Bernhard Goldhammer[1], European Central Bank

# A Measure for the Similarity of Time Series

## 1. Problem

With the advent of multilateral methods being used in computing price indices from scanner data, the question of choosing the “correct” index formula, window length and splicing method has gained unprecedent importance. This means that different time series are generated and compared, be it against each other (De Haan, Hendricks, & Scholz, 2021; Chessa, Verburg, & Willenborg, 2017), a set of axioms (Eurostat, 2022) or a benchmark index (Chessa, 2019; Bialek & Bobel, 2019). However, the comparison of different timeseries is normally done using eyeballing; a measure of describing the degree of index similarity is missing.

Unfortunately, standard textbooks seem not to deal with the problem of measuring the similarity of time series in in a way that it matches the demands of a similarity measure for economic indices. However, there is literature about identifying “similar price and quantity structures” for interregional comparisons (Diewert, 2006). This is not the same, but a related problem, and the axioms used by Diewert (2006) can be translated into some that can help in testing a similarity measure for time series. They should be complemented by axioms that take into account the special nature of economic indices: especially in communicating them, the absolute values of the indices are only playing a minor role, while the focus is on change rates. This needs to be considered by the design of the similarity measure as well.

The paper is organised as follows: in chapter 2, traditional approaches that may point to similarity of time series are shown and discussed. Then, in chapter 3, the similarity measure is developed by setting up axioms, proposing some measures and testing them against the axioms. In chapter 4, I will apply the preferred similarity measure to some comparison problems of the literature. Contributions and limitations are discussed in chapter 5. Chapter 6 concludes.

## 2. Traditional approaches and their drawbacks

Before developing a new measure, it makes sense to evaluate existing measures that may point to time series similarity. As there are plenty of them,[2] I will only discuss selected measures.

### 2.1. Descriptive statistics: Mean and Variance

In descriptive statistics, mean and variance are the most common metrics being used. One could argue that two time series are similar when their mean and variance are similar. For example, sine and cosine have the same mean and variance. However, being put into the context of economic indices, one would not describe them as being similar as their gradients – or, in the world of discrete measures like price indices – their period-on-period percentage changes are different. In addition, given the economic growth over the last centuries and the long-term trend of inflation, the assumption of a stable mean and also a stable variance fails. Hence, we would get different results of

[1] Bernhard Goldhammer, Principal Economist-Statistician, European Central Bank, Directorate General Statistics, Division MOE, e-mail bernhard.goldhammer@ecb.europa.eu. Disclaimer: The views expressed in this paper are those of the author and do not necessarily reflect the views and policies of the European Central Bank or the Eurosystem.

[2] For a brief overview, see Ropella, Nag, & Hunt (2003).

a similarity analysis, if different time ranges are being investigated. In addition, mean and variance have been developed for probability statistics. This means that the order of periods does not play a role. However, for economic indices, it plays a decisive role, as it determines period-on-period percentage changes. In conclusion, mean and variance are not suited as similarity measures for economic indices. The demand of having the same dynamics at the same point in time for a measure of index similarity also rules out other possible similarity measures that allow one series lagging behind the other, like ARIMA models or cross-correlations, or tests on time series stationarity.

## 2.2. Bravais-Pearson correlation coefficient

The Bravais-Pearson correlation coefficient, for two time series **X** and **Y**, is defined as follows:

$$r = \frac{cov(\boldsymbol{X},\boldsymbol{Y})}{sd(\boldsymbol{X})sd(\boldsymbol{Y})}$$

It measures the strength and direction of a linear relationship between **X** and **Y.**[3] So, the correlation coefficient is invariant of linear transformations.[4] However, this is again not in line with the idea of a similarity measure for time series in an economic sense, e.g. with the intention that similarity is present if period-on-period change rates are identical, as linear transformations will change the change rates. In addition, as it is independent of the order of periods (applying the same permutation to the members of **X** and **Y** will not change the correlation coefficient), also does not account for the specific property of time series that a natural order of the observation values exists. Therefore, the correlation coefficient is also not a suitable measure for time series similarity.

[3] Vogel (1999), p. 75: „Der Korrelationskoeffizient… mißt die Stärke und die Richtung linearer Beziehungen zwischen zwei metrischen Merkmalen X und Y.“
[4] Vogel (1999), p. 76.

# 3. Development of a similarity measure, especially for price indices

## 3.1. Useful properties and axioms

As we have seen, classical measures show weaknesses if being used as a similarity measure for time series. Therefore, I will follow the axiomatic approach for the derivation of a similarity measure: beneficial properties will be defined, against which some possible measures of time series similarity will be tested. This is done against the special background of economic indices that relate to a certain base period (whose index value is 100).

There is not much literature on similarity of time series. However, Diewert (2006) provides some similarity axioms, which have been developed for several cases like one or many variables and relative and absolute dissimilarity. While his work was on the similarity of price and quantity structures for interregional comparisons, it is possible to transfer his axioms to time series. For time series like price indices, where relative differences are more important than absolute differences, his axioms for "relative dissimilarity indexes in the N variable case" (Diewert, 2006, pp. 11) can be applied. Let $S(\boldsymbol{X}, \boldsymbol{Y})$ denote the similarity measure for the time series $\boldsymbol{X}$ and $\boldsymbol{Y}$ with elements $x_t$ and $y_t$. Both time series are indices that have a base period whose index value is 100. Then, the following axioms can be deducted that represent beneficial properties of measures of time series similarity:[5]

A. *Continuity*: $S(\boldsymbol{X}, \boldsymbol{Y})$ is a continuous function defined for all time series $\boldsymbol{X}$ and $\boldsymbol{Y}$ (with $x_t, y_t > 0$).
B. *Identity*: $S(\boldsymbol{X}, \lambda\boldsymbol{X}) = 0$ for all time series $\boldsymbol{X}$ with $x_t > 0$ and scalars $\lambda > 0$.
C. *Positivity:* $S(\boldsymbol{X}, \boldsymbol{Y}) > 0$ if $\boldsymbol{Y} \neq \lambda\boldsymbol{X}$ for any $\lambda > 0$.
D. *Symmetry:* $S(\boldsymbol{X}, \boldsymbol{Y}) = S(\boldsymbol{Y}, \boldsymbol{X})$ for all time series $\boldsymbol{X}$ and $\boldsymbol{Y}$ (with $x_t, y_t > 0$).
E. *Proportionality*: $S(\boldsymbol{X}, \lambda\boldsymbol{Y}) = S(\boldsymbol{X}, \boldsymbol{Y})$ for all time series $\boldsymbol{X}$ and $\boldsymbol{Y}$ with $x_t, y_t > 0$ and scalars $\lambda > 0$.

