COLUMBIA | ECONOMICS
Program for Economic Research

Transcript

LEVERAGING BIG DATA TO MANAGE EXTREME WEATHER RISKS?

Welcome Remarks, Miguel Urquiola (Columbia)

Okay everyone. Why don't we get started? My name is Miguel Urquiola, I'm the Chair of the Economics Department. It is my distinct pleasure to welcome you all for this evening, for the panel on leveraging big data to manage extreme weather risk. This event is co-sponsored by Capital Fund Management and the Program for Economic Research, which we call PER. The Program for Economic Research at the Department of Economics is the main outfit or organization we use to interact with donors and to interact with multiple organizations that are basically interested in economic issues. We interact with really many types of places. Recently we have also been interacting with places that might let us use alternative data to think about economic issues. And the partnership with the Capital Fund Management, CFM, is a shining example of this and it's something that we look forward to doing more of in the future. You'll hear more about this collaboration basically from Harrison Hong in one second. He is the Director of the Program for Economic Research, and later on also from José Scheinkman who is the Chairman of the Steering Committee of PER. We'll also be hearing from Charles-Albert LeHalle who is the CEO of Capital Fund Management. Finally let me also mention that we are obviously speaking about climate this evening and I thank one more of our sponsors, the Center for Environmental and Economic Policy at SIPA and we also look forward to basically collaborating with them.

Without further details let me turn it over to Harrison for details on what we're doing.

CFM-PER Data Initiative, Harrison Hong (Columbia)

Thanks Miguel for your remarks and for taking the time to be here. Welcome everyone to our event this evening. This event is really an outgrowth of this new data initiative that Miguel mentioned between Capital Fund Management, CFM for short, which is a very large and successful quantitative asset manager based in Paris, and the Program for Economic Research here at Columbia University that's based in the Economics Department.

As part of this data initiative students on campus now will now be able to submit thesis research proposals to use CFM's vast database that covers – it's very impressive, everything from mobile payments, supply chain, to satellite imaging. We think this is a terrific opportunity for the students here and for the intellectual life of our campus. To complement these research projects we're going to host twice annually events such as this, which is the inaugural one, focused on important questions where we think alternative datasets might be very useful.

The topic of this evening's panel seems sort of like a perfect start for this initiative since we live in a time of climate change when extreme weather such as the fires in California or the droughts in Africa pose fairly severe risks for life and wealth. We also live in a time of big data, as most of you probably know, when significant developments in imaging along with massive developments in machine learning have really transformed the industry. So the topic of this evening, Can We Leverage Big Data to Manage these Extreme Weather Risks is the question at hand. The one thing I want to emphasize is most panels typically involve topics where the researchers have already done a ton of research on that topic.

Obviously this is not the case this evening, so I'm particularly grateful to our speakers and panelists for speculating, chiming and providing their insights on the question at hand.
So why don't we get started? Let me introduce Charles-Albert LeHalle of CFM to say a few words and he'll be the one supervising and organizing the speakers that's the first part of the evening.

Introduction, Charles-Albert LeHalle (CFM)
So thank you Harrison. So thank you to all of you to come to listen to this probably wonderful evening and a lot of interesting presentations. So at CFM we are more and more using these kind of ?? data. We see every day a lot of providers looking at ?? with a lot of fancy data, and I think that we wanted to understand more the information it contains of these datasets, and what could be better than discourse with people like José or Harrison to understand the information it contains in terms of economics meaning of all that is into these datasets? So that's why we are very, very happy to make this partnership with PER data initiative to try to have the insight of people like you on what is in all these datasets that we have in our databases, and so this kind of event is a perfect occasion to put the focus on one of the aspects. And here it will be weather data, so we see more and more data providers coming with datasets about weather. We have a special insight on that with Bobby who is from Bloomberg here and will be at the panelists. So here's a challenge today is to understand, at least for us, what is the economic meaning that is hidden inside all these data?
So maybe we can start with the first speaker, with Chris Small.

Presentation # 1 - Christopher Small (Columbia), "Advances in Remote Sensing"

Thank you. Thank you to Harrison and colleagues for inviting me here this evening. This is the first time I've given a talk in this room.
Now moment of truth, right?

Harrison Hong (Columbia)

I'll say a quick word about Chris. I was looking for – obviously this is big data, we were looking for the main expert on remote sensing everywhere I turn. The name Chris Small always came up so we're really thankful that you could be here tonight.

Christopher Small (Columbia)

Thank you.
00:06:39
[Technical problems.]

00:08:59

So this evening Harrison asked me to talk about the intersection of remote sensing, machine learning, and extreme weather. So I'll start off by saying a little about how machine learning is currently being used in weather and climate forecasting, and then I'll say a bit about how it's been used in land surface remote sensing over the past four decades, and then I'll introduce the idea of combining imagery archives of land surface properties with meteorological parameters both from satellites and in situ observations to possibly be able to say something about future loss distributions based on characteristics of past events and past effects on the ground. And then I'll give a very simple example of that based on my colleagues' and my own work in Bangladesh with cyclones, cyclone impacts. And then finally I'll say just a bit about some of the new

technologies that re in the pipe now that will almost certainly influence what we're talking about this evening in the near future.

Okay, first I'd like to start with some bibliometrics, some Scope search results. And you can see the search terms that I used in the upper left, and you see the number of documents as a function of the year that Scope has returned. Notice that it's on a log scale because an order of magnitude difference in terms of number of documents in different disciplines here. And so the idea here is that the simplest search term should reflect mainly the use in the computer science and engineering communities where the machine learning algorithms have been developed. And sure enough that's what you see here if you look at the breakdown.

In remote sensing, primarily land surface remote sensing, machine learning and neural networks were adopted almost as soon as they became available. Very, very quickly after the development of the algorithms people began using them to attempt to do land cover classifications. And so we see a smaller number but a parallel rise, a parallel increase in the number of publications. Now the application into weather forecasting of extreme weather came in significantly later and the number of works is significantly smaller. And if I were to use my own experience I would speculate that the reason for this is that in the physical sciences – my background's in geophysics – in the physical sciences, particularly in the late '80s and early '90s when I was in graduate school we were very strongly discouraged from using what were called black box algorithms. Even though we were using computing very heavily it was all in the form of process modeling, using known physical principles. But a fundamental difference is that back then we weren't drowning in data and model output the way we are now, so times have changed.

Here are the same plots on a linear scale, and so you can see things have really taken off in all three fields after about 2015 in terms of the number

of publications going ballistic. And you also notice a different breakdown in terms of categorized disciplines of the journals.

The other thing I found very interesting is this rose diagram here showing the relative use of machine learning in different disciplines in different countries. And notice here that the use in natural sciences is relatively low for all the different countries, in fact it's comparable to maybe even a little lower than social sciences. If I had to speculate I would attribute to the same kind of disciplinary bias against black box algorithms particularly artificial neural networks.

Okay. So just very briefly, most of the use of machine learning for the weather prediction and extreme weather prediction that I've come across has been primarily the use to identify spatial patterns, to identify features like cyclones and fronts in either past observations, what's called reanalysis data, or in model simulations going forward.

And here's another example, a more recent example, and pretty much the same thing. using convolutional neural networks to identify particular features that are associated with extreme events in model output. The model output now from climate and weather forecasting models is so voluminous that it's almost the only way to do it is with machines, there just aren't enough people.

