

COMPARING SOVEREIGN DEBT RATING CONSISTENCY WITH SOVEREIGN AML RATING ACCURACY

Charles Littrell¹ & Danae O'Brien

Central Bank of The Bahamas

December 2019

Abstract

Many stakeholders in the global anti-money laundering (AML) space, including correspondent banks, rely upon explicit or de facto AML sovereign rating services. This reliance is similar to that reposed by fixed income investors in sovereign debt ratings.

In this paper we demonstrate that sovereign debt rating agencies provide highly consistent and reasonably accurate ratings. Sovereign AML ratings, by contrast, are usually inconsistent and lack any empirical basis to judge their accuracy. Users of sovereign AML ratings are therefore cautioned that, whatever value they place upon these ratings, they are unreliable measures of the actual incidence of money laundering in any given jurisdiction.

Introduction

In a globalized world, cross border ratings become increasingly important. There is considerable interest in, among a great many examples, the relative and absolute rankings of the world's universities, airline safety, living quality in cities, and a wide range of economic and social variables.

Ratings quality can be assessed in several ways. We consider that the gold standard for sovereign ratings are the debt ratings assigned by the major international agencies, particularly Fitch, Moody's, and Standard & Poor's. These services have been in place for roughly a century, are broadly accepted, and subject to rigorous statistical validation processes mandated by their regulators.

At the other end of the scale we see ratings that exist more for the sake of fun and (hopefully) friendly debate. Examples include lists of the world's best beers, tastiest foods, or most attractive people.

In between these extremes, we see a range of cross border ratings deserving more or less respect for their rigour and usefulness. These ratings and rankings are often prepared or assembled by multilateral bodies, and in many instances rely upon a weighted index approach. The United Nations Human Development Index (HDI)² is a prominent case in point. Neither the United Nations nor any sensible user claims that the HDI is perfect, but over time it has demonstrated broad acceptance and usefulness.

It is important that users of cross border ratings and rankings understand the degree to which they can rely upon them. In this paper we compare and contrast the approach and outcomes associated with the three sovereign debt rating agencies, with the approach and outcomes associated with five sovereign AML ratings services.

¹ Corresponding author: cwlittrell@centralbankbahamas.com

² <http://www.hdr.undp.org/en/composite/HDI>

In particular, we address the question: should users of the sovereign AML ratings and rankings services treat them with the respect due sovereign debt ratings, or with the respect due to a ranking of the world's best beers?

We use rank correlation metrics to establish the degree to which sovereign debt ratings are consistent with each other, and the degree to which sovereign AML ratings are consistent. We use publicly available data to observe the degree to which users can rely upon the relative and absolute rankings of sovereign debt ratings.

There is a particular feature in sovereign AML services that some ratings use other ratings as inputs to their own indices. We examine statistical consistency on the basis of each full rating, and also with the cross-listed elements stripped out.

PART 1: SETTING UP THE DATA

Sovereign debt ratings data

Under United States Securities and Exchange Commission regulatory requirements, organisations can register as Nationally Recognized Statistical Ratings Organizations (NRSROs). The three most prominent NRSROs are Moody's Investors Service³, Standard & Poor's⁴, and Fitch Ratings⁵. All three agencies (and several others) are also subject to European Union and other national regulation, but the NRSRO framework imposes a common and substantial regulatory reporting obligation on debt ratings agencies.

NRSROs are required to publicly disclose a great deal of data, from which the authors have extracted:

- 1) Current sovereign ratings for the 75 jurisdictions that have completed the Financial Action Task Force's (FATF's) fourth round mutual evaluations;
- 2) The most current NRSRO reporting on sovereign issuer 10 year transition and default rates, from Exhibit 1 of each agency's NRSRO's filings; and
- 3) The most current NRSRO reporting on U.S. public finance issuer 10 year transition and default rates, also from Exhibit 1 of each agency's NRSRO's filings.

AML ratings services

Recently⁶ we provided a field guide to the sovereign AML ratings services. There are two services that rely upon original analysis:

- The FATF, which assesses 40 elements of technical compliance, and 11 elements of effectiveness. We will aggregate these into two ratings series designated FATF 40 and FATF 11 respective; and
- The U.S. State Department, which issues an annual International Narcotics Control Strategy Report (INCSR). Part 2 of the INCSR gives a binary listing of jurisdictions which are considered high risk for U.S. AML purposes⁷.