In addition, Diewert (2006) also proposes axioms for "invariance to changes in the unit of measurement" and "invariance to the ordering of commodities". They have not been transferred to our case, as they are either directly fulfilled or violated by the nature of economic indices and time series:

- "Invariance to changes in the unit of measurement" is already fulfilled, as we are looking at indices. Indices in economic statistics are already transformations of the original measurement unit in such a way that they enable index continuity despite changes of the underlying measurement units. Good examples for that property are price indices, where the index continues even when the underlying currency of prices is changed.[6]
- "Invariance to the ordering of commodities" is *per se* violated, as one fundamental characteristic of a time series is that the order of its values is important, as it represents the development over time.

When looking into the properties of time series in general and price indices in particular, we can derive additional important axioms:

F. *Correction for the number of time periods.* The similarity measure should not depend on the number of time periods included in the time series. Let $\boldsymbol{X}, \boldsymbol{Y}, \widetilde{\boldsymbol{X}}, \widetilde{\boldsymbol{Y}}$ be time series, $\boldsymbol{X}$ and $\widetilde{\boldsymbol{X}}$ with the same value $x_t = x$ for all periods, $\boldsymbol{Y}$ and $\widetilde{\boldsymbol{Y}}$ with the same value $y_t = y$ for all

[5] Cf. Diewert (2006), p. 12.

[6] A recent example is Croatia accessing the euro area: while the Croatian currency changed from Kuna to Euroa, the Croatian HICP has continued and enables price level comparisons without a break.

periods. $\boldsymbol{X}$ and $\widetilde{\boldsymbol{X}}$ are defined for periods 1,…,T, as well as $\boldsymbol{Y}$ and $\widetilde{\boldsymbol{Y}}$ for periods 1,…,T+1. Then, if this axiom is kept, $S(\boldsymbol{X},\boldsymbol{Y}) = S(\widetilde{\boldsymbol{X}},\widetilde{\boldsymbol{Y}})$ .

G. *Focus on change rates.* For economic indices, the index level itself is rather an arbitrary decision, depending e.g. on the choice of the base year. Therefore, $S(\boldsymbol{X},\boldsymbol{Y}) = 0$ if, for all periods *t>1*, $x_t/x_{t-1} = y_t/y_{t-1}$. This means that indices are defined as being similar if they have identical period-on-period change rates. This axiom is similar, but more restrictive than identity axiom B, as it does not only define the outcome for a given input of two time series, but also specifies that all time periods later than t=1 need to be considered. As we will see, this has an impact on the performance of the proposed similarity measures in testing the axioms.

After having decided upon the properties a similarity measure for time series should have, I will now propose three measures of index similarity, that are then tested against these properties.

## 3.2. Proposal for similarity measures for price indices

This paper is not meant to propose and test an exhaustive number of similarity measures; instead, it is a first approach of raising awareness for the problem and proposing first similarity measures. Further measures, with properties that may be better, can be derived by future research. Therefore, I will focus on three measures of the family of average quadratic distances that by their very nature already satisfy axiom A:

A. *Average quadratic distance of indices.* This is based on the absolute index numbers. Let there be *n* time periods $t = 1, \dots, T$, and two indices (time series) ***I*** and ***J***. Then, similarity measure $S_1$ is defined as follows:

$$S_1 \equiv \frac{\sum_{t=1}^{T}(I_t - J_t)^2}{n}$$

B. *Average quadratic distance of annual change rates.* This is based on annual change rates. Therefore, it needs time series that are at least 13 months long (under the precondition that we have a monthly time series). Using the same notation as for $S_1$, $S_2$ is defined as follows:

$$S_2 \equiv \frac{1}{n-12}\sum_{t=13}^{T}\left(\frac{I_t}{I_{t-12}} - \frac{J_t}{J_{t-12}}\right)^2$$

C. *Average quadratic distance of monthly change rates.* This similarity measure is based on monthly change rates. Therefore, it needs time series that are, at least, two months long. Using the known notation, $S_3$ is defined as follows:

$$S_3 \equiv \frac{1}{n-1}\sum_{t=2}^{T}\left(\frac{I_t}{I_{t-1}} - \frac{J_t}{J_{t-1}}\right)^2$$

These three similarity measures will be tested against the axioms in the next chapter.

## 3.3. Testing of the measures against axioms: derivation of the preferred similarity measure

Testing the similarity measures against the axioms has been carried out; the proofs are shown in the appendix. Table 1 shows the results.

*Table 1: Result of axiomatic tests for proposed similarity measures.*

| Axiom | S1 - indices | S2 – annual change rates | S3 - monthly change rates |
|---|---|---|---|
| **A. Continuity** | ✓ | ✓ | ✓ |
| **B. Identity** | 💣 | ✓ | ✓ |
| C. **Positivity** | ✓ | ✓ | ✓ |
| D. **Symmetry** | 💣 | ✓ | ✓ |
| E. **Proportionality** | 💣 | ✓ | ✓ |
| **F. Correction for number of periods** | 💣 | ✓ | ✓ |
| G. **Focus on change rates** | 💣 | 💣 | ✓ |

As we can see, S3 satisfies all axioms, while S1 fails most of them. S2 fails only the focus on change rates. So, with regard to the desired properties, S3 seems to be superior and will be used in the remainder of this paper by applying it to some problems of time series comparison from the literature, focusing on price indices.

### 3.4. Further considerations: additional metrics to report when using $S_3$

While $S_3$ seems to be a promising measure for time series similarity, it should not be used without context. Therefore, when reporting it for two or more timeseries, some additional information should be provided as well:

- *Benchmark index (if more than two series are investigated).* It can be shown that $S_3$ is not transitive, so the choice of the benchmark index has an impact on the results. Therefore, it needs to be reported.
- *Length of time series.* While the measure is designed to abstract from the time series length by averaging over the number of time periods, in order to enable the same interpretation of the value of $S_3$ for long and short time series, the length itself is an important additional information to the user. It tells us, if the shown degree of similarity was observed during a long or a short period, and so if the measure is more or less robust to adding an additional observation period.
- *Absolute time range (if available).* Especially when applied to economic indices, the similarity measure may yield different results depending on for which time period it is actually calculated. Conjunctural developments may be a driving force for this issue.
- *Root $S_3$.* The interpretability of $S_3$, as an averaged sum of squares, might be rather limited. In addition, for economic indices in general and price indices in particular, the differences found in period-to-period change rates will be in the dimension of 0.01 (i.e. 1 p.p.), which, when squared, amount to 0.0001. That looks very small, while representing a 1 percentage points difference. Therefore, to get more meaningful numbers, we may calculate the root of $S_3$:

$$RS_3 = \sqrt{S_3}$$

  An interpretation of $RS_3$ may draw upon its twin, the standard deviation, as standard difference between the change rates of the two series.