So another interesting application I came across was the use of neural networks for post-processing ensemble weather forecasts, meaning weather forecasts based on either a number of different models or a model with a number of different parameter settings to give you a distribution of outcomes. And these ensemble forecasts have to be post-processed after being generated in order to reduce some of the dependencies. And these authors here found a significant benefit to using neural networks not necessarily as black boxes but in place of regressive-based tools that are parametric and assume distributions that aren't always the case. And they actually claim, I think convincingly, that

they can actually learn something about the physics of the process by identifying some of the relationships that neural networks identify.
So the way that we have used machine learning in remote sensing is related to the way that the remotely sensed data are used. And there's kind of a dichotomy in the population in the community of people who use remotely sensed data where starting with the same data, where here you see a series of images. This is generally intended to represent the different spectral bands, the different visible and infrared spectral bands that the satellites typically capture. And so if you think of a normal digital camera it images visible red, visible green and visible blue and uses those three primary colors to combine to make all the other colors that we see. And so if you relax your conception of color and extend it to the infrared the way we use most of these sensors is to try and identify the infrared analog of color and relate that to physical properties on the land surface. And so if that's done in a purely physical way using physical process models the result is generally a map of continuous variables, and here are some examples of things that people routinely estimate using optical imagery. And this stands in contrast to using for mapping purposes where the objective is to produce maps of discrete thematic classes or categorical variables like these. And here we use physical process models, and here the approach has generally been statistical and more machine learning-based going all the way back to the 1970s.
So just to show you what I think are interesting results about the results of a large number of studies. In this case the authors of this meta-analysis looked at almost 7,000 publications between 1976 and 2012, and looked at the self-reported accuracy of these studies and tried to relate that to different characteristics of the classifications. And so here you see a number of different classification approaches. Almost all of these would be considered some form of machine learning by most people. There are a couple here that aren't actually classifications that

shouldn't be included but the vast majority of these are, and those of you who are familiar with machine learning approaches probably recognize some of the names here, most of them old now. And you notice in terms of the number of studies by far the dominant method is maximum likelihood classification which has been around since the 1970s at least, and followed closely by object-oriented classification and this general term supervised. Most of these are actually supervised, so I guess this is a case where they didn't specify. But when you look at the self-reported accuracies the distribution of self-reported accuracies of these methods what you find is with the exception of K-means most of these have fairly similar overall accuracies when looked at over a large number of studies with significantly varying dispersions about their means. So we see the typical accuracy of these studies' return is somewhere in the mid-80s percent correct.

Even more interesting is that if you look at how the self-reported accuracies of these studies have changed there's been no significant increase. What you see is that more and more studies are done, the distributions get larger, and the low end of the accuracy distribution drops. The upper doesn't hardly change. And part of the reason for that is that the method of determining the accuracy in these studies is very subject to confirmation bias. And so if you see accuracies close to 100% either you're looking at a very simple classification task or you're looking at accuracies that were gamed a little bit maybe.

So the other thing that we see is as the – this is intended to show that the relationship between the size of the study area and the accuracy, but as you can see there's really no relationship. But what you do see is as the number of studies increases again the lower tail drops. And the closest thing to a strong relationship that they found was looking at the complexity of the classification system relative to the accuracy, and there they found is as the classification task becomes more complicated the accuracy drops, not surprising.

So now I'd like to go and say a little bit about a potential application or class of applications that may go beyond the use of machine learning as it's currently being used in weather forecasting, and that is to try and establish relationships between past events, between the meteorological characteristics of past events, and their impacts on the ground. And so I'll use Bangladesh as an example because my colleagues and I have been working there for the past fifteen years or so, and because Bangladesh is very close to sea level, very densely populated, and gets hit with a lot of cyclones off the Bay of Bengal.

So what you see here are a set of distributions of casualties as a function of different variables related to the cyclones. So maximum wind speed, hour of landfall, residual which is actually the height of the storm surge so the change in the water level as a result of the storm – this is what we've focused on in this paper – and population density of the area where the cyclone made landfall. And as you can see there's not a significant correlation with any of these variables. The closest thing I can see to a correlation here is inverse with population density, which doesn't really make sense, does it?

And so what I want to compare specifically now are two of the more destructive cyclones that have happened in Bangladesh in recent years, and those are Cyclone Sidr in 2007 and Cyclone Aila in 2009. And you see that they have very different values for some of these factors and very similar for others, and about an order of magnitude difference in terms of casualties with Sidr killing more people but Aila doing more infrastructure damage. So what we take away from this is while all of these factors can contribute to casualties and damage it's clear there's not one that's dominant. In other words, all good weather days are the same and all cyclones are unique in their own way in terms of the factors.

So Cyclone Aila in 2009 the cyclone made landfall to the west of the lower part of the delta, and therefore because of the cyclonic wind

direction it drove up a higher storm surge. And what you see here are satellite images collected just before and just after the cyclone struck. The areas that are black are standing water, and the areas that are black before the cyclone, these are aquaculture ponds, most of this area is below sea level and so the islands are embanked and so if the water level comes up too high you get massive flooding, but it also makes it easy to flood these islands to produce shrimp which is a major export for Bangladesh. And so we have a mix of aquaculture and agriculture in this area. This is a mangrove forest down here. And you see after the storm much larger areas flooded, and so that was flooding, embankments breaching on some of these islands and rapid flooding.

In the case of Cyclone Sidr the storm surge wasn't as high but the wind speed was much higher, and what you see here, this is the mangrove and these are the areas where the agriculture and aquaculture happens. Immediately before every place that's not under shrimp cultivation is green, and immediately after you can see the effect of the defoliation and the damage on the ground, the crop damage in the agricultural areas. And so the idea being that if you combine both meteorological observations of the storms with the damage extents and the type of damage that you have going back through the satellite archive for the last forty years we may be able to identify relationships that are not easy to predict using the physical processes that we use to simulate the storms now.

So what's coming up in the future is – actually this is happening now – the one technology that will have a big effect on all of this is what we call hyperspectral imaging. What you're looking at here is what we call hyperspectral cube. The front face of the cube shows the infrared color of the landscape from which we infer the land cover. The sides show how the reflectance is a function of wavelength changes. This is what we use to distinguish different land cover materials. Effectively they're infrared color. These data are collected by aircraft now, but within the next few

years both NASA and the European Space Agency will have satellites up collecting these data from space globally. And the other is high frequency revisit satellite constellations, of which now we have the Landsat constellation plus the Sentinel constellation gives us coverage nominally every five days as opposed to every sixteen days, which is a major benefit for dealing with clouds. This enables us to look at temporal processes as well as the physical characteristics of the land surface. So here you see a time series of vegetation abundance. The different colors indicate different temporal patterns. We can use this kind of information to tell us more about what the physical – like in this case what type of crops are on the ground by how often they're harvested. So I'll stop there.

Charles-Albert LeHalle (CFM)
Thank you. Maybe we can open to one or two questions or we can save the questions for the panel.

Question from Audience: I have a question. When you look at the different factors between beginning and the end, now I know they're not perfect correlates.

Christopher Small (Columbia)
You mean the casualty plots.

Question from Audience: Yes. ??? So you look at them as being one-by-one and you mentioned they are not perfectly correlated.

Christopher Small (Columbia)
I would say they're not correlated at all.

Question from Audience: ??? correlated, that is are there cases in which could you find the combination of those factors and have them come together and then look at how these combination tend to come together, how do they affect?

Christopher Small (Columbia)
That's exactly why I used that as an example. This is the kind of thing when you have more data, largely and more different types of data, you can potentially do exactly what you're suggesting. So look at the relationships among larger numbers of variables that may be nonlinear where the relationship – we may not understand the physical basis well enough to derive it from first principles, but there may be relationships nonetheless that maybe be nonlinear and harder to detect otherwise. In this case we didn't even bother because as you noticed the number of data were very small so we didn't even consider multiple regression or anything. But that's a relatively expensive measurement to make, number of casualties. We can't measure with a satellite. So being able to improve classifications both using temporal and spectral information together hopefully will give us a wider range of variables that may be related, although maybe in a complicated way, but for which we can derive first principles relationship we have to rely on some kind of more stochastic relationship.

Charles-Albert LeHalle (CFM)
So thank you again.
And now with Francesc Ortega we will switch from Bangladesh to New York because you will talk to us about Sandy casualties.

00:29:32
[Technical issues.]

