6. The JRC Statistical Audit of the Financial Secrecy Index 2018

William Becker and Michaela Saisana, European Commission, Joint Research Centre

Summary

The construction of the Financial Secrecy Index (FSI) from 115 variables to 20 Key Financial Indicators (KFSIs) that are grouped further into a Secrecy Score and aggregated together with a global scale weight across 112 jurisdictions worldwide inevitably entails both conceptual and practical challenges. The statistical audit discussed in this chapter constitutes the first collaboration between the Tax Justice Network and the European Commission's Joint Research Centre (JRC). The statistical assessment carried out by JRC aims at enhancing the transparency and reliability of the FSI and thus to enable policymakers to derive more accurate and meaningful conclusions. Prior to undertaking this statistical assessment, the Tax Justice Network and JRC engaged in previous discussions during spring 2016 and fall 2017, whereby earlier versions of the FSI were assessed by the JRC. Preliminary JRC suggestions were taken into account by TJN for the final computation of the FSI scores and rankings.

The intentions of the audit are to:

- Investigate the characteristics of the underlying data and check for eventual errors in calculation
- Assess the associations between indicators and see to what extent they agree with the conceptual framework
- Review the methodology used to treat, weight, and aggregate data
- Assess the impact of modelling assumptions (uncertainty and sensitivity analysis) on the FSI ranks
- Eventually recommend modifications based on the conclusions of the above.[434]

In particular, the JRC analysis complements the reported FSI ranks for the 112 jurisdictions with estimated confidence intervals, in order to better appreciate the robustness of these ranks to some modelling choices (such as choice of the variable to capture the global scale weight, the normalisation the weighting scheme and the aggregation formula).

Importantly, the construction of a composite indicator is a balance between statistical "rigour" and conceptual considerations, which can not infrequently contradict each other. This audit aims to investigate and analyse the statistical side of the equation, but does not aim to offer conceptual suggestions (which are better left to experts in international finance and regulations), or suggest where the balance should be struck between statistics and the concept of financial secrecy.

6.1 Construction of the Financial Secrecy Index

While the making of the FSI is described in more detail in the previous chapters of this report, a brief description of the index is helpful to put the audit in context and to allow the present chapter to be read independently if necessary.

The Financial Secrecy Index aims to *measure a jurisdiction's contribution to global financial secrecy in a way that highlights harmful secrecy regulations*. The FSI 2018 covers 112 jurisdictions, which have been selected according to their importance in international financial

[434] The JRC statistical audit was based on the recommendations of the OECD & JRC (2008) Handbook on Composite Indicators, and on more recent research from the JRC. Generally, JRC audits of composite indicators and scoreboards are conducted upon request of their developers, see https://ec.europa.eu/jrc/en/coin and https://composite-indicators.jrc.ec.europa.eu/

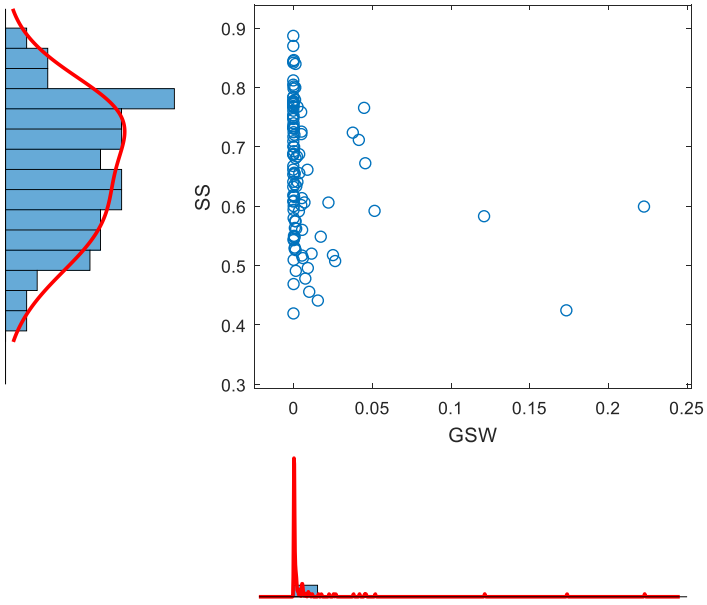
services. In fact, these 112 jurisdictions cover 99.3% of all global exports of financial service, as reported by the index developers. The number of jurisdictions has increased in successive editions of the FSI, with the present edition adding nine new jurisdictions. The FSI does not offer full global coverage because of gaps in data coverage, and the significant research effort of gathering data (much of which is based on original research). For more information on jurisdictions see Chapter 2, Jursdictions Covered.

The FSI is constructed (for each jurisdiction, indexed by *i*) as the product of a *secrecy score* (SS) and a *global scale weight* (GSW) as follows:

$$\mathrm{FSI}_i = \mathrm{GSW}_i^{1/3} . \mathrm{SS}_i^3 \qquad (1)$$

Figure 6-A shows the distributions of the GSW and SS: while the distribution of the SS is roughly normal, the GSW distribution is highly skewed to the left (reflecting the fact that some few jurisdictions have very large GSWs, while the large majority have very small GSWs). Highly skewed distributions are problematic when aggregating indicators, because the variability of the indicator is only due to some very few points, with the remainder having (relatively) almost no variability.

Figure 6-A: Scatterplot and marginal histograms of GSW and SS



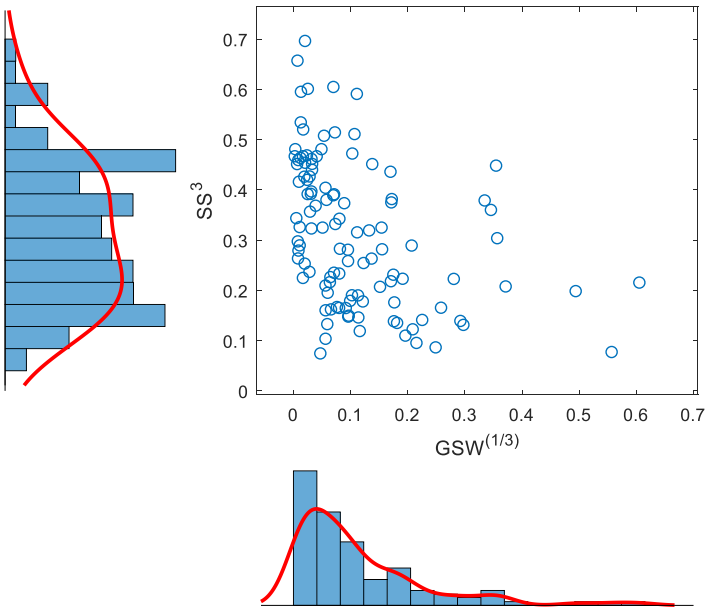
Source: European Commission, Joint Research Centre (JRC), 2017

The TJN acknowledge these problems by illustrating the imbalance in percentiles (see previous chapter). Their solution is to take the cube root of the GSW, and the cube of the SS—this results in a much better balance in percentiles between the two variables. Figure 6-B shows the distributions after this transformation. The distribution of the GSW is much improved, although still slightly skewed to the left. The SS now has a slight skew to the right.

The JRC tested different approaches to combine the SS and GSW into one number, all of which have different statistical and conceptual properties, and are presented later in this chapter. Yet, all aggregation methods involve different compromises between achieving statistical balance and not distorting the measured reality of the GSW (as well as conceptual considerations), therefore none of them are recommended per se. After careful

consideration, the TJN decided to retain the original formula for calculating the FSI scores, for a number of reasons that are discussed in the previous chapter. Nevetheless, the different approaches tested by JRC are retained here as a useful discussion of the statistical properties of the FSI.

Figure 6-B: Scatterplot and marginal distributions of SS^3 and $GSW^{(1/3)}$



Source: European Commission, Joint Research Centre (JRC), 2017

6.1.1 Secrecy Score

Underlying the Secrecy Score are 20 *Key Financial Secrecy Indicators* (KFSIs). This data is compiled by original desk-based research of TJN analysts, and comes from analysis of reports published by international agencies and organisations, country level original legislation, and a questionnaire that is sent to the ministries of finance and anti-money laundering "financial intelligence units" of each reviewed jurisdiction. Each KFSI is itself comprised of one or more questions which are posed to the experts, so in a sense each KFSI already represents an aggregation of sorts. Overall 115 questions have been selected by the TJN for the calculation of the twenty KFSIs. For each of these sub-indicators (questions), data that was not forthcoming from the questionnaire is assigned the most secretive score. After accounting for this assumption, the KFSI data does not have any missing data for the 112 jurisdictions covered.

The original sources of data for each KFSI are all referenced in detail on the website, and the definitions of each KFSI are given in depth on the individual KFSI fact sheets. This transparency and detail in the source information lends considerable credibility to the FSI and opens the data for use by stakeholders, as well as other researchers and analysts.

The KFSIs are grouped according to four conceptual themes as shown in Table 6-A, on the following page.

Table 6-A: Definitions and grouping of Key Financial Secrecy Indicators

Grouping	Number	Definition
Ownership registration	1	Bank Secrecy
	2	Trust and Foundations Register
	3	Recorded Company Ownership
	4	Other Wealth Ownership
	5	Limited Partnership Transparency
Legal entity transparency	6	Public Company Ownership
	7	Public Company Accounts
	8	Country-by-Country Reporting
	9	Corporate Tax Disclosure
	10	Legal Entity Identifier
Integrity of tax and financial regulation	11	Tax Administration Capacity
	12	Consistent-Personal-Income-Tax
	13	Avoids Promoting Tax Evasion
	14	Tax Court Secrecy
	15	Harmful legal vehicles
	16	Public Statistics
International standards and cooperation	17	Anti-Money Laundering
	18	Automatic Information Exchange
	19	Bilateral Treaties
	20	International Legal Cooperation

To obtain the Secrecy Score, SS for country *i* is obtained by taking the arithmetic average of the 20 KFSIs (indexed here by *k*) for each jurisdiction:

$$\mathrm{SS}_i = \frac{1}{20}\sum_{k=1}^{20} \mathrm{KFSI}_{k,i} \qquad (2)$$

6.1.2 Global Scale Weight

The global scale weight aims to measure each jurisdiction's share of offshore financial services activity in the global total. To do this, the TJN considers several alternative possible variables, which are discussed in Annex G. A description of the alternative GSW measures are as follows:

GSW-A	Trade in financial services (IMF Balance of Payments data)
GSW-B	Foreign direct investment (UNCTAD Foreign Direct Investment statistics)
GSW-C	Derived liabilities (IMF Coordinated Portfolio Investment Survey data)
GSW-D	Trade in services (UNCTADStat statistics)
GSW-E	Trade in goods (UN Comtrade data)
GSW-F	Bank deposits (Bank of International Settlements)
GSW-α	GSW-A, GSW-B, GSW-C

GSW-β	GSW-B, GSW-C, GSW-D

After an analysis of these alternatives, the TJN decided to use GSW-A, the *Trade in financial services* (TFS), from the IMF Balance of Payments data[435] as it is conceptually the closest to the definition of the GSW and the aims of the FSI. In this audit, this version of the GSW will be used in the analysis (although some analysis on the alternatives is performed in the following section, and an alternative is considered in the uncertainty analysis at the end of this chapter).

The TFS is scaled for each country *i* by dividing country's TFS by the sum of the TFS for all jurisdictions. This results in a GSW that represents the share of the global total (neglecting jurisdictions for which no data was available):

$$\mathrm{GSW}_i = \frac{\mathrm{TFS}_i}{\sum_{i=1}^{112} \mathrm{TFS}_i} \qquad (3)$$

GSW data was directly available for 85 of the 112 jurisdictions (76%). For the remainder, GSWs were estimated using data on stocks of internationally-held financial assets with which there is a strong correlation (see Chapter 4).

An important conceptual difference between the SS and the GSW is that the GSW is a measurable quantity which, for each country, can be reasonably interpreted as the share of the global total of offshore financial activity. The SS is more subjective: although it is based on objective indicators, the choice of which indicators to include and which scores to assign for various responses is necessarily subjective. The SS represents the extent to which each jurisdiction is secretive in its financial activity.

[435] http://www.imf.org/external/datamapper/datasets/BOP

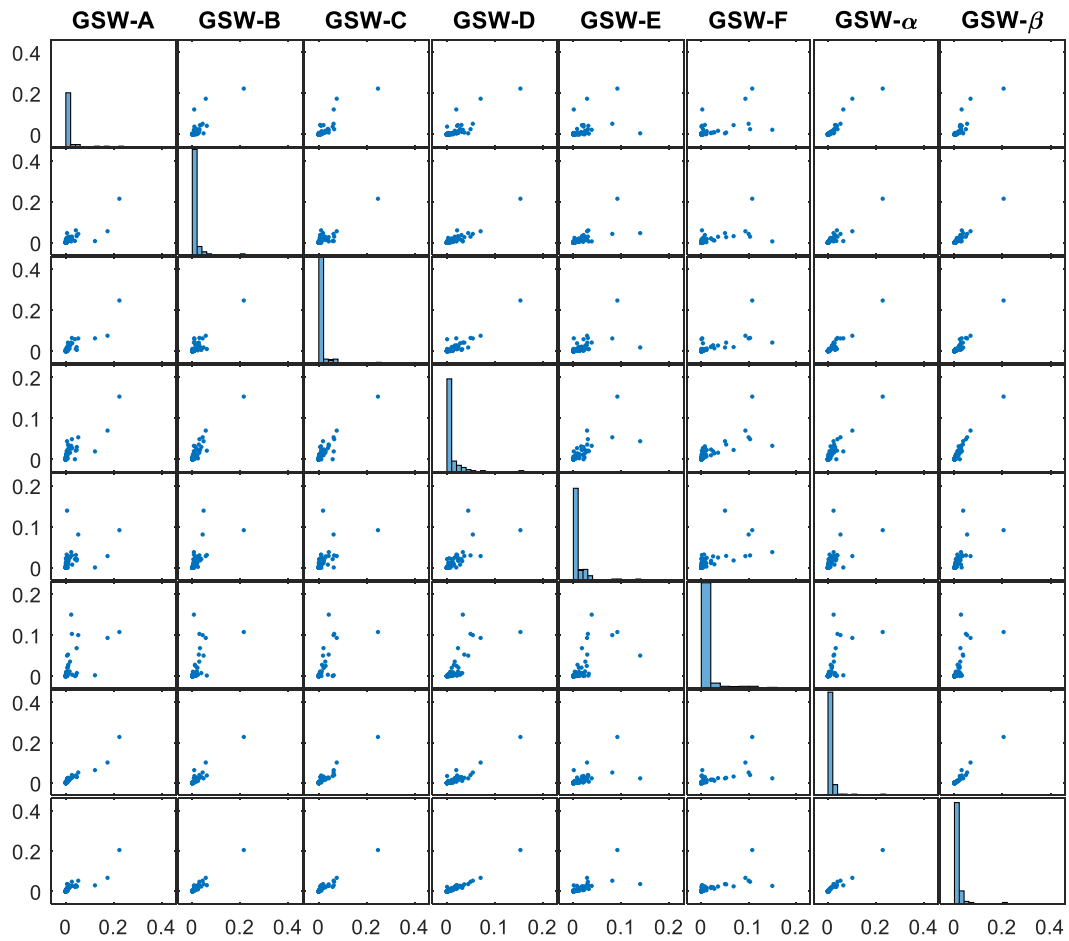
6.2 Exploring the data

This section comprises an exploratory analysis of the data at the indicator level (i.e. within the two FSI components, the global scale weight and the secrecy score). For the global scale weight, this comprises mostly an analysis of eight alternative GSW measures. For the secrecy score it examines the correlations between the twenty KFSIs and the links between individual KFSIs and the overall SS.