We use the 75 (as of mid 2019) jurisdictions subject to the FATF 4th round mutual evaluation report as our data set for this paper. Each of these jurisdictions has complete or near-complete evaluations on the FATF 40 and FATF 11. For the INCSR, we assume that any FATF jurisdiction not listed has not been designated high risk by the U.S. State Department. It is possible that some of these jurisdictions have simply not been considered.

In addition to the FATF and State Department, several groups produce proprietary AML sovereign ratings, which in each case are based upon weighted indices of other assessments, including the FATF and INCSR ratings, but also inputs such as Transparency International's ranking of countries by perceived corruption, and the Tax Justice Network's assessment of jurisdictional tax evasion facilitation.

³ <https://www.moodys.com/Pages/reg001002.aspx>

⁴ https://www.standardandpoors.com/en_US/web/guest/regulatory/form-nrsro

⁵ <https://www.fitchratings.com/site/regulatory>

⁶ *A Field Guide to Sovereign AML Ratings Agencies*, ACAMS Journal, January 2019

⁷ The State Department does not promote the INCSR as an AML sovereign ratings tool—but it is commonly used that way by many stakeholders.

In this study we include three providers:

- The Refinitiv (formerly Thomson Reuters) Country Risk Ranking⁸;
- The Basel AML Index from the Basel Institute on Governance⁹; and
- The Country Report listing from KnowYourCountry.com¹⁰.

The authors thank each of the above groups, who have provided both publicly available and in some cases subscriber-only data, including the raw data underlying rankings. Without this data provision, our paper would have been much less robust.

All the above three ratings providers rank many more than the 75 countries in the FATF group. With one minor exception, all three ratings providers rank all 75 of these countries.

Moving from ratings to rankings

The ratings agency case is straightforward: a AAA rating outranks a AA rating, which outranks an A rating, and so forth. We have converted the three ratings scales to non-parametric numeric scales (AAA = 1, AA+ = 2, etc.) to generate a numeric rather than a symbolic ranking.

The only complexity associated with the ratings agency rankings was that not all jurisdictions were rated by all three agencies. From 75 candidate combinations, we had 49 Moody's/Fitch dual ratings, 54 Moody's/S&P, and 47 S&P/Fitch dual rated jurisdictions.

Refinitiv, Basel AML, and KnowYourCountry all produce explicit rankings by jurisdiction. For each of the three ranking providers, we used the provider's stated rankings. We also disaggregated ("stripped") the FATF and INCSR elements from each rating service, and re-ranked the jurisdictions on the basis of each provider's other index elements.

For the INCSR, we have assigned a binary score: 0 for countries not listed, and 1 for listed countries.

This leaves the FATF assessments. Each of the 40 and 11 compliance and effectiveness ratings is scored on a four point scale, ranging from "essentially perfect" to "essentially non-existent"¹¹. The descriptions of the rating imply a non-linear scale. After experimenting with various options, all of which produced similar results, we elected to map the FATF scale into the following squared algorithm, and rank separately for the FATF 40 and FATF 11:

Highest rank = 0

Second highest rank = 1

Third highest rank = 4

Lowest rank = 9

⁸ https://www.refinitiv.com/en/products/country-risk-ranking/?utm_content=Refinitiv%20Brand%20Product-CRBN-AMER-B-EN-BMM&utm_medium=cpc&utm_source=bing&utm_campaign=68832_RefinitivBAUPaidSearch&elqCampaignId=5917&utm_term=%20+refinitiv%20+country%20+risk

⁹ <https://www.baselgovernance.org/basel-aml-index>

¹⁰ <https://www.knowyourcountry.com/copy-of-country-reports>

¹¹ <http://www.fatf-gafi.org/publications/mutualevaluations/documents/fatf-methodology.html>

We then summed each of the 40/11 rankings for each jurisdiction, to develop an ordinal consolidated table, with lower scores producing better ranks.

We emphasise that this is not an official ranking, but an arithmetic convenience to allow us to convert 51 disparate FATF assessments into two rankings of jurisdictions. Among other things, this approach assumes that each ranking element is equivalently important.