With these pre-conditions, I will now apply the similarity measure to problems from the literature.

# 4. Application of the similarity measure to problems from the literature

By thinking about possible applications of the similarity measure $S_3$ to the analysis especially of price indices, three use cases emerge:

- Comparison of alternative price indices with a benchmark
- Comparison of two time series
- Finding imputations for missing observations for a time series by comparing it to other ones and choosing the most "similar" index development, with which the missing observation can be imputed.

## 4.1 Comparison of several time series with a benchmark

Especially with the advent of multilateral methods in the domain of consumer prices as a means of aggregating transaction data, it has become a major task of comparing time series aggregated from the same microdata using different index formulas, comparison windows and splicing techniques.[7] However, a comparison for its own sake does not help for deciding on which compilation method to choose, if no benchmark exists. The most likely best method is then the one that behaves most "similar" to the benchmark index.

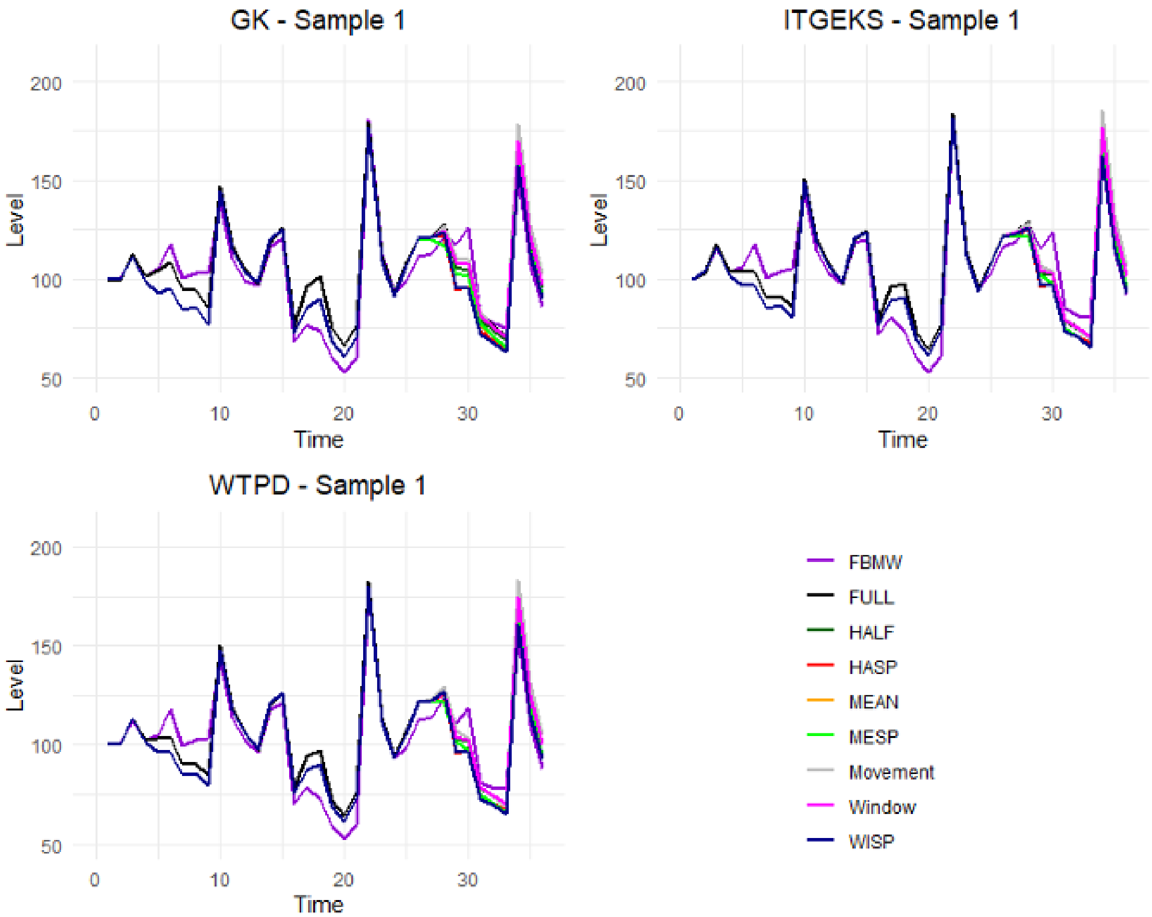
### 4.1.1 Many indices, differences look small – is that the case?

A nice example can be found in Radjabov & Van Loon (2022). The authors compare several index compilation methods for seasonal products. While they show a preference for the GEKS index, they do not discuss which of the splicing methods should be preferred; they rather show the sensitivity of several multilateral price indices to splicing methods by providing the range of percentage changes that they compute with 9 different splicing methods. They conclude that "… GEKS price indices are less sensitive to a usage of different splices rather than GK and WTPD price indices …" and that "… FBMW price indices significantly deviate from benchmark price indices." (Radjabov & Van Loon, 2022, S. 13). Applying the similarity measure $S_3$ might provide additional insights and would help in deciding which splicing method should be preferred. I have carried out such calculations for their "sample 1" for Geary-Khamis (GK), ITGEKS, and Weighted Time-Product Dummy (WTPD) indices, as they are shown in their figure 2 (Radjabov & Van Loon, 2022, p. 14; see below).[8]

[7] For an overview, see (Eurostat, 2022).

[8] Many thanks to Lucien May, Botir Radjabov and Ken Van Loon for sharing the numbers with me.

*Figure 1: The spliced Price Indices of Geary-Khamis, ITGEKS and WTPD Methods - figure 2 in Radjabov & Van Loon (2022, p. 14)*[9]




From looking at the figures, one could get the idea that the results are rather similar. However, the scale points to significant differences between the methods, as already the benchmark index FULL has values in the range of 60 to 185 (depending on the index formula). So, also differences that look small in the figure may point to very dissimilar indices.

Let's have a look at the similarity measure $S_3$ for this example. The calculations of the similarity measure use the FULL index as benchmark.[10] The results are shown in the following table and figure.

