**XIA Le： Big Data and Media Sentiment in Economic and Financial Analysis**：

**Some Chinese cases**

**The following transcript is based on Dr. Xia’s live stream lecture on Xiaoetong from 19:00-21:00 pm, Nov 6, 2020.**

Transcript:

As introduced by Yuanyuan, now I'm leading a research department at PingAn Tech, and our research department is called Digital Economic Research Center. We are focusing on research of digital economy. But in fact, we try to combine this kind of advanced technologies like AI and big data with traditional research in Economics, in Finance, and even in some business areas. So, this is a quite interesting job. Before my current position, I worked for BBVA (Banco Bilbao Vizcaya Argentaria), a Spanish bank, for 10 years, and I was in charge of this micro research.

But in the past several years, I noticed that even in this micro research in traditional areas of economics, we can still feel the strong trend of advanced technology. We can see revolution or evolution because now people use these tools to look further. That's why today I propose that we can use several interesting small cases which are very simple. Through these cases, you can see that big data do help us look further. But as I said, they are very simple ones. That's why when talking about this, some people think there are a lot of mysteries inside, and they think “I have no relevant backgrounds, I cannot understand these sort of things”, but in fact, it’s very simple so that you can easily understand.

On the other hand, you need to know where demand is. Now if you talk to tech people, you’ll see that they know a lot of things and they try to create very good products. But unfortunately, sometimes this kind of things need to be demand-driven. As far as I know, many of audiences are students from School of Finance. Maybe for the moment you don't have any background of AI or big data, but don't be scared, because you know where demand is. According to my experience, you must know how to combine this kind of AI technologies and the real demand in finance, in economic areas, and then you can find very sensible, reasonable research, or you will have a very good product. That's my feeling of doing this kind of research in the past several years.

Today, I’ve listed some Chinese cases. In fact, big data is widely applied in other countries as well, and I will mention some of them later. But at least these research are from Chinese perspectives.

Another thing that I’d like to talk about before starting this presentation is the history of this kind of technology. In fact, AI has a long history, but the thing is, in the past 20 years, they have made some great breakthroughs. That's why we see a lot of new things related to AI and people get very excited about these AI things. But in fact, they are not new things. If you think about why we have such a big breakthrough, there are several important reasons.

The first one is internet. Now we have a lot of data, or we say information, free online, that's very important. By the way, I remember that when we talk about the recent AI revolution, in fact, they start from human language, that is, the machine starts to know how to listen to human language and how to interpret it. That's why one of my friends who used to be an AI expert but now is doing the hedge fund. Everyone is doing hedge fund because they can make a lot of money. He told me that in 1980s, they try to use some kind of AI things, these machines, to listen to people. But they don't have enough data, especially free data. So they have to buy this kind of data from some people who can make phone calls, especially for international students in the United States because they made international calls. But usually international students are poor, so they don't mind their privacy. After making the call, they can sell this kind of records to the AI people who will use their machine to train their algorithm to listen to the records of phone calls. That's very interesting, but also very expensive. But nowadays, we have a lot of data online, such as voice, video, audio data so that AI people can train their algorithms at a very low cost. So, that's one of the reasons why AI technology made a very big breakthrough in the past 20 years.

Another reason is that now we can see computer has a very strong capacity to calculate. Decades ago, for example, when you speak for a while, the machine did not have enough time to react because they're too slow. But now it’s totally different. People has made some great breakthroughs at methodology level as well. So all these are factors combined together. Now we have AI revolution. So, AI is everywhere, and at the same time, because of AI revolution, we are able to get more data. For example, when you worked in a Credit Card Center, you used to have a lot of customer data, because when you try to issue your credit card, you would send a survey, ask them to fill it, of course some people have made reservations, and you cannot get 100% sure of good data, but anyway, if they want to apply their credit card, they need to fill these forms including name, age, income, tax, and something like that.

That's what we called as the structural data. So with this data, we can do a lot of things. We can assess the credit worthiness of this person and then decide whether we are going to issue credit card to this guy or not. There are some technologies and they use different techniques. This is quite similar to this traditional econometric things that are quite useful. But even now, I think these econometric things have not been out of date yet. In many machine learning models, we have to use this one as the starting point. And in some cases, because it depends on the data, these traditional econometrics can be more efficient than the fancy machine learning technologies. Anyway, we had a lot of data before, but we called it as structured data because you can make them structural, such as age, income, tax, and their preference, and you can easily access this kind of data, process them, and to draw some conclusion you want.

