ADVISORY PANEL ON CONSUMER PRICES – STAKEHOLDER

Proposed framework for shortlisting appropriate index number methods

Status: FINAL
Expected publication: Alongside minutes

Purpose

1. A framework that assesses index number methods against predetermined quality criteria has been developed to ensure that the methods chosen for the production of UK consumer price statistics best meet user needs.

Actions

2. Members of the Panel are invited to:
   a) advise on the number of methods that they would deem suitable for use in production of consumer price indices
   b) advise on the relative importance of each criterion used to assess the appropriateness of index number methods
   c) note their invitation to an ESCoE workshop being hosted (with a provisional date of 25th October 2019)

Introduction

3. We are investigating the use of new data sources to ensure it provides improved statistics and more efficient ways of working. Globally, National Statistic Institutes (NSIs) are beginning to recognise the impact of big data in the production of statistics. For consumer price statistics, web-scraped and scanner data are the data sources that are being most widely investigated and used.

4. We plan to introduce these sources into our headline measures of consumer price statistics by 2023. But with the introduction of new data sources comes the opportunity to use new index number methods. For example, scanner data contains quantity information meaning that previously unusable bilateral methods, such as Fisher and Törnqvist, can now be investigated and used.

5. Bilateral index number methods have traditionally been used in consumer price statistics at the lowest level of aggregation (elementary aggregate). These methods, such as Jevons and Dutot, compare prices for goods and services in a current period (t) with prices for comparable goods and services in a base period (0) to construct an index.

6. In contrast, multilateral methods compare interactions between prices in three or more time periods simultaneously to construct an index. The international price statistics community has reached a consensus (UNECE 2016) that these are generally the most appropriate methods to produce elementary aggregate price indices using web-scraped and scanner data. There are different ways of extending these methods to better suit temporal price indices, meaning that there are many variations of each unique method.

7. The introduction of scanner data and multilateral methods, along with different extension methods (used to ensure that the multilateral indices are sensible and not revised over a given time period), has increased the choice of index number methods for use at the elementary aggregate level to over 50. (Annex B, an introduction to index number methods, provides the list of elementary aggregate methods currently being investigated.) However, there is currently no consensus among NSIs as to the best choice of index number method to construct these data into price indices.

8. In the absence of consensus, we are developing a framework to assess the index number methods against predetermined criteria to control the number of possible methods for use in production and to ensure that the methods shortlisted for use are those that best meet our users' needs. This framework was first introduced in a paper presented to the Technical Advisory Panel on Consumer Prices in May 2019.

9. The chosen criteria have now been assigned weights to reflect their relative importance in the framework. The intention is to use these weighted criteria to produce a shortlist of approximately 10 index number methods for use in consumer price statistics[1]. Once a shortlist has been selected, methods will be further tested against a range of datasets with different price behaviours (for example, datasets with high product churn, high volatility and seasonal pricing). This should give an indication as to what method may be most appropriate for each unique item. For example, an item with highly volatile price movements may be more suited to method A in the shortlist, while an item with high product churn and seasonal pricing behaviour may be more suited to method B. A visualisation of the process for choosing an appropriate index number method for each unique item is provided in Annex A.

10. In the CPIH and CPI, currently only a few methods are used at the elementary aggregate level. Most items use Jevons or Dutot, while some items also make some use of expenditure information to produce a "Laspeyres-type" index. The new approach would see the potential introduction of substantially more methods being used at this level of aggregation. Users may prefer that we choose the most appropriate methods based on the relative statistical and behavioural properties of the method applied to each unique item, conversely, users may prefer only a small number of methods are used to enable a better ease of understanding.

11. **We are therefore seeking guidance on a) what characteristics users' value in determining the quality of an index number method, and b) how many methods are appropriate for use at the elementary aggregate level.**

12. As well as requesting feedback from APCP members, we are also hosting an ESCoE workshop to discuss these issues in greater detail with a wide range of stakeholders and academics. **The workshop is provisionally planned for Friday 25 October, and panel members are encouraged to attend if possible.**

13. The remaining sections of this paper discuss the framework for assessing the different index number methods, as it is currently proposed.

Framework for assessing index number methods

[1] Note that both bilateral and multilateral methods will be assessed against this framework.

