to the same CS

- Two brands and models of 40" TVs can be used in similar situations and normally have many other specifications in common
- Two pairs of men's jeans can be used in similar situations and as well have many specifications in common
- Two hammers made by different producers can be used in the same situation

'Quality adjustment' means the procedure of making an allowance for an observed quality change by increasing or decreasing the observed current or reference price by a factor or an amount equivalent to the value of that quality change

So a CS can, in general, be seen as a set of heterogeneous products.

With this in mind the following proposals for recommendations are put forward:

1. **Products which clearly have different uses and cannot be substitutes to each other belong to different CS and should be given fixed expenditure weights.**

2. **Product offers, which most consumers would consider as alternatives/substitutes on the basis of price and quality (characteristics) do belong to the same CS.**

These two principles provide an upper and a lower limit to the setting of thresholds for any product group. These limits give room for judgement depending on local circumstances. We give two examples here of how the proposal could be understood:

- Within Consumer Electronics it would be possible to define all TVs as a CS. It would likewise be possible to divide up TVs into several subgroups on the basis of size, since size determines the place where it would be suitable to have the TV. But too many size classes would seem questionable. Decomposition into technologies (LCD/LED etc.) seems questionable since the use is the same. Brands are generally alternatives for most consumers and should not distinguish two CS but certain low or high level brands may possibly be put into different CS.
- Within clothing, function determines a CS. Shirts, trousers, jackets, socks, underpants etc. for men and women separately are obviously different CS. Within each of these categories several CS may or may not be defined. Decomposition, if any, should be made on the basis of more or less timeless functional differences rather than short-lived fashions. But it would likewise be ok to make each of these categories as a whole a CS.

Aggregation from homogeneous product to consumption segment

This step is the one where the greatest divergence is visible today. Some countries do not at all use the CS concept in their approach.

The Dutch approaches, both the previous one based on monthly chaining (also used in Norway) and the newly proposed QU-GK method (a special case of a quality adjusted unit value index), are designed to cover the full universe of products and GTINs within consumption segments. No one-to-one replacements or imputations are needed in these approaches. We call this the *dynamic approach* to scanner data. Dynamic approaches use methods that include new GTINS immediately and aggregate heterogeneous products taking quality differences into account. Other examples of such approaches are other forms of quality adjusted unit value indexes and the Time Product Dummy method, which are well described in many papers, for example in de Haan et al (2016).

In several other countries, for example Sweden and Switzerland, the approach is to use reference period GTINs and outlets as a basis for defining low level entities for price comparisons. We call this the *static approach* to scanner data.

With the dynamic approach it is possible to assign algorithms that are well-defined functions of, in principle, all transactions. On the contrary, the static approach necessarily involves replacements and imputations that are more or less impossible to define rigorously and therefore give room for subjectivity and arbitrariness. The static approach does not move us away from the non-transparent and non-comparable procedures that plague today's practices.

All replacements and imputations involve assumptions and rough judgements that are difficult to validate, this is in our opinion a great disadvantage of the static approach. Furthermore, coverage of the current transactions universe is radically better with a dynamic approach. These are fairly obvious and important advantages of the dynamic approach. Why is this approach then not accepted by all other countries, which instead opt for a static approach? Are there balancing arguments in favour of a static approach?

The following reasons can be imagined in favour of a static approach:

- The static approaches are similar to traditional methods and as such easier understood
- Early versions of the dynamic methods were susceptible to chain drift (more below). (The new Dutch method are however demonstrated to be free of chain drift.)
- The static approach could be seen as applying Laspeyres-type indexes at as low aggregation level as possible.
- The dynamic approaches could be seen as being more complicated in a mathematical sense.

We do not find any of these reasons convincing.

Traditional approaches can in no way be seen as ideals or benchmarks. Out of practical necessity they are based on purposive and not random selection in all or most dimensions; variety, outlet and product-offers. Replacements and imputations with their inherent lack of rigour are necessary. It is not possible to give any assurances about coverage of the universe.

Since the rate of attrition at the product-offer level is very high, Laspeyeres applied at this level means a very substantial loss of information relating to new varieties of products. Indeed, for many products outside the food area, product cycles have been shown to be 4-20 months (Dalén and Tarassiouk, 2013). This length of a product cycle is much too short to support a methodology that uses one such detailed product as the reference product. Instead many replacements will be needed.