But nowadays things changed with AI because, as I said, the machine now can capture a lot of information from people's talks in other ways. And now they're not only able to watch, but they are also able to listen, and then they got a lot of unstructured data. So, that's why we have big data, when we put them together, we can do a lot of things, such as what I'm going to talk about.

Now let's start my presentation. The topic is **Big Data and Media Sentiment in Economic and Financial Analysis：Some Chinese cases**. Because we all know that this media sentiment is a very important one. In China, you can use Weibo, WeChat to read people's sentiment and we are doing some research now.

Another important source is media sentiment. We have all these newspapers, and now we have many news media websites. People used to read newspapers, but now they don't need to print them out. That's why they have lots of media sentiment. So, how are we going to use this media segment? I'm going to talk about this later. Here, I’d like to give some heads-up. If you look at the financial markets globally, I think many of them are driven by people's media sentiment. Everyone in financial markets claim that “I know the fundamentals better than other people, that's why I can keep making money”. But unfortunately, if everyone claims that, it means no one really can beat the market. So if you look at the markets, it’s up and down, it will not be difficult to draw such a conclusion that they are driven by these sentiment, at least to a large extent.

But who create this sentiment? As I said, everyone has his or her sentiment. If you read WeChat or Weibo, and even Twitter outside China, you can read other people's sentiments. Some people even did research about how to read or investigate people’s sentiment via social media. Another way to have this sentiment is reading newspapers, because newspapers, with a lot of news, can lead the trend. They can shift people’s focus from one event to another. At the same time, you can also take media sentiment as a kind of mirror, which can mirror other people's thinking. That's why media sentiment is very important for economic and financial analysis. I'm going to give examples later.

As you can imagine, this media sentiment is unstructured data to some degree. So you cannot give a form to people and ask them to fill it every day. The only thing you can do is to collect the data by yourself. Of course machine can help you do that part, and then you need to clean the data to make them structured ones. Then you can use this kind of media data to do something. And also, we’ve read these articles, if you think about this sentiment, you can feel it. But how can we define whether it is a positive or negative one and to what extent they are positive or negative? We can even give different grades to them. I think many database about news media have already been doing these sort of things, they graded the tone of these articles. And then you can use their grading to do many things.

So now let’s move to the next slide, China vulnerability sentiment index. I think it is one project that I did four or five years ago, that's my first time to apply this kind of sentiment data to real things. It’s very interesting and also I will take these times when I introduce this case.

I'm going to take this chance to introduce a very powerful and useful database by Google, a global news sentiment database. In fact, almost all the cases I'm going to introduce today are related to these GDELT data, which are Google's new sentiment database. But I would start from this first case. Now let's move to page 3. I’ll show the big picture of these GDELT database. So what's GDELT? Now they already have two generations. The first generation is called GDELT One. And now they have GDELT V2. Basically you can imagine that this GDELT database collect all the news media. You can see the newspaper, magazines, books, also other publications, the things you can read around the world. That's very powerful because they're made by Google, and Google will know that they have a very powerful searching engine. So starting from this point, they made a lot of derivative products.

Basically, I’d like to simplify the structure of this one. In fact, we know that in every newspaper, we have a lot of articles. But in fact, how can you make these articles useful? As I said, it looks like unstructured data. You cannot directly use this article because there are many letters in the article, or in Chinese, you have many characters in the article and how you are going to use this one. So some people find a very interesting idea and just use the length of this article. They calculated the letters, the characters of this article to try to show something.

And also I think this idea is quite similar. We all know that Google have this index in China, Baidu have this index, which is called Search Index. They just check how many people are tracing or searching one related topic. For example, recently the one important example is US president election. So if you go to Baidu or Google, and you search this kind of keywords, it will pop up a lot of results of this one. And then Baidu and Google know you're searching and they make a calculation to see how many people are doing the similar things, and this can form index. But still the index has a limitation. I did another research, but today I'm not going to introduce it.