14. There are five key criteria that we are currently considering in our assessment of index number methods (with reference to the ESS guidelines for measuring statistical quality):
    1. resource (timeliness and punctuality),
    2. theoretical properties (accuracy and reliability),
    3. flexibility (relevance),
    4. ease of understanding (accessibility and clarity),
    5. cohesion (coherence and comparability)

Criterion 1: Resource, 10%

15. The resource criterion questions whether: the method can be maintained without regular human input; the information processing requirements are reasonable and manageable; the method makes efficient use of all the information available; and the amount of information used can be scaled up within the infrastructure.

16. Within resource, a primary filter is used to see if the information processing requirements are reasonable and manageable. If the processing requirements are not currently reasonable or manageable, this would put the timeliness of publication at risk and therefore the method cannot be shortlisted.

17. The other resource requirements currently have a relatively small weight within the framework (10%), as resource should not come at the cost of statistical accuracy provided that the resource requirements do not hinder the timeliness and punctuality of consumer price statistics production.

Criterion 2: Theoretical properties, 70%

18. The theoretical properties criterion considers the accuracy of each method based on its inherent properties. The properties of each method are being tested against a number of different approaches including the axiomatic approach (for both bilateral and multilateral methods), the economic approach and the stochastic approach. The characteristicity and transitivity of each method are also considered in their own right. For details of the different approaches and the transitivity and characteristicity concepts, please refer to the May APCP-T paper.

19. Currently, we are in the process of completing a literature review to assess each method against these approaches. However, it may be considered appropriate to also get an independent review of these methods against the approaches to ensure that we introduce no bias in our assessment. This is currently being arranged.

20. Given that we are primarily concerned with producing accurate and reliable statistics, the theoretical properties criterion currently has the largest weight within the framework, at 70%. Within this however, the economic approach is given a negligible weight owing to the fact that the we do not currently consider a Cost of Living Index (COLI) our primary target[2].

Criterion 3: Flexibility, 10%

[2] The economic approach considers how well an index number method approximates the consumers' cost of living. The CPIH and CPI are not intended to approximate cost of living indices.

21. The flexibility criterion concerns how well the measure can adapt to new information. The key questions are how well the method adapts to the addition or removal of products, and whether the method is suitable for data both with and without quantity information.

22. While flexibility is inarguably an important criterion, it has a relatively small weight within the current framework (10%). This is because if the method itself cannot adapt to the addition or removal of products, there are additional options to ensure that these behaviours can be accounted for. For example, we can look at following groups of products over time rather than following individual products over time. This would allow product churn to be controlled for without necessarily needing a flexible index number method.

23. It also is not a primary concern if the method is not flexible enough to be used for both scanner data (including quantity information) and web-scraped data (excluding quantity information) as different methods can be used for different data sources and weighted together.

Criterion 4: Ease of understanding, 10%

24. The ease of understanding, or interpretability, criterion concerns whether: the method is recognised and well documented in international literature; the price movements are understood by those who are involved in the implementation and/or those who use the output data; product level contributions can be calculated, allowing index movements to be easily interpreted; and the method can be explained to a non-technical audience.

25. The price movements being understood by those who are involved in implementation and/or those who use the output data acts as a primary filter for the methods. If the index number method does not produce sensible results that cannot be understood by the compilers or users, it should not be assessed within the framework.

26. Once this primary filter has been applied, the ease of understanding criterion currently has a relatively low weight within the framework (10%). This is because users understanding of a method should not come at the cost of statistical accuracy. We will, however, endeavour to make sure any methods used in the statistical production of consumer price statistics are explained in a transparent way, to allow users the best chance of understanding them. As an example, we have provided an introduction to index methods, particularly multilateral methods, in Annex B.

Criterion 5: Cohesion, 0% (used as secondary filter)

27. The cohesion criterion covers two elements, *internal* and *external* cohesion. Internal cohesion concerns whether the method is used elsewhere within the construction of consumer prices and external cohesion concerns whether the method is used in consumer price indices internationally.

28. Within the current framework, cohesion is applied as a secondary filter. If internal cohesion were to be included in the framework it would reduce the allowance for any new methods to be included. The international community have already agreed that multilateral methods may be more appropriate for new data sources than current methods, so this criterion is not considered appropriate currently. Similarly, given other NSIs do not appear to have carried

out a comprehensive review of the index number methods before introducing them into their price statistics, external consistency is also not a primary consideration at this stage.