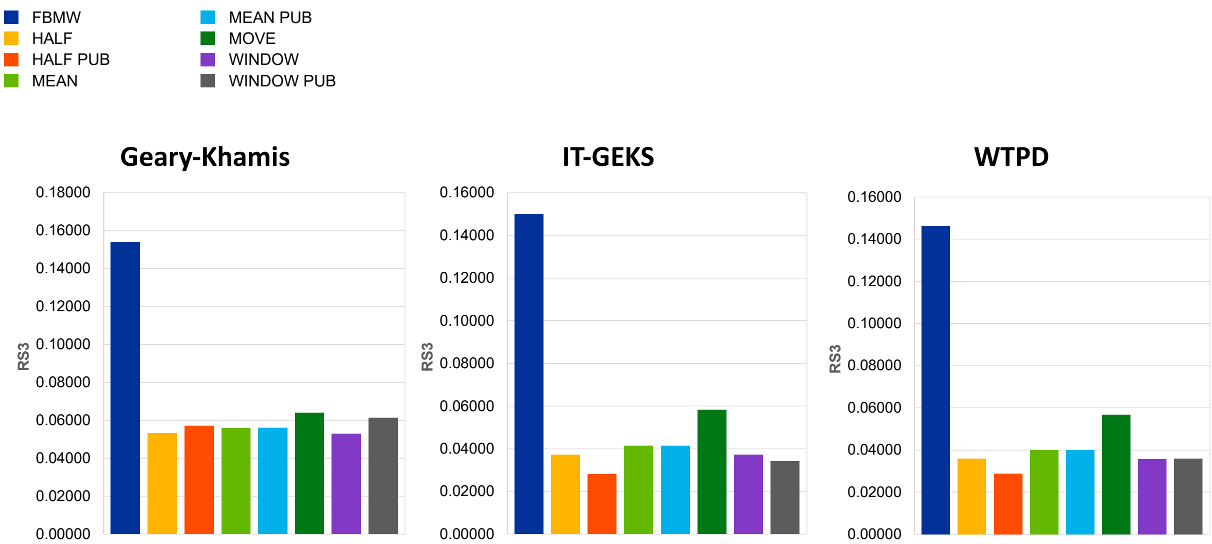
[9] Meaning of the abbreviations of the splicing methods:
FBMW: fixed base moving window; FULL: benchmark index (covers full period); HALF: half splice; HASP: half splice on published price indices; MEAN: mean splice; MESP: mean splice on published price indices; MOVEMENT: movement splice; Window: window splice; WISP window splice on published price indices. See the explanations in Radjabov & Van Loon (2022), p. 12.

[10] As suggested by Radjabov & Van Loon, p. 13: "Moreover, Figure 2 also provides full window (FULL) price indices, which are seen as benchmark price indices since they are "chain drift" free."

*Table 2: S3 and RS3 for different splicing methods and different multilateral methods (source: ECB calculations based on Radjabov & Van Loon, 2022, sample 1).*

| **Index** | | **FBMW** | **HALF** | **HALF PUB** | **MEAN** | **MESP** | **MOVE-MENT** | **WINDOW** | **WISP** |
|---|---|---|---|---|---|---|---|---|---|
| **Geary-Khamis** | Number of periods | 35 | | | | | | | |
| | S3 | 0.02369 | 0.00281 | 0.00326 | 0.00312 | 0.00314 | 0.00410 | 0.00280 | 0.00376 |
| | RS3 | 0.15390 | 0.05300 | 0.05712 | 0.05585 | 0.05607 | 0.06401 | 0.05295 | 0.06136 |
| **IT-GEKS** | Number of periods | 35 | | | | | | | |
| | S3 | 0.02251 | 0.00139 | 0.00078 | 0.00172 | 0.00171 | 0.00338 | 0.00139 | 0.00117 |
| | RS3 | 0.15004 | 0.03726 | 0.02797 | 0.04142 | 0.04137 | 0.05815 | 0.03728 | 0.03414 |
| **WTPD** | Number of periods | 35 | | | | | | | |
| | S3 | 0.02141 | 0.00129 | 0.00083 | 0.00159 | 0.00159 | 0.00321 | 0.00128 | 0.00129 |
| | RS3 | 0.14631 | 0.03591 | 0.02874 | 0.03985 | 0.03991 | 0.05662 | 0.03573 | 0.03585 |

*Figure 2: RS3 for different splicing methods and different multilateral methods (source: ECB calculations based on Radjabov & Van Loon, 2022, sample 1).*

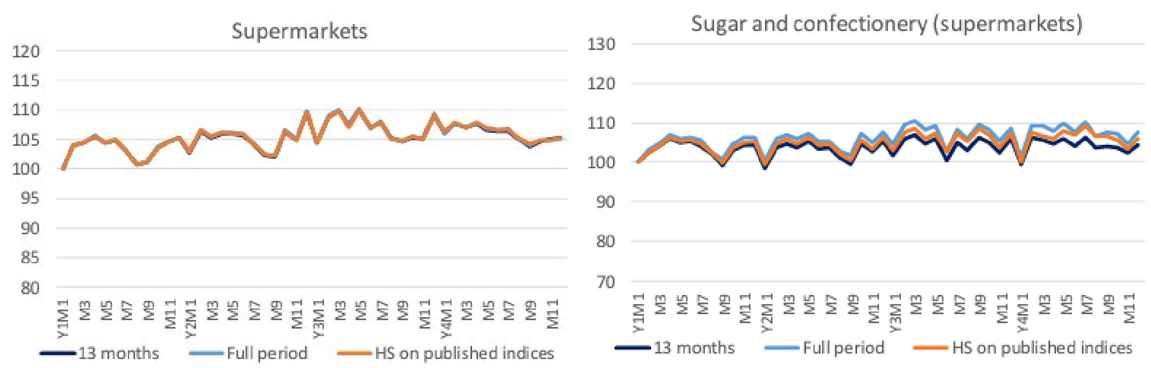



First, it can be directly seen that their claim on the FBMW splicing method is correct, as it is least similar to the benchmark index. However, when it comes to the sensitivity regarding the splicing method, the finding is less clear for sample 1: IT-GEKS' and WTPD's similarity measures closely match for almost every splicing method, most pronounced differences are shown for the worst performing methods (FBMW and MOVE) only. However, the use of the similarity measure also enables us to draw a conclusion about which splicing method matches best the benchmark index: for IT-GEKS and WTPD, it is HASP; for Geary-Khamis, WINDOW and HALF show almost the same degree of similarity, but, in general, the level of similarity is worse compared to IT-GEKS and WTPD. So, it may be not only about the choice of the splicing method, it may be about the GK method itself.