But on the other hand, it produces a reasonable basis for comparison; the authors are happy to entertain suggestions on better arithmetic transformations.

Resultant rankings tables

We are left with a 3 x 3 matrix (but with differing underlying dimensions) on the debt ratings agencies, and a 9 x 9 matrix for the AML agencies. The nine AML elements comprise:

- 1) FATF 40
- 2) FATF 11
- 3) INCSR
- 4) Refinitiv full
- 5) Refinitiv stripped
- 6) Basel AML full
- 7) Basel AML stripped
- 8) KnowYourCountry full
- 9) KnowYourCountry stripped

PART 2: RANKINGS CORRELATIONS AND SIGNIFICANCE VALUES

For both debt and AML rankings, we use Spearman and Kendall rank order correlations. In the Kendall case, we use the tau-b method, which better accounts for tied ranks¹².

Debt ranking correlations

Inspecting a comparison list of country ratings by the three debt ratings agencies often generates the observation: “they are essentially the same.” The correlations demonstrate this point.

Table 1: Debt ratings agency rank correlations

	Spearman	Kendall tau-b
Moody’s vs. Fitch	0.982	0.936
Moody’s vs. S&P	0.976	0.911
S&P vs. Fitch	0.987	0.947

P-values are not shown in the above chart: they are all zero to many significant digits.

The authors in this case have demonstrated that the major ratings agencies generate near-identical ratings for sovereign debt default risk. We do not assert that this is a novel or interesting finding, but include it to establish a benchmark for the more interesting results on the AML correlation matrices.

The sovereign debt ratings are consistent, but are they correct?

Because the sovereign debt ratings are so consistent, they must all be more or less equally right or wrong. “Correct” in this context means that a well-rated exposure (or portfolio of exposures) will default less often than a poorly rated exposure or portfolio. Each agency uses very similar labels (AAA/Aaa down to D) to assign ratings opinions.

A look at the three agency NRSRO Exhibit 1 filings gives the following ten year default rates. Table 2 gives average default rates across the AAA, AA, A, BBB, BB, and B and lower ratings groups.

¹² For more background see (among other examples): Agresti, A. (2010). *Analysis of Ordinal Categorical Data* (Second ed.). New York: John Wiley & Sons.

Table 2: Ten Year Default rates (%) by rating category, Sovereign exposures

	Moody's ¹³	S&P ¹⁴	Fitch ¹⁵
AAA	0	0	0
AA	6.67	0	0
A	5.3	21.7	5.0
BBB	7.1	0	0
BB	0	5.0	5.0
B and lower	20.0	7.4	17.4

We can see that broadly but imperfectly, default rates increase for lower-rated jurisdictions. In statistical terms, we are working with very few defaults from a small population, so some randomness creeps into the results. In Table 2, Moody's, S&P, and Fitch rate only 109, 122, and 127 jurisdictions respectively, among which 5 to 6 defaults are distributed.

Table 3: Default rates (%) by rating category, U.S. public sector exposures

	Moody's	S&P
AAA	0	0.04
AA	0	0.06
A	0.15	0.15
BBB	0.51	1.89
BB	5.77	4.43
B and lower	28.2	11.1

¹³ <https://www.moody.com/sites/products/ProductAttachments/Exhibit1.pdf>, page 22

¹⁴

https://www.standardandpoors.com/en_US/delegate/getPDF?articleId=2193561&type=COMMENTS&subType=REGULATORY, page 35

¹⁵ Fitch link available via free subscriber sign-up

As a check on Table 2, we show Table 3, giving default rates on public sector exposures¹⁶. The agencies collectively rate over 10,000 exposures in this sector, considerably reducing statistical randomness. We see here that the ordinal results are somewhat more robust than in the sovereign exposures. Because the agencies use quite similar ratings methodologies for their international and American public sector exposure, it is reasonable to conclude that, in broad terms, their ratings “work”. That is, investors can rely with some confidence on the ordinal value of ratings, when measured against default rates across credit portfolios.

In summary: the sovereign debt rating agencies produce a highly consistent and reasonably accurate outcome. Subscribers to these services can engage them with some confidence in the product they are using.

We now apply the same analysis to the nine AML ratings sources.