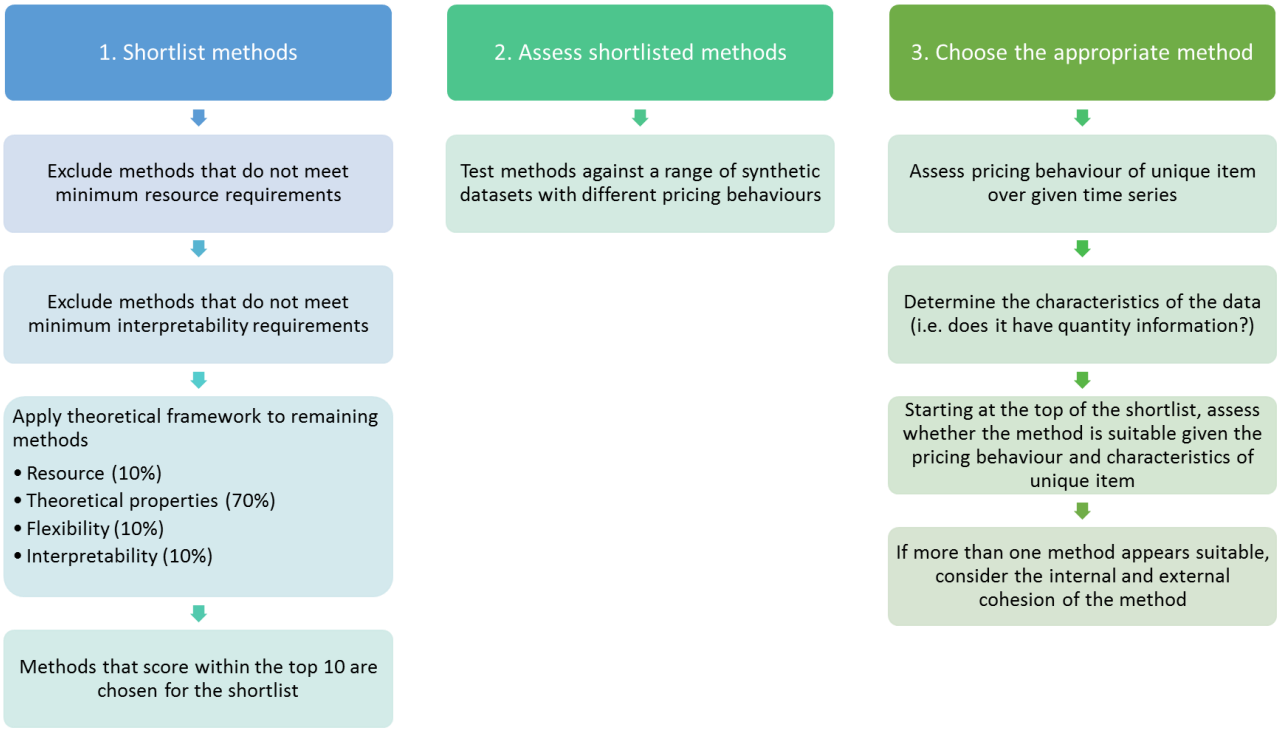
29. Coherence may be considered when deciding between two index number methods for a unique item. For example, if both index number method A and C are appropriate for unique item *X*, then the method that is most cohesive internally and externally can be chosen.

Helen Sands
Prices, Office of National Statistics
September 2019

List of Annexes:

Annex A	**Stages for choosing an index number method for each unique item**
Annex B	**An introduction to index number methods**

Annex A: Stages for choosing an index number method for each unique item



Annex B: An introduction to index number methods

Background

Bilateral methods have traditionally been used in consumer price statistics and compare prices for goods and services in a current period (t) with prices for comparable goods and services in a base period (0) to produce an index.

Multilateral methods compare three or more time periods simultaneously to produce an index, and have advantages over bilateral methods, particularly in that they maintain transitivity when reweighting and chaining frequently. This means that if a relationship holds between a first and a second time period and for a second and a third, it must also hold between the first and the third period, no matter how these periods are linked. These methods also have a practical advantage in that automated processes allow a greater sample of products to be used to produce indices.

There are a number of methods that sit under this "multilateral" umbrella term. Here, we provide an introduction to three of these methods (GEKS, Geary-Khamis and hedonics indices), and how they can be adapted to suit CPI construction purposes.

Geometric Mean

Before explaining multilateral methods, it is useful to understand the difference between the geometric and arithmetic mean.