### 4.1.2 Unveiling differences invisible in the figure

Chessa (2019) compares different extension methods for multilateral methods, in order to find out which extension method is superior for the calculation of scanner-data based indices. A special attention is given to finding chain-drift free extension methods. A main finding is that both methods with a fixed base and methods that use splicing on published indices are to be recommended. Out of the many comparisons shown by him,[11] the two shown in the following figure will be examined using $S_3$.

*Figure 3: two cases from Chessa (2019, ppt, slide 28, left panel, and slide 30, right panel).*




The benchmark, in both cases, is the "full period" index. From eyeballing, one would say that the series on the left panel are almost identical; in the right panel, the "HS [half splice] on published indices" seems to match the "full period" index more closely than the "13 months" index.[12]

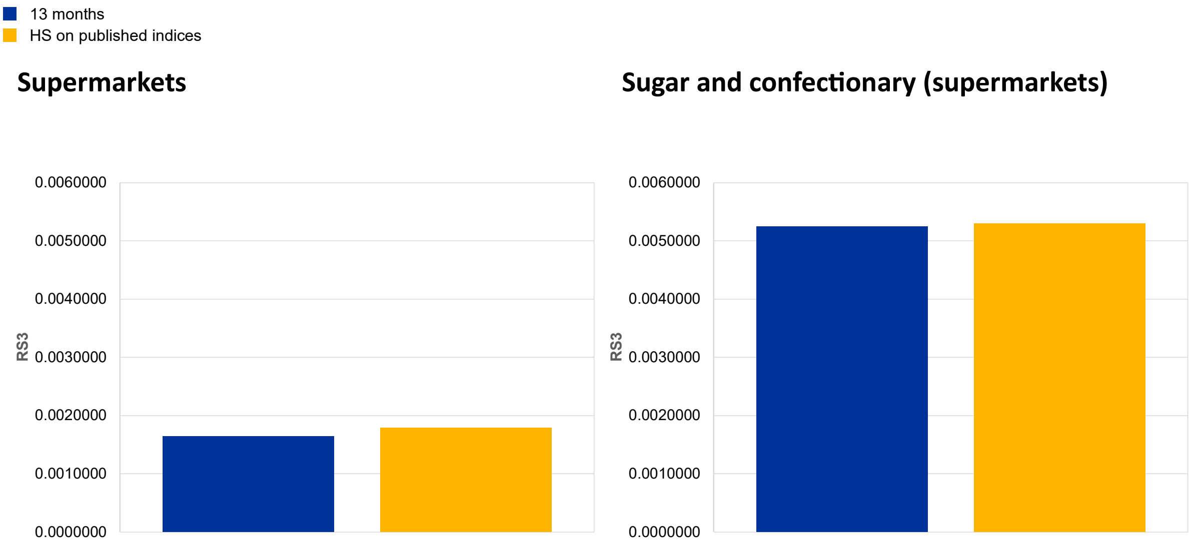
Let us apply $S_3$ to these two cases, also using the "full period" index as benchmark.[13] The results are shown in the following table and figure.

*Table 3: $S_3$ for two cases from Chessa (2019, ppt, slide 28, left panel, and slide 30, right panel; source: ECB calculations).*

| **Case** | | **13 months** | **HS on published indices** |
|---|---|---|---|
| **Supermarkets** | Number of periods | 47 | |
| | S3 | 0.0000027 | 0.0000032 |
| | RS3 | 0.0016390 | 0.0017914 |
| **Sugar and confectionery (supermarkets)** | Number of periods | 47 | |
| | S3 | 0.0000275 | 0.0000281 |
| | RS3 | 0.0052435 | 0.0052993 |

[11] Not only in the paper. In the presentation, he was showing even more comparisons, one of which is used for this paper. See Chessa (2019, ppt).
[12] For the sake of completeness, it should be mentioned that Chessa used the Geary-Khamis index for these comparisons.
[13] Many thanks to Tony Chessa for making the data available.

*Figure 4: $RS_3$ for two cases from Chessa (2019, ppt, slide 28, left panel, and slide 30, right panel; source: ECB calculations).*



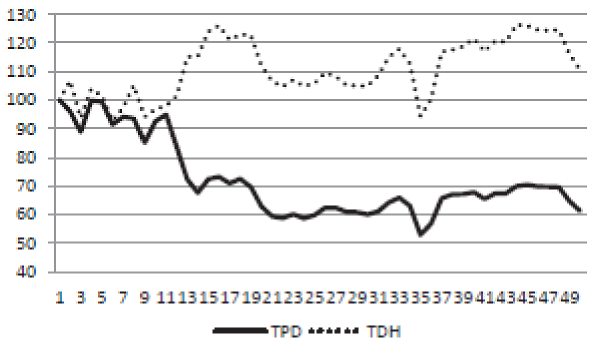

Not surprising, the similarity measures show that the series for “supermarkets” are more similar than those for “sugar and confectionary”. What is more surprising is that, for “sugar and confectionary”, the differences in similarity between 13 months and HS on published indices are only marginal, despite the divergence visible in the chart. And they are even smaller than those for supermarkets. This is an example of how relying on visible evidence only may be misleading.

It should be noted that, in the paper, Chessa (2019, p. 9, p. 16) uses not only visualisation, but also several statistics to judge the similarity of two indices, notably mean, minimum and maximum of the differences between the year-on-year indices. That may point to further research on having a comprehensive set of indicators for judging index similarity.

## 4.2 Comparison of two time series

In the paper “A Comparison of Weighted Time Dummy Hedonic and Time-Product Dummy Indexes”, De Haan et al. (2021) show an example for the comparison of two time series (see Figure 5). They describe the different results obtained by using different index formula on the same raw (scanner) data. This is a quite interesting case, as the two indices diverge fully after some periods, even with different signs of the change rate; for later periods it is then hard to judge if they still behave in a similar manner, as the magnitudes of the indices are substantially different.

*Figure 5: TPD and TDH Index - Figure 2 in De Haan et al. (2021), p. 406.*




Next to calculating the similarity measure for the whole time range, it might be interesting to calculate it for some parts of the time range only:

- The beginning – periods 2-11: rather similar trend of the time series, but deviating period-on-period change rates.
- Time of divergence – periods 12-14: Different sign of the p-o-p change rates, and increasing difference in absolute index numbers, leading to different index levels.
- Parallelism at different levels – periods 16-50: small differences in the p-o-p change rates, but already very different index levels should still lead to a small similarity measure $S_3$.

The results for these three time ranges as well as the whole time range have been calculated[14] and can be found in Table 4.