00:31:33
Presentation #2 - Francesc Ortega (CUNY), "Impact of Hurricane Sandy on NYC Property Prices"
Okay. So thank you the organizers for giving me the chance to present my work here.

Okay, three years ago I didn't know anything about this so I've been learning as we've been working on this project, so I'm sure I can learn much more from the audience than maybe what you're going to learn from me.
All right, so I'm going to talk a little bit, give you a very brief overview of my take on the economics literature and then I'm going to talk a little more in detail about the economic effects of Hurricane Sandy on New York City, which is where I've done some work, and then if there's time I may talk about a couple of papers that go beyond New York just to highlight the kind of datasets that they use because I think they are interesting applications and they are in the spirit of what you are trying to hear today, then I will conclude.
All right. So let me start with – it's helpful to start talking about the literature that's not exactly my literature but it kind of sets a nice benchmark here. So there's a very large literature studying the economic effects of hurricanes and flooding events. People have looked at many outcomes ranging from economic activity, measured using night lights [?] to housing values, employment, household income, so on and so forth. There are hundreds of papers if you put them all together. Consistently these papers find that it's bad news to be hit by a hurricane so they find negative effects on economic activity, employment, so on and so forth. However those effects are very short-lived, so in most of these papers

within a couple of years everything is gone, we're back to normal. So if you keep that in mind I'll come back to that idea.

Now we move to a different literature, so this is the literature that speaks more directly to sea level rise. The first group of papers – I'm not citing here all the papers that are relevant because we would have a million slides, I'm just citing very selectively. So the first paper here with Neumann et al. is one of the papers that estimates what could be potentially the economic cost of sea level rise and they estimate that it could be a trillion dollars of economic losses for the US through year 2100. And then some people have been modeling a bit more potential responses from businesses and households in terms of relocation and investments and have said that actually if businesses and households relocate then actual costs could be lower, but the range of estimates is very large and there is an interesting kind of using data to see whether we see any of that relocation going on. I think that's really my main interest here.

Second group of papers have been looking at the question of sea level rise in the context of home prices. Recent papers are starting to show evidence that sea level rise can be measured already in different ways. I'll talk a little bit about some of these papers later on in more detail. But for instance Bernstein co-authors estimate that the properties that are exposed to sea level rise, their definition is that they are within six feet of being inundated, they find that those properties sell at a 7% discount relative to properties that are less exposed than that.

And then some studies have also been showing that elevation within the flood zone makes a difference, and more elevated properties within a flood zone sell at higher prices, again consistent with saleable rights being factored into housing prices.

The third group of papers that I want to talk about, and this is where me and my co-authors have been working, is specifically about Hurricane Sandy and the effects on the New York City area. Now you may be

wondering why these papers which are about a hurricane whether or not in the previous slide which was all about hurricanes and the reason is that what we tend to find here is that Hurricane Sandy had negative and fairly persistent effects on property values and also more recent work on outcomes related to business establishments, such as employment and wages at the establishment level. So then the natural question is what's different, why the previous hurricane studies found very short-lived effects and here we find much more persistent effects? So Hurricane Sandy was seven years ago so persistence means seven years.

So this is our current interpretation of this. We think that it has to do with new news versus old news. If we're in Florida and we're expecting, we know hurricanes happen all the time, flooding happens all the time, if we just get flooded in a common way then that's not going to update our beliefs, that's going to have short-lived effects. However, we think that Hurricane Sandy brought abnormal levels of flooding to New York so then that really was new news. And then people updated their beliefs and now there is a change in fundamentals that could be expected to have persistent effects, and that's our interpretation of what's happening. I'll give you a bit more detail on that.

So now let me talk a little bit about the work that we've been doing on Hurricane Sandy and I'll talk about our work on property values and a little bit about our more recent and still preliminary work on using data on business establishments. Probably in this crowd I don't have to talk much about Hurricane Sandy but there are a few numbers here that help frame what we're going to do later on.

So this hurricane had an enormous impact on the City. So 90,000 residences were damaged by flooding and almost two-thirds of those were one- or two-family homes. So when we come to property sales and different types of homes, that's really where the action is. It's not the twenty-floor high rises, it's the one-family and two-family homes.

All right, now the titles on all this don't show, but that's okay. So the first dataset that we've been using, and after Chris' talk that's obvious to anyone who knows the data but not so obvious to us, to economists who started working on this a few years ago, but it's been very useful for us to use the damage point estimates that FEMA puts out after hurricanes, so this type of data identifies which buildings suffered flooding and suffered damages, and a little bit of information about how much, how much damage.

So damage is estimated there combining aerial imagery and field-verified inundation assessments. For instance, talking about New York City's inundation zone about 20% of the buildings suffered major damage during Sandy. All right, so that type of data, it's a natural dataset to use.

So here is just a quick map – again this is New York so maybe you all know this – but this is a quick map from FEMA about the Sandy inundation area. And you can see that the action is going to be around Staten Island and then the lower parts of Brooklyn, Coney Island there, and the lower part of Queens with the Rockaways. A little bit of downtown Manhattan but downtown Manhattan in all the studies that we've been doing never really plays a role, and that has a lot to do with the type of building structure, much more resilient buildings, it's not single family homes there.

Okay, now I'm going to compare this to the data on damage points. This is very similar. So the damage variable takes on a few categorical values, ranging from less damage to more damage. You can see the orange dots, those are properties that were affected somehow, and then minor damage, major damage all the way to destroyed. And you see that the major damage and the destroyed buildings are found again in Staten Island and then bottom part of Queens and Brooklyn.

All right, so what do we do with this data? Well in our first paper on housing values we take these data and then we merge it with property sales data from the New York City Department of Finance. To do this

what we use is a dataset that was very useful to us and I'm sure to other people working on New York City, is a dataset called Pluto which provides the shape files for all the building footprints in New York, and then for each building you have the associated parcel number, the BBL, which allows you to merge the two datasets and then also to merge to a lot of other information that you can get from the City, like related to schools, related to crime, related to whatever you want, so it opens up a lot of possibilities in social science research.

Okay, so in this merged data for instance we see that 5% of all sales in our period in New York City took place in buildings that had been damaged by Sandy. Okay, so let me just show you maybe a couple of pictures. So this picture we're trying to look at the effects of Hurricane Sandy, so that's the vertical red line on 2012, on sale prices. What we look at now is just generally properties that were affected by Sandy in some way. So here being affected means the property is located in the hurricane ?? zone AB, so close to the water. Now some of these were damaged, others were not damaged, but they had damaged neighbors. What you see is that up to 2012 housing prices evolved similarly for these affected properties to unaffected properties elsewhere in the City. However, after 2012 you see a drop in sales prices that has remained there. So this is about a 10% price drop.

Now let's go a little bit deeper and let's use the damage data. So some of these properties were damaged, others were not. Now here the dash line, we label these damage zero, these are properties in the damage zone, let's say, but not damaged themselves. And for this you see that after the hurricane hit there is a gradual drop in housing values. Now compare that to the red line that's the houses that experienced moderate levels of damage, and then down to very damaged, but that's a very small group so it's not really precisely estimated. But the red line shows that right after Sandy sale prices drop a lot in damaged buildings. Those buildings couldn't be used for a while, they needed to be fixed. But then after that

there was sort of a gradual recovery and a convergence towards this 10% penalty that we observe for the buildings that were not damaged. So here is where we think that really beliefs have changed and even though seven yeas down the road, six years down the road, houses have been fixed, something has changed and we think that's the beliefs about flood risk. So now let me say a few words about the continuation project where we are turning to businesses and we want to see whether businesses that were affected by the hurricane whether we see anything there. So now we need to switch datasets and the dataset that we switched to was something called the Quarterly Census of Employment and Wages. This is a dataset produced by the Bureau of Labor Statistics. And what you can get from there is all business establishments in the City, their exact address and then a couple of outcomes like employment and wages. So we got this data, we geocoded that and then we merge again with the FEMA damage point data about Sandy, again going through the Pluto crosswalk. So what we are finding here is that businesses that were in buildings that were damaged up to this day, so that's basically seven years after Sandy we see that employment is lower and we see that the wages are also lower.