It may be true that the mathematics involved in some dynamic approaches appear mystical at first sight. But it is a one-time effort to grasp and understand it. Moreover, they carry a great potential for harmonisation across Europe or wider still. Harmonisation leads to comparable indexes which lend themselves to advanced analytical work. The subjectivity involved in today's traditional methodologies is the real black box.

For all these reasons, it is judged that presented dynamic approaches are clearly superior to traditional approaches that try to retain elements of the earlier purposive sampling approaches.

As for choosing the best dynamic approach we leave this issue to the very competent researchers in the Netherlands, New Zealand and other countries and to their practical experiences.

Criteria for the choice of method

We finally put forward a number of criteria that should be relevant for the choice of method to use for transactions data.

Free of chain drift

This is the most important criterion, since chain drift can give rise to huge errors.

Chain drift can be defined as the opposite of transitivity, which is defined as price comparisons between any time periods being independent of the choice of base period.

Chain drift results from a failure of an index formula to meet the so-called circularity test, which in simplified terms says that a specific index formula I_{0t} applied to three periods 0, 1 and 2 should meet the criterion $I_{01} \times I_{12} = I_{02}$.

The left-hand side, which can be extended to any number of links, corresponds to a *chain index* whereas the right-hand side is the *direct index* from the first to the last period. Index formulas that meet this criterion can be called *transitive* formulas.

Fixed base index formulas like Laspeyres and Paasche do not satisfy this test but neither do superlative formulas like Fisher and Törnqvist. Among elementary aggregate formulas, Jevons and Dutot satisfy circularity but Carli does not.

Non-transitive formulas, when chained, usually tend to trend away from the direct index, either upwards (Laspeyres, Carli) or downwards (Paasche). In statistical practice this could lead to catastrophic biases. CBS-NL, in their first experiments with scanner data, found that also superlative formulas showed unacceptable chain drift when applied for monthly links. This was due to the extreme price bouncing behaviour in the supermarket data they analysed.

For chaining annual links, the HICP uses Laspeyres-type indexes that in principle suffer from some chain drift. This is because of practical necessity and also because the chain drift effects can be expected to be small, since it applies only to multi-year periods and not a single year.

The issue with transactions data is index calculation within a CS and, in the HICP setting, from December to a month next year, *the intra-year index*. If some form of monthly chaining is used, chain drift can have devastating effects on the index, since a year consists of twelve monthly links. A transitive index formula s therefore needed or else a formula directly comparing December to the current month without chaining.

Therefore**: Index formulas for consumption segments have to be free from chain drift. This can be achieved either by using fully transitive index formulas (if monthly chaining is used) or by using a direct index from December to the current month without chaining.**

Methods that have so far been proposed for transactions data are generally designed so that they are free of intra-year chain drift. The introduction of multilateral methods, previously used for spatial comparisons, is done for this purpose. Such methods are for example the QU-GK method as well as time dummy methods and the GEKS/RYGEKS indexes.

Other considerations

There are also other considerations when using transactions data. While being of lesser importance than the ones discussed above, they are still worth mentioning.

Minimisation of adhockery (imputations etc.). Methods should be generic in the sense that they are applicable and applied to as large segments of the index as possible without extra manual interventions.

Good interpretation and justification. Methods should be possible to give a good interpretation, with reference to generally accepted economics. For example, the Chessa's new QU-GK method uses multiple overlaps over a long period as quality adjustment factors. The justification for this method falls back on the traditional justification for the overlap method, for example as in the CPI Manual 1.236:

"If consumers are well informed, have a free choice and are collectively willing to buy some of both at the same time, economic theory suggests that the ratio of the prices of the new to the old quality should reflect their relative utilities to consumers. This implies that the difference in price between the old and the new qualities does not indicate any change in price."

The snag that often occurs with the overlap method is that the timing of a replacement matters a lot. It is often applied when one of the products is outgoing with a temporarily low price and the other one is new with a higher introductory price. The ability of the QU-GK method to consider a longer period for the overlap is a great advantage and it would therefore be possible to use the traditional overlap justification with much greater confidence.

Statistical efficiency. Maximum coverage/representation of the full transactions universe. This follows from the general statistical error minimisation criterion, in this case by minimising undercoverage. Again, the dynamic approaches hold great promise in this regard.

