Linda
Hello everyone and welcome to Theory Meet Practice, our video series about academic research into sustainable finance and the lessons investors can take from it. In 2019, a working paper published by three finance professors from MIT caught the attention of investors with an examination of differences among ESG ratings by six providers, including MSCI. In it, the paper shows that ESG ratings from competing providers had low correlation to each other.

And the paper dissected the sources of those differences. So the paper, Aggregate Confusion, the Divergence of ESG Ratings, which was eventually published in 2022 and won the 2023 Pagano and Zechner Prize for the best non-investment paper published in the Review of Finance. And so it's my privilege today to be in conversation with one of the co-authors of that paper, Roberto Rigabon. He is an applied economics professor at MIT Sloan School of Management and founder of the Aggregate Confusion Project.

We are thrilled to have him with us. Welcome.

Roberto
Thank you so much. And thank you for the kind introduction.

Linda
Well, you know, the number one question I've gotten over the years over and over again, actually, is, you know, why is your ESG rating, the MSCI ESG rating, for company X so low when, say, Sustain Analytics gives it a high rating? So I think clients, you know, keep asking us that question, but it's actually really hard for us to answer for the very practical reason that I don't actually subscribe to other ratings. And so I don't know the finer points of their methodology. But you and your colleagues at Gloria and Berg and Julian Kerbo were able to access the scores and ratings of multiple providers. So explain to us what actually accounts for these differences.

Roberto
Actually, there are three reasons why the ratings tend to differ. So one is when you go to the relatively narrowly defined attributes, let's say that you are thinking about labor treatment. And by the way, this can go even finer. That is labor treatment of women, for example.

Data providers use these different proxies for that. And in fact, actually what we have found is that the proxies are very culturally dependent. So in France, for example, a lot of the big jurors, which is one of the rating agencies that give us the data, they concentrate on labor complaints to a judge or to a labor court, which is something very rare in the United States. In the United States, if you are mistreating a worker, they typically quit. So the turnover is a better indicator.

And what is very interesting is that if you think about it, you would like to have both sources of these proxies to construct a good measure of labor treatment, but each rating agency is very much looking at their own markets to decide which of these should be used. So that produces a measurement error, is that we use different proxies. And that explains more than half of the discrepancy, like 55%. Then there's a second part, which is all the discussion about materiality.

So some rating agencies will consider, again, let's say labor treatment probably is material in almost all the companies, but that will not necessarily be true, for example, with human trafficking. So clothing, probably I will pay attention to human trafficking, but the financial system, I will pay less attention to human trafficking. That notion of what is material is very different across the data providers.

And what that implies is that you know, Sustained Analytics is measuring some things that you are not measuring, and then MSCI is measuring other things. And we call that a scope differences. Uh, so, I mean, there's no chance you will ever have the same score if, if you do not measure, no? So that explains about 40%. And the other 5% is what I thought was important. That's why the paper is called aggregate confusion, because I thought that the biggest, you know, I thought that the biggest issue was going to be that people aggregate these numbers with different procedures. That explains only 5%. So this just tells you how much I contribute to the paper. My idea was only 5%.

Linda
Well, I really like the framework you put around these different types or categories of differences, right? The scope and the measurement and I think the weights. But I guess from a practitioner's perspective, when I'm kind of looking across different providers, I differentiate between the different historical roots and the methodologies and which ones actually capture financially materialistic companies versus those whose historical roots come from more of an ethical investing base, right? And so this is actually reflect maybe more of a philosophical difference. I mean, so do you think that this philosophical difference that I'm talking about is reflected in one of your concepts or more of kind of a meta level framing?

Roberto
No, it will be reflected in all three of them.
It will be reflected on the measurement for sure. For example, if you want to understand the risk, you're thinking about, well, how, for example, see what arise impacts my business. So that's a risk approach. That's very different from the moral point of view or the ethical point of view, but how much I'm contributing to see what arises. So the indicators will be completely different. For sure, they will be different. So again, from the ethical point of view, you know, I think that everything that is related to a crime should be there in the index.

So lying about your marketing, products that are unsafe, you know, human trafficking, you know, money laundering. All these activities are illegal activities. So from the ethical point of view, if I want to know, is MIT doing something illegal, then I want to measure all of those issues, regardless if this is risky or not, regardless if this is material or not. But when I'm analyzing the impact of any of these issues into MIT as a risk profile, then there are some that I will just disregard. Not because they are not morally important, it's just because they are just financially not relevant.