6.2.2 Global Scale Weight

The GSW data, for all the alternative variables considered, is classical log-normal data. This means that it is heavily skewed to the left, i.e. most jurisdictions have very small GSWs, while a small number have very large values. To emphasise this point, Figure 6-C shows scatter plots of all alternative GSW variables against each other, with histograms of each variable on the diagonal. From this figure it is difficult to understand the extent to which the variables are related to each other, because of the heavy skew of the distributions.

Figure 6-C: Scatter plots and histograms of untransformed GSW data.



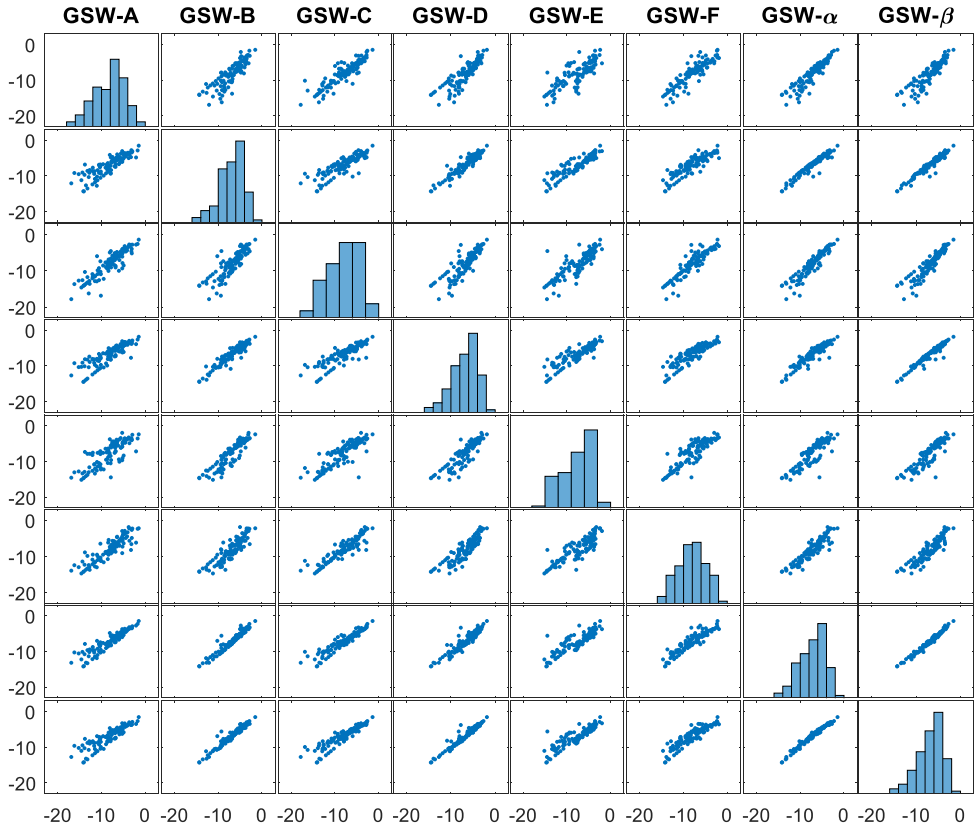
Note: The rows are in the same order as the columns.

Source: European Commission, Joint Research Centre (JRC), 2017

In order to see the relationship between the eight alternative measures to capture the global scale weight, the log transform is taken of all variables. The new scatter plot matrix is shown

in Figure 6-D. Now the relationships become easily visible: there are strong linear relationships between all of the variables considered. Since correlation is a linear measure of dependence between two variables, the log-transformed correlation values are much more representative of the relationships between the variables. These relationships support the TJN's approach of using regression to estimate the 24% of the missing values for GSW-A.

Figure 6-D: Scatter plots and histograms of log-transformed GSW data.

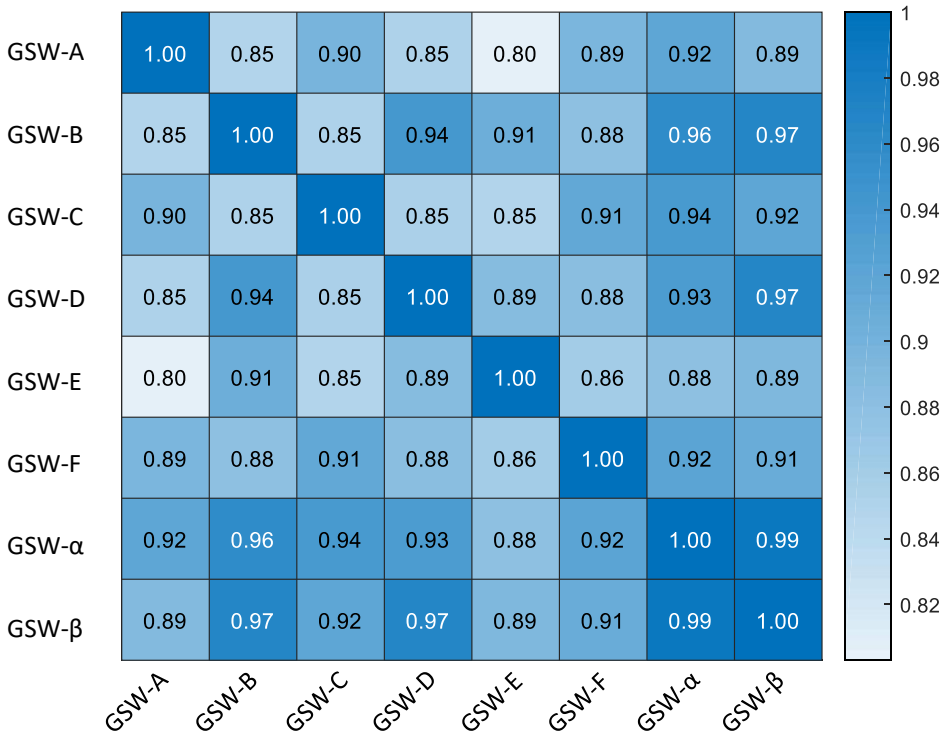


Note: The rows are in the same order as the columns.

Source: European Commission, Joint Research Centre (JRC), 2017

Figure 6-E shows a heat map of the correlation values (after log transforms) of the various GSW measures. The lowest correlation between any pair of variables is around 0.8: this means that the GSWs all appear to measure a similar concept. If the GSWs are taken as independent possible GSW measures, the GSW variable with the highest overall correlation with other variables is GSW-α (an average of GSW-A B and C), with an average bivariate correlation of 0.94. This is in fact no surprise since it is constructed from three of the other indicators. The lowest average bivariate correlation is found with GSW-E, namely the trade in goods (0.88). Despite the strong correlations between the eight alternative measures, the choice of the variable to represent each jurisdiction's global share of offshore financial services activity may have a noteworthy impact on the FSI ranks. For this reason, the impact of this assumption is assessed in the uncertainty and sensitivity analysis in Section 6.5. Overall, GSW-A has an average correlation of 0.87 with the other GSW alternatives.

Figure 6-E: Pearson correlation coefficients between GSW variables after log transforms.

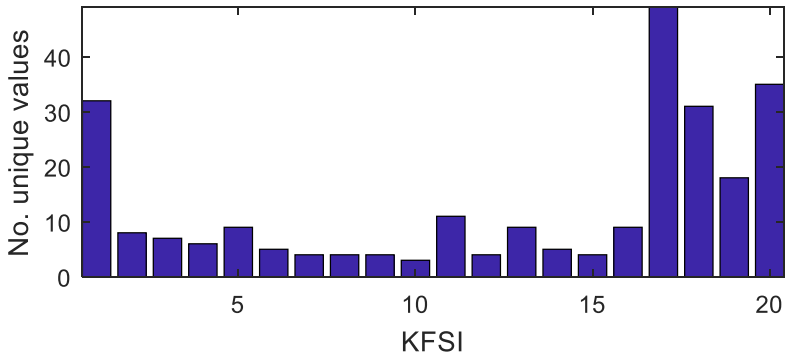


Note: Shading relates to strength of correlation. All values are significant at the 1% level.
Source: European Commission, Joint Research Centre (JRC), 2017

6.2.3 Secrecy Score

A similar exploratory analysis of the KFSI data can also be performed. In this case however, the context is different because the KFSIs are all used in the final FSI, and are aggregated together using an arithmetic average—see Equation (2). A heat map can again be generated which shows the relationships between the twenty KFSIs. However in this case, the data is largely discrete, with many indicators having only a small number of unique values—see Figure 6-F below.

Figure 6-F: Number of unique values in each KFSI over all 112 jurisdictions



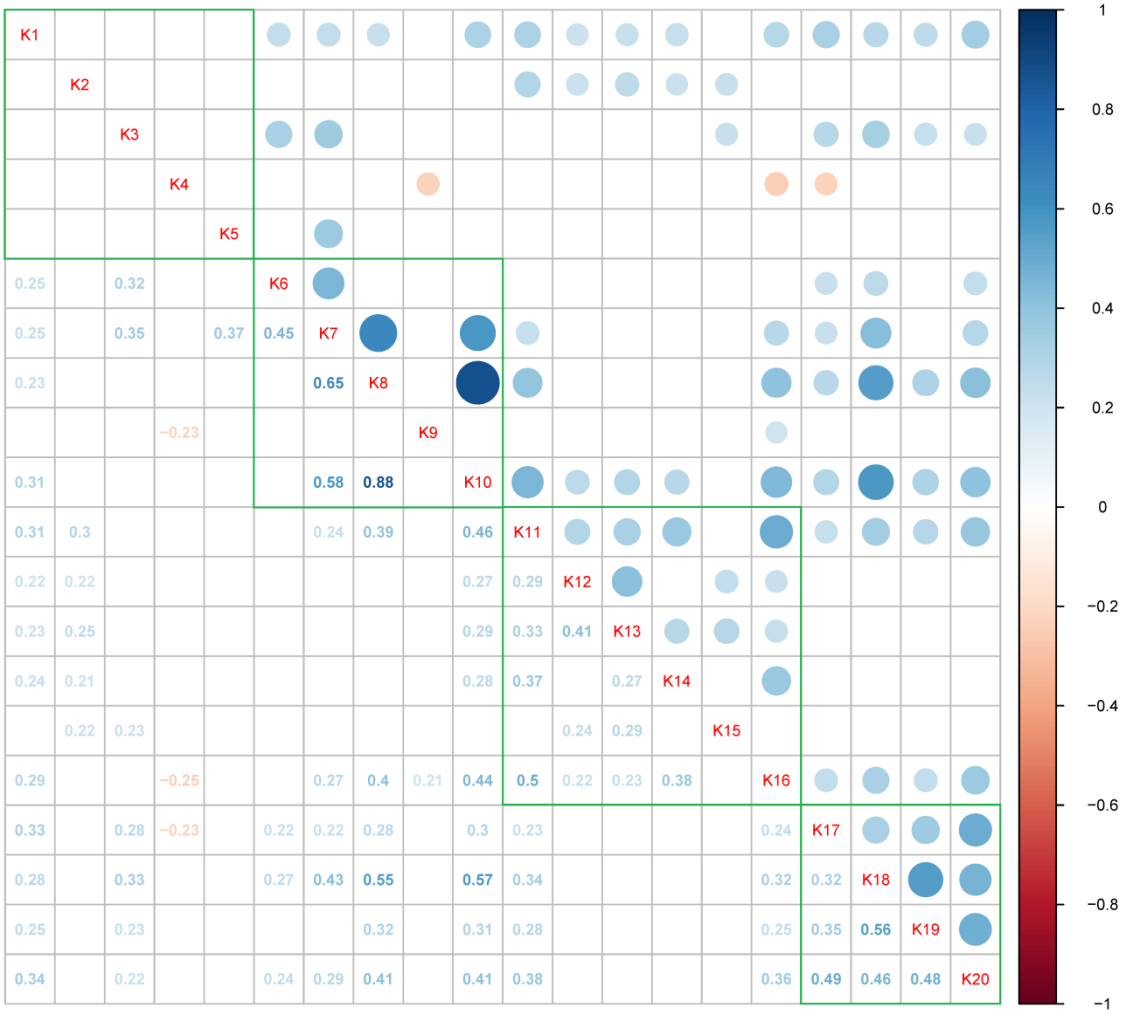
Source: European Commission, Joint Research Centre (JRC), 2017

In the case of the KFSIs, the data cannot be interpreted by linear regression, and log transforms do not help either. To understand the data structure, the Kendall-Tau rank correlation is used: this is essentially a measure of how similar the ranking is between pairs of KFSIs. An alternative measure would be to use the Spearman rank correlation, which is also a measure of rank similarity. Figure 6-G shows the heat map of Kendall-Tau correlations: evidently the large majority of KFSIs are positively correlated, although they are not in general strongly correlated.