I did another research using this GDELT data to see whether they can forecast the stock market return. I checked the literature because when you do research, you cannot depend on yourself, you must check other people's work. And then based on the other people's work, you can make further efforts and progress. So when I check other people's work, I realized that some people have already tried to do it, but they didn't use GDELT data. Maybe they don't have access to it, or they missed it. But anyway, they use Baidu index, because that's a research of Chinese stock market. They use Baidu search engine index to see whether they can predict the stock market. I think this idea is very straightforward. We check how many people are searching the terms like “stock market”, and then we can see whether the stock market is attracting a lot of people's attention or not. When people are keeping eyes on the stock market, we can speculate that the stock market will go up.

But unfortunately, this research is very rigorous, and it didn't find any evidence of that correlation. So you cannot link this Baidu search index, for example the term can be “stock market”, with the real stock market performance. But why? Because as I said, some people may care about this one for good reasons while others for bad reasons. Even when you have a market crash one day, there are still many people checking the stock market. That's a very interesting phenomenon.

Another reason why they cannot find strong evidence is that this kind of information are too broad. If you just look at the number of the people who are searching for a certain term, it may not be enough. You must make a difference, to differentiate those kind of people who just try to search the term for fun because they noticed a stock crash recently, or who are really interested in this kind of topics like stock return and really want to invest their money in the market. That's a why we need to make granular matrix or approach to use this kind of information, but fortunately GDELT help us to achieve this. How? Because GDELT database collects all the articles around the world, in publications, on newspapers, websites, etc. But it not only contains articles, because articles are difficult to process. You must find a way to transform these articles to some structured data. We have a lot of experience to do that.

Also, if you're familiar with this database concept, you can understand that every article is a record in GDELT database. It means that there are a lot of variables for this record. For example, if you read an article predicting who's going to win in the United States president election, it must talk about Joe Biden and Donald trump. Now we have two people, and then in this record of GDELT, they were on the list of people in this article. It will list the name of Joe Biden, Donald trump, or all the people who made comments in the article, maybe vice president Pence or Nancy Pelosi, they made a statement in the article and then they are included in the database as well. So that is one variable it identifies who is the person appeared in the article. At the same time, who wrote this article? Or in which location it happened? They are all very straightforward. Apparently, I think this article is going to be written in the United States because it is a US present election, then we will have a location of this event. They have all this information, such as time, affiliation like Financial Times, Wall Street journal, People’s Daily who published this article and all the information are there.

Most importantly, it uses some algorithms to rank or to grade this article as positive or negative. But they do it in a very granular way. So it means that's not just positive or negative. It will show different grades of positive or negative. It may have “positive 5” as the highest one, and “negative five” as the lowest one. So, every article will have distribution. You can think about this, in every article, they have been transformed to theories of the variables, and these variables have all the information that we can make as structured one, so that we can use it at a later stage. That's the good part of GDELT database. As you can see, now it only includes codes, names, amounts, and even videos, the social media, they put all information together. It is thus a quite comprehensive database. As you can imagine, every article you read corresponds to one record in this database.

Another thing I’d like to talk about is the language issues solved by Google. In China, we speak Chinese and in Japan, people speak Japanese. How they are going to process different languages? The way they adopted is to translate all the other languages back to English, and then they make assessment based on the English version. I think that is a very good practice, but I will talk about this point later. I will show you some interesting results related to this kind of translation. People always think there will be a lot of things missed in translation. In fact, there must be some, but not very much. When we try to process this kind of language now, as I said, articles or letters, we cannot directly process even using Chinese. For example, if there is a Chinese article and we try to process it directly using Chinese, in fact, we have to do it word by word, right? That's why although some sophisticated nuance parts might disappear after translation, those important words are still there. It starts from the words.

Now let's move to the next stage. Here you can see different ways of how to rank or grade articles. Firstly, there must be emotions, such as angry, disappointed, happy, excited, etc. Until now, about 2300 emotions have been classified. And in fact, every article will lead some topics. For example, as I mentioned, if there is an article discussing about the US president election, the topics could be “president”, the “US politics”, “election”, etc., and the right topics must be found to classify this article. We cannot just put articles there because it’s impossible to process them like that. As they claimed, they have 30,000 topics in total. With the topics and emotions, we can grade or give a score to one emotion. For example, if you are happy, you’ll get one in positive. If you are super happy, maybe you can get three. But if you are sad, you’ll get one in negative, and if you are very sad, you’ll get two in negative and so on. We can use all the emotions because they are expressed through words and we can grade these words in different categories. By doing so, we can calculate how many negative words or positive words are in the article, because there are different negative words and positive words.