A geometric mean is a different type of average. It is defined by multiplying together the numbers you want to average and taking the nth root (n is amount of numbers you want average).

For example, the geometric mean of 1,2,3,4,5,6 is

$$\text{Geometric mean} = (1 \times 2 \times 3 \times 4 \times 5 \times 6)^{\frac{1}{6}} = 720^{\frac{1}{6}} = 2.99 \text{ rounded to 2 decimal places}$$

This is different to the arithmetic mean (normally called the mean or the average) which is the sum of the numbers divided by the amount of numbers you have (n).

$$\text{Arithmetic mean} = \frac{(1+2+3+4+5+6)}{6} = \frac{21}{6} = 3.5$$

The geometric mean is often preferred, as it is less influenced by very large values in a skewed distribution.

List of bilateral index methods being investigated

The table below shows the eight main bilateral index methods we are investigating. Most of the indices have two variants. The fixed-base variant occurs where product prices for each month are compared to a consistent base month (January). The chained method occurs where product prices are compared to the previous month and the month-on-month comparisons are linked together to form an index. These chained methods have the potential to solve instances of high product churn (where products leave and enter the market regularly).

For the unit value index, the fixed-base and chained methods are equivalent since no product matching occurs. Similarly, the fixed-base and chained forms are equivalent for the Rothwell index.

Base method	Variants			Uses quantity (or expenditure) data
	Fixed-base	Chained	Other	
Paasche	✓	✓	×	✓
Laspeyres	✓	✓	×	✓
Fisher	✓	✓	×	✓
Törnqvist	✓	✓	×	✓
Jevons	✓	✓	×	×
Dutot	✓	✓	×	×
Unit Value	×	×	✓	✓[3]
Rothwell	×	×	✓	✓

Multilateral Methods: GEKS; Geary-Khamis; Hedonics

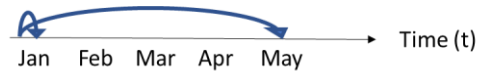
GEKS

To calculate the GEKS index between a base period and a current period, we take a geometric mean of comparisons made between the base and an intermediate period, and an intermediated period and the current period, for all the possible choices of an intermediate period, linking them together.

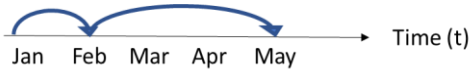
The comparisons made between the base and the intermediate period, then the intermediate period and the current period use any bilateral index method.

For example, if we want to calculate the price change between January and May using the GEKS index we do the following comparisons:

- January-January and January-May



- January-February and February-May

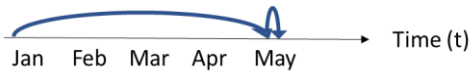


- January-March and March-May

- January-April and April-May

[3] Note that a form of the Unit Value index can be used without quantity/expenditure data

- January-May and May-May



The GEKS index is then calculated by taking a geometric mean of the price changes calculated using all five routes between January and May.

- January to May

Geary-Khamis

The Geary-Khamis Index is based on the concept of a "Unit Value". A Unit Value is a way of saying how much does someone pay for a group of products, on average. For a given period, it is calculated by taking the total amount spent over all products within the group and dividing it by the total quantity of the products sold.

A Unit Value index is then defined as:

$$\text{Unit Value Index} = \frac{\text{Unit Value in the current period (t)}}{\text{Unit Value in the base period (0)}}$$

As the unit value is calculated from all products within the group (for example, all brands or varieties within "chocolate bar"), it may be affected by the different mix in the quality of the products over time. The Geary-Khamis adjusts the unit value to account for this quality mix to ensure the index is calculated with the products at constant quality. The calculation for this is as follows:

$$\text{Quality Adjusted Unit Value} = \frac{\text{Total amount spent on all product in the group}}{\text{Sum of the quantity} \times \text{quality adjustment factor for all product in the group}}$$

The Geary-Khamis index is then defined using the following equation:

$$\text{Geary-Khamis Index} = \frac{\text{Quality Adjusted Unit Value in the current period}}{\text{Quality Adjusted Unit Value in the base period}}$$

The Geary-Khamis index is sometimes referred to as the Quality-Adjusted Unit Value.

Hedonic indices

Hedonic indices are a different way to quality adjust the prices before calculating an index. We are introducing two different hedonic methods.

1. Time-product dummy hedonics

Time-product dummy hedonics decomposes the price collected into two parts. The first part is the amount of the price that is because we collected it in a certain period – a time specific effect. The second part is the amount of the price that is unique to the product collected – a product specific effect.