*Table 4: $S_3$ measures based on TPD and TDH series from de Haan et al. (2021, p. 406); source: ECB calculations..*

| Time range | 2-11 | 12-14 | 15-50 | Total |
|---|---|---|---|---|
| **Length of time series** | 10 | 3 | 36 | 49 |
| **$S_3$** | 0.002613 | 0.033528 | 0.000101 | 0.00266 |
| **$RS_3$** | 0.05112 | 0.183107 | 0.010079 | 0.05158 |

The results show, in total, a substantial dissimilarity between the two series. The interpretation of $RS_3$ as some kind of average distance in p.p. of growth rates would see 5.1 p.p. "standard distance", which is a lot. And, as the analysis by time ranges show, this is not only due to the time of divergence ($RS_3$=0.18) and the already diverging beginning, but also to the parallelism at different levels: while eyeballing suggests rather similar growth rates, $RS_3$=0.01 suggest 1 p.p. differences in growth rates per period, which is a lot. It can be concluded that the series are not similar and may yield very different inflation rates – and that being applied to the same underlying data.[15]

## 4.3 Selecting a time series for the imputation of missing values

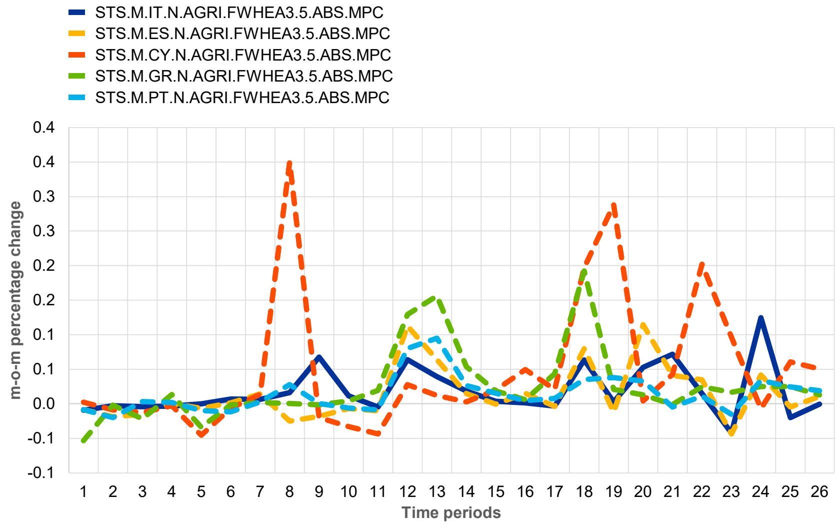
Another use case for the application of $S_3$ is the imputation of missing values. If an observation value for a time series in a certain time period is not available, the development of which other time series should be taken for imputation? This is a question that frequently occurs in official statistics, be it because of late reporting of respondents or because of an exhaustion of the sample frame for a particular stratum.

As an example, the ECB gets data from the European Commission on prices of agricultural commodities [16] and calculates agricultural price indices out of it. As the price data are often incomplete, imputations are needed. Currently, this is done using the correlation coefficient of the period-to-period percentage changes of the series of absolute prices.

In the example below, for the Italian series on prices for durum wheat, a series for imputation is sought. Figure 6 shows the period-on-period percentage changes of the absolute prices[17] for durum wheat for Italy and other countries in southern Europe. Note that these are from a time range between August 2004 and March 2024; they show the periods in which data was available for all countries in (at least) two consecutive months.

[14] Many thanks to Jan de Haan for providing the numbers underlying Figure 5.
[15] The reasons for that behaviour are well explained in De Haan et al. (2021).
[16] Data source: European Commission (2024).
[17] As no quality and quantity adjustment is done, the series of absolute prices per ton are pretty similar to a price index, with the exception of not referring to a particular base period. The month-on-month percentage changes, however, are the same as they were for a price index based on these absolute prices.

*Figure 6: monthly percentage changes of prices for durum wheat for southern European countries. Benchmark series: Italy. Source: European Commission, ECB calculations.*




While the current implementation in the ECB's production system uses the correlation coefficient, whose drawbacks as a measure for similarity have been discussed in chapter 2.2, I have also calculated $S_3$ and $RS_3$ for the example.

*Table 5: Correlation coefficient, $S_3$ and $RS_3$ (source: ECB calculations).*

| **Country comparison** | **IT vs. ES** | **IT vs. CY** | **IT vs. GR** | **IT vs. PT** |
|---|---|---|---|---|
| **Number of time periods** | | 26 | | |
| **Correlation coefficient** | 0.7262 | -0.12559 | 0.24096 | 0.40496 |
| **$S_3$** | 0.000670 | 0.007147 | 0.001875 | 0.001057 |
| **$RS_3$** | 0.025893 | 0.084542 | 0.043297 | 0.032512 |

In this example, both $S_3$ and the correlation coefficient detect the same series (the one from Spain) as most suited for imputation. While the correlation coefficient also shows that the series from Cyprus may even move in the opposite direction than the Italian one, the $RS_3$ measure provides an indication about the magnitude of differences in month-on-month change rate that can be expected when imputing with the respective country series.

## 5. Discussion

From the calculation examples above, it can be seen that the similarity measure $S_3$ proposed in this paper enables a better judgement about the similarity of time series than previous methods. It avoids the pitfalls of visualization and also helps to judge the similarity of time series at different numerical levels. A main advantage for its application to economic indices is that it takes into account the period-on-period change rates, which are often in the focus of communicating these numbers, instead of the absolute index levels. As such a measure did not exist before, it may contribute to the discussion about preferred index formulas and imputation methods.

However, some limitations should be acknowledged: the measure is based on differences in p-o-p change rates, that might be in permille dimensions mostly. As these numbers are being squared, the resulting measure might be a very small, which complicates the interpretation of the similarity measure. I have proposed $RS_3$ as a first remedy, also with a possible interpretation of "root mean squared difference". Further research may strive for a standardised measure in the spirit of the correlation coefficient, who is a benchmark in this regard.

Another issue is that this "calculated" similarity may not be sufficient to finally judge about the suitability of the "most similar index" to approximate a benchmark index. It may be advisable to use additional metrics, as, for example, shown in Chessa (2019). Further research may be aimed at developing a consistent framework of metrics for this purpose.