Now here it's important to think about – we didn't think about this too much at the beginning of the project but then later on it really hit us in the face – it's important to think about what's the unit of observation here? So the data's about business establishments, an establishment is a company in a given location. So if you think about the company and you track the company over time, clearly you're never going to have a persistent negative effect on the company. If you were good at making pizzas you're going to relocate and you're going to be equally good, it's not going to have a persistently negative effect. However the buildings, that's what we track here, and the buildings and their income-generated potential, that's what's going to be negatively affected in a persistent manner. That's basically what we show in the paper.

Now when we look across boroughs we find that being damaged really negative effects on businesses that were in Brooklyn and Queens, not really Manhattan. That's where we're doing more work now but we think it has to do with the type of building. Given the time I won't go into detail.

Now let me just say a few words about papers that go beyond New York City so that you see the type of datasets that they use. So this is a paper on sea level exposure around the country and the effects on property values. This is the paper by Bernstein and co-authors, that's the title you don't see. So they use Zillow data which they merge with the NOAA sea level rise calculator, and they use that to show that more exposed properties to sea level rise sell at a discount. Now interestingly they find these effects not on all properties but mostly in non-owner-occupied housing, which they interpret as this is about sophisticated buyers, these are the ones that sort of update their beliefs or see projections and then can figure out what's going to happen and then that's the sub-market where you find the effects.

The other paper I wanted to highlight in just a few seconds is a recent paper that looks at Hurricane Harvey and how households are coping with financing the fix-up of their houses. So this paper what they do is they take the same damage point data that we used for Sandy but for Harvey and then they merge that with a dataset with the beautiful name of See Car Y 14M [phonetic] which has data about credit card accounts and then what they see is people are borrowing on those cards and for a few months that's what they use to survive and sort of fix their houses until the FEMA insurance money comes in. So that's another clever way of merging datasets.

And if I have a minute, so I can throw out there a couple of guesses on what are we going to see going forward. So the first message was probably we're going to see a continuation of the trend of falling housing prices. Studies show that financial markets do respond to climate change

in different ways, so Harrison's papers on droughts shows that and then Schlenker and Taylor, if I remember correctly, about sea level rise. But among homeowners people who have surveyed that population they say that there's a lot of inattention and belief heterogeneity so people don't really update their beliefs automatically or the way we think about in economic models. Here is where we think that direct experience makes a difference, and our papers on Sandy and people who were affected, that's what's going to bring these homeowners to update beliefs. I'm going to stop here. Thank you.

Charles-Albert LeHalle (CFM)
Okay, so maybe we can save questions for just after the panel. And so now Joséphine Gantois. So maybe we can try to do better than the two other ones.

Presentation #3 - Joséphine Gantois (Columbia), "Applications of Remote Sensing for Studying Socio-Economic Outcomes"
Good evening everyone. Thank you for being here, and thank you Harrison for the opportunity of joining that conversation. So I'm a PhD student here in sustainable development and I work across economics and ecology. And my work falls under the very broad umbrella of how humans and ecosystems interact. But part of my research agenda is more methodology-oriented and is about using alternative sources of data along with either a causal or predictive framework to monitor the climate and ecosystems, and so I thought I would just describe briefly two of the projects that I work on that are related to the topic, and then some thoughts for the discussion.
So the first project that I wanted to mention and I don't have titles here, but I guess it doesn't matter, so it's work with Wolfram Schlenker and in

that work we try to use tree ring data to reconstruct high temperature extremes in the past, so it's not forward-looking, it's backwards-looking. The reason why we focus on high temperature extremes is that this is something that has been proven to negatively affect crop yield, and so if we can reconstruct in the past those temperature extremes then we can potentially dig into the relationship between this type of weather extreme and societal development through the channel of food production. And so here the map is just showing you the locations of the raw tree ring data. And right now we have some sort of a proof of concept in the sense that if you look at the impacts of temperatures, so here on the X axis, the range is between 0 and 40 degrees Celsius, so if you look at the impact of temperature on annual tree growth you see that at high temperature you start having strong negative impacts on increasing temperatures on tree growth. And to give a sense of magnitude if you look at a year and you replace a day felt at 26 degrees by a day felt at 32 degrees then you get an average a decrease of 1% in average tree ring width. I forgot to mention this curve is only for the southeast. We looked separately at different climactic zones.

So this is interesting because then we show as well that we do see a strong signal of this type of high temperature extreme in the tree ring data. And it has similar predictive power compared to other climate variables that are usually reconstructed from the tree ring data, like average maximum temperature over the growing season, ?? as much as drought which we would expect physiologically from the point of view of the trees. We're now working on the reconstruction itself and so reverting the relationship that we estimate here in order to reconstruct the temperature measure from the historical records in the tree ring data. So that's for the first project.

The second one is maybe a little bit closer in nature to – I had everything in the titles – is a little bit closer in nature to the topic of the panel, and it's a project where I try to monitor flowering phenology using satellite

data and deep learning. What I'm trying to do here is that I want to see if I can capture year-to-year variations in the onset of flowering at the large scale – I'm looking at the entire eastern half of the US – at a fine spatial resolution. Why I think this is interesting, I think that this is interesting to capture this type of variation in wild areas because I'm interesting in things like the impact of climate change on the synchronization between plants and pollinators. It's interesting in agricultural areas as well because flowering is the time during which a crop is most sensitive to climate, so for all the people who try to understand that impact of climate change on food production it's really important to be able to track this sort of window of sensitivity over the years and to include it in the food impact of climate change. And it's interesting in urban areas because there are links between flowering and allergies and things like that.

And here I'm showing you a map from the National Phenology Network. It's a spring index that represents first bloom during the year 2018, so the redder the area the earlier the bloom occurred during the year, the greener the area the later the bloom occurred. And this map is purely a temperature model, so basically what it captures is the time when enough warm days accumulated that flowering was triggered. And what I want to try to do is get a large scale measure that captures more of the local variation in flowering phenology and that's why I turned to satellite data. And the reason why I bring in deep learning is that satellite data gives you access to many features, depending on the satellite data you have access to a number of different spectral bands, but more important for me it gives me a time series. During a year I have several snapshots of each area. And the intuition for flowering is that at the time when the onset of flowering occurs you have a sharp temporal transition in what a landscape looks like but it's very difficult to specify that transition a priori in particular because it's very varied. And here I have just one very obvious example. This is planet data over the national ?? monument in

California. This is at the end of the growing season in February and then this is during the super bloom of 2017 in March. And this kind of transition in the most visible, but in most places it's not going to be as obvious. And so I'm turning to deep learning in the hope that the model can differentiate between different types of transitions in different areas in different years.

Specifically I used a neuron network that has a convolution in time because I want my model to detect temporal patterns that are particularly relevant for detecting that transition, that flowering transition. It's really a work on progress so I'm still fine tuning the structure of my model, still protesting [?] part of the data, but it's been extremely useful in particular for me to get a better sense of all the potential usefulness but also all the limitations of the remote part and the deep learning part.

I just wanted to make a couple of extra points, more related to the topic of the panel, so using big data and machine learning to manage extreme weather events in real time.

Part of the underlying question today I feel is how do we prepare our financial and economic systems for all the coming climate change-related extreme weather events? And because I work in sustainable development and achieving sustainability is inherently a long-term goal I tend to interpret that question not so much as how can we better predict and monitor the next big weather event in order to smooth out the impacts on tomorrow's portfolio return, I tend to interpret it more as how do we build resilience into our financial and economic systems in the face of climate change? And there are interesting initiatives like the ESG Sustainable Impact metrics, so the type of metric that tries to tell you for each company how much of its revenue comes from activities that promote sustainable development goals, that promote good governance. This type of initiative is interesting but I think that even more relevant for today's discussion is the question of quantifying a company's exposure to

climate risks, which is potentially very different because obviously it's not because a company works towards sustainable development, that its assets and people and operations are at risk or not with regard to civil rights, hurricanes, etcetera. So the question becomes how do we use big data and machine learning in order to both characterize the expected distribution of extreme weather events and also characterize the climate risk of individual companies?