Only a few strong correlations (above 0.6) are present: between KFSI-8 ("Country-by-Country Reporting") and KFSI-10 ("Legal Entity Identifier") or KFSI-7 ("Public Company Accounts").

Furthermore, many variables also do not have statistically significant correlations (using p=0.01 as a threshold for statistical significance). In particular, KFSI-5 ("Limited Partnership Transparency") and KFSI-9 ("Corporate Tax Disclosure") have no statistically significant association to any of the other KFSIs, except for a moderate to low association to one KFSI.

Figure 6-G: Kendall-tau rank correlations between KFSIs.



Note: Size and colour of circles relates to strength of correlation according to the colour scale on the right. Correlations that are not significant at the 1% level are left blank. Green boxes show conceptual grouping of KFSIs.

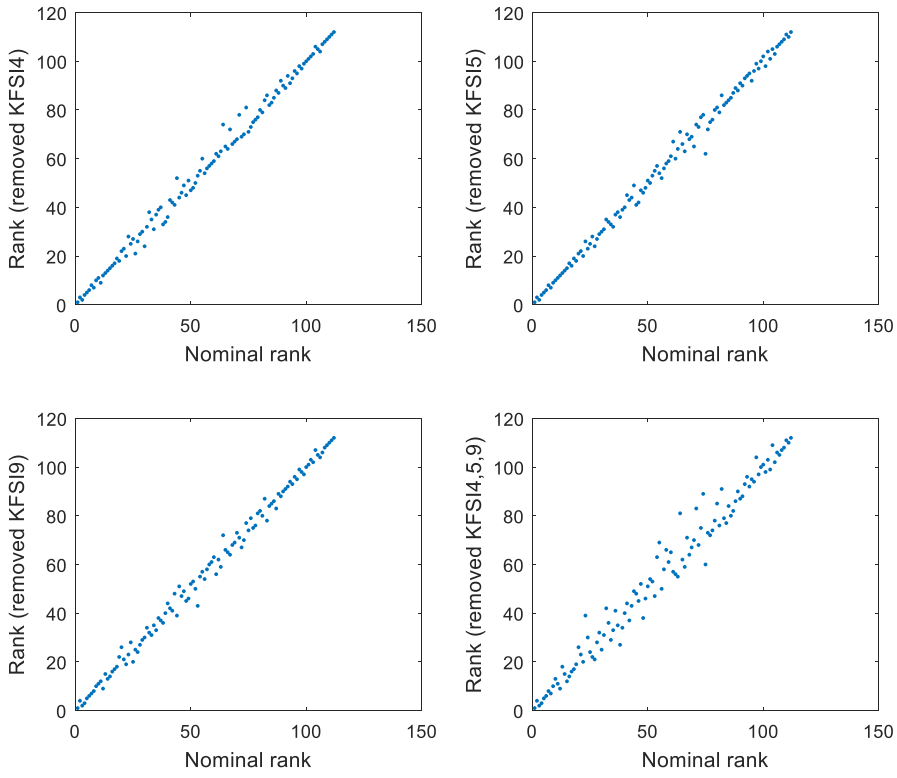
Source: European Commission, Joint Research Centre (JRC), 2017

There are some undesirable negative correlations, all of which are associated with KFSI-4 ("Other Wealth Ownership") and which help to flag possible conceptual issues with this indicator. In this case, KFSI-4 is negatively correlated with 3 of the 19 other KFSIs and it is not

statistically related to any of the remaining 16 KFSIs. This might be a concern, because (purely from a statistical point of view) "Other Wealth Ownerhsip" seems to be measuring a type of secrecy that goes against the trend of other indicators and is entirely different from the secrecy aspects captured in the framework as a whole.

A first recommendation from this type of analysis is therefore to review KFSI-4, KFSI-5 and KFSI-9 to make sure that they indeed have an added value in the framework. From a purely statistical point of view, it is possible to check the rank changes which occur when removing these KFSIs. Figure 6-H shows the rank plots which result when removing each of these indicators one by one, and all three at once. Visually, there is the greatest rank change when removing all three indicators simultaneously, whereas when each is removed individually, the impact is relatively modest. The average absolute rank shifts are, respectively, 1.8, 1.5, 1.7 and 3.7 for removing KFSI 4, 5, 9, and all three simultaneously. In all four cases, Portugal is the country with the greatest drop in rank as a result (-10, -7, -8 and -17 places respectively), whereas the countries that gain the most ranks are Turkey, Tanzania, Belgium and Tanzania, with increases of 6, 13, 10 and 15 places respectively. Given the effort that is put into collecting KFSI data, it may be useful to consider the added value of each KFSI for future versions of the FSI.

Figure 6-H: Scatter plots of nominal FSI ranks against ranks obtained after removing KFSI4 (top left), KFSI5 (top right), KFSI9 (bottom left) and KFSIs 4, 5 and 9 together (bottom right).



Source: European Commission, Joint Research Centre (JRC), 2017

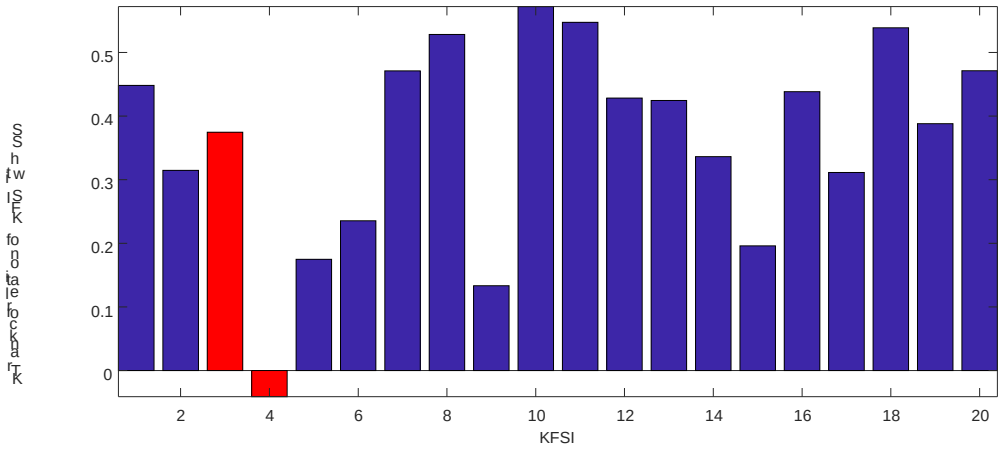
Another type of analysis looks into the cross-correlations between indicators belonging to different conceptual groupings. The expectation here is that the KFSIs should be in general more related to the indicators in their own group than to any of the other three groups. Indeed the six indicators (KFSI-11 to KFSI-16) capturing *Integrity of tax and financial regulation* and the four indicators (KFSI-16 to KFSI-19) summarising *International standards and cooperation*

fit well together. This is not the case for the other two conceptual groups on *Ownership Registration* and *Legal entity transparency*. The statistical analysis suggests that indicators from the two groups have a greater statistical association with the indicators under the other two groups. More specifically, KFSI-2 ("Trust and Foundations Register") and KFSI-10 ("Legal Entity Identifier") fit well together with the six indicators under *Intergrity of tax and financial regulation*. Instead, KFSI-1 ("Bank Secrecy"), KFSI-3 ("Recorded Company Ownership"), KFSI-6 ("Public Company Ownership"), KFSI-7 ("Public Company Accounts") and KFSI-8 ("Country-by-Country Reporting") fit well together with the four indicators under the *International standards and cooperation*.

Hence, a second recommendation to the FSI developing team is to review the grouping of indicators and eventually consider two groups instead of four, if this latter can be justified on conceptual grounds on top of the statistical findings. This adjustment should be seen more as a refinement. It is not expected to have a noteworthy impact on the overall secrecy scores that are calculated as the simple average of the 20 KFSIs (without taking into account any grouping). Yet, this fine-tuning is expected to add to the coherence of the framework and to building sounder narratives based on the two conceptual groupings that may be renamed to encompass elements from the additional indicators.

Following on from this analysis, Figure 6-H shows the correlation (again using the Kendall Tau rank measure because of the discrete nature of the data) between each KFSI and the aggregated Secrecy Score (arithmetic average of the 20 KFSIs as per the FSI methodology). Here, the effect of the negative correlations of KFSI-4 is visible, because the rank of KFSI-4 is negatively correlated with the SS. However, this is not statistically significant, so does not provide any evidence of a problem in statistical consistency. The remaining KFSIs are all positively correlated with the SS, which means that higher values of the KFSIs mean higher values of the SS. There is some variation in the degree of correlation, but the fact that they are all positive (apart from KFSI-4) is reassuring.

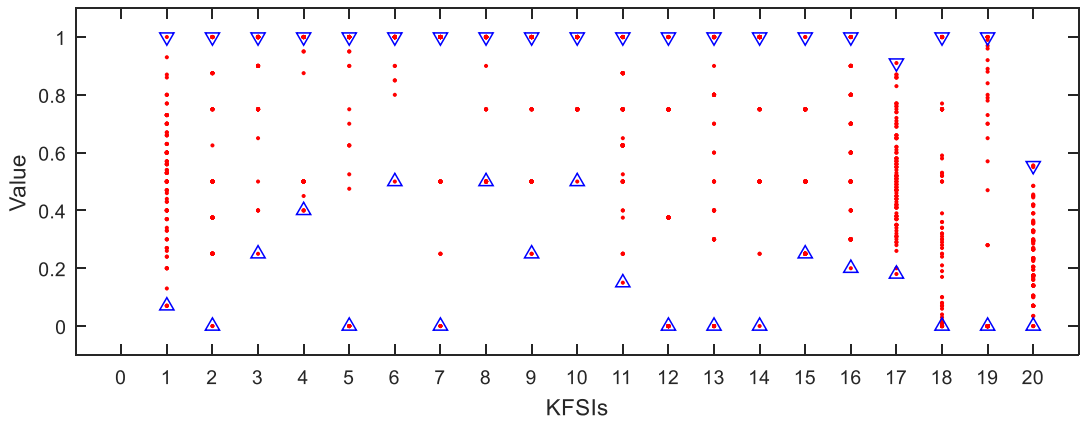
Figure 6-H: Kendall-Tau rank correlation of KFSIs with SS. Correlations that are not statistically significant are marked in red.



Source: European Commission, Joint Research Centre (JRC), 2017

The variation of the correlations of the KFSIs with the secrecy score may be in part attributed to the differing ranges of variation between the KFSIs: see the distributions of the KFSIs visualised in Figure 6-I. While most KFSIs have maximum values of 1, the minimum values vary substantially. However, if the distributions of each KFSI are transformed so that the minimum values are all 0, and the maximum values are all 1 (i.e. the min-max transformation), the resulting correlations are almost identical to Figure 6-H.

Figure 6-I: Visualisation of the distributions of the KFSIs



Note: Red points are data points and blue triangles represent maximum and minimum values.

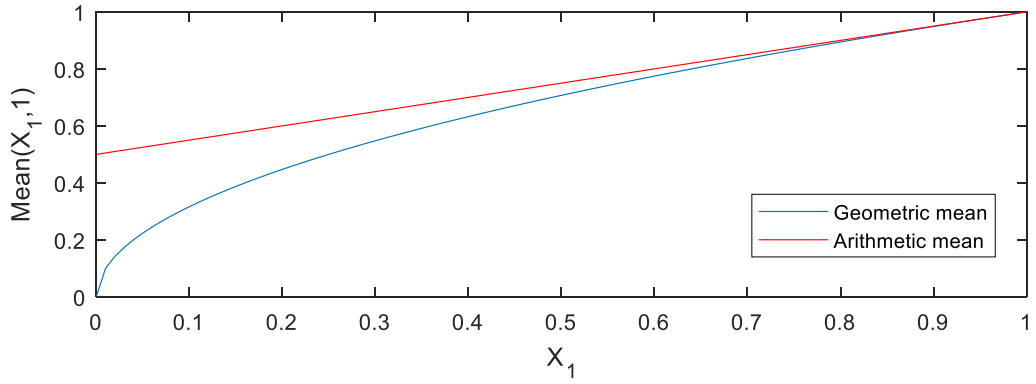
Source: European Commission, Joint Research Centre (JRC), 2017

In general, it is good practice to scale indicators to the same scale to ensure that they contribute more or less equally to the composite. In Section 6.5 the uncertainty analysis will include this normalisation as an alternative assumption.

A second reason – the most plausible in the FSI case – for the variation of the correlations of the KFSIs with the overall secrecy score are the correlations among the KFSIs. In fact, KFSI-10 ("Legal Entity Identifier"), KFSI-11 ("Tax Administration Capacity"), KFSI-18 ("Automatic Information Exchange") and KFSI-8 ("Country-by-Country Reporting") have the highest average bivariate correlations with the indicators in the framework, and consequently, they are more influential in the secrecy scores.