## 6. Conclusion

So far, judging the similarity of time series in general and economic indices in particular relied on visualization and some descriptive statistics, methods that are prone to error or not suited for the matter. The proposed similarity measure $S_3$ has desirable properties, and its application to use cases from the literature or current practice shown in this paper already helped gaining new insights. It might therefore contribute to the discussion about economic indices and, in particular, help to identify the most suitable index formula and splicing method for the use of scanner data. Being a first proposal, it leaves room for future research, like the formulation of different similarity measures with even better properties, for example regarding the dimension of the measure, and identifying a set of indicators that, being shown together, enable a true comparison of time indices: relying on just one indicator may always fall short for describing complex relationships. Altogether, the similarity measure $S_3$ enables more evidence-based comparison of time series and also imputation of missing values, and I hope to see it more often in the future than charts overloaded with lines, from which a conclusion is hard to draw.

## Acknowledgements

Many thanks to Erwin Diewert for pointing me to his article on price and quantity similarity, Martin Eiglsperger for advice on the correlation coefficient and time series analysis, Eduardo Leite Kropiwiec Filho for information about the imputation procedure for agricultural prices, and to Antonio Chessa, Jan de Haan, Lucien May, Botir Radjabov and Ken Van Loon for providing me with the data underlying their papers.

## References

Bialek, J., & Bobel, A. (2019). Comparison of Price Index Methods for the CPI Measurement Using Scanner Data. *Paper for the 16th meeting of the Ottawa Group*. Rio de Janeiro.

Chessa, A. (2019). A comparison of index extension methods for multilateral methods. *Paper for the 16th meeting of the Ottawa Group*. Rio de Janeiro.

Chessa, A. (2019). Extension methods for multilateral index series. *Presentation of the paper for the 16th meeting of the Ottawa Group, 2019-05-09*. Rio de Janeiro.

Chessa, A., Verburg, J., & Willenborg, L. (2017). A Comparison of Price Index Methods for Scanner Data. *Paper for the 15th meeting of the Ottawa Group*. Eltville, Germany.

De Haan, J., Hendricks, R., & Scholz, M. (2021). Price measurement using scanner data: time-product dummy versus time dummy hedonic indexes. *The Review of Income and Wealth*, Vol. 67 No. 2, pp. 394-417.

Diewert, W. (2006). *Similarity and Dissimilarity Indexes: An Axiomatic Approach.* Vancouver: Discussion Paper No. 02-10, Department of Economics, University of British columbia.

European Commission, D. G. (30. 04 2024). *Monthly Market Prices*. Von https://agridata.ec.europa.eu/extensions/DashboardPrice/DashboardMarketPrices.html abgerufen

Eurostat. (2022). *Guide on Multilateral Methods in the Harmonised Index of Cnsumer Prices.* Luxembourg: Publications Office of the European Union.

Radjabov, B., & Van Loon, K. (2022). Seasonal Products and Multilateral Method. *Paper for the 2022 meeting of the Ottawa Group*. Rome.

Ropella, G., Nag, D., & Hunt, C. (2003). Similarity Measures for Automated Comparison of In Silico and In Vitro Experimental Results. *Proceedings of the 25th Annual International Conference of the IEEE, Vol. 3*, (S. 2933-2936).

Vogel, F. (1999). *Beschreibende und schließende Statistik. Formeln, Definitionen, Erläuterungen, Stichwörter und Tabellen.* München: Oldenbourg.

# Appendix

## Proofs for axiom testing

### Axiom B: Identity

S1:

$$S_1(\boldsymbol{X}, \lambda\boldsymbol{X}) = \frac{\sum_{t=1}^{T}(x_t - \lambda x_t)^2}{n} = \frac{\sum_{t=1}^{T} x_t^2(1-\lambda)^2}{n}$$

As the index values are positive, this term is 0 only in the special case when $\lambda = 1$, so the property is not kept by S1. 💣

S2:

$$S_2(\boldsymbol{X}, \lambda\boldsymbol{X}) = \frac{1}{n}\sum_{t=13}^{T}\left(\frac{x_t}{x_{t-12}} - \frac{\lambda x_t}{\lambda x_{t-12}}\right)^2 = \frac{1}{n}\sum_{t=13}^{T}\left(\frac{x_t}{x_{t-12}} - \frac{x_t}{x_{t-12}}\right)^2 = 0 \text{ q.e.d.}$$

S3:

$$S_3(\boldsymbol{X}, \lambda\boldsymbol{X}) = \frac{1}{n}\sum_{t=2}^{T}\left(\frac{x_t}{x_{t-1}} - \frac{\lambda x_t}{\lambda x_{t-1}}\right)^2 = \frac{1}{n}\sum_{t=2}^{T}\left(\frac{x_t}{x_{t-1}} - \frac{x_t}{x_{t-1}}\right)^2 = 0 \text{ q.e.d.}$$

### Axiom C: Positivity

As S1, S2, and S3 are constructed as an averaged sum of square numbers, which are by definition positive or 0. The case "0" is achieved for cases of the identity axiom for S2 and S3, and for S1 only if **X = Y**. So, S1, S2, and S3 all keep the positivity axiom.

### Axiom D: Symmetry

Let us define a and b in the following way, with $x_t, y_t$ being strictly positive members of the time series **X,Y**:

- For S1: $a = x_t, b = y_t$
- For S2: $a = \frac{x_t}{x_{t-12}}, b = \frac{y_t}{y_{t-12}}$
- For S3: $a = \frac{x_t}{x_{t-1}}, b = \frac{y_t}{y_{t-1}}$

So, without limitation, we can write:

$$S_{1,2,3}(\boldsymbol{Y}, \boldsymbol{X}) = \frac{\sum_{t=1}^{T}(b-a)^2}{n} = \frac{\sum_{t=1}^{T}(-1)^2(-b+a)^2}{n} = \frac{\sum_{t=1}^{T}(a-b)^2}{n} = S_{1,2,3}(\boldsymbol{X}, \boldsymbol{Y}) \text{ q.e.d.}$$

### Axiom E: Proportionality

S1:

$$\begin{aligned} S_1(\boldsymbol{X}, \lambda\boldsymbol{Y}) &= \frac{\sum_{t=1}^{T}(x_t - \lambda y_t)^2}{n} = \frac{\sum_{t=1}^{T}({x_t}^2 - 2\lambda x_t y_t + \lambda^2 y_t^2)}{n} \\ &= \frac{1}{n}\sum_{t=1}^{T}({x_t}^2 - 2\lambda x_t y_t + 2x_t y_t - 2x_t y_t + \lambda^2 y_t^2 + y_t^2 - y_t^2) \\ &= \frac{1}{n}\sum_{t=1}^{T}({x_t}^2 - 2x_t y_t + y_t^2) + \frac{1}{n}\sum_{t=1}^{T}(-2\lambda x_t y_t + 2x_t y_t + \lambda^2 y_t^2 - y_t^2) \\ &= S_1(\boldsymbol{X}, \boldsymbol{Y}) + \frac{1}{n}\sum_{t=1}^{T}(-2\lambda x_t y_t + 2x_t y_t + \lambda^2 y_t^2 - y_t^2) \end{aligned}$$