Table 4: AML rank correlations, Spearman

(Correlation lower half; p-values upper half, p-values over 0.05 omitted)

	FATF 40	FATF 11	INCSR	Refinitiv	Refinitiv Stripped	Basel	Basel Stripped	KYC	KYC Stripped
FATF 40	1.00	P - 0.00	P- 0.02	P - 0.00	P - 0.01	P- 0.00			
FATF 11	.62	1.00	P- 0.02	P- 0.00	P- 0.00	P- 0.00		P- 0.00	P- 0.00
INCSR	-0.27	-0.27	1.00					P-0.00	
Refinitiv	0.32	0.49	0.07	1.00	P-0.00	P-0.00		P-0.00	P-0.00
Refinitiv Stripped	0.28	0.44	0.17	0.98	1.00	P-0.00		P-0.00	P-0.00
Basel	.49	0.71	0.13	0.71	0.70	1.00		P-0.00	P-0.00
Basel Stripped	-0.09	-0.04	0.13	0.05	0.09	0.15	1.00		
KYC	0.18	0.39	0.40	0.66	0.67	0.61	0.18	1.00	P-0.00
KYC Stripped	0.20	0.41	0.16	0.72	0.71	0.58	0.16	0.81	1.00

¹⁶ Fitch is much smaller in this rating category so is excluded.

Table 5: AML rank correlations, Kendall tau-b

(Correlation lower half; p-values upper half, p-values over 0.05 omitted)

	FATF 40	FATF 11	INCSR	Refinitiv	Refinitiv Stripped	Basel	Basel Stripped	KYC	KYC Stripped
FATF 40	1.00	P-0.00	P-0.02	P-0.01	P-0.02	P-0.00			
FATF 11	0.48	1.00	P-0.02	P-0.00	P-0.00	P-0.00		P-0.00	P-0.00
INCSR	-0.22	-0.23	1.00					P-0.00	
Refinitiv	0.21	0.34	0.06	1.00	P-0.00	P-0.00		P-0.00	P-0.00
Refinitiv Stripped	0.19	0.28	0.14	0.89	1.00	P-0.00		P-0.00	P-0.00
Basel	0.34	0.51	0.11	0.51	0.50	1.00		P-0.00	P-0.00
Basel Stripped	-0.05	-0.01	0.11	0.04	0.05	0.12	1.00		
KYC	0.12	0.27	0.33	0.49	0.49	0.44	0.12	1.00	P-0.00
KYC Stripped	0.14	0.28	0.13	0.53	0.52	0.42	0.11	0.63	1.00

The results from Tables 4 and 5 look nothing like the results from Table 1:

- The correlations range from substantially negative (INCSR vs. FATF) to near unity (Refinitiv vs. Refinitiv stripped).
- The two “original analysis” ratings sources are negatively correlated. This is perverse: the better the FATF 40 or FATF 11 rank, the more likely the jurisdiction will feature on the State Department’s list of jurisdictions of concern for money laundering. This negative correlation also suggests that index raters using both FATF and INCSR face worrisome internal contradictions.
- By comparing the correlations between the full and stripped versions of Refinitiv, Basel AML, and KnowYourCountry, we can infer that the Refinitiv index places little reliance upon FATF and INCSR, Basel quite a lot, and KYC somewhere in between.
- The third major index feed for the Basel AML model is the Tax Justice Network’s Financial Secrecy Index.¹⁷ This index tends to rate large countries such as the United States and Switzerland badly, opposite the FATF results, so the correlation between the full Basel AML index and the stripped index is quite low.
- The correlation between the FATF 40 and FATF 11 is reasonably high. As will be discussed later, however, there are some remarkable outliers between the two rankings.