So, then we can calculate a weighted average of emotion for every article, which means we calculate how many “happy” appear in the article. If there is one for “very happy”, you’ll have five points. And then you can also check how many negative words are in the article. If there are three “very sad” and each is counted as two in negative, then it will be six in negative. Then we combine it together, it will be a negative one because we have five “happy” in positive and three “very sad” in negative. Then this article was graded to be negative one. This is the final point of this article, which reflects the general emotion of it. So, that's why we transform from this unstructured article to structured data, i.e., negative one. This is the grading and the tone of the article.

Of course we have other things to do. For example, we can calculate how many articles related to specific topics, just like the search engine index. We check how many people are searching for the stock market, and we can apply the similar thing to this article as well. We just calculate how many articles in China, for example, mentioned the US president election in a day. That provides another metrics for us to make a further exploration. So, it is a powerful and useful tool. Here, if you look at the slide, on the right side, you can see we have the economic debt tone, basically, the tone means that every article has a final point and that's the tone of the article. And then, we can combine all the articles with the same topic together and get an average tone of this topic. For example, here the economic debt tone is listed, you can see up and downs, which means people's tone on this economic debt issue. Next to this one, there is housing prices tone. At the bottom, there are economic debt and Cosco tone. So, you can combine all these topics together and you will find it. It is very interesting and powerful how we are using this kind of data.

As I said, you must be very careful when you use it, because some people may think that we have all the articles, and I met this sort of things before, so I can directly key in some words line by line. This is very hard work now, even harder a couple of years ago. But if you directly key in one word, you cannot get results because it’s not in the dictionary of the database. So, if you want to use this database, you need to know their dictionary. As I said, you need to know all the topics you can find in this database. Sometimes, it happens that the thing you think about doesn't appear in the dictionary of the database. That means even you key in this word, and even this word is widely used globally, you cannot find relevant records from this article. Because as I said, they are in the form of the database. If you want to check the database, you must know these are kinds of keywords and indexes included in the database. That's why we need to be careful when we use it.

Now let's move to the next slide. I think I have discussed quite a lot about the usefulness of using sentiment data. In fact, Pigou (1927) is the first person to talk about this. He thinks that there's a business cycle fluctuation driven by expectations and people's optimism and pessimism. I think that's a very good starting point for people. And then in the case in 1936, Keynes (1936) highlighted the importance of the changes of expectations are not necessarily driven by rational probabilistic calculations, but by “animal spirits”. It means that even about 100 years ago, people had already realized the importance of this kind of sentiment, or sometimes we can say “animal spirits”, which play a very important role in business cycle in the economy. But at that time, even being aware of it, they didn't know how to use it, how to get relevant data to measure this kind of sentiment. The only the thing that they know or they see is the ups and downs of stock market, but they have no idea about the sentiment behind it due to the lack of structured data to do that. And then Shiller(2017) showed that “narratives” can explain aggregate fluctuations through epidemic models.

With new data coming in, we have the empirical literature. You can see that Angeletos, Collard and Dellas (2015) find that sentiment shocks account for around 1/2 of GDP variance and 1/3 of the nominal interest rate variance at business-cycle frequencies. I think that's quite a big figure. The ups and downs of GDP are not that obvious in China before, but after this COVID-19 flaring up, we find the GDP fluctuation in China is very big, too. So, you can see the half of these fluctuations are coming from sentiment, rather than fundamental things or other factors. And 1/3 of the nominal interest rate is quite a big figure, not to say 1/3, even 10%, the variance of this nominal interest rate can make sure that you can make a lot of money, because all these financial indicators are very sensitive and useful. So, you can expect to make good money from that.

Then Barsky and Sims (2012) found that this informational component forms the main link between sentiment and future activity in international business cycles. And more recently, Shapiro, Sudhoff, and Wilson (2017) show how the news sentiment measures outperform the University of Michigan, they have this kind of Michigan survey, which can reflect the sentiment of the people, especially the investors. But if you look at news sentiment, maybe they already reflect this important sentiment. In fact, recently we see a lot of research works with this sentiment. And most of the results have the same implication that the sentiment drives many fluctuations in many important indicators, like GDP and interest rate, which are all important. But as I said, I did some research in this area. You can also find something related to this stock market performance. So, sentiment is a very important thing.