A final consideration that might be relevant to the SS is the way that the KFSIs are aggregated to give the SS. The current method is to take the arithmetic mean: this is a *compensatory* statistic that allows poor values in one KFSI to be "compensated" by good values in another, i.e. two indicators with values 0.1 and 0.9 would have an average score of 0.5. An alternative aggregation is to use the geometric mean, which compensates much less—in fact, the geometric mean of 0.1 and 0.9 is 0.3. To illustrate this relationship a little further, Figure 6-J shows the arithmetic and geometric means of two values: X_1, which varies between 0 and 1, and a second value, which is always 1. This illustrates the conceptual difference: when one of the two values is low, the geometric mean is lower than the arithmetic mean. To interpret this in the context of indicators, the geometric mean requires that all indicators have high values to give a high geometric mean. For example, if one were to try to measure quality of life, one might reason that even if a country has a high GDP, it is meaningless if there is no personal freedom. This is not to say that the geometric mean is necessarily the best choice for aggregating the KFSIs, but only to mention that it is an alternative option, depending on the intended meaning of the SS. If the geometric mean were to be considered, the scale of the KFSIs would also have to be adjusted to avoid all zeros and adjusted as the higher the better (less secrecy). After calculating the geometric mean, the overall scores would then be brought to the intended direction as the higher the worse (more secrecy) and subsequently aggregated with the global scale weight.

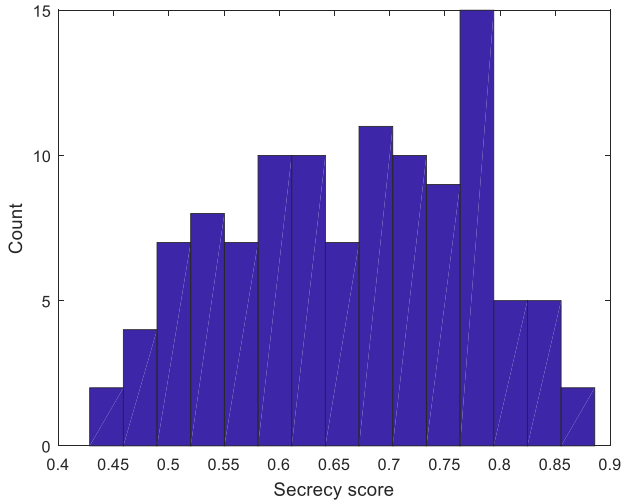
Figure 6-J: Comparison of arithmetic and geometric means of the set {X1, 1}



Source: European Commission, Joint Research Centre (JRC), 2017

As a final point of investigation, Figure 6-K shows the distribution of secrecy scores as a histogram. The distribution is roughly normal, which means that no treatment of outliers or skew/kurtosis is necessary.

Figure 6-K: Histogram of Secrecy Scores, using original arithmetic mean aggregation of KFSIs



Source: European Commission, Joint Research Centre (JRC), 2017

6.3 Transformation and Aggregation

In order to arrive at a single score and rank for each jurisdiction, it is necessary to aggregate the secrecy scores and the global scale weights. Two questions that arise in are therefore:

1. Should the GSW and/or SS be transformed in any way?
2. How should the GSW and SS be aggregated together?

Both decisions will have a significant impact on the final results. As with most decisions in building composite indicators, the choices should be made given a full understanding of the implications of alternative methodologies, and how this relates to the concepts that are meant to be conveyed.

The FSI 2018 uses the cube/cube-root aggregation formula given in Equation 1. This section first explores the implications of this formula, and then tests a few possible alternative transforms which have different statistical properties. These alternatives were carefully considered by the TJN in the preparation for the FSI 2018, but on balance the original formula was retained. The reasoning is summarised in the concluding remarks of this section and given in more detail in the previous chapter.

As with many aspects of composite indicators, there is no objectively "right" way of aggregating variables together. Instead, it is important to understand the statistical properties of the aggregation, and balance them against conceptual considerations. The fundamental aims of the index must therefore be accounted for: in the case of the FSI, the core objective is to "measure a jurisdiction's contribution to global financial secrecy in a way that highlights harmful secrecy regulations". On the other hand, the FSI should ideally reflect a balanced contribution from both the SS and the GSW: this can be analysed statistically.

A recurring them in investigating aggregations for the FSI has been that "statistical balance" (which is based on the dependence of the FSI on the GSW and SS) comes at the expense of distorting measured reality. Because the GSW is a highly skewed variable, it is very difficult to ensure that it has an equal contribution with the SS to the FSI, without applying strong transformations which significantly distort the fact that the GSW of jurisdictions consists of some few "giants", with the large majority having relatively very small values.

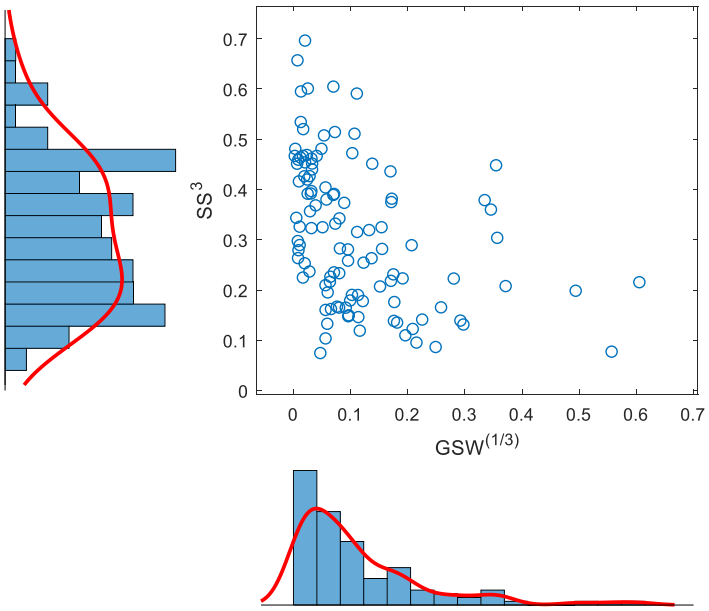
A final important issue is that, if an alternative formula is to be used, it implies discontinuity with previous years and risks sending mixed messages. These considerations are all taken into account in the following analyses.

6.3.1 FSI 2018

The FSI 2018 (and previous versions of the FSI) use the formula shown in Equation 1, in which the secrecy score is cubed, and the global scale weight is cube-rooted. The two quantities are then multiplied together. The reasoning for the cube/cube-root transformations is that it largely removes the skew in the distribution of the GSW, and results in similar percentile ratios between the GSW and SS, as opposed to the untransformed variables which are substantially different. As noted by the TJN, if the variables were not transformed, they would be extremely unbalanced in terms of their correlations with the final FSI scores and ranks.

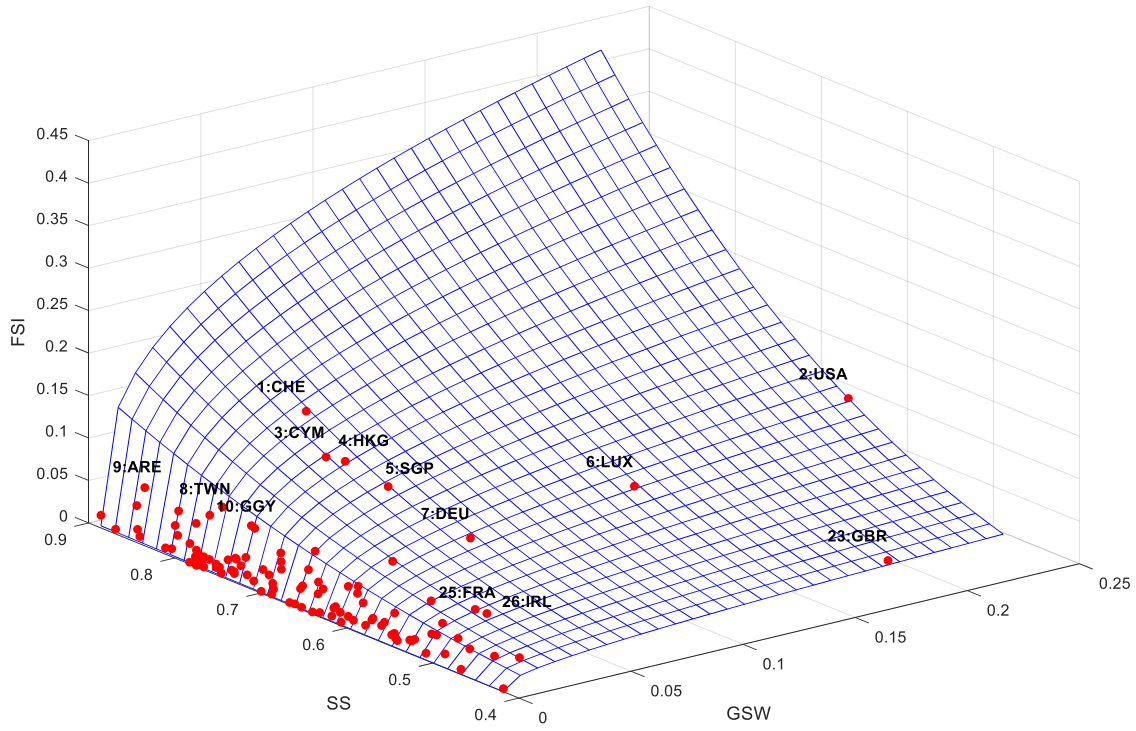
The effect of the cube/cube-root transformations is shown in Figure 6-L—evidently the result is a roughly normal distribution for the SS (although still slightly skewed to the left) and a still quite skewed distribution for the GSW.

Figure 6-L: Scatterplot and marginal distributions of SS^3 and $GSW^{(1/3)}$



Source: European Commission, Joint Research Centre (JRC), 2017

Figure 6-M: Surface plot of FSI against GSW and SS.



Note: Countries labelled with highest 10 FSI scores, and highest 10 GSW scores. Numbers indicate FSI ranks.

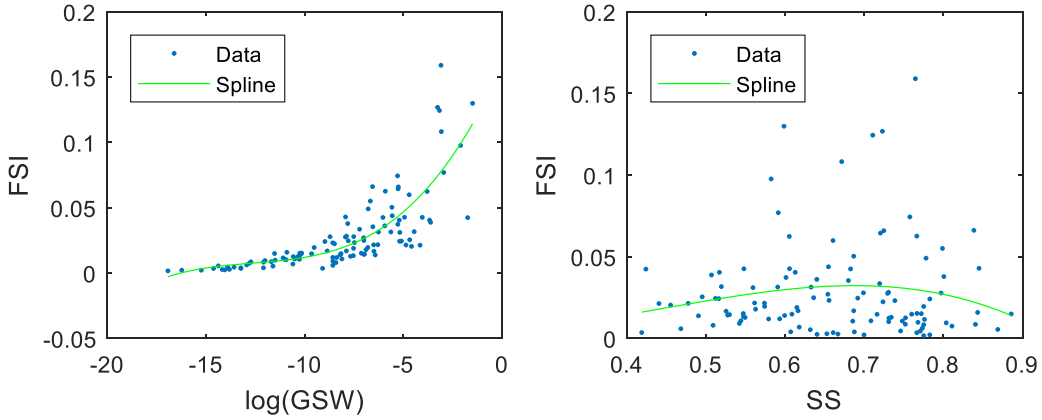
Source: European Commission, Joint Research Centre (JRC), 2017

The cube/cube-root transformations applied to the SS and GSW are *nonlinear*, which can mean that the relationship between the FSI and its constituents is not immediately obvious. In order to better understand this relationship, Figure 6-M shows FSI plotted against the SS and GSW, with jurisdictions plotted as red dots. The blue surface interpolates between the points and allows us to see the "functional form" of the FSI. The countries with the ten highest FSI scores are labelled, as well as those with the ten highest GSWs.

This plot reveals a number of features. First, the relationship of the FSI with the GSW and SS is nonlinear and slightly complicated. The gradient of the surface varies quite substantially over the space of countries—for example, the gradient is quite high in corner of high SS and low GSW, meaning that in this area, a small increase in GSW results in a very sharp increase in the FSI. The implication is that countries that have a similar SS can have markedly different FSIs as a result in relatively small differences in GSW. On the other hand, countries with low SS and low GSW will only experience a small increase in FSI if the GSW were to be increased. Overall, for countries with small GSW, their FSI is driven much more by their GSWs than by their SSs. The opposite is true for countries with large GSW: here countries are differentiated mainly on their secrecy scores.

To investigate the relative influence of the GSW and SS on the FSI, two measures are used. Due to the nonlinearity of the FSI with respect to the GSW and SS, the nonlinear *correlation ratio* is used: this is a nonlinear extension of the correlation coefficient, which measures the dependence of two variables on one another. Further, the Kendall-Tau rank correlation measure is used to compare the similarity of the FSI rank with its constituents.

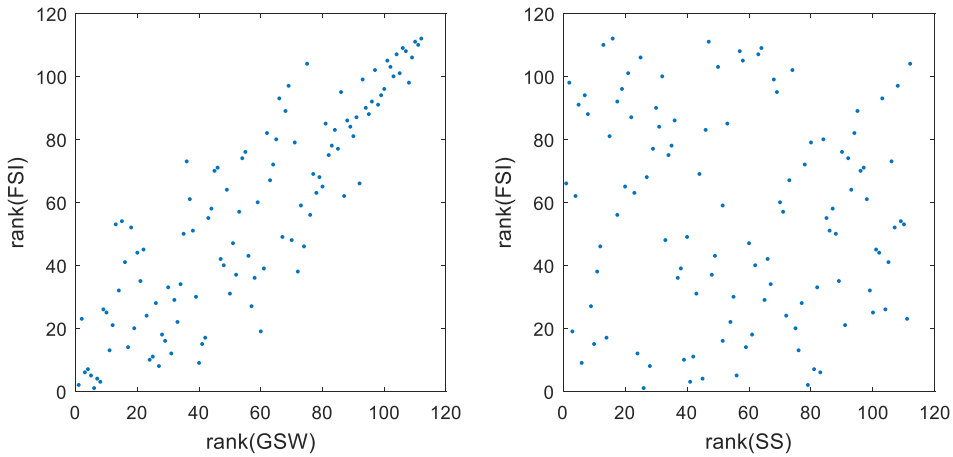
Figure 6-N: Scatter plots of FSI (original version) scores against log(GSW) and SS (used to calculated correlation ratio)



Source: European Commission, Joint Research Centre (JRC), 2017

Figure 6-N shows the nonlinear regression fits used to calculate the correlation ratio: here there is a clear visual indication that the GSW dominates the FSI: it shows a strong nonlinear relationship with the FSI. The SS plot is quite diffuse and shows that it only has a weak effect on the FSI.