And there's a very interesting project that tries to do exactly that by Julian Nyarko, a former post-doc here at the Law School and Eric Talley, a professor at the Law School, where they put together climate exposure data with asset returns from public companies to try to gauge exactly for individual companies what is their sort of financial resilience in the face of climate risks.

And the last point that I wanted to make is about the choice of a spatial scale. I tend to me let's say obsessed by whether the choice of my spatial scale is relevant for the problem that I'm dealing with, in part because if I want to be a little bit ?? cultural I work in between ecologists who for many if not all of them work at the very local scale trying to create detailed local expertise in particular because of data measurement constraints, and in between economists who again I want to be ?? cultural tend to work at larger scales because it's when you have variation in your data, when you have numbers in your data, that you can take advantage of ??? models and extract interesting patterns.

And so I generally like when you have a clear political [?] because then the scale of the policy can dictate very naturally the scale of your analysis. But when I think about extreme events it gets quite complicated because of the correlated nature of those events, and in particular thinking about the spatial correlation.

There was a very interesting event at Columbia in the spring about correlated extreme weather events organized by Colin Raymond, Radley Horton and others. And it was fascinating but also a little bit

overwhelming to get a sense of the extent of the temporal correlation and the spatial correlation for those extreme weather events.
And why I'm talking in the first place about the question of the spatial scale is because as soon as you're working with satellite data as a proxy measurement and with supervised model the big sort of ghost in the room is the ground truth data that you need to connect your proxy measurement with what's actually happening on the ground and so if you need – I mean the wider the scale, the more diverse the area that you're looking at is going to call for more extensive ground truth data, so that's something as well to keep in mind when you're using these kinds of tools.
That's it for me and I'm looking forward to the rest of the discussion.

Charles-Albert LeHalle (CFM)
Thank you. We've got a little bit of time for questions.

Question from Audience: Yes, I have a question, but sort of between Chris' talk ?? about do you think it's possible to collect enough events that – this is ?? Chris ?? is there kind of some metrics that would predict or explain losses to property, losses to business? ??? the Bangladesh cyclone, do you think there's a possibility ??? these events across the globe and to kind of come up with some prominent [?] features that might explain the types of losses and kind of flesh out a little more the conditional loss ???

Christopher Small (Columbia)
I think in the case of cyclones that maybe in a place like Bangladesh where they have a lot of cyclones ?? the impact of cyclones is something you can easily see from space. But say in the case of New York we've been hit so rarely there's not a lot of information in the past. So I guess

it's how extreme the extreme event is. If it's truly extreme by definition then you'll have a small sample no matter what.

Question from Audience: ?? across the world?

Christopher Small (Columbia)
In the case of cyclones I think that the location is so important because it's a function of the circulation of the ocean and the depth of the shelf water, the river circulation, and I don't know that there can be a generalized model that would apply to every place, I think you'd almost have to tune it by location.

Charles-Albert LeHalle (CFM)
So maybe I've got a question for Joséphine. So when you took ?? to using a neural network, so even on the last slide supervised learning or unsupervised learning, did you ??? what should be the input for doing like supervised learning and should it be like multitask because you want to predict – you want to extract information that will be for a lot of different ??? so did you ?? about the good architecture that will enable to extract something from satellite images but with different targets in mind for different ???

Joséphine Gantois (Columbia)
You mean for predicting different outcomes?

Charles-Albert LeHalle (CFM)
Yes.

Joséphine Gantois (Columbia)
Right now I'm starting with one.

Charles-Albert LeHalle (CFM)
Sometimes it helps, yeah.

Joséphine Gantois (Columbia)
?? right now choose to focus on my outcome ?? choose ?? architecture. ?? exactly the same architecture ????

Charles-Albert LeHalle (CFM)
For instance, the NLP [?] I mean ?? processing doing multitask is far better, more robust ???ized.

Joséphine Gantois (Columbia)
Maybe it's ?? but I don't want to complicate – lose a sense of the theoretical intuition about what's happening, which may not ?? performance. Is ?? understanding why I'm getting the pictures that I'm getting, why am I ???

Charles-Albert LeHalle (CFM)
Okay, thank you. So now we can switch to the panel, so José.

Panel Discussion

José Scheinkman, Moderator (Columbia)

The structure of the panel is going to be that we're going to have some people sitting here but the speakers and eventually the whole audience should also intervene.

Our members of the panel all come from financial markets one way or the other. So we have James Lucier from Capital Alpha Partners, and then Bobby Shackelton from Bloomberg, you're here, I've seen you before, and Kévin Sin Ronia from CFM. Please to give each one of the panelists maybe three minutes, maybe five minutes if you have a lot to say and to start.

Panelist, Bobby Shackelton (Blomberg LP), (presentation)

All right, thank you. Welcome everybody. My name's Bobby Shackleton, I'm from Bloomberg, I manage all their mapping applications, alternative data and data products, so anything that has a location is kind of in my wheelhouse. I've been working at Bloomberg for about four and a half years. I was at another financial technology firm before that, built the first web mapping application for Wall Street about twelve or thirteen years ago. And I'm now at Bloomberg really shifting my focus to weather and climate and impact on markets. And I think some of the conversation that we heard earlier was spot on, where the impact of weather in the financial markets is similar to ESG which we heard but actually the complete opposite. ESG is the company's effect on the world. What I'm trying to do and I think what we're talking about is the world effect on the company or on the markets. So it's kind of yin and yang but they have a lot of overlap.

The challenge is this isn't really been done today in the financial markets. There might be a shop or two, one off [?], that are trying this, but they're really aren't a lot of methodologies or practices doing this yet.

But I've integrated with the UN, many of the biggest banks and many of the hedge funds and quants around the world and we are now developing methodologies, white papers and tools to do just this. So I'm excited to say that we have started to quantify the weather to companies where we're actually able to correlate that to stock price impacts, whether it's against snowfall, cyclones, and we're testing and back testing against other variables.

I won't dig into too much into this. You guys kind of know the state of the climate, but I think from the financial marketplace what I wanted to explain was the way that the players that I've been speaking to, a lot of hedge funds and quants and banks and others, they kind of bucket weather and how they want to consume this data in three different ways. One way is by company, they want to understand the stock price effect kind of in real time from an investment standpoint. That's kind of in the near term. In the medium term they want to look at risk, risk against their portfolio of what companies they have, maybe what assets they even have if they're a bank and they have loans out too. So that's kind of more in the medium term of the quarter. A hurricane could strike and an earthquake could trigger. Let's simulate or let's understand the impact of my portfolio if that event happened. There's no solution in the world for this right now. This is a golden opportunity in the marketplace for a fixed income sort of credit risk type of tool to be integrated for environmental risk. This is something that we're trying to get towards, but it's going to take years of studying and truly understanding how much an impact does, what the impact actually is, so that we can quantify. So I'd say at this stage we're more in the statistics realm and now starting to study so that we can do that risk. Kind of the third area is the really forward-looking climatology risk. This is that 2030, 2040, 2050 kind of outlook of climate risk. Many of the banks are starting to look at this now, and actually Michael Bloomberg chairs something TCFD, the Task Force for Climate Related Financial Disclosures, where 900 companies, banks,

utilities, around the world are volunteering to report their impact or the weather in the world's impact of climate on their holdings and loans on a forward-looking basis so that they can manage physical risk and transition risk. This has only been going for a year and a half now, this is a new organization, but I'll touch a little bit more on this here at the end.
So I traveled the globe several times over the last two years talking to hundreds of hedge funds and quants and insurance agencies, ESG agencies, and I've got a lot of intel from that. You see a lot of buzzwords around here. What these hedge funds want to do is they want to put weather as a systematic – that means kind of an automatic feed of data that can go into a black box to execute trades or decisions. This is a real time kind of tool to make money off of, that's really it, so it's not really managing risk, it's leveraging the opportunity to make money.
I touched on some of the climate risk in TCFD. This is a really growing initiative. There's white papers. This went from last year 400 companies to this year 900. There's more on that a little bit later.
The term alpha is a very financial term. Investors care about weather right now because it's not really a product or quantified as a signal out there. So alpha means it's an opportunity to make money, it's new, it's unique. And so they want to seize on any alpha alternative data to make the most money possible, and they usually keep that to themselves and not tell anyone. That's the challenge I have because I want to democratize it all for everyone.
And stock price. What we've been talking about here a lot doesn't speak to the financial community. Understanding the impact of kind of a high large macro region doesn't really affect a stock price or a bond, maybe a currency but that's not like the big trading area. So quantifying this to equities and stock price is the golden nugget that we're trying to achieve here.
And then obviously disruptions in supply chains and commodities are kind of known entities kind of within the financial community, but what