Let’s move on. I already talked about how to measure the sentiment. As I said, it starts with the words. First you can see they translate all the other languages to English, and then they process more. Then they do the Shallow Parsing, Deep Parsing, that means from this sentence, it will go to the level of the words to make sure these words are related to the previous one or the later one. I'm not going to talk too much about this one because they have a lot of tricks.

In fact, there's a very big literature or very big direction now because of the AI revolution, natural language processing. I have some colleagues who are doing that, and every day they try to make their own dictionary, as I said. And then they use their own dictionary to evaluate or to grade different articles. Because we all know that Chinese is a very difficult language for many foreigners, and in fact, even for Chinese, that's a difficult language, too. That's why it tries to interpret one article, as people always say in Chinese, you need to read between the lines because they have lot of implications. So, I’d like to say, if you want to do natural language process, it is much more challenging to do that in Chinese rather than in English. That's why Google translate all the languages back to English and then do the similar things. But they must take the risk of the loss of translation.

The last part is Sentiment Aggregation. As I said, there are positive one and negative one, and negative one minus positive one divided by total one, that's one sentiment. And more important thing is, you can grade very positive word and negative word, and then you have another different tone. So, if one person wrote an article, but he used very strong negative words, and then in the rest of the part, he used mild positive words. Then you can see the grading is different. It makes a lot of difference. So, you can just directly calculate the number of negative words and positive words, and make a calculation. Another way is to give every word a point or a score, and aggregate all these scores together, then you can get a different result for the same article. But anyway, the measurement of sentiments is like this, translate and then parse (shallow parse, deep parse), and the last part is to calculate. That's why at last we will see that one article corresponds with one emotion or one tone, the negative or positive tone, may be negative 3.5, because after aggregation, the final point may not be an integer. So, you can measure, or you can rank different articles by this metric.

But now we start to apply. Let's move to the next slide, “Tracking China Vulnerability in Real Time: Value Added through Big Data”. Now we try to apply the big data using GDELT database. From our perspective, we care about the financial market, we care about micro economy, so the way we are using this one should be sophisticated. Some people think if we have this sentiment measurement, why don't we just directly use it? Put a measurement to show. For example, we can choose a topic like economic debt, and we find the tone of the order related articles, we can find the index for that one. But the point is, if you directly use that one, it could be very volatile, because of the fluctuation of people's sentiment.

Also, most of them are newspapers and reports. But we know that these journalists or the financial reporters always try to show the importance of their news (I'm not exaggerating). For example, if you look at currency, I remember that when Chinese Yuan has depreciation, all these financial reporters would focus on this point. For example, every day the RMB depreciates by maybe five basic points, very small one, they would report like RMB has already reached the historic low. But that's only 5 points, not a big deal. And on the other day, RMB appreciates quite a lot, maybe 10 points, and then they said, okay, RMB appreciates just point one or less than 1%. That’s very tricky. It means the first day it depreciated by five basis points. On the second day, it appreciated by 10 basis points. If directly compared to this figures, apparently, the second day is much more important than the first day. But if the journalists have a very strong view that for the short term, I must write something bad about are the RMB exchange rate to get a lot of attention, they will use a different way to report this. They would say the first day it hits the historic low, and on the second day, they would say, okay, it just rebounds mildly, maybe less than 1 percent of this one, because they use a different matrix, they play the games.

But that could be a very relevant thing if we want to use this kind of big data information in our research, that's why we cannot just simply rely on the sentiment data and index to reflect the real economy. So, if we want to use it, we need to incorporate more data. That means we have hard data. What's the hard data? It means the statistical data. We all know that the statistical bureaus, that is, authorities of every country, they publish a lot of data. I think this kind of data should combine with sentiment data to give the big picture to the economy. And we also have a lot of financial market data. We have exchange rate, interest rate, we have stock market, and ups and downs of different stocks every day. If we can manage to combine all this data together, I think that can reflect the real situation of the economy. In the first place the reason why I try this kind of project is because it's related to this currency depreciation. We all know that on August 11, 2015, China had a devaluation or depreciation of the RMB, which made the financial market not that stable. Although the financial market made very strong reaction, the Chinese GDP data still showed a very stable situation. In that case, we worried about that the headline figures, like GDP inflation in China at that time, cannot reflect the real situation in China, that's why we try to find the vulnerability sentiment index. We try to combine the hard data, the financial market data and sentiment data together to find a new index of Chinese economy. That's a motivation of this project.