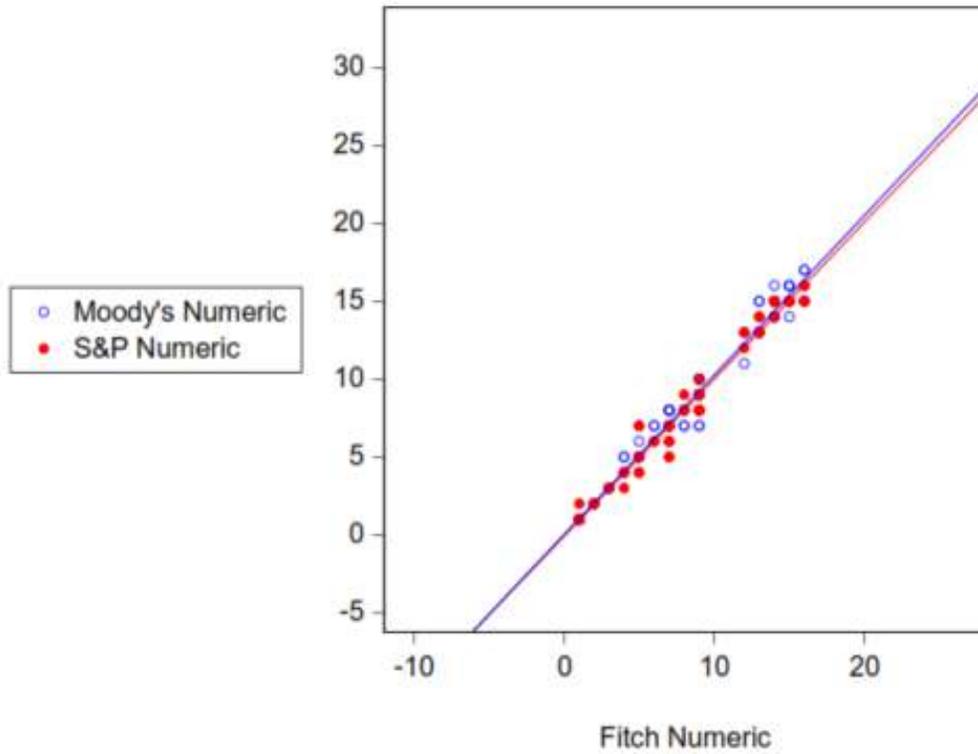
¹⁷ <https://financialsecrecyindex.com/en/>

- The Refinitiv and KYC correlations to the FATF rankings are reasonably good, particularly for the FATF 11 rankings, and even on a stripped basis. This suggests that the FATF effectiveness ranking and the Refinitiv and KYC rankings are broadly looking at the same thing. Such a result does not hold elsewhere in the tables.
- The stripped Basel AML ranking correlates poorly with the stripped KYC and Refinitiv rankings. The KYC and Refinitiv rankings, by contrast, are strongly correlated. The correlations increase slightly when the FATF and INCSR inputs are removed.

Two pictures are worth a thousand suspicious transaction reports...