Linda
So I want to move on to talk about credit ratings. So this paper had found a correlation ranging from 0.38 to 0.71 with an average of about 0.5, 0.54 I think. So you know, whether you think that's high or whether that's low kind of depends on the benchmark that you actually compare it to. And of course, the benchmark that you compared it to in the paper is credit ratings where the correlations are something more like 99%. Right. So what was your rationale for using credit ratings as a benchmark?

Roberto
So let me actually tell you what is what I do in classes. I ask the students to rate the snacks. So I give them chocolate, almonds, you know, potato chips, you know, goldfish. In fact, here I have it just for advertisement. This is what I actually bring in the class. And so I tell them, I tell them, I want you to rate these snacks by your preferences, which one you like the most and so on. And then also I want you to tell me different categories, which is, you know, yumminess, sweetness. I like it, I'm allergic or not. I mean, things like safety and so on. And I tell the students to rate all these different chocolates.

And let me tell you, they are less correlated than the ESG rated agencies. So there's more discrepancy on the evaluation of snacks than there is on the ESG rated agencies. So there's an important difference here, is that when you have different preferences on snacks, we are all allowed to buy different snacks. So I mean, so I don't like chocolate in particular, but you might, you know, you might like sweet snacks and you might like M&Ms. So.

See, in a market like that, we are making individual choices. So different preferences will be represented in different choices. But when we are trying to actually, for example, a regulator, let me just extend this all the way to regulator. I know that's not your intention. Let's assume I want a regulator to look into unethical behavior of firms. Well, now it's a single choice for the whole country. So this is almost like choosing which of the single snacks we want to actually eat for all of us. The ESG is kind of in the middle of this too, okay?

That the investors are trying to find some guidance about how do I make a better decision? And that tends to be kind of collective values without them actually knowing exactly what their values are. So it's kind of worse than a regulator in that sense. But it's kind of collective. We want to

know who is a Brown firm, and then which firms will be subject to transition risk. We want to equate these across different investors in some sense, and that complicates the analysis.

Linda
Well, I'll tell you, I get two types of questions about credit ratings and ESG with regards to your paper. So once the questions actually come from non-users of the ratings. So, you know, if you're a journalist, you're a policymaker, people who aren't actually subscribing to the full ratings and all the underlying data, and they don't use them in their day to day work of allocating capital, because of this comparison in your paper, they assume that ESG rating should be like credit ratings. So they ask about whether ratings diverge because companies pay for higher ratings.

And then they are surprised when we say, you know, ESG ratings are actually paid by investors and not companies. Yeah, yeah. Something you pointed out in your paper. The other thing that you pointed out in your paper is that while credit ratings is measuring one thing, so this probability of default, ESG ratings aim to capture multiple dimensions, and those dimensions can be quite different between industries. So it's always hard to answer this question of whether E or S or G factors, you know, just like your yumminess factors, the most important because it also sort of depends.

A different set of questions though I get is from users of ratings, so investors who are actually using this day to day. And the question that I get is one I'm going to ask you, are ESG ratings more like credit ratings or are they more like sell-side ratings?

Roberto
It's interesting. I actually think that this depends on the rating agencies. Sell rating agencies will mean more, they would like more to be like a credit rating agency in the sense that they want to measure the likelihood that you will make something, you know, immoral or illegal. So from that perspective, they're trying to see, are you going to be on the newspapers because you did something illegal? So that one sounds very much like a credit rating score, not like, so you did something wrong. If I'm a firm, I can do many dimensions wrong. So default is only one, I default, no? So the credit rating is just super simple.

I have to measure your willingness to pay and your ability to pay. But it still is extremely hard. I'm not saying that it's trivial, but it's only one issue. There's one issue. Here you have hundreds, hundreds of different issues that you, where the unethical behavior of the firm can be shown. So again, the proliferation of mad behavior that you could in principle have is just tremendously large. And that's all. So trying to make an assessment of that is an order of magnitude harder.