Figure 6-O: Rank plots of FSI (original version) against ranks of GSW and SS



Source: European Commission, Joint Research Centre (JRC), 2017

A further useful visualisation is to see how the FSI ranks compare to the GSW ranks and SS ranks (Figure 6-O). This also gives a measure of the strength of the signal between each variable and the composite. Again, we see that the GSW has a very dominant contribution to the FSI compared to the SS. This is confirmed numerically by the values of the correlation ratio (measuring the [nonlinear] similarity in scores between FSI and constituents), and the rank correlation (similarity of ranks) in Table 6-B: from a statistical point of view, the GSW and SS provide quite unequal contributions to the FSI. This might seem like a contradiction of sorts when re-examining Figure 6-M, because the effect of SS is clearly visible in the shape of the surface. However the analysis of correlation focuses on the *average* association of the *sample points*, not the surface of the function itself. From Figure 6-M we see what almost all of the points (jurisdictions) are located in the low-GSW area where the effect (partial gradient of the FSI) of the GSW is very dominant over the SS. This is why, on average, the SS has a much lesser influence on the FSI than the GSW.

Nevertheless, the secrecy scores are responsible for putting the spotlight on some jurisdictions that may have gone unnoticed had only the global scale weight been considered. More specifically, the following six jurisdictions are classified in top 30 positions of the FSI owing to their high secrecy scores: United Arab Emirates (Dubai), Panama, Thailand, Bahrain, Bahamas and Kenya.

Jurisdiction	Rank FSI	Rank Secrecy Score (SS)	Rank Global Scale Weight
United Arab Emirates (Dubai)	9	6	40
Panama	12	24	31
Thailand	15	10	41
Bahrain	17	14	42
Bahamas	19	3	60
Kenya	27	9	57

For these jurisdictions, except for Bahamas and Thailand, the TJN provides special narrative reports exploring the history and politics of their offshore sectors on their dedicated website.

Table 6-B: Correlation ratio and Kendall-Tau rank correlation of FSI (original version) with GSW and SS

Measure	Global Scale Weight (GSW)	Secrecy Score (SS)
Correlation ratio	0.64	0.02
KT rank correlation	0.69	-0.05

Source: European Commission, Joint Research Centre (JRC), 2017

In summary, the main advantages of the original FSI methodology are as follows:

- Partial treatment of skew of GSW
- Continuing using this methodology would cause minimal upheaval because ranks of jurisdictions would change very little (it is following the status quo)
- To some extent, it strikes a balance between the pursuit of statistical balance against the distortion of the GSW distribution

On the other hand, the disadvantages are arguably as follows:

- By transforming GSW, there is a departure from the measurable reality
- From the perspective of correlation, the influence of the SS is much less than that of the GSW, when averaged over all jurisdictions.
- By transforming both SS and GSW, the resulting measure risks being difficult to interpet

6.3.2 Different aggregations

Here three alternative aggregations are tested (called Alt 1, Alt 2 and Alt 3). The aim is to investigate the statistical properties of different approaches to aggregate the secrecy scores and the global scale weights. These alternatives were presented in greater detail in correspondence between the JRC and the TJN, and after extensive discussion, the original FSI formula was retained. Here, a summary of the properties of these alternatives is given mainly

because it helps to shed more light on the methodology of the FSI, and serves as brief record of the alternative possibilities that were tested.

Alternative 1

Alt 1 tries to follow the logic that global financial secrecy is a quantity which is the sum of the contributions of each country (one of the stated aims of previous versions of the FSI). Following this logic, this would imply *not* transforming either of the variables, because the reality is distorted. However, since GSW and SS are on very different scales, it is at least necessary to rescale them onto the same interval. Here [0,10] is used for both variables. The FSI-Alt1 is therefore defined as follows:

$$\mathrm{FSI}_{\mathrm{Alt1},i} = \mathrm{GSW}_i . \mathrm{SS}_i; \quad \mathrm{GSW}, \mathrm{SS} \in [0,10] \tag{4}$$

As already discussed, while the secrecy scores have a fairly normal distribution, the global scale weights are very heavily skewed. This is what led the TJN to originally consider transforming the GSW.

Table 6-C shows the nonlinear correlation ratio and the Kendall-Tau rank correlation of the SS and GSW with the FSI-Alt1. The effects are clearly very unbalanced, and the SS even has a slight negative rank correlation with the overall FSI. However, the objective of the FSI-Alt1 is not to try to balance the GSW and SS, but to attempt to treat the FSI as a "physical quantity" that can be measured and added together. From this perspective alone it is arguably the most suitable.

Table 6-C: Correlation ratio and Kendall-Tau rank correlation of FSI (Alt.1) with GSW and SS

Measure	Global Scale Weight (GSW)	Secrecy Scores (SS)
Correlation ratio	0.66	0.03
KT rank correlation	0.88	-0.24

Source: European Commission, Joint Research Centre (JRC), 2017

Since the FSI-Alt1 is a departure from the current FSI methodology, Figure 6-P shows the rank of the FSI-Alt1 plotted against the FSI using the original methodology. While there is some scatter, the "upheaval" is not very huge and rank shifts are fairly modest. The rank correlation of this plot is 0.81: if we subtract this from 1, we can get a loose measure of the upheaval[436] of this option: 19%.

To summarise, the main advantages of this approach would be:

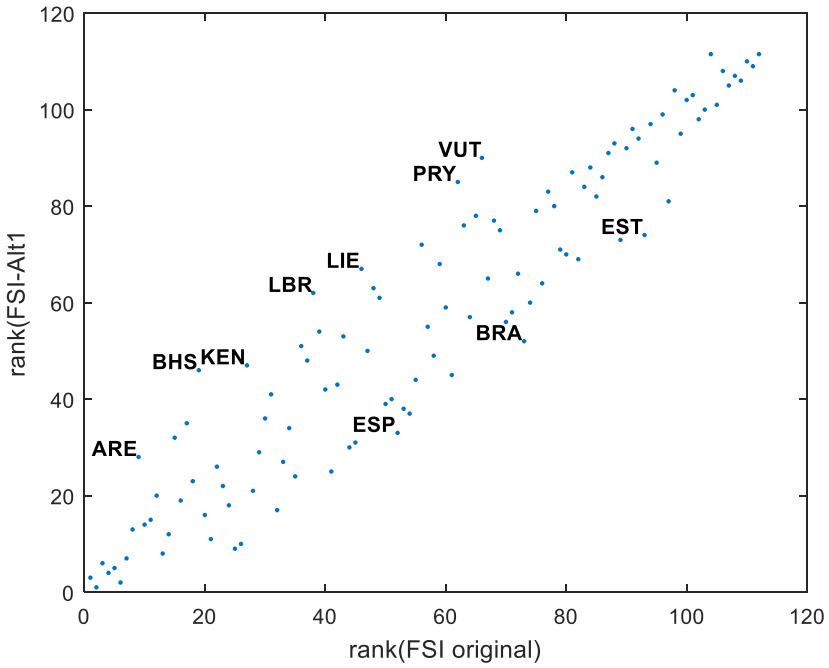
- No transformation means a more faithful representation of reality: jurisdictions with a huge financial sector are held more strongly to account because even a small amount of secrecy is applied to a large volume of financial activity.
- It is arguably the easiest formula to interpret (because no nonlinear transformations are involved)
- It has only a modest upheaval score

[436] An upheaval value of 0% would mean no change in rank for any country (compared to original methodology). A value of 100% would mean that the ranks of all countries change significantly.

The disadvantages are:

- The GSW and SS are very unbalanced: the SS ranks have effectively no relation to the FSI ranks.

Figure 6-P: Plot of original FSI ranks against FSI-Alt1 ranks, with top ten greatest rank shifts labelled



Source: European Commission, Joint Research Centre (JRC), 2017

Alt 2: Log transform

As observed in previous sections, the GSW data is a typical log-normal distribution. By taking the log of a log-normal variable, the variable becomes normal. So, if the objective is to correct the skew of the GSW, the log transform is the best choice. Note that both the log, cube and cube root transformations are *monotonic*, which means that they will not change the ranks of the GSW, but only the scores. However the choice of transformation will inevitably change the FSI ranks.

The FSI-Alt2 is therefore constructed by taking the log of the GSW, and then scaling both log(GSW) and SS to the [0,10] interval:

$$\mathrm{FSI}_{\mathrm{Alt2},i} = \log(\mathrm{GSW}_i).\,\mathrm{SS}_i; \quad \log(\mathrm{GSW}), \mathrm{SS} \in [0{,}10] \tag{5}$$

By taking the log transformation, the GSW is very significant for small jurisdictions, such that small changes in GSW will have a large change in FSI. For the large-GSW jurisdictions, the FSI scores are much more separated from one another by the secrecy scores. Clearly the log transform represents a departure from the reality for the global scale weight. However the secrecy score is arguably easier to interpret than the original FSI methodology because it is untransformed (i.e. no *nonlinear* transformation is used, such as cube-root or log transform).

Table 6-D shows the correlation ratio and KT rank correlation: the GSW and SS are indeed much more balanced than in the original FSI and FSI-Alt1. Yet, now the SS is actually more

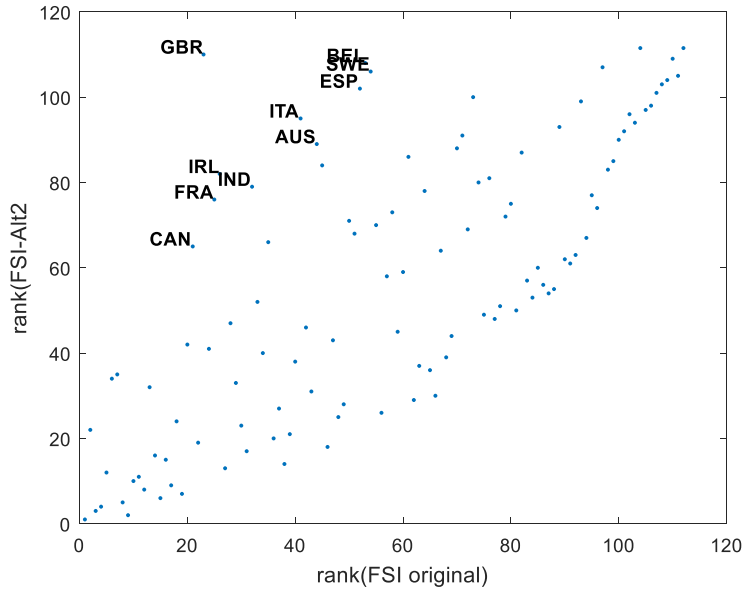
influential in the index. This could be corrected by further adjustments, however the balance of an easy-to-communicate formula must be kept in mind. Further transformations to match percentiles would probably over-complicate the message.

Table 6-D: Correlation ratio and Kendall-Tau rank correlation of FSI (Alt.2) with GSW and SS

Measure	Global Scale Weight (GSW)	Secrecy Score (SS)
Correlation ratio	0.19	0.34
KT rank correlation	0.25	0.39

Source: European Commission, Joint Research Centre (JRC), 2017

Figure 6-Q: Plot of original FSI ranks against FSI-Alt2 ranks, with top ten greatest rank shifts labelled



Source: European Commission, Joint Research Centre (JRC), 2017

The rank plot comparing FSI-Alt2 with the original FSI is shown in Figure 6-Q. Evidently there is a more significant departure from the original FSI ranking than with FSI-Alt1. The biggest rank shifts are from jurisdictions with fairly large GSW scores such as the UK. The rank correlation of this plot is 0.56, which gives an "upheaval score" of 44%.

To summarise, the main advantages are:

- The GSW distribution is properly "corrected" (if that is desirable).
- The GSW and SS are much more balanced compared to either the original FSI or FSI-Alt1.
- Although it involves a log transform, it is still reasonably easy to communicate the formula.

The disadvantages are:

- The log transform strongly distorts the reality of the size of the financial sectors of the jurisdictions. Therefore there is no longer the possibility to interpret the scores as shares of offshore financial services activity in the global total.

- The upheaval score is quite large and some very visible jurisdictions will experience large changes in the overall index rank (e.g. UK).

Alt 3: Log transform and arithmetic average

A final alternative that is studied here is simply to take the arithmetic mean of the GSW and SS (rather than the product). In this case it makes sense to also transform the GSW to correct for skew, and seek to balance the influence of the two components in the index. The FSI-Alt3 is essentially the same as the FSI-Alt2, but uses the arithmetic mean instead of the product:

$$\mathrm{FSI}_{\mathrm{Alt3},i} = (\log(\mathrm{GSW}_i) + \mathrm{SS}_i)/2; \quad \log(\mathrm{GSW}), \mathrm{SS} \in [0,10] \tag{6}$$

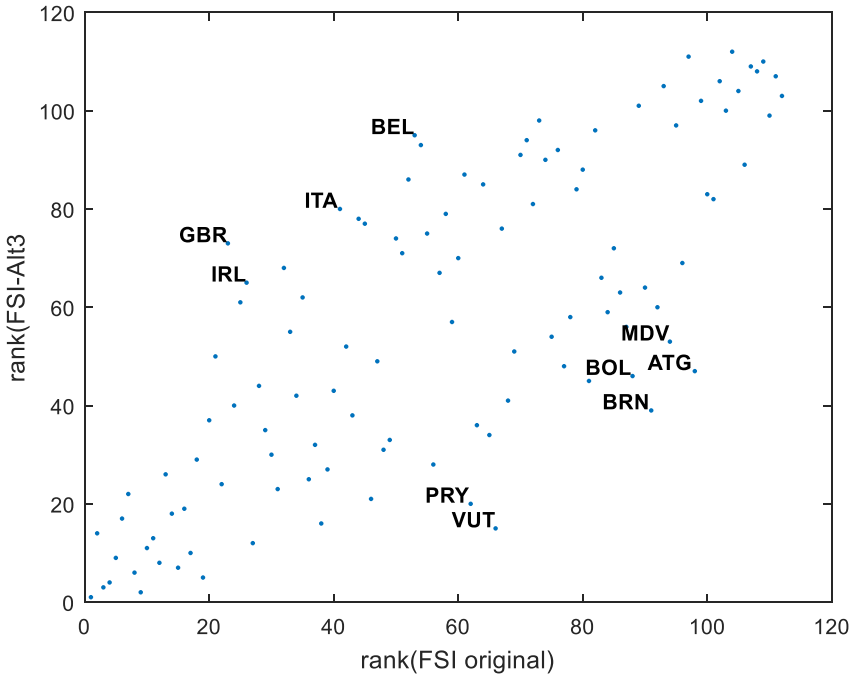
Table 6-E shows the correlation ratio and rank correlation values. This shows that in terms of statistical dependence, the GSW and SS are actually quite balanced. They could be further balanced by assigning weights, but again this does not seem to be worth complicating the formula.