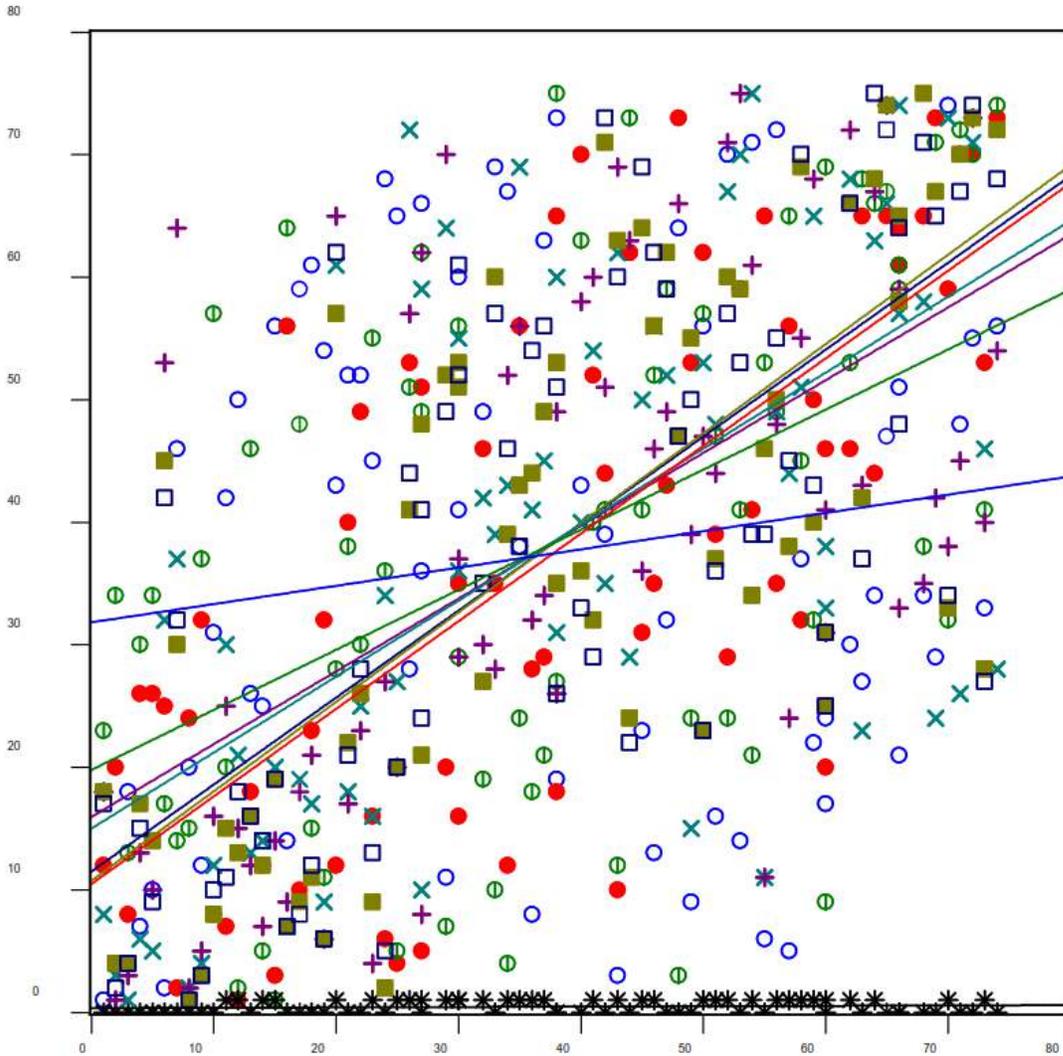
The first chart aggregates the three rank vs. rank outcomes for the debt rating agencies. It demonstrates the near-unity of outcomes across the three ratings sources.

Chart 1



The Second chart gives the same pictorial transformation for the nine AML ratings sources.

Chart 2

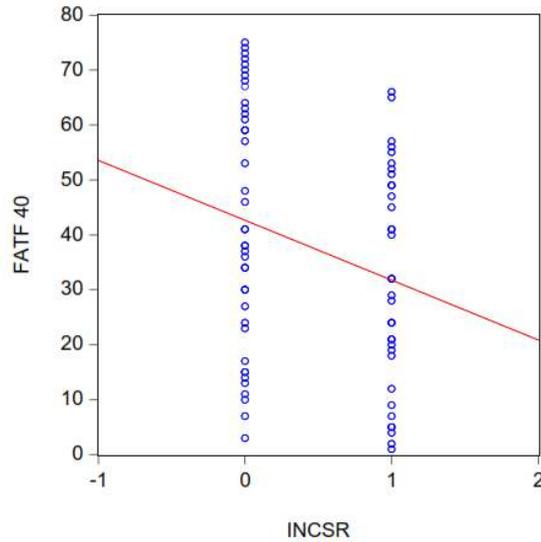


Broadly and perhaps simplistically, Chart 1 represents a highly correlated world, and Chart 2 a much less correlated world. In fact, in many cases an uncorrelated or negatively correlated world.

The curious case of INCSR vs. FATF

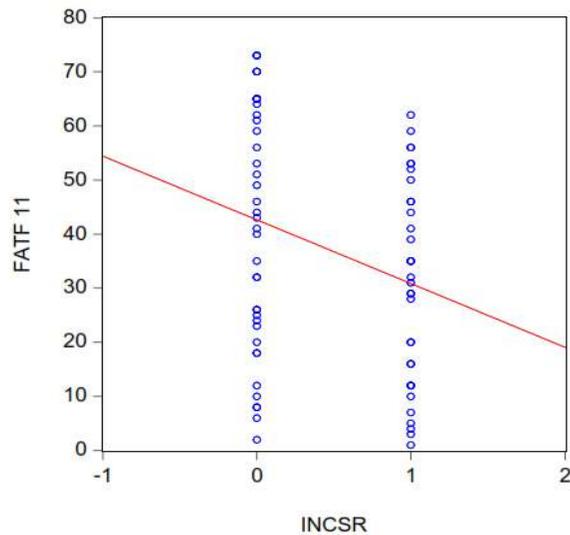
Chart 3 matches INCSR findings (0 = not on the list, 1 = on the list) with FATF 40 rankings. We see the counterintuitive result that the worst ranking FATF technical compliance jurisdictions are under-represented on the INCSR list, and the best-rated somewhat over-represented.