The time specific effect is estimated over all products collected in the same period, while the product specific effect is estimated within a product over all periods in the dataset and is fixed. Because the product specific estimate is "fixed" in each period, this method is also called fixed-effects hedonics. The index is calculated by taking the time specific effect from the current period.

2. Time dummy hedonics

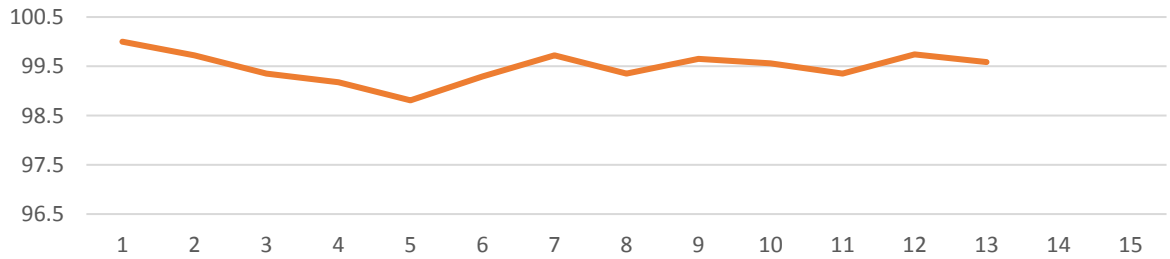
Time dummy hedonics decomposes the price collected into more than two parts. This assumes that the characteristics which make up a product affect the product individually, for example the price of a laptop is made up of the price of the Hard Drive, the Memory etc. The characteristics have the same effect regardless which product they appear in. The index is calculated by taking the time specific effects from the current period.

Multilateral Extension Methods

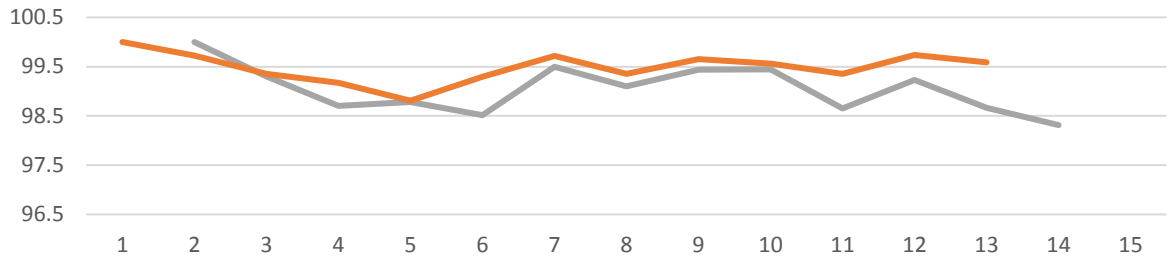
When a multilateral method is used to produce an index, each index depends on prices observed in other periods of the multilateral comparison window. As a result, joining a new period into the multilateral comparison window could alter the index values of previous periods causing the need to revise previously published data. This is both undesirable and impractical.

To avoid this, multilateral methods are paired with extension methods which "links" on the index. One example is the window splice. The window splice links on the movement at the start of the series. An example of this is provided in the following.

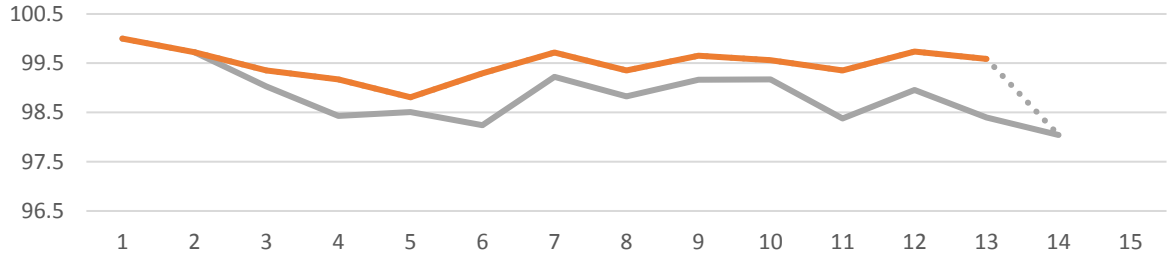
Let's start with the following series between periods 1 and 13:



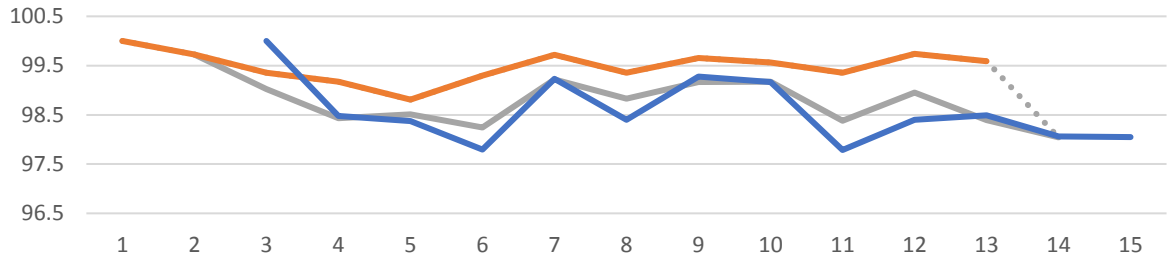
We now have period 14's data available, and we re-calculate the multilateral method using data from period 2 to 14. This revises the index between periods 2 and 13:



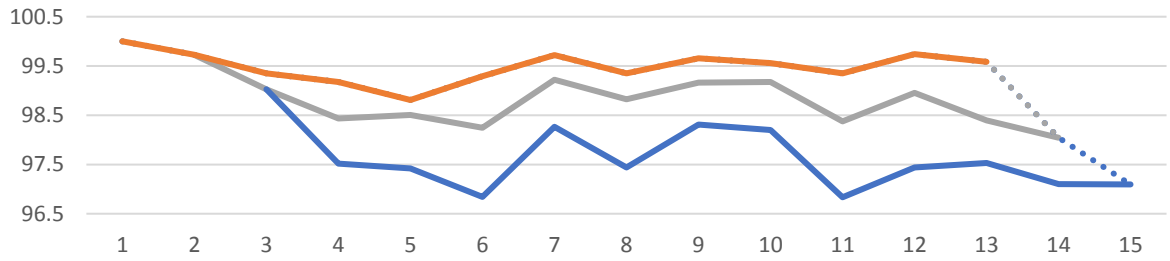
We can now rescale the index at period 14 using the ratio of the two values at period 2: the originally published value and the revised value. This keeps the already published data the same, so it does not need to be revised:



When period 15 data are collected, we re-calculate the multilateral method using data from period 3 to 15:



Again, we can rescale the index at period 15 using the ratio of the two values at period 3: the originally published and latest revised value. This gives us our final period 15 index:



The published series would be the original estimates from periods 1 to 13 then the rescaled estimates in periods 14 and 15 (the orange and then dashed lines).

Different window lengths can be chosen, although it is typical for the window length to cover a multiple of a year (one year, two year, three years etc.) to account for seasonality.

Note also that this window splice was created by taking the ratio at the start of the window. If the ratio was taken in the middle of the window, this would be a half-window splice, and if taken at the end of the window, a movement splice.

Aside from the varying splices, there are other available methods too, such as the Fixed Base Monthly Expanding Window.

List of multilateral index methods being investigated

Any bilateral index method can be paired with the GEKS method, but international research suggests that it is best paired with a superlative index (Fisher, Tornqvist). We are also investigating the GEKS-Jevons for use when we lack quantity data. As mentioned, we are also investigating the Geary-Khamis method, and two regression-based methods: Time Product Dummy Hedonics and Time Dummy Hedonics.

Between six base methods and six extension methods we are considering, this leads to 36 shortlisted multilateral methods.

Base method	Variants (extension methods)						Uses quantity data
	Direct Extension	Movement Splice	Window Splice	Half-window Splice	Fixed Base Monthly Expanding Window	Geometric Mean Splice	
GEKS-Törnqvist	✓	✓	✓	✓	✓	✓	✓
GEKS-Fisher	✓	✓	✓	✓	✓	✓	✓
GEKS-Jevons	✓	✓	✓	✓	✓	✓	×
Geary-Khamis	✓	✓	✓	✓	✓	✓	✓
Time Product Dummy	✓	✓	✓	✓	✓	✓	✓[4]
Time Dummy	✓	✓	✓	✓	✓	✓	✓[5]

[4] Note that an unweighted regression model would allow the Time Product Dummy and Time Dummy models to be used without expenditure data

[5] See 4.