I'm trying to do is connect that to the equity kind of security level that's the much bigger wedge of the pie of the financial community.
Here's an example of some of the work that I did with the UN and the banks last year. This was published, you can Google it, TCFD physical risk, you'll see this white paper, it's 80 papers. 16 pilot banks got together, co-authored this along with me and two other climatologist consultants. And this was kind of the feature case study in methodology and we brought a tool into the Bloomberg terminal, the financial tool, where I took climate data, forward-looking production loss data, and was able to connect that to physical assets like power plants. And then doing merging techniques of spatial merges, and most importantly having the ownership of the company to the physical asset, so more than the footprint which the buildings are to do a New York analysis knowing that specific.dot is owned by Con Edison, Southern Company, etcetera, so that we can roll that up and make that company relevant. So this was a first step where we started to quantify the weather to climate, outlook towards the 2040, 2050, based on different scenarios.
And what I wanted to say on this is I think this is the precursor for regulations that will ultimately force other investors and companies to actually report this and then that actually becomes something that's really apparent. Once the climate is being evaluated and it follows the money I think you'll start seeing this be more widely adopted around the world, less from the government and policy, the money will drive this.
And then lastly this is my area. You can see down there in the bottom right that's like a little snippet of my mapping application, you can't see much there. But this is one of the products, a data feed in business that I am quantifying weather and not just extreme weather, also standard weather, because we're seeing impacts based on snowfall, rainfall, temperature, soil moisture, things like that. So the weather goes from non-extreme to extreme to long-term climates. And so I'm trying to deliver this and these feeds the second the weather changes via an alert,

via news feed. And then from that news feed or alert you then can go into the map and see the picture and tell that story or understand it more, or deliver that data by company raw directly to the black box or to that hedge fund so that they can build systematic trading tools. So that's kind of the topic or frame that I wanted to put out there is there's questions about the market and how investors are looking at weather, the type of actions that are happening in the marketplace around regulations or TCFD and climate. And then the quantification of weather to companies and things that Bloomberg and other companies are trying to do.
That's it, thank you.

José Scheinkman (Columbia)
Thank you Bobby. We're going to continue and leave the questions for later.

Panelist, Kévin Sin Ronia (CFM)
So I'm Kévin Sin Ronia. I'm working at CFM as vice-president in the alpha team and I'm basically the guy who is relieved [?] you're making money with weather as Bobby was talking about. So basically at CFM we use satellite data and weather data to do trade weather uncertainty in all the possible ways. We've used it for a long time to actually trade commodities, whether it's oil, grains or even some soft and metals. And of course now the challenge as Bobby was saying is actually how you use this to trade stocks these days.
So I wanted to actually give you is a bit of an overlook of what kind of challenges we're facing today and what are basically the next interesting things to look for in that field, and maybe a bit more what is it to actually find alpha signals out of this kind of data?

So basically if you think about weather it's a really main risk factor in the market. For instance, if you look at grains it's like 30% of price changes that can be explained by weather, so this is something really important for us. And actually it impacts – if you look at weather patterns there are really like three different scales at which we're looking. So the first one is really the everyday changes that actually have an impact on the prices, so this regular price movement that can be explained by regular weather movement. And then you can move to bigger patterns, like for instance extreme weather events, hurricanes or snowfalls or even like fires that can be seen from the satellites, and that really needs to be seen and predicted to actually account for huge ?? risk. And there are even people now who are actually looking at solar eruption data to actually predict long-term weather cycles. So it seems that long-term changes in the seasons can be explained by those data.
So if I go back to the challenges we are facing, what's the big challenge in predicting assets in the business with weather is that we're trying to predict a macro factor which is a commodity or even index or maybe a currency out of really diverse data, that is weather in many different locations. And the question is how you aggregate all this information across the globe to actually predict a single product is the main challenge.
So the idea how you geomap local information on the weather or on some cornfields or some local events to actually predict one single product. So this geomapping problem is really important. For instance, for stock you have to know where the stock is located, which means where all the plants for a given company or which are the regions that are going to be impacted to actually be able to retrospectively see an impact.
And probably I don't have time left so the last thing I want to talk about is the way we do learning on those data is quite special. I think Joséphine talked about that. There is this problem of spatial and time dependence. And the weather factors are pretty special because you have

to understand what is the special area that again actually impacts an asset, but also the fact that local weather is going to impact a given asset only a certain period in the year, and therefore you don't have factors that are stable in time the way you do learning. So you really have to account for this variability over time to do predictions, which ends up doing some kind of learning that is way different from the regular algorithm you're using.
So I think I can stop there.

José Scheinkman (Columbia)
Thank you very much Kévin. James Lucier is our last speaker from the panel. I'm going to ask for the people and we're going to have questions too. James.

Panelist, James Lucier (Capital Alpha Partners, LLC)
Well thanks very much to Columbia and the Maison Française for hosting this terrific event.
I'm actually going to do something very dangerous, which is that I'm going to completely throw out the window the comments I had come prepared to say. I'm more of a specialist in climate policy and trends in climate policy in Washington and state capitals around the country. However, I'm certainly not a data scientist. But it occurred to me that information is actually data in context. And much of what we've been talking about this evening is the problem of taking data such as tree ring data or locations of industrial facilities or things like that and putting it into a context that is meaningful. And so it's difficult to find a way to turn this data into information where the context tells you how to interpret it precisely.

Now there's one area where this happens to be much, much easier than studying the world in general, that is municipal finance, because in the muni finance world we're talking about specific geographic areas, cities and even neighborhoods, or we're talking about specific facilities. And these are systematically tracked by fixed income investors who have tremendous databases about everything there is to know about the city, starting with population, with tax collection, with economic activity, with really everything you would want to know from geospatial mapping, you know it about this city. So what we're seeing is actually a tremendous amount of work being done right now where the context is clear. What happens in Miami-Dade County, what happens in New York City, what happens in other parts of the country where you have exposure to severe weather? So I'd suggest that you look to muni finance as an area where because of the simplicity of the problem where you can identify the city and all the other relative criteria right away and also get an immediate market valuation, you have a laboratory for testing this. What we're seeing in muni data right now is that for the shorter term durations it's really not making much difference yet, it's mostly a longer term effect. But you would think that cities that are heavily exposed to sea level change or to other climate threats might have a lot to worry about, but investors at the same time are actually evaluating cities not only on their outright exposure in terms of economic risk, employment, businesses, etcetera, but also on their resilience plans, their resilience spending, and what they're actually doing to solve the problem. So it's interesting to see that both sides of the equation, the risk and the risk management strategies are beginning to get very close attention in the financial markets.