Balanced inclusion of time periods. In the HICP, as well as in many national CPIs, December is taken as a price reference period. To a smaller or greater extent this means that December products and prices are given an unproportional importance in the index. As long as the HICP is defined as now with annual Dec-Dec links, European countries will have to adapt their index designs to this also with regard to scanner data methods. Chessa's variant of QU-GK method does this albeit with some loss of "time period balance". For the foreseeable future, it is probably necessary to accept this imbalance.

Considerations on production efficiency and cost. The acceptance of a method in practice heavily depends on how convenient and inexpensive it is to apply. Detailed manual, monthly intervention in the data at the micro-level are obstacles for a successful method for both cost and timeliness reasons.

Potential for harmonisation. A generic method that can be expressed by programmable algorithms is also great for harmonisation, which is one of the basic objectives of Europe's HICP.

Issues for an empirical research program

We cannot expect that ideal solutions will always be possible to implement, mostly because of limitations in the available data. So, if NSIs will have to accept less than ideal scanner data sets, it is of importance to establish how large differences would result from various "second best" approaches. For example:

1. Grouping GTINs into not quite homogeneous products for applying unit values
2. Not grouping GTINS into larger groups when relaunches occur

3. Using unit values across outlets with different service levels or in different regions
4. Using unit values across various subperiods in a month with possibly different price levels (before/after a holiday season, weekends, late hours)
5. Applying equal weights for products/outlets with different expenditures

Chessa (2016) has demonstrated the large biases that can be the result of the second item in this list. But the other aspects have not yet been much studied and more needs to be done, since we will likely have to live with many different methods across countries for a long time that are less than ideal in many respects.

Summary

As more and more countries adopt transaction-based index methods, the need to agree on what are the best methods become more and more pressing. But the present tendency is towards highly divergent approaches.

In this paper we have provided some criteria for the choice of best approaches when dealing with scanner data. They have dealt with

- The delineation of homogeneous products, where the unit value is to be applied
- Setting the threshold for consumption segments, within which dynamic approaches are desirable.
- Arguments for choosing a dynamic approach rather than static
- The need to ensure absence of chain drift

The most important choice that countries are faced with, is whether to use the opportunity that transactions data offers to devise methods that attempts to capture the dynamics of the market in the most profound and complete way possible.

Three main recommendations:

- Definitions of homogeneity should be such that relaunches are neutralised, both identical and similar relaunches. Avoiding unit value bias is a second priority by comparison but should nevertheless be addressed to the greatest extent possible.
- Dynamic methods, which cover the evolving universe of transactions as close as possible are superior to static methods, which try to assign low level reference units from a historic period. (When choosing the best dynamic method, listen to research, which is still evolving!)
- Ensure that methods are guaranteed free of intra-year chain drift.

References:

Chessa, A.G. (2013), Comparing scanner data and survey data for measuring price change of drugstore articles. Paper presented in a conference on scanner data, Portugal 26-27 September 2013.

Chessa, A.G. (2015), Towards a Generic Price Index Method for Scanner Data in the Dutch CPI, Paper for the fourteenth Ottawa Group Meeting, 20-22 May 2015, Tokyo, Japan.

Chessa, A.G. (2016): A new methodology for processing scanner data in the Dutch CPI. EURONA, 1/2016, Eurostat, p 49-70.

Dalén, J. and Tarassiouk, O (2013): Replacements, quality adjustments and sales prices. Paper for the thirteenth Ottawa Group Meeting, May 1-3, 2013, Copenhagen, Denmark

Dalén, J. (2014): The Use of Unit Values in Scanner Data, Paper presented at the Scanner Data Workshop, 2-3 October 2014, Vienna, Austria.

De Haan, J, Willenborg L. and Chessa, A.G (2016): An Overview of Price Index Methods for Scanner Data. UN-ECE, Conference of European Statisticians, Meeting of the Group of Experts on Consumer Price Indices, 2 – 4 May 2016, Geneva.

Ivancic and Fox (2013): Understanding Price Variation Across Stores and Supermarket Chains: Some Implications for CPI Aggregation Methods. Review of Income and Wealth. Volume 59, Issue 4, Pages 629–647.

SNA (2008): System of National Accounts. New York, 2009.