Table 6-E: Correlation ratio and Kendall-Tau rank correlation of FSI (Alt.3) with GSW and SS

Measure	Global Scale Weight (GSW)	Secrecy Scores (SS)
Correlation ratio	0.21	0.30
KT rank correlation	0.28	0.37

Source: European Commission, Joint Research Centre (JRC), 2017

Figure 6-R: Plot of original FSI ranks against FSI-Alt3 ranks, with top ten greatest rank shifts labelled



Source: European Commission, Joint Research Centre (JRC), 2017

Finally, the plot of rank shifts compared to the original FSI is shown in Figure 6-R. While there are significant changes in ranks, the largest shifts in rank are less than in the FSI-Alt2, and the big ranks shifts are in both directions, so it seems like a more "balanced" change in methodology. The rank correlation of 0.58 gives an upheaval score of 42%.

The main advantages here are:

- It is reasonably easy to explain: first we take the log of GSW to correct the skew. Then we scale each variable on a [0,10] scale and take the average.
- The fact that it is linear in SS increases its interpretability.
- The GSW and SS are statistically fairly well-balanced.

The disadvantages are:

- It is the furthest away from the idea of measuring a share of a global total of financial secrecy.
- It is quite a significant departure from the existing FSI methodology, both in terms of the transformation and the aggregation.

6.3.3 Summary

No particular transformation is recommended here. Instead the approach is to show the properties of a number of differing alternatives. These are summarised in Table 6-F.

Table 6-F: Summary of properties of alternative FSI aggregation and transformation approaches.

	Original FSI	FSI-Alt1	FSI-Alt2	FSI-Alt3
Formula	$GSW^{(1/3)}.SS^3$	GSW.SS; GSW,SS ϵ [0,10]	log(GSW).SS; log(GSW), SS ϵ [0,10]	(log(GSW)+SS)/2; log(GSW), SS ϵ [0,10]
Statistical balance	Low	Very low	Fair/good	Good
Simplicity	Fair	Good	Fair/good	Fair/good
Upheaval	0%	19%	43%	41%

Source: European Commission, Joint Research Centre (JRC), 2017

To conclude, there are two reasons to continue with the present methodology for combining the global scale weight with the secrecy score. The first is that the cube/cube-root aggregation, in some sense, is a compromise between statistical balance (in terms of correlation) and distorting the distribution of the GSW. The second reason is simply to minimise disruption. On the other hand, if one were to pursue the goal of interpreting the FSI as a summable quantity of two measurable variables, FSI-Alt1 seems the best option (because no nonlinear transformations are used), but comes at the price of a very heavy imbalance. If one were purely interested in balancing the correlations of the GSW and SS, FSI-Alt2 and Alt3 are both alternatives with better statistical properties, and are arguably simpler than the original FSI in that only one variable is transformed. FSI-Alt3 has the best statistical balance and also implies less upheaval than FSI-Alt2. Of course, all of the statistical considerations

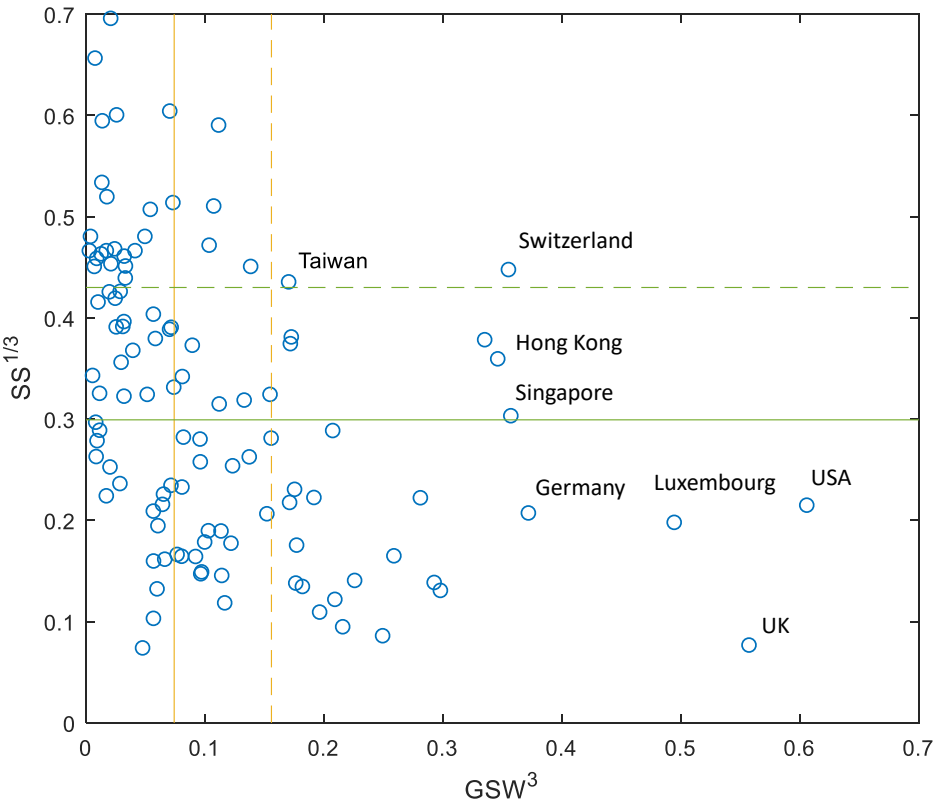
presented here have to be balanced against the conceptual considerations, and this is a matter left to the developers.

6.4 Communicating the FSI results

The challenges in identifying a suitable formula to combine the two FSI components into an overall index stem from the negative association between the two FSI components. As shown in Figure 6-L above, the global scale weight has a significant and negative correlation to the secrecy score (Spearman rank correlation: -0.52). When it comes to monitoring financial secrecy aspects, this finding is reassuring. It suggests that on average jurisdictions with high global scale weights tend to be less secretive and vice-versa. Had juridictions with high global scale weight (high share of offshore financial activities) been the most secretive ones would have been particularly worrying. Yet, from a methodological point of view, this negative association between the two components poses the challenges discussed and illustrated above.

The JRC recommendation would be not to aggregate the GSW and SS into an overall index but to focus the communication of the FSI results using a plot of the two components (see below). Jurisdictions at the right hand side and top left side should be carefully monitored.

Figure 6-S: Scatter plot of SS^3 and $GSW^{1/3}$



Note: Solid lines represent median values of the transformed variables. Dashed lines represent 75th percentiles.

Source: European Commission, Joint Research Centre (JRC), 2017

Despite these suggestions, the aggregation into a single number of financial secrecy and an overall ranking thereafter would undoubtedly seem irresistible to some. An overall

classification may also better serve as advocacy tool by helping to put the spotlight on certain jurisdictions. To this end, the FSI developing team, alongside the FSI ranking could also provide special narrative reports for those jurisdictions that arrive at the top 30 positions of the financial secrecy classification, when alternative aggregation approaches are considered. Besides the top 30 ranked jurisdictions in the overall FSI, and staying with three approaches tested above, one should carefully monitor the offshore financial activities in the following jurisdictions:

Table 6-G: Jurisdicitons in the alternative top 30 other than those in FSI top 30.

FSI-Alt-1	FSI-Alt-2	FSI-Alt-3
Australia	Anguilla	Anguilla
Austria	Barbados	Bermuda
India	Bermuda	Liberia
Italy	Liberia	Liechtenstein
South Korea	Liechtenstein	Malaysia
	Malaysia	Marshall Islands
	Marshall Islands	Paraguay
	Mauritius	Vanuatu
	Paraguay	
	Saudi Arabia	
	Vanuatu	

Source: European Commission, Joint Research Centre (JRC), 2017

6.5 Robustness

An important part of a composite indicator audit is to check the effect of varying assumptions inside plausible ranges. In this section, the question of how to aggregate the GSW and SS is not included, because that is largely a conceptual decision and it has already been discussed in the previous section. Instead, three assumptions are tested that have plausible alternatives, and can be easily varied. The assumptions are as follows:

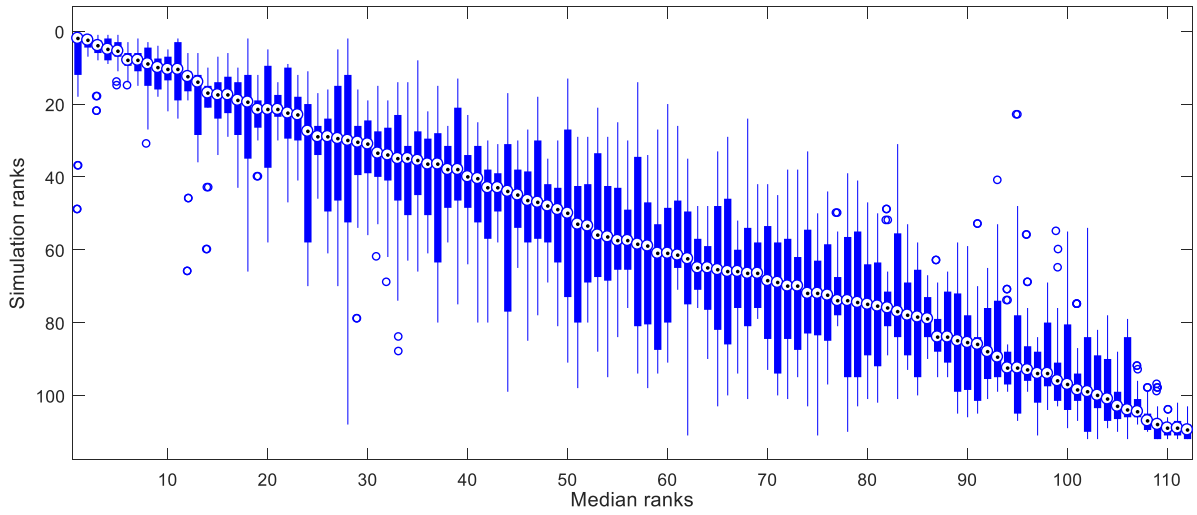
1. **Choice of GSW:** Although as shown, the GSW alternatives are strongly correlated, it is worth checking the effect of the plausible alternatives. Here, we take the GSW alternative that is least correlated with the default GSW-A, which is GSW-E, and use this as an alternative measure. This should serve as a plausible upper bound on the uncertainty in this respect.
2. **Normalisation of KFSIs:** as noted earlier in the report, the KFSIs are not all strictly scaled on to the same [0,1] scale, as is more common in composite indicator practice. Here the alternative assumption is tested where all KFSIs are scaled exactly to [0,1] before aggregating. There are a total of two alternative assumptions here, including the nominal.
3. **Aggregation of KFSIs:** an alternative method is tested where the median of the KFSIs is taken, rather than the mean. The geometric mean is not tested here. There are a total of two alternatives here, including the nominal.

The total number of alternative procedures tested for building the FSI is all combinations of the above alternatives, which is 8 x 2 x 2 = 32.

To first visualise the uncertainty of all these assumptions varied simultaneously, Figure shows the distributions of ranks over the 32 alternative simulations, for each jurisdiction, ordered by median rank. Essentially, the height of these boxes represents the uncertainty in their ranks, given the assumptions tested. We see that although there is some uncertainty, as expected, the magnitude of uncertainty is manageable and the ranking is relatively robust. Instead, for

those juridisctions that present wide confidence intervals, ranks should be analysed within those intervals instead of being taken at face value. Furthermore, one should be careful in attributing great significance to small changes in ranking. It is important to note however that the alternative assumptions tested here might be less plausible than the nominal assumptions, in which case the uncertainty here would be reduced. This argument is particularly relevant for the choice of the variable to capture the global scale weight, which is discussed next.

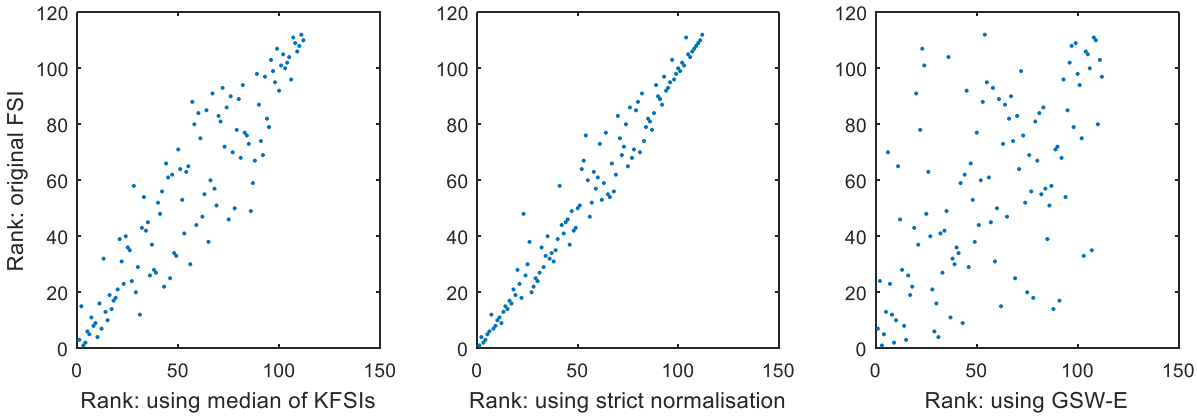
Figure 6-T: Box plots of rank distributions on Monte Carlo analysis, against median ranks observed



Source: European Commission, Joint Research Centre (JRC), 2017

To give a cursory idea of sensitivity, Figure shows rank plots of the original FSI rank against three alternatives: the first which is the same except the median of KFSIs is used; the second is the same but strict normalisation is used (in the sense defined previously); and the third where GSW-E ("Trade in goods") is used instead of GSW-A ("Trade in financial services"). With regard to the use of GSW-E, this was chosen because it has the lowest correlation with GSW-A (see Figure 6-E above), hence it represents the limit of uncertainty for this assumption.