Two further things I'd throw out are studies I'd like to mention. In particular my friend John Mousseau who is the Director of Research at Cumberland Investments. He happens to be in Sarasota, Florida, a place where you think about hurricanes quite a bit, not to mention sea level

rises. He's done a study of bond values comparing treasury ten year yields with those of municipal bonds, and he's actually found an interesting effect that in many of the more populous areas you get a result that's the opposite of what you would expect, that is bond yields decrease after hurricanes, which means the capital value, the price of the bond, is increasing. He doesn't do detailed work explaining why this might be the case but perhaps it has to do with the stimulus that comes from federal aid to rebuild or just the fact that places like Sarasota, Florida seem to attract people no matter what, regardless of the storms. I'd also throw out the issue of insurance and the property and casualty insurance companies are also really experts in assessing context and risk. You would think that extreme weather phenomena would discourage these companies from investing, but the reality is that a hurricane can actually be a terrific profit driver for an insurance company because even though they have to make a significant payout for the hurricane or the disaster they have more pricing power afterward and they make it up over time. And so what we're seeing is that bigger buildings are being built, more people are moving in, and that these coastal areas are becoming more and more attractive to people that want to live in fancy buildings.

So one thing to look at in your research might be whether or not this trend is sustainable. Can you have this arms race where insurance companies rebuild in the places where the catastrophe ?? before, they rebuild bigger and better, perhaps more resiliently, and then the next disaster comes along. Can this cycle of insurance companies make more and more money off the insurance to sustain the overbuilding issue?

So I'll leave you with those three things, the muni bonds, my friend John's study of hurricanes, and the role of insurance companies in this situation.

José Scheinkman (Columbia)

Thank you James. All three interventions were extremely good and I think we learned a lot about it. I want to invite the previous speakers if you have something to add to the topics that were raised here or we can go to the questions.

Francesc Ortega (CUNY)

I want to ask James if he had any thoughts on the National Flood Insurance policy and the role of the tablet [?] insurance, because it seems to me that for most residential properties that's the most relevant one, if you have any thoughts to share with us.

James Lucier (Capital Alpha Partners, LLC)

Yeah, I don't have a lot of thoughts on that. Often public policy supports people in their bad behavior and that's probably where we are right now. There are a lot of debates in Washington about how the National Flood Insurance Program really ought to take into account that some areas are more dangerous and that rebuilding ought to be discouraged. I should have a better answer for you. My impression is that we're not really seeing a decisive change in that debate yet. Likewise with insurance regulation most insurance companies are state chartered, fifty different states, fifty different regulation centers, and so there doesn't seem to be a national trend yet. I mean the companies and the states are all working toward better climate disclosure, but the fragmented nature of that regulatory structure, fifty different states, is impeding progress.

José Scheinkman (Columbia)
So I think we're ready for questions. Bernard?

Bernard Salanie (Columbia)
???? insurance companies ?? insurance ???? they tend to place the matter ?? climate ?? satellite companies ???

José Scheinkman (Columbia)
Of containing the risk, right?

Bernard Salanie (Columbia)
Well at least segregating it.

José Scheinkman (Columbia)
Yeah, you segregate your risk, yeah.

Bernard Salanie (Columbia)
And not subject to prevention regulations.

José Scheinkman (Columbia)
Anyone wants to talk about that, or are you aware of that?

John Lucier (Capital Alpha Partners, LLC)
No, not really.

José Scheinkman (Columbia)
That's a good remark. There was a question out there.

Question from Audience: Yeah, for Bobby, so in tandem with alternate data platters [?], things like credit card data, ??? data, how does weather fit on top of that in terms of predicting what's going to happen ???

Bobby Shackelton (Bloomberg LP)
Yeah, that's a good question. So I think they're enhancers. Stay tuned for a white paper that's being published on Bloomberg.com where we've quantified the impact of snowfall to retailers. And so for example, Macy's is here in New York City. And my hypothesis was we could predict quarterly earnings based on snowfall. And what we found though was that the sensationalism of an event like snowfall, everyone runs to the grocery store. The same thing happens, the stores don't – well, the investors think that the people won't go to the stores, so you start to see those stocks slide and change. And then you see the day of the storm or the day after it flattens out. And so from an investor you're able to use that signal, mostly sentiment and fear of perception of people that don't have quantified data that are already pricing the weather impact to stocks. So we're already seeing this effect kind of go up and down from storm to storm and we're able to ?? test that with sharp ratios and like a real trading strategy from there. And so that you can do in isolation with locations, weather, information. If you layer on top, say, credit card transactions and other things like that, which are normally anonymous, like we can't really find as a data provider a specific company's transactions usually, it's usually rolled up to a zip code, but you could again join that to see some sort of generic attribution to that store, possibly join that…

José Scheinkman (Columbia)
Macy's must have their own zip code, right?

Bobby Shackelton (Bloomberg LP)
That's actually a good point, that's a very good point. But there's a few companies that maybe… But I would say that would enhance the signal. So imagine if weather with stores, with sentiment, you had a 1.47 sharp [?]. If you're able to attribute credit card transactions, foot traffic, you would hopefully have another layer of signal to sharp that signal and to add like a level deeper of delta that you'd have an advantage on somebody else that's just doing weather by itself, or to sticking their finger up to the wind or just watching the news. So I see the more sources that you can layer in on top of that the higher the value. But there's potential value of just a single source of alternative data, but I think that alpha moment often will start with one source and then that'll become a beta because everyone will use it, and then you have to find another source and layer that in, and then that is kind of vicious cycle of investors chasing alpha and alternative data.

James Lucier (Captial Alpha Partners, LLC)
I can certainly attest from my own experience that I've been on many financial markets morning calls where the retail analyst comes in and talks about snow or rain or projected lousy weather and talks about how this is going to hurt the stock in this quarter. So it's very interesting to see that no one has really recognized they ought to be betting against that play.

Kévin Sin Ronia (CFM)

?? implement the answer. Yes, we tend to see some relationship between not even extreme weather but regular weather and consumption. And actually there seems to have a relationship between macro variables like consumption, inflation, GDP, and regular weather, the way it impacts the mood of people. Of course you can easily relate that to some sectors, some sectors are clearly exposed to weather and there it is quite easy to do prediction, but even in a global scale it seems to have an impact on consumption, therefore you can relate that to credit card data where you see the cycle of consumption of people when your dataset is big enough.

Bobby Shackelton (Bloomberg LP)

Just real quick. One thing to add. I think a stock trader doesn't actually care if that store per se, that specific store, is profitable, they could care less. The corporation that owns the stores they actually might care about those operations. So the investor, it's all about sentiment or the way the markets are reacting to a certain thing, that's the betting signal. And then at the end of the quarter when that earnings call is on and they report what's happened, now the totality of the earnings or the revenue sales from that might come into effect, but for three months until that very day there's no understanding of what the revenue that's coming in at that given company or store.

José Scheinkman (Columbia)

Thank you. I'm going to return to the audience, but I'm going to use my privilege as a chair to ask a question.

So we talk here about a two-way street. One is how companies' policies affect weather, and that's what interests a lot of economists and a lot of

people get interested in that, and the other is how the weather affects the companies' performance. But I think that the company policies can also affect the profitability depending on the dynamics of the weather, if the weather's getting worse in terms of climate change and so on. And I think that at least two avenues for that. One is consumers. So I come from Brazil and I actually am connected to a company which does transportation of agricultural output, I'm on the board, and we're very worried about the fact that if people in Europe start saying this is contributing, agriculture in Brazil is contributing to climate change, which may or may not be true, but that's just a question of sentiment, that it will affect exports and that's going to affect the amount we transport, because we basically transport stuff from the interior of Brazil to the coast so it can be exported. So that's an example. Another example is some work that Harrison and I have been doing, we try to see how companies which have a good reputation in terms of their effect on climate change do better, in this case with legal procedures but also with regulators, I'm sure the results are there too. So there are at least two ways in which companies' policies, what they do, how much they affect weather, can then affect their own performance.
So what I'd like to know is that something you guys look at, to see whether companies that have certain particularly, let's suppose, scores in terms of environmental policy, how they get more affected when certain bad things happen, or…

Kévin Sin Ronia (CFM)
We haven't looked at that yet. We looked at the policy of companies separately and see if that has an impact on their performances, but we for the moment don't see any relationship between this and weather.