Chart 3



A very similar result obtains for INCSR vs. the FATF 11 effectiveness rankings, as depicted in Chart 4.

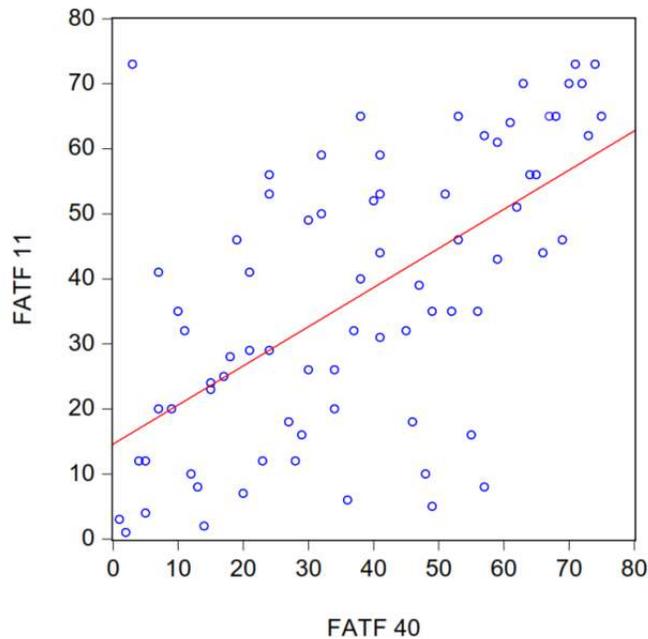
Chart 4



If we split the FATF 40 into top and bottom halves, then 51 per cent of the better-ranked jurisdictions are on the INCSR list, but only 38 per cent of the worst-ranked FATF 40 are on the list. The corresponding numbers for the FATF 11 are 51 and 36 per cent.