Figure 6-U: Effect on ranks of individually varying assumptions.



Note: Right: original ranks vs. ranks using median of KFSIs; centre: original ranks vs ranks using strict normalisation of KFSIs; left: original ranks vs. ranks using GSW-E instead of GSW-A.

The plots show that the greatest scatter is found in the alternative using GSW-E, with the least found for the alternative normalisation method, and the use of median being somewhere in the middle. The fact that the choice of normalisation is fairly inconsequential is often observed

in composite indicator sensitivity analysis. The median produces a moderate amount of uncertainty as an alternative. The GSW-E produces a fairly substantial change in rankings, despite being well-correlated with GSW-A. The conclusion here is to re-examine the choice of GSW-A, and see whether the other GSWs can be discarded on conceptual grounds. If this is the case, the uncertainty will be reduced because the number of plausible alternative models will be reduced and because the choice of the GSW, of the three analysed here, is the most influential in the FSI development. The median is a viable alternative which could also be examined. If there is no good reason to neglect this alternative then it will have to be accepted as a plausible alternative which comes with associated uncertainty.

For full transparency and information, Table 6-I reports the FSI 2018 ranks together with the simulated 80% confidence intervals in order to better appreciate the robustness of the results to the choice of the variable for calculating the global scale weight, the normalisation and aggregation approaches for the twenty KFSIs. While in some cases these confidence intervals are wide, it should be noted that the inclusion of an alternative GSW in the analysis might be discarded on conceptual grounds (i.e. GSW-A is simply a better measure of international financial activity than GSW-E). In that case, the intervals would become narrower.

Table 6-I: FSI rankings with 10% and 90% percentiles in brackets (according to uncertainty analysis)

Country Code	Rank [interval]	Country Code	Rank [interval]	Country Code	Rank [interval]
CHE	1 [1, 6]	MHL	39 [26, 60]	SYC	77 [56, 83]
USA	2 [3, 22]	PHL	40 [25, 54]	GTM	78 [20, 77]
CYM	3 [1, 37]	ITA	41 [34, 77]	HRV	79 [69, 92]
HKG	4 [2, 18]	IMY	42 [43.7, 78]	GRC	80 [51, 93]
SGP	5 [5, 19]	UKR	43 [9, 41.6]	WSM	81 [67.4, 101]
LUX	6 [8.8, 68]	AUS	44 [29, 74]	MEX	82 [42, 95.2]
DEU	7 [9.1, 36]	NOR	45 [52, 99]	GIB	83 [41, 91]
TWN	8 [3, 15]	LIE	46 [23.7, 61]	CUW	84 [13, 80.5]
ARE	9 [1, 9]	ROM	47 [35, 91]	VEN	85 [29, 70.9]
GGY	10 [4, 17]	BRB	48 [33.7, 62]	VIR	86 [43.1, 103]
LBN	11 [16, 65]	MUS	49 [13, 54]	TCA	87 [52, 102]
PAN	12 [9.7, 46]	ZAF	50 [46, 91]	BOL	88 [18, 74]
JPN	13 [7, 44]	POL	51 [44, 70]	BGR	89 [74, 98]
NLD	14 [5, 13]	ESP	52 [51, 76]	BLZ	90 [72, 87]
THA	15 [2, 11]	BEL	53 [69.1, 98]	BRN	91 [20, 92]
VGB	16 [10, 29]	SWE	54 [62.4, 111]	MCO	92 [61.4, 104]
BHR	17 [14, 30]	LVA	55 [72.7, 103]	EST	93 [95.1, 107]
JEY	18 [15, 39]	AIA	56 [42, 93]	MDV	94 [24, 92.6]
BHS	19 [14, 43]	IDN	57 [22, 74.7]	GHA	95 [33, 97]
MLT	20 [34.3, 86]	NZL	58 [58.7, 100]	DMA	96 [85, 110]
CAN	21 [22.4, 47]	CRI	59 [31, 65]	LTU	97 [98.7, 112]
MAC	22 [20.8, 78]	CHL	60 [32, 73.3]	ATG	98 [35, 94.3]
GBR	23 [55, 106]	DNK	61 [53, 95.3]	MNE	99 [94, 109]
CYP	24 [38, 86]	PRY	62 [19, 42.8]	COK	100 [85.4, 108]
FRA	25 [25, 57]	KNA	63 [55, 110]	GRD	101 [71, 101]
IRL	26 [34.4, 71]	PRT	64 [57, 92]	MKD	102 [75, 103]
KEN	27 [17, 40]	PRI	65 [36.4, 90]	BWA	103 [44, 95.1]
CHN	28 [7, 41]	VUT	66 [48, 82]	SVN	104 [104, 111]
RUS	29 [11, 25.9]	URY	67 [67, 94]	ADO	105 [102.5, 110]
TUR	30 [16, 25.9]	ABW	68 [38, 79]	GMB	106 [64, 107]
MYS	31 [3, 18]	DOM	69 [21, 60]	TTO	107 [37, 103.2]
IND	32 [34.8, 56]	CZE	70 [79.1, 103]	NRU	108 [106.7, 111]
KOR	33 [18, 57]	FIN	71 [55, 79.6]	SMR	109 [105.4, 112]
ISR	34 [28, 54]	ISL	72 [78, 102]	LCA	110 [45, 104.5]

AUT	35 [35.7, 62]	**BRA**	73 [50, 91.2]	**VCT**	111 [84, 110]
BMU	36 [4, 104]	**HUN**	74 [52, 87]	**MSR**	112 [83, 112]
SAU	37 [11, 40]	**TZA**	75 [17, 73.6]		
LBR	38 [23, 54]	**SVK**	76 [69, 97]		

Note: Nominal ranks that fall outside the 10/90 percentile interval are highlighted red.

Source: European Commission, Joint Research Centre (JRC), 2017

6.6 Conclusions

The JRC statistical audit has delved around in the workings of the Financial Secrecy Index to assess the statistical properties of the data, and the methodology used in its construction. Overall the FSI is a well-constructed index into which a lot of thought has clearly been put. One of the greatest strengths is the amount of original research into financial secrecy, and the transparency and detail of all data associated with the index, as well as the extensive documentation on the methodology. The KFSIs framework is also coherent within two of the four groups, namely *Integrity of tax and financial regulation* (KFSI-11 to KFSI-16) and *International standards and cooperation* (KFSI-16 to KFSI-19). The FSI 2018 version is already an improved version of the FSI 2015.

Nevertheless, a few recommendations or points for discussion have been raised.

First, the issue of combining the global scale weight with the secrecy score into an overall FSI score is quite crucial and must be decided on a best understanding of the alternative possibilities and the implications. Of the alternatives aggregation methods tested here, three paths seem to be possible:

a. *Keep the aggregation as it is.* The main advantage of doing this is minimising disruption in terms of changes in ranking and communicating the new methodology. Moreover, it looks for a middle-ground between balance of correlations and distortion of measured data. At the same time, this aggregation formula puts the spotlight on some jurisdictions that may have gone unnoticed had only the global scale weight been considered. Six jurisdictions are classified in top 30 positions of the FSI owing to their high secrecy scores: United Arab Emirates (Dubai), Panama, Thailand, Bahrain, Bahamas and Kenya. On the other hand, the present aggregation results in imbalance between Global Scale Weight and Secrecy Score, such that the Secrecy Score is very uninfluential in the ranking of the FSI for the majority of the jurisdictions. Moreover, the transformation is somewhat hard to communicate.

b. *Use no transformations*: this path is in line with the philosophy of treating the FSI as a global quantity to which each jurisdiction contributes its own share. This is simple to communicate, however because of the huge skew of the GSW the global scale weight will dominate the secrecy score, such that the FSI will essentially be an alternative measure of global scale.

c. *Use log transformation*: Statistically, the log transformation is the "correct" way to normalise the GSW distribution. Then the SS and GSW can be aggregated either by multiplying or by taking the mean (after scaling). The advantage of this approach is that the GSW and SS are statistically well-balanced in the calculation of the overall index. On the other hand, it is a significant departure from the original methodology and heavily distorts the distribution of measured data.

Second, the inclusion of three KFSIs in the framework merits reconsideration. The KFSI-4 ("Other Wealth Ownership") is negatively correlated with three KFSIs and bears no statistical relevance to the remaining indicators. The KFSI-5 ("Limited Partnership Transparency") and KFSI-9 ("Corporate Tax Disclosure") bear no statistically significant association to any of the

other indicators in the framework. This is simply a statistical flag which calls for a second look at the KFSI framework.

Third, the framework could be simplified from four to two groups of indicators, if this can also be justified on conceptual grounds on top of the statistical findings. In this case, KFSI-2 ("Trust and Foundations Register") and KFSI-10 ("Legal Entity Identifier") fit well together with the six indicators under *Intergrity of tax and financial regulation*, whilst KFSI-1 ("Bank Secrecy"), KFSI-3 ("Recorded Company Ownership"), KFSI-6 ("Public Company Ownership"), KFSI-7 ("Public Company Accounts") and KFSI-8 ("Country-by-Country Reporting") fit well together with the four indicators under the *International standards and cooperation*. This adjustment should be seen more as a refinement, which is not expected to have a noteworthy impact on the overall secrecy scores.

Fourth, the aggregation of the KFSIs could also be done by the median, as opposed to the arithmetic mean, or indeed the geometric average. These should be checked as alternatives, based on conceptual reasoning. The normalisation could also strictly map each variable onto the [0,1] interval.

Fifth, the sensitivity analysis shows that the choice of the variable to calculate the global scale weight is the most significant uncertainty of the three tested (the other two being the normalisation and aggregation method for the twenty KFSIs). The GSW-alpha variable (trade in financial services) has the highest overall correlation to all other variables tested, and hence provides a the most suitable variable from a statistical viewpoint. However, as a composite measure, it is (arguably) conceptually further from the intended concept than GSW-A.

Sixth, the uncertainty analysis shows that the rankings are reasonably robust. Yet, for the majority of the jurisdictions the FSI ranks should be analysed within their expected confidence intervals instead of being taken at face value. The intervals presented here might be refined on further study (excluding or including plausible alternatives).

Finally, the JRC recommendation is not to aggregate the global scale weight and the secrecy score into an overall index, the reason being the negative correlation between the two FSI components. While this negative association is desirable from a conceptual point (jurisdictions with high global scale weight are on average less secretive and vice versa), it poses numerous methodological challenges. Hence, the JRC suggestion is that the communication of the FSI results should mainly be done using a plot of the two components, where jurisdictions at the right hand side and top left side should be carefully monitored. At the same time, arriving at a single number of financial secrecy would undoubtedly seem irresistible to some because an overall classification may better serve as advocacy tool by helping to put the spotlight on certain jurisdictions. To this end, it is recommended that the FSI developing team, alongside the FSI ranking could also provide special narrative reports for those jurisdictions that arrive at the top 30 positions of the financial secrecy classification, when alternative aggregation approaches are considered. In the FSI 2018, besides the top 30 FSI ranked jurisdictions, this would imply additional reporting and careful monitoring of the offshore financial activities in sixteen jurisdictions: Anguilla, Australia, Austria, Barbados, Bermuda, India, Italy, Liberia, Liechtenstein, Malaysia, Marshall Islands, Mauritius, Paraguay, Saudi Arabia, South Korea and Vanuatu.

Overall, the FSI 2018 offers an extremely detailed analysis of the concept of financial secrecy based on a wealth of original research. While the aggregation (or not) of the secrecy score and global scale weight still calls for further discussion and investigation, no objectively "right" solution exists, and the methodology of any composite indicator, as necessarily subjective instruments, is always open for debate Nevertheless, a number of recommendations are offered herein as food for thought order to help the Tax Justice Network to bring the FSI reach its full potential as a monitoring and benchmarking tool that can guide policy formulation.

References

Becker, W., Saisana, M., Paruolo, P., Vandecasteele, I. (2017) Weights and importance in composite indicators: Closing the gap, Ecological Indicators 80: 12–22.

Munda, G. 2008. *Social Multi-Criteria Evaluation for a Sustainable Economy.* Berlin Heidelberg: Springer-Verlag.

OECD/EC JRC (Organisation for Economic Co-operation and Development/European Commission, Joint Research Centre). 2008. *Handbook on Constructing Composite Indicators: Methodology and User Guide.* Paris: OECD.

Paruolo, P., M. Saisana, and A. Saltelli. 2013. 'Ratings and Rankings: Voodoo or Science?' *Journal of the Royal Statistical Society* A 176 (3): 609–34.

Saisana, M., B. D'Hombres, and A. Saltelli. 2011. 'Rickety Numbers: Volatility of University Rankings and Policy Implications'. *Research Policy* 40: 165–77.

Saisana, M., A. Saltelli, and S. Tarantola. 2005. 'Uncertainty and Sensitivity Analysis Techniques as Tools for the Analysis And Validation Of Composite Indicators'. *Journal of the Royal Statistical Society* A 168 (2): 307–23.

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M., Saisana, and S. Tarantola. 2008. *Global Sensitivity Analysis: The Primer*. Chichester, England: John Wiley & Sons.