But that's why I think ESG rating and data providers are so important. If I ask any individual for them to rank these issues, just to rank them, forget about putting weights, which one is more important? Poverty, environment, civil rights, product safety. Most individuals will not be able to rank. But if I give them four, they might rank. If I give them 50, they have no clue. Investors that

are not terribly sophisticated rely on… on organizations exactly like yours, where you provide some guidance about how will you aggregate that? And it's absolutely crucial. By the way, we do this with almost everything in finance and in our lives. When we look at a building, it's not like the building reports the windows, the ventilation, the HVAC, and the carpet, and the screen, whether or not it's automatic or not, they don't report that. They tell you a ranking. They give you a number.

And we trust that to reflect a combination of things that are really hard for non-technical people to digest. And then, you know, we get a ranking. Well, I think one reason we get this question about sell-side ratings is that investors are used to sell-side ratings. And that means that they're used to a dispersion of opinions, right? And that dispersion is actually what gives them a fuller set of information. So, you know, what do you think about the fact that the divergence that you have found in your studies, that in itself might actually be providing more information for investors that they're using in their investment decisions.

Okay. So that also depends on the rating. For example, MSCI is super open and super transparent. So you tend to explain more or less how you compute things. So the dispersion, I can make sense of that dispersion when I look at you. But there are some rating agencies that are just incredibly opaque. So it's almost like giving me a number. There's no rationale behind it. So those numbers, I actually, I don't think that they're that useful.

Linda
Well, I think it's interesting that you found this rater effect, right? Because many of our clients who really know the methodology well, they understand the specific features of our methodology that they disagree with so that they will systematically adjust those components to better reflect their view. And so part of the reason to be transparent is that we can have these long-standing discussions where some of the clients do want to underweight or overweight certain elements. And so I think that that radar effect that you found is it's definitely there, but it's also a systematic difference that in some ways the predictability of it is what makes us sort of useful for clients who want to reflect their own views.

Roberto
Yeah, I agree. So what this means is that there's no substitute to thinking from the client's point of view, yeah, it's exactly what you're saying.

Linda
Well, one question I also always get is, you know, will there be more convergence between ratings? Do you think that there will be in the future? And actually, should there be more convergence toward ratings?

Roberto
Again, it depends to whom you talk. You know, an investor that would like to have one number, so I make my own decision whether or not I buy or not a stock or I till my portfolio or I engage. So like, imagine that I just need one number to decide what to do, then that investor will like to have convergence. But from the perspective of information provision, I actually think that divergence is good. And let me explain why in a little bit. There are many aspects that are there that we really don't know how to measure.

I mean, I really don't think that we know how to measure discrimination properly. I don't think we know how to measure happiness properly, sadness. We don't know how to measure this. We measure the extremes. So we need a lot of innovation on thinking about how to measure this. And that is only going to happen only if we are competing against each other. I mean, like the data providers are competing. So that you find a better data source that actually is more meaningful than a particular index.

I want you to have the ability to say to a competitor or to the client, you know, I have a better measure of discrimination because I invested in the R&D to produce this. So in this process of measuring better, I think divergence is good. In all my work in inflation, I spent 17 years measuring inflation rate. This also shows that we academics have a lot of free time. Okay. So anyway, I spent 17 years measuring inflation.

And I have been spending eight years measuring unethical behavior. And in those 17 years, I can tell you, I am absolutely certain that there's not one inflation rate for the United States, that there's thousands, thousands and thousands and thousands. And these are all great theories and judgments that you make at every point in time about how you deal with the missing data. How do you deal with the fact that you didn't find the product in the store? I mean, millions and millions and millions of assumptions that are taken in the construction of one index and then move $73 million.

So I always tell people, you're taking this a little bit too seriously. You know about that is maybe you should actually be more careful. Like this is not this is not like your weight when you go to the doctor. That's way better measure. So ESG is an order from my middle heart. So I the discretion to me is it could also reflect this competition that we truly need,

Linda
Another question we get is, how can we tell if a rating works? Because when it's something like credit ratings, we can measure how good they are at predicting defaults. But although let's not forget that they tab 99% correlation at failing to predict the last major defaults. And for sell-side ratings, I guess we can measure forecast accuracy. But what about for ESG ratings? Because our clients certainly ask us to validate that the ratings capture what it says it aims to capture, which is financial performance.

So we do that, we've published our results, we just put out actually a new study on the relationship between ESG ratings and financial performance over the past 11 years. But as we said at the outset, not all ESG ratings in your study are certainly aiming to capture that financial performance. So how do you think practitioners should be evaluating then the effectiveness of any ESG rating that they're considering?