Then, let's move to the next slide. Here I’ll show you a very big table, which shows the secrecy of this model. In fact, this index is Chinese Vulnerability Sentiment Index. It has four sub-indexes, including SOE Vulnerability Index (I don't know why but my colleagues called it SOE Index, but I think we should call it Debt Index. The only reason why we call it SOE is at that time, Chinese SOE had the debt problem relatively, that's a reason why we have a such a name. The second one is very important, Housing Index. Because even now many people are talking about the Chinese housing price, they think it has a serious bubble, or some people think it'll have a serious bubble. But anyway, that means housing market is a very important part of China's economy. If housing market has any problem, I think that through different ways, it will have strong spin over to other sectors of the real economy.

The third one is Shadow Banking Index. I think recently there's still some people talking about shadow banking, maybe I shouldn't say that, but we all know the Ant Financial problem, I think that means the supervision in China, they think some practices of Ant Financial company have very similar features with shadow banking. But in China, a few years ago, shadow banking problem is very acute, I’d like to say many financial institutions are using their off-balance-sheet business or channels to fund the shadow banking. I can show you some figures later. But I think starting from 2017, Chinese authorities started to clamp down the shadow banking sector. That's why recently we didn't hear a lot of shadow banking argument until this financial thing happened.

The last part is FX speculative pressure index. In fact, the exchange rate currency is a reason why we have this index, because at that time, people worried about currency depreciation quite a lot. And then the first part is we find this sub-index and then we put these four sub-indexes together to form the final index. This final index is a quite comprehensive one, just like GDP is one indicator, which is published every quarter and reflects the status of the economy, that's why we tend to have the final one to match the GDP and other important indicators of the economy. But in fact, these sub-indexes are very interesting as well. So, I would spend some time to introduce these sub-indexes.

But from this table, you can see for example, SOE index. They have 25% and two variables and there are hard data and financial data, as you can see, that means the total profits and liabilities account for about 20% of the final sub-index of SOE, and the rest of 75% are sentiment data. So, here we can see a lot of words, such as state owned enterprises, resources allocation, failure, resources allocation, SOE, these are in the dictionary of GDELT database. So, basically we just use this kind of words to do some queries, and then we find all the tones related to this kind of words. Every word is a topic. We find that for every topic there are average tones and we formed a time series for every topic. And then, we have several time series. We have eight time-series. They jointly account for 75% of the final sub index, and then we have some hard data. So, the similar thing applied to other sub-indexes. They have different weights, for example, for the housing bubble index, we have 5 hard data index like mortgage loans, housing index, housing price, new constructions, real estate, and something like that. They jointly account for 45%, and for the rest, they have 55%. The sentiment indicators and all these things apply to the same one. So, let's move to the next part.

Here, the next slide shows the result for the importance of the sentiment. In fact, we do two-step estimations. The first one, as I said, for every sub-index SOE, HB (housing bubble), SB (shadow banking) and FX. Then they aggregate up all the relevant indicators I showed in previous slides and get 4 sub-indexes. In the second step, we use combination, almost 25% for each indicator. As you can see on the right side, the SOE vulnerability index account for around 30%, but FX speculative index accounts for 22%. So, we use this one to form the final one. So, let me show you the final one and I’ll explain.

Here I show the Chinese Vulnerability Sentiment Index. Because they can differentiate the English one and Chinese one, some articles are written in Chinese while others in English. So, many people always think that the foreign media always report some bad things, negative things about China. But the news media in China always report the good part of the economy. So, we did such a calculation and tried to find whether it helps or not. But we found a very interesting result that there's no systematic bias in English and Chinese language media. It's quite to many people's surprise because all the topics are economic related topics, that's why we didn't get a lot of buyers from foreign or from domestic news media. But anyway, let's move to the next slide.

Here. This slide I showed this CSVI. I didn't update this one to today’s lecture due to some copyright problem. But anyway, I think I can show this one. You can see usually this final one includes 4 sub-indexes, but this is the final one. The positive one means that they have a better situation of the economy. The negative one means not that good situation. Now we find some turning points. For example, they start from 19 June, 2015, you can see they have very high level one, and then they dropped down quite a lot. What happened? If you go back to history, we’ll find about five years ago, there was a very serious stock market crash in China.