Bobby Shackelton (Bloomberg LP)

It falls a little out of my wheelhouse at Bloomberg where I'm focusing on the world affecting the companies. There's another ESG group that's really focused on kind of that side of it. But what I'll say is Bloomberg and other financial companies and data companies are creating scores and statistics that measure, say, the governance and the policies and they could give them an F or a B or an A or some sort of quantitative value to try to resemble that. So I'd say it's happening. I can't speak to say the exact effect necessarily on the stocks or its profitability, but I think the general sense that I get from my travels and talking with clients is from that higher executive management there's a real need to become more of an ESG-friendly company or investment shop, meaning you'll have pools of stocks or companies that you invest in that are more socially and governance better, and younger generations, younger investors, tend to gravitate towards wanting to be in those pools and ETFs more. So I think you're seeing a society kind of shift that's also driving that that's kind of from the ground up, and then you're seeing the executives from the very top wanting to look good on paper at the same time. They just want to make money but they're also kind of doing a good PR stunt. But now I think that's converging where they're able to do both.

Question from Audience: You had said that you want to democratize these models that you're working on. And Mr. Ronia, from the other point of view, you're providing this for your own business. Can you describe the world in which either happens, democratization or the opposite? How does that affect business and how in terms of that effect ?? sustainability in its investing ?? people?

Kévin Sin Ronia (CFM)
First I can answer I'm not against democratization of these models, as soon as I always have an edge on the rest of the people. I think that would my answer. What was the second part of your question?

Question from Audience: If there is democratization or there isn't, let's just say one or the other, how would that affect businesses? And presuming that these models are going to make business be more sustainable, because that's profitable, that in turn affects the environment and our life on Earth, what are the different worlds that we live in those different contexts with democratization ???

Kévin Sin Ronia (CFM)
Weather models have been used by companies for years actually. I mean if you think about it marketing models of ?? or beverage companies, I mean they include accurate weather data for more than twenty years. So at a level of businesses they are clearly impacted but…

Question from Audience: ???? models don't exist yet you said that the things you're working worked on really don't exist. If you develop them ??? or given to everybody what's the difference?

John Lucier (Capital Alpha Partners, LLC)
Well the point is not everyone can use the model, not everyone can use the tool. A hammer does no good to someone who needs a screwdriver. So you could make all of these exotic models available to the general

public but if they don't have trading strategies available or simply can't understand them it doesn't work. From a company perspective, companies are in something of an arms race with the financial markets themselves. They're using big data too, especially logistics or utilities or energy companies, people with big networks. They want to get ahead of the curve, they want to beat the financial markets, so they're developing internal big data models that are totally unlike the wonderful things we're going to see from Bloomberg, but it shows companies using a tool that is applicable to them. So the question is what's a tool that the ordinary person might use? That might mean following ESG ratings, it might mean using common sense, but the fact is we're talking about very specific, specialized tools that simply aren't going to be useful to everyone.

Bobby Shackelton (Bloomberg LP)

The way I see it, I see is there's three paths that can live in parallel and that's the beauty of our country. So I think from the investors' standpoint we're capitalists, they want to make money. If they find an edge that's to them to do that. I think from a distribution company like Bloomberg, or an information company, we also want to make money, and the way we make money is to sell it to as many people as possible, and so that's what we want to do. And then while that's happening you have policies like that TCFD or other things that are being enacted maybe by volunteering first, and then at some point will be mandated. And then what'll happen is the data companies like us will create a product, everyone will have it, and then it's like an academic science practice where how do we up that democratization to the next level? Because the client will now try to find another delta edge on top of that. So it actually helps build and I think make it a better like innovative platform, and that what follows right behind that after that's all

happening – it always comes behind it – is a regulation. And then once a regulation comes in place it's fully democratized but we'll still look for advantages. I'll try to democratize advantages in like the scientific and academic community. Hopefully that just continues to build.

José Scheinkman (Columbia)
Charles ?? who was ahead of you, Harrison, I'll take first him and then you.

Charles-Albert Lehalle (CFM)
I have a question about ??? and you are talking about alpha. I would like to know if ?? because I heard Joséphine ?? saying okay, in fact weather is something that is highly like a ?? machine. So given a place in the world and given ?? the way weather is affecting ??? different. So maybe you can infer some combinations between ??? that now ??? by weather. Do you see using that not only to say it's kind of alpha ??? but more in terms of ??? a layer that is more contextual and ???

Bobby Shackelton (Bloomberg LP)
I see it as a journey and I see that insurance markets are already doing this, they have risk models and they're able to kind of predict. But it's usually kind of that large generic macro kind of level. I think some of the instruments that the financial community cares about at the company and security level, like the products that I'm building and others are probably starting to go towards, we are kind of in the statistic phase right now, so we've quantified the data. But the tale [?] to get to that risk product, to manage risk with those statistics, is we actually have to understand when that stat means impact. And so that will take a lot of

study, a lot of back testing, quant analysis, and hopefully white paper and kind of buy-in from a larger community of the world. So what I see is some of this quantification alpha investment weather impact, that fizzles down, alpha fades away, and then what happens is we understand the impact through now years of modeling and understanding that impact. And now you have a risk product that it's Bloomberg for portfolio management and risk management. And now that's what we call beta product that has been democratized that many people use, every portfolio manager factors in, and now we're at a true fixed income or managing risk type of product. In my assessment and what I've seen we're nowhere close to that.

José Scheinkman (Columbia)
Harrison, I'll give you the last question.

Harrison Hong (Columbia)
?? questions ?? comment ??? about ??? judging the resilience of ??? extreme weather events. Can you elaborate a little bit more about how would you measure those ??? like what would go into this?

John Lucier (Capital Alpha Partners, LLC)
Well, we're still early days. And as I also said, we're not really seeing an effect yet in the shorter maturities, but longer term there may be an effect. People argue about it. But the point is that when you're issuing bond ratings you consider things like the ability to pay. Is this a general revenue bond, what is securing the bond? But you can also add criteria such as have they published a climate risk disclosure? What can they tell us about flood management systems? What other investments are they

making? In other words, towns that want to keep their financing costs as low as possible are going to be pretty proactive in making the necessary improvements but also publicizing them, and the bond rating agencies will try to collect that information in a systemic way. So at the moment I think that the effects are probably small, but still the value of mitigation is such that investors are going to be looking at both and taking both into account. And I'm pretty sure that in time we'll see that localities with a good risk mitigation program are able to offset some of the increased costs they would otherwise face.

Closing Remarks, José Scheinkman (Columbia)

Thank you. I think we were right on the time we're supposed to end. I'll give the mike to Harrison in a moment but I want to also tell that I was very impressed with this event, from our panelists, from our speakers, but I also found the quality of the questions and the discussion, and I see many of our students around here. This is an incredible area for you to do research in economics. I think at this point it's one of the most interesting areas. The fact that there is all this data around makes it something that you actually can do something that matters as opposed to just theorize, like mostly what I do. So I think it's something that it's tremendous and I want to thank Charles ??? for arranging this event and we're going to have others like this. There's going to be other data. I think it'll be hard to match the excitement of something like climate data, but I'm sure they'll be also very interesting. Thank you. Thanks, José, thanks for moderating a great panel.

Harrison Hong (Columbia)

So I think we're going to have a reception, so this is a great time for you guys to network and to find the speakers. We're going to hopefully post –

we should be posting all of the slides because they looked really terrific with lots of resources on them. So we'll post them on the same webpage for the PER webpage. And also there's going to be a transcript that we're going to also process of the panel as well. So there should be lots of material. Thanks for everyone for showing up, thanks for giving us the full house. So let's go and network and have some drinks. Thanks.
01:46:50