Roberto
So, when I look at the equivalent of credit rating agencies, the equivalent is sovereign default. Not the default of individuals like NeofiqoScore or the firms is a sovereign debt. And credit rating agencies on sovereign debt are lousy. The best predictor that they actually will get something correct is, I don't know, Argentina. Because Argentina is always on default. But except for that, it's very difficult for them to nail the default. And so you are correct. The credit rating agencies make very big mistakes for aggregate shocks because they are very unusual. So when I think about, let's say, risk of sea water rise or temperature, these are very rare events. So how do you validate this data? We don't want to wait until the event takes place, you know? So that would be one.

So when you think about it from the perspective of what is the role of what you're trying to accomplish by providing this information, is that we want not a single one to ever be validated. And I actually explain to investors all the time. This is not a default. A default is a tiny event in comparison to what we are measuring on ESG. I said, find any ESG that is not dramatically important. The other part is a performance I think is simpler to do.

But this depends on the objectives. For example, the MSCI index, I think, performs really well. I mean, we have a paper also on evaluating the stock return implications of getting a better score in MSCI on the E, and the S, and the G. I mean, it's a new paper that we have. You know, clearly, you provide alpha in our regressions. Very clearly. And we correct for noise. We correct for the things that probably most practitioners don't do very significant effects and but not all the rating agencies are the same.

So the second question I think is simpler. Yeah, does it matter for my portfolio decisions? I think the answer for that clearly is yes. Okay, but you know, should I validate the quality of information giving humankind kind of shocks? I don't know. I don't want to validate anything. I really hope that we're effective in stopping this behavior.

Linda
Yeah, no, that's a philosophical problem, right? So in terms of trying to measure something that you then either don't hope to happen or there is no actual practical way to measure it in a very valid way. So I wanted to kind of get to a little bit of the recommendations that you give in your paper. I mean, one of them you recommended is to increase the transparency, right? So it's calling for greater transparency from USG ratings providers. And I think that's obviously one that we take to heart because we certainly publish our ratings methodology. But I guess one issue

that I have, and I love to your perspective on this, is that you can make all the methodology transparent and actually put it out there. But I think that they're still quite complex.

So many people just don't want to deal with the complexity. So it's easier to say that it's a black box, but the box is a clear box. It's just got a lot of different stuff in it that most normal people aren't going to be able to impact and understand how it works. So how do we think about that in terms of transparency?

Roberto
So there's one source of transparency that you did not mention that I would like to add to this, which is that you specify in each not really defined category what were you pretending on to measure, what are you trying to measure? So that part of the transparency, I mean, you make it very clear, but that's not true in all the rating agencies. And so we actually found some of them that they didn't know exactly what was their measure. They knew it was something about labor. I said, well, do you know, but this is, you know, risk or this is, you know, the violation? They didn't know. So somebody collected the data earlier and then, you know, legacy, they keep the index, but it's not clear that what they do. I think the transparency is such that each rating agency can learn from each other.

Linda
That does make sense, because I hadn't realized that when you call for transparency it was more about learning from each other as raters as opposed to for the investors, because of course, I think the buyers of the ratings themselves do have access and full transparency to the methodology. So that's a really good clarification. So I'm going to ask you one final question.

And this is my desert island question. So you're on a desert island and you can only have one paper on sustainable investing other than the ones that you have actually authored. What would it be? Okay. And, and you can either choose one that you want because it's just so intellectually interesting to you and close to your heart or the opposite, which is that it's one you'd be really happy to use as a fire to, you know, on your desert island.

Roberto
I actually have, I cannot select. On the second one, on the burning, I have like 2000. How much you want to burn over them? No, no, I like the Schuart-Hartzman paper a lot, the one on engagement. This is a paper that talks about this notion of exclusion. So I don't want to invest, I want to sell the stocks of the brown firms, for example, and that that actually might produce more pollution, not less.

Linda
That's fantastic. Thank you so much. I really, really enjoyed our conversation.

Roberto
Same here. I love it.

Linda
So we'll definitely have you for the follow up if you're if you're updating the paper at the end.

Roberto
You will be the first one I'll send it.

Linda
Really looking forward to it. Thank you.

Roberto
Thank you so much for the chance to speak and thank you to your audience as well.