The second point I highlighted here is in August of the 2015. As I said at that time, the Chinese authorities decided to depreciate, or devalue its currency, 3% a day, they said they were doing this market liberalization things. But apparently, the market didn't buy the story, that's why we can see further job over this vulnerability index until Black Monday, they had a lot of problems. And then you can see the up and downs, when good news happened, there would be an upward trend, and when bad things happened, it will go down. If you look at the events I highlighted here, China was downgraded in 2017. The sovereign rating was downgraded by rating agencies. Then you can see the index dropped quite a lot because they include other hard data apart from the sentiment data.

Let’s move on. Then we had the rising trade war threat which stopped at June, 2019. At that time, the trade dispute escalation had been very bad, but the good thing is that if you look at the fluctuation, compared to 2015, the change of this index wasn't that dramatic during the period of the trade war. So, to some degree, I think it shows the Chinese economy is resilient to this kind of external shock. But for domestic shock, like what happened in 2015, maybe that's comparatively more serious.

Several slides have been skipped just now, so let's go down to the second part, SOE index. I’ll talk about this sub-index, but I think we don't need to spend too much time on this, and I can save some time to talk about more interesting topics. The SOE index, in fact, as I said, is an index of debt which stopped at June, 2019. You can see up and downs during the period of 2018, and there were a lot of downs. Why? Because at that time, China was implementing financial deleverage, which means all the enterprises with high debt were living a hard time. That's why we had such a negative one. The condition was improved a lot in 2019, but still negative due to the trade war problem. And this trade war problem will disproportionately have impacts on prices with a high debt level.

In the following several pages, I'll show you other data from other perspectives, the real data to show the situation of that one, but it's not that important to this lecture. Then let’s move on further to the third part, the Housing Bubble Index. In this index, you can see housing bubble up and downs in 2018. There was some very good time because people were worrying about the housing bubble, at that time they implemented restrictive policies on the housing market. You may remember the Chinese idea of “房住不炒” was raised, which means the houses are used for living but not for speculating. That's a national policy, so that's why we had some positive feedbacks during that period. But in 2019, we can see some price fluctuations in big cities. I'm not going to go to the detail of this one, but it's quite in line with the real economy and people's feeling.

The fourth part is the Shadow Banking Index. Here you can see during the period of the 2016, this index had a very bad performance. But quite recently, I mean until 2019, you can see they all in the positive territory, which means after the government started to clamp down the shadow banking activities, things were improved quite a lot. That's very good, but we don't know. For the financial regulations, people always play different kinds of games. Even there is a new regulation, financial institutions can always find ways to do arbitrage. That means we need to be careful about this for longer term.

Let's go to the 5th part, the FX Index. As you can see here, because I stopped at June, 2019, during the period of August, 2015, they had the famous devaluation of RMB. During that time, they had a very negative index, and then it shows ups and downs. During the period of the 2018 and 2019, China had the trade war problem. So, usually we can think that if you have trade disputes with very big US trade partner, you will have this kind of problem. There was a lot of pressure on currency. But the good thing is, if you look at this index, the fluctuation during 2018 and 2019 was not that dramatic or not that big compared to 2015. So, at that time, when we look at these figures, we can draw strong conclusion that RMB will depreciate, but it will not be out of control. This is one important conclusion we drew from that index. So, at that time, we don't think the currency value of RMB will have a very big or dramatic depreciation. Instead, the depression will happen gradually over the period of the two years. That's a strong conclusion which helped us maintain a relatively stable portfolio. I do think it is a useful tool. If you think they will be a very sharp depression again, then you need to make some adjustments, such as selling the currency in your hand, or making more hedging. If you want to do this, you need to pay more money because hedging is costly. But if you think this kind of depreciation will be mildly and gradually, then you can have more levers to adjust your portfolio. So, there are quite different actions you can do. That's why I said big data help us to make decisions in investment and risk management. In all these areas, it will be a useful new tool for us. So, that's why I show this FX Index to you.