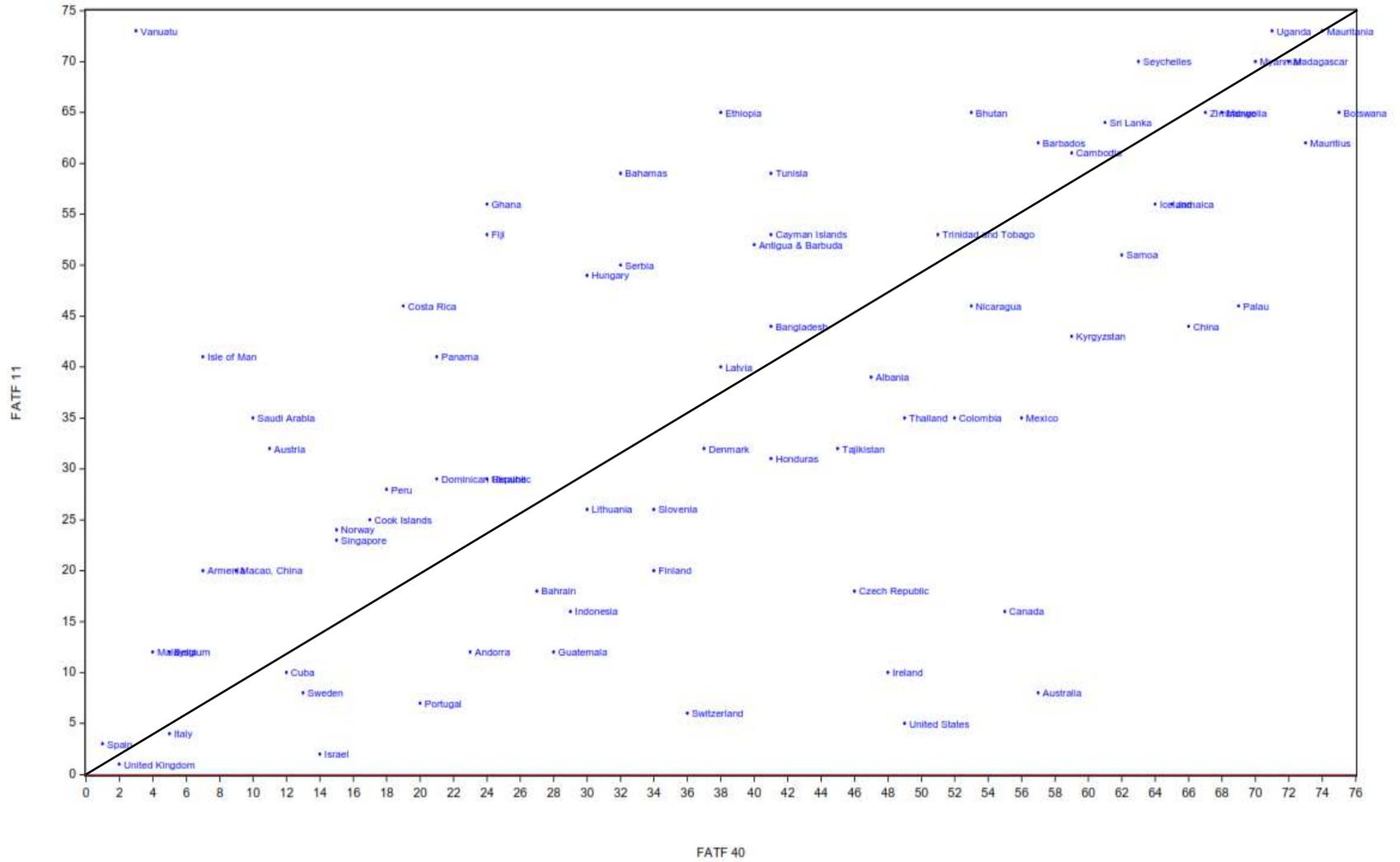
FATF 40 vs. FATF 11

Chart 5 (a)



The two FATF ranking scales are somewhat but not fully correlated. Vanuatu is a massive outlier, ranking very near the top on technical compliance, but very near the bottom on effectiveness. More substantively, the United States is also a substantial outlier, lying towards the middle of the pack on technical compliance, and ranking second on effectiveness. The authors editorialise that the FATF 11 does not measure effectiveness, in the sense of financial crime suppression, but “effectiveness” in the sense of lots of demonstrable punishments. We note that the two best-ranked jurisdictions, the United States and the United Kingdom, register thousands of ML prosecutions and confiscations per year. Perhaps to a lawyer this represents effectiveness. To an economist, such an outcome signals regimes that are awash in dirty money.

Chart 5 (b): Repeat of Chart 5(a) with Country Names



Sovereign AML ratings are inconsistent. Are any of them correct?

This is the trillion-dollar question. Debt ratings agencies have a well-developed dependent variable: credit default rates. There is no equivalently available dependent variable in the AML ratings and rankings space. It is not even clear what is being measured. The INCSR purports to measure the degree to which countries are a money laundering risk to the United States. The FATF 40 and FATF 11 assess the degree of compliance with the FATF 40 and FATF 11. The three index services aggregate mainly opinion-based ratings from many sources.

What no provider in the AML sovereign space is measuring, or even defining, is the actual incidence of financial crime. Until the ratings providers create this dependent variable, there is no way to statistically assess the relative or absolute quality of any AML sovereign rating opinion.

Conclusion

This paper presents in rank correlation terms the well-known proposition that the three major sovereign debt rating agencies generate highly consistent ratings and rankings. Using publicly available regulatory filings, we also observe that sovereign debt ratings have a reasonable claim to ordinal accuracy, in that portfolios comprising well-rated sovereigns are generally less likely to default than portfolios comprising poorly rated sovereigns.

Using the same rank correlation approach, we then demonstrate that there is little consistency among the five sovereign AML ratings providers considered in this paper. Furthermore, without an objectively defensible database of financial crime outcomes against which to measure these ratings, they will never be empirically robust.

In summary: if any users of sovereign AML ratings consider that they can rely on them equivalently to the way users of sovereign debt ratings rely on those opinions—these users are dangerously in error.