=> S1 does not satisfy axiom E. 💣

S2:

$$S_2(\boldsymbol{X},\lambda\boldsymbol{Y}) = \frac{1}{n}\sum_{t=13}^{T}\left(\frac{x_t}{x_{t-12}} - \frac{\lambda y_t}{\lambda y_{t-12}}\right)^2 = \frac{1}{n}\sum_{t=13}^{T}\left(\frac{x_t}{x_{t-12}} - \frac{y_t}{y_{t-12}}\right)^2 = S_2(\boldsymbol{X},\boldsymbol{Y}) \text{ q.e.d.}$$

S3:

$$S_3(\boldsymbol{X},\lambda\boldsymbol{Y}) = \frac{1}{n}\sum_{t=2}^{T}\left(\frac{x_t}{x_{t-1}} - \frac{\lambda y_t}{\lambda y_{t-1}}\right)^2 = \frac{1}{n}\sum_{t=2}^{T}\left(\frac{x_t}{x_{t-1}} - \frac{y_t}{y_{t-1}}\right)^2 = S_3(\boldsymbol{X},\boldsymbol{Y}) \text{ q.e.d.}$$

### Axiom F: Correction for the number of time periods

The axiom can be tested by calculating $S(\widetilde{\boldsymbol{X}},\widetilde{\boldsymbol{Y}})$, as defined in the axiom. These time series have *n+1* time periods $t = 1, \ldots, T+1$.

S1:

$$S_1(\widetilde{\boldsymbol{X}},\widetilde{\boldsymbol{Y}}) = \frac{\sum_{t=1}^{T+1}(x-y)^2}{n+1} = \frac{(n+1)(x-y)^2}{n+1} = (x-y)^2 = \frac{n}{n}(x-y)^2 = \frac{\sum_{t=1}^{T}(x-y)^2}{n} = S_1(\boldsymbol{X},\boldsymbol{Y})$$

q.e.d.

S2:

Without limitation: *n*, in this case, is the number of periods from $t = 13$ to $T$.

$$S_2(\widetilde{\boldsymbol{X}},\widetilde{\boldsymbol{Y}}) = \frac{1}{n+1}\sum_{t=13}^{T+1}\left(\frac{x}{x} - \frac{y}{y}\right)^2 = 0$$

$$S_2(\boldsymbol{X},\boldsymbol{Y}) = \frac{1}{n}\sum_{t=13}^{T}\left(\frac{x}{x} - \frac{y}{y}\right)^2 = 0$$

So, $S_2(\widetilde{\boldsymbol{X}},\widetilde{\boldsymbol{Y}}) = S_2(\boldsymbol{X},\boldsymbol{Y})$ q.e.d.

S3:

Without limitation: *n*, in this case, is the number of periods from $t = 2$ to $T$.

$$S_3(\widetilde{\boldsymbol{X}},\widetilde{\boldsymbol{Y}}) = \frac{1}{n+1}\sum_{t=2}^{T+1}\left(\frac{x}{x} - \frac{y}{y}\right)^2 = 0$$

$$S_3(\boldsymbol{X},\boldsymbol{Y}) = \frac{1}{n}\sum_{t=2}^{T}\left(\frac{x}{x} - \frac{y}{y}\right)^2 = 0$$

So, $S_3(\widetilde{\boldsymbol{X}},\widetilde{\boldsymbol{Y}}) = S_3(\boldsymbol{X},\boldsymbol{Y})$ q.e.d.

### Axiom G: Focus on change rates

S1:

We already know that S1 fails to satisfy axiom B (identity). However, for every pair of time series $(\boldsymbol{X},\lambda\boldsymbol{X})$ with $\lambda > 0, \lambda \neq 1$, the period-on-period change rates are identical:

$$\frac{\lambda x_t}{\lambda x_{t-1}} = \frac{x_t}{x_{t-1}}$$

As $S_1(\boldsymbol{X},\lambda\boldsymbol{X}) \neq 0$ for $\lambda > 0, \lambda \neq 1$, axiom G is not satisfied by S1. 💣

S2:

It is a well-known fact from seasonal adjustment practice that time series, that have the same annual change rates, do not necessarily have the same monthly change rates. Here is an example:

For this proof, the notation is the following: Let $x^{mt}$ be the (strictly positive) value of time series $\boldsymbol{X}$ in month m (1,…,12) of year t (0,…T). Then, we can construct a time series $\boldsymbol{Y}$ with members $y^{mt}$ by using monthly factors $\lambda^m > 0, \neq 1$, that are the same for every year, but different for every month:

$$y^{mt} = \lambda^m x^{mt}$$

In such a situation, S2 would give us

$$S_2(\boldsymbol{X},\boldsymbol{Y}) = \frac{1}{n}\sum_{t=1,m=1}^{T}\left(\frac{x^{mt}}{x^{m,t-1}} - \frac{y^{mt}}{y^{m,t-1}}\right)^2 = \frac{1}{n}\sum_{t=1,m=1}^{T}\left(\frac{x^{mt}}{x^{m,t-1}} - \frac{\lambda^m x^{mt}}{\lambda^m x^{m,t-1}}\right)^2$$
$$= \frac{1}{n}\sum_{t=1,m=1}^{T}\left(\frac{x^{mt}}{x^{m,t-1}} - \frac{x^{mt}}{x^{m,t-1}}\right)^2 = 0$$

However, the monthly change rates are different by the ratio of $\lambda^m$ and $\lambda^{m-1}$:

$$\frac{y^{mt}}{y^{(m-1),t}} = \frac{\lambda^m x^{mt}}{\lambda^{m-1} x^{(m-1)t}}$$

So, axiom G is not satisfied by S2. 💣

S3:

S3 is already defined using the monthly indices:

$$S_3(\boldsymbol{X},\boldsymbol{Y}) \equiv \frac{1}{n}\sum_{t=2}^{T}\left(\frac{x_t}{x_{t-1}} - \frac{y_t}{y_{t-1}}\right)^2$$

So, S3 will only be equal to 0 if, for every time period *t=2,…T*:

$$\frac{x_t}{x_{t-1}} = \frac{y_t}{y_{t-1}}$$

This is exactly the definition of axiom G. q.e.d.