Appendix

Table 6-G: Kendall-Tau correlation of KFSIs with SS

Indicator	KT Correlation	p-value
KFSI1	0.456	0.0000
KFSI2	0.295	0.0000
KFSI3	0.347	0.0000
KFSI4	-0.054	**0.4769**
KFSI5	0.183	**0.0149**
KFSI6	0.257	0.0009
KFSI7	0.451	0.0000
KFSI8	0.521	0.0000
KFSI9	0.225	0.0028
KFSI10	0.568	0.0000
KFSI11	0.546	0.0000
KFSI12	0.417	0.0000
KFSI13	0.431	0.0000
KFSI14	0.336	0.0000
KFSI15	0.198	0.0072
KFSI16	0.459	0.0000
KFSI17	0.316	0.0000
KFSI18	0.528	0.0000
KFSI19	0.385	0.0000
KFSI20	0.471	0.0000

Table 6-H: Rankings of aggregation alternatives

Rank	FSI 2018 Country	Alt 1 Country	Alt 1 Rank shift	Alt 2 Country	Alt 2 Rank shift	Alt 3 Country	Alt 3 Rank shift
1	CHE	USA	1	CHE	0	CHE	0
2	USA	LUX	4	ARE	7	ARE	7
3	CYM	CHE	-2	CYM	0	CYM	0
4	HKG	HKG	0	HKG	0	HKG	0
5	SGP	SGP	0	TWN	3	BHS	14
6	LUX	CYM	-3	THA	9	TWN	2
7	DEU	DEU	0	BHS	12	THA	8
8	TWN	JPN	5	PAN	4	PAN	4
9	ARE	FRA	16	BHR	8	SGP	-4
10	GGY	IRL	16	GGY	0	BHR	7
11	LBN	CAN	10	LBN	0	GGY	-1
12	PAN	NLD	2	SGP	-7	KEN	15
13	JPN	TWN	-5	KEN	14	LBN	-2
14	NLD	GGY	-4	LBR	24	USA	-12
15	THA	LBN	-4	VGB	1	VUT	51
16	VGB	MLT	4	NLD	-2	LBR	22
17	BHR	IND	15	MYS	14	LUX	-11
18	JEY	CYP	6	LIE	28	NLD	-4
19	BHS	VGB	-3	MAC	3	VGB	-3
20	MLT	PAN	-8	BMU	16	PRY	42
21	CAN	CHN	7	MHL	18	LIE	25
22	MAC	GBR	1	USA	-20	DEU	-15
23	GBR	JEY	-5	TUR	7	MYS	8
24	CYP	AUT	11	JEY	-6	MAC	-2
25	FRA	ITA	16	BRB	23	BMU	11
26	IRL	MAC	-4	AIA	30	JPN	-13
27	KEN	KOR	6	SAU	10	MHL	12
28	CHN	ARE	-19	MUS	21	AIA	28
29	RUS	RUS	0	PRY	33	JEY	-11
30	TUR	AUS	14	VUT	36	TUR	0
31	MYS	NOR	14	UKR	12	BRB	17
32	IND	THA	-17	JPN	-19	SAU	5
33	KOR	ESP	19	RUS	-4	MUS	16
34	ISR	ISR	0	LUX	-28	PRI	31
35	AUT	BHR	-18	DEU	-28	RUS	-6
36	BMU	TUR	-6	PRI	29	KNA	27
37	SAU	SWE	17	KNA	26	MLT	-17
38	LBR	BEL	15	PHL	2	UKR	5
39	MHL	ZAF	11	ABW	29	BRN	52
40	PHL	POL	11	ISR	-6	CYP	-16
41	ITA	MYS	-10	CYP	-17	ABW	27
42	IMY	PHL	-2	MLT	-22	ISR	-8
43	UKR	IMY	-1	ROM	4	PHL	-3
44	AUS	LVA	11	DOM	25	CHN	-16
45	NOR	DNK	16	CRI	14	WSM	36
46	LIE	BHS	-27	IMY	-4	BOL	42
47	ROM	KEN	-20	CHN	-19	ATG	51
48	BRB	SAU	-11	SYC	29	SYC	29
49	MUS	NZL	9	TZA	26	ROM	-2
50	ZAF	ROM	-3	WSM	31	CAN	-29
51	POL	BMU	-15	GTM	27	DOM	18
52	ESP	BRA	21	KOR	-19	IMY	-10
53	BEL	UKR	-10	CUW	31	MDV	41
54	SWE	MHL	-15	TCA	33	TZA	21
55	LVA	IDN	2	BOL	33	KOR	-22

Rank	FSI 2018 Country	Alt 1 Country	Alt 1 Rank shift	Alt 2 Country	Alt 2 Rank shift	Alt 3 Country	Alt 3 Rank shift
56	AIA	CZE	14	VIR	30	TCA	31
57	IDN	PRT	7	GIB	26	CRI	2
58	NZL	FIN	13	IDN	-1	GTM	20
59	CRI	CHL	1	CHL	1	CUW	25
60	CHL	HUN	14	VEN	25	MCO	32
61	DNK	MUS	-12	BRN	30	FRA	-36
62	PRY	LBR	-24	BLZ	28	AUT	-27
63	KNA	BRB	-15	MCO	29	VIR	23
64	PRT	SVK	12	URY	3	BLZ	26
65	PRI	URY	2	CAN	-44	IRL	-39
66	VUT	ISL	6	AUT	-31	GIB	17
67	URY	LIE	-21	MDV	27	IDN	-10
68	ABW	CRI	-9	POL	-17	IND	-36
69	DOM	MEX	13	ISL	3	DMA	27
70	CZE	GRC	10	LVA	-15	CHL	-10
71	FIN	HRV	8	ZAF	-21	POL	-20
72	ISL	AIA	-16	HRV	7	VEN	13
73	BRA	BGR	16	NZL	-15	GBR	-50
74	HUN	EST	19	DMA	22	ZAF	-24
75	TZA	DOM	-6	GRC	5	LVA	-20
76	SVK	KNA	-13	FRA	-51	URY	-9
77	SYC	ABW	-9	GHA	18	NOR	-32
78	GTM	PRI	-13	PRT	-14	AUS	-34
79	HRV	TZA	-4	IND	-47	NZL	-21
80	GRC	GTM	-2	HUN	-6	ITA	-39
81	WSM	LTU	16	SVK	-5	ISL	-9
82	MEX	VEN	3	IRL	-56	GRD	19
83	GIB	SYC	-6	ATG	15	COK	17
84	CUW	GIB	-1	NOR	-39	HRV	-5
85	VEN	PRY	-23	MNE	14	PRT	-21
86	VIR	VIR	0	DNK	-25	ESP	-34
87	TCA	WSM	-6	MEX	-5	DNK	-26
88	BOL	CUW	-4	CZE	-18	GRC	-8
89	BGR	GHA	6	AUS	-45	GMB	17
90	BLZ	VUT	-24	COK	10	HUN	-16
91	BRN	TCA	-4	FIN	-20	CZE	-21
92	MCO	BLZ	-2	GRD	9	SVK	-16
93	EST	BOL	-5	BGR	-4	SWE	-39
94	MDV	MCO	-2	BWA	9	FIN	-23
95	GHA	MNE	4	ITA	-54	BEL	-42
96	DMA	BRN	-5	MKD	6	MEX	-14
97	LTU	MDV	-3	ADO	8	GHA	-2
98	ATG	MKD	4	GMB	8	BRA	-25
99	MNE	DMA	-3	EST	-6	LCA	11
100	COK	BWA	3	BRA	-27	BWA	3
101	GRD	ADO	4	TTO	6	BGR	-12
102	MKD	COK	-2	ESP	-50	MNE	-3
103	BWA	GRD	-2	NRU	5	MSR	9
104	SVN	ATG	-6	SMR	5	ADO	1
105	ADO	TTO	2	VCT	6	EST	-12
106	GMB	SMR	3	SWE	-52	MKD	-4
107	TTO	NRU	1	LTU	-10	VCT	4
108	NRU	GMB	-2	BEL	-55	NRU	0
109	SMR	VCT	2	LCA	1	TTO	-2
110	LCA	LCA	0	GBR	-87	SMR	-1
111	VCT	MSR	0.5	MSR	0.5	LTU	-14
112	MSR	SVN	-7.5	SVN	-7.5	SVN	-8

7. TJN's Response to JRC Audit

The Tax Justice Network are grateful to the JRC for their excellent statistical analysis, and for discussions over the last two years which have proceeded alongside our stakeholder survey and rounds of deep, expert engagement on the substantive content and structure of the index. As leading global experts on index evaluation, we warmly welcome their overall assessment that the FSI is a well-constructed index, and their appreciation of the depth and originality of the underlying research.

Perhaps inevitably, the statistical and substantive analyses are not always in agreement. Where experts identify additional areas of financial secrecy that pose a global risk, the resulting variable will not necessarily have all the desirable statistical properties, for example. But we firmly believe that the index is strengthened by ensuring rigorous, open and ongoing evaluation of this work and the choices made. We record here the main points of our response, and look forward to continuing these discussions over the following two-year cycle of the FSI.

The JRC statistical audit raises seven points. The first and sixth jointly form the most important issue, which is a criticism of the FSI's method of combination of secrecy and scale. While the JRC do not recommend moving away from the current method, they highlight two points: first, that the role of secrecy is dominated by the role of scale (compared to alternatives designed around the 'statistically correct' log transformation of the scale variable); and second, that because the components of secrecy and scale have a negative statistical relationship, it would be better to present them separately and to focus more on narrative reporting for individual jurisdictions.

We welcome the analysis here, although naturally we do not share it fully. The FSI plays a valuable role, in contrast to any existing 'tax haven' or 'non-cooperative jurisdiction' list, in setting a level playing field for all jurisdictions, by assessing them each against the same, objectively verifiable criteria. Highlighting the secrecy score separately makes sense for this reason, and in the 2018 release we give it more prominence accordingly. We also provide narrative reports for more jurisdictions than any previous release of the index.

We do, however, maintain the view that the overall index ranking is valuable. In addition to the statistical analysis, we cannot neglect the substantive issues being addressed. In particular, the combination of secrecy and scale reflects the key insights of our approach. Firstly, because no jurisdiction is completely transparent, all jurisdictions pose some risk: and so we should think of a secrecy spectrum on which all jurisdictions sit, rather than a list of jurisdictions of concern – where all others are by definition of no concern.

Secondly, the jurisdiction's share of the global provision of offshore financial services is crucial – not (only) as a separate indicator, but as a measure of the degree to which their secrecy should be of concern. It matters that a jurisdiction with near-zero financial service exports is secretive – but not nearly as much as does the secrecy of a major financial centre. As the JRC assessment notes, the current approach reflects this logic well (emphasis added):

The implication is that countries that have a similar SS can have markedly different FSIs as a result in relatively small differences in GSW. On the other hand, countries with low SS and

low GSW will only experience a small increase in FSI if the GSW were to be increased. *Overall, for countries with small GSW, their FSI is driven much more by their GSWs than by their SSs. The opposite is true for countries with large GSW: here countries are differentiated mainly on their secrecy scores.*

Once we recognise that there is important meaning to a ranking based on combined secrecy and scale, the question is how that combination should be made. The JRC alternatives analysed make very clear that there is no obvious answer. A multiplication with no transformation eliminates the impact of secrecy. A statistically 'correct' log transformation of scale results in a ranking which, per JRC, "distorts the reality of the size of the financial sectors of the jurisdictions… and some very visible jurisdictions will experience large changes in the overall index rank". Given the choice between statistical correctness and correct representation of the actual economic phenomenon, we retain a preference for the former. However, we continue to believe that further analysis may allow the development of an alternative which is statistically 'cleaner' than the current transformation, without distorting the substantive meaning of the two components.

The JRC's second point is that three of the 20 KFSIs exhibit a negative or zero association with the others. While this may be statistically unattractive, we are confident in the substantive importance of the issues reflected, and see the absence of association as positive confirmation that additional information is being conveyed. Similarly, the third point involves a suggestion to rework the groupings of KFSIs to reflect their statistical associations, rather than their substantive meaning. We see the statistical analysis as shedding interesting new light on the types of strategy pursued by secrecy jurisdictions. Potentially, further analysis on this point may demonstrate a substantive reason to rework the groups – but current correlations alone do not rise to this level.

The fourth point made is to suggest using geometric rather than arithmetic means, to combine the KFSIs into a single secrecy score. We do not see a strong case here, and reflect that it adds minor complexity in understanding the compilation, and a significant deviation, without any great benefit. Such a change is mostly preferred where a compensatory statistical feature is not intended. However, compensatory aggregation appears to make sense for the KFSIs, because none of the KFSIs are clear substitutes for another KFSI. If there was full substitutability between any of the KFSIs, it might make sense to decrease the "compensatory" aspect of the arithmetic aggregation by moving to a geometric aggregation. This option is set aside for now, but could be reconsidered should future changes to KFSIs shift the balance of costs and benefits.

The fifth point concerns the 'uncertainty' of the rankings around the choice of scale measure. As part of the expansion of FSI analysis, we created a set of alternative scale measures, different by design, to consider the case for change and to give options for more bespoke risk analysis in different areas of economic and financial activity. Inevitably the range of measures compiled (see Annex H) would lead to a range of rankings, and these are likely to be of value in country-level risk analysis in particular. While the main focus of the FSI remains on the risks associated with secrecy in offshore financial services, however, the substantive case for remaining with the current GSW is clear – and 'uncertainty' seems not

to be the right expression for the fact that rankings differ if based on different types of activity (i.e. trade in goods rather than financial services).

The sixth point made by JRC is that the uncertainty analysis shows the rankings to be relatively robust; and that it would be worthwhile to publish confidence intervals alongside the ranking. This makes sense, and in addition to publishing the JRC analysis this year we will examine a process to construct more precise confidence intervals for future releases of the index.

Once again, we extend our great thanks to the JRC, and to William Becker and Michaela Saisana, for their excellent contribution.