Let's move on to the sixth part and look at more cases. here, in the first case, we used the GDELT database and this one is simple. Because China has a lot of outward investments in United States, Australia and many other countries, we try to know what kind of views the local people and local newspapers have toward Chinese investment in their own country. So, we made this one, used this similar GDELT database, and in fact, we don't think the absolute level will matter. The most important thing is change. For example, in 2015, everyone loves you, but then in two years, 2017, people started to change their minds. So, changes are more important, especially when you see some dramatic changes, significant changes, that could matter. We made this kind of change, and then in every country we found the local news medium views of Chinese outward investments. Then we calculate the change, one interesting thing we found in this graph is that in Australia, there are views of China is almost in red, which means they have strong negative changes, green ones stand for positive changes. If you look at the map at that time, there were green ones in Latin America, North America, and most of the western Europe. That means during that period, people in these countries became more welcome to Chinese investments. But in other countries like Japan and Australia, there were negative views against the Chinese investments. In Africa, there were different colors, some countries showed a welcome attitude while other countries had reserved attitude toward Chinese investments in their countries.

Clearly, it helped us follow what happened around the world, what people’s views toward China. Some of the audience may already have this kind of thinking, and I think they can apply this to other things, especially after the US president election, what is the dynamic of Chinese relations with foreign countries because of the COVID-19, and some countries may have worse relationship with China recently. That would be a very interesting project.

Now let's move to the next slide. Here we can see a comparison, which is another interesting story. The story happened in 2018, Chinese Present Xi Jinping visited Spain, and in March of 2019, he visited Italy. There are about 20 European countries, which are quite similar to some degree. If you look at the performance of the COVID-19 this year, they are very similar, too, they hit very hard at the beginning of this year. But the thing is in November of 2018 when President Xi visited Spain, from the local news mediums’ attitudes towards Chinese investments, we can see during that period, there's no significant rise of these news because here we showed the times of the theme “investment” mentioned in their newspapers. And we also showed average tones toward Chinese investments. So, we see during that time, there's no significant change. Even President Xi visited Spain. Why? Because they didn't sign the MOU of One Belt, One Road with China. At the beginning, President Xi wanted to sign it with Spain, but Spain got some pressures from the European Union and at last, they didn't sign it. But in the following year, Italy decided to sign it in March when Present Xi Jinping visited Italy. You can see on the right side of the chart, during that period of President Xi’s visit, the tones, which means the red lines have risen quite a lot. And at the same time, the coverage of this topic in the bar chart rose quite a lot as well. Compare the charts of two countries, we can see how the One Belt One Road MOU made a difference in their news media, that's another interesting application.

Now I think the time is almost up, let's go to the last one, the sentiment change of client portfolio. Here we use heat map which is often used to reflect risk or to monitor something. This heat map shows the sentiment change of important clients. Every financial institution has a very big client portfolio, and inside the portfolio they may have several important big corporates. But how are you going to monitor it? One way is that if something happened, you can send the information to your manager or someone who are managing risks, and you follow the news of this corporate every day. That's one way, but they must make sure that they're not going to miss the important news of these corporates. So, I think that's a very stressful work because you need to follow the company every day.

Another way is to use this news medium to follow companies because we can know the news media tone against one corporate. So, we make such a heat map. If dark blue happened, it means the news media tone toward this corporate has a very dramatic deterioration. In that case, we can directly Google it or use other things to find what happened to this corporate, which makes your monitor smarter. For example, in this case, we have more than 40 corporates, and if you use the first method, each person covers five to ten corporates, then you need four to eight people working on that task, but with this chart, you just download the data and feed in data into that one, one person is enough. One person can monitor more than 40 corporates. I'm not saying this AI things but using big data can save you a lot of time. That's my idea about how we can better use this kind of big data to our advantage, you can do this job more efficiently. I think this is a good idea, so, we have been using this one for quite a long time. If you are interested in this kind of topics, you can make more exploration.

I think now time is up and we can stop here. Last but not least, I like to make some introduction of my school because I know some of our audience are not from School of Finance in Renmin University of China. Our School of Finance has made a lot of efforts to improve the quality of their education products and provide them to young people. It’s always on the front frontiers of this job. When they contacted me to start this lecture, I was told that they are going to launch a new program in this form for the master program of FinTech. I think this is a very good one, because according to my work experience, it's a trend, AI and big data will be gradually applied in many areas. Today I showed some cases, but the application scenarios are much richer than what I showed to you. And it also requires a lot of the young people to join these areas. If you look at the samples I showed, you can see it's not that difficult. Right? It's not complex at all. But with them, we can draw some insightful conclusions to directly help us not to lose too much money, just like the FX Index. Some of them can make our workflow more efficient. For example, the portfolio monitor heat map things. So, I think that young people should know this trend, and they are welcome to join this area, or they can start from this program at Renmin University.