Project: Estimation of Intraday Volatility based on Big Data Analytics

Abstract:

"Communities of active investors and day traders who are sharing opinions and in some case sophisticated research about stocks, bonds and other financial instruments will actually have the power to move share prices […] making Twitter-based input as important as any other data to the stock.", TIME, 2009.

The recent digital revolution reformed our societies' economic and financial models as well as our access to information. The availability and the different types of information broaden exponentially. Nevertheless, the main barrier to build up financial applications from this massive amount of unstructured information is the combination of sophisticated technological architecture, adapted analytics to those new set of data and financial experiences.

QuantCube Technology enables to overcome this barrier and specializes in Global Big Data analytics for more than two years. QuantCube Technology has been the first company in the world to launch a Big Data Financial index in December 2013 that currently overperforms the S&P 500 index significantly.

Whereas most of financial institutions struggle to improve their predictive models with traditional market data such as prices, volumes and fundamental reports, we take into account complementary sources of information aiming at generalizing these models. Indeed, QuantCube Technology covers a wide range of data sources including social and professional networks, online retailers, blogs, auction markets, meteorological data, etc.

One of our expertises consists in aggregating and analyzing data coming from social media, professional networks and blogs to gauge the market sentiment in real-time such that we become a reference in this field.

In order to keep our leadership in Big Data analytics for financial markets, we will develop a new generation of volatility estimators that take account the micro-impacts of incoming news coming from new sources of information, including different types of non-structured data.

Identification of project

Name and contact information of head scientist:
Thanh-Long HUYNH, CEO at QuantCube Technology

Project description

Project Outline:

1. Identifying relevant sources of information
2. Structuring and Automating the real-time Data-Mining Process
3. Treating and Standardizing the data extracted (Noise/Spam Removal)
4. Setting up NLP algorithms and Sentiment Analysis to give the data extracted a sense
5. Aggregating all this information to model a volatility estimator
6. Calibrating the model
7. Backtesting the model

Methods:

Thanks to a multicultural data scientist team - 13 languages covered -, we are actively involved on the main markets, ranging from the US to the Chinese market.
We have already developed individual and global real-time sentiment indexes on these stock markets by structuring, aggregating and analyzing the large amount of information available among social media and local blogs.

By building these sentiments using different types of NLPs, we learned one common thing:
Information credibility is asymmetric.
Indeed, negative news and reviews proved to be way more reliable than positive ones.

Taking this into consideration, creating a very short-term volatility estimator based on incoming information makes sense, especially as stock markets react sharply to surprising news, especially to negative news.

Volatility estimators generally depend on market data such as prices and volumes. But prices and volumes only reflect how market actors react to information they receive.

What about estimating volatility directly by quantifying and giving a sense to information spreading?

From our point of view, we have already built systems and algorithms to systematically capture, quantify and analyze a wide range of information flowing in real-time using NLP processes.
Therefore, it seems now relevant to build volatility estimators based on the real-time analytics of this raw information and not on prices as it is usually done. It also requests some mathematics research to build up those volatility estimators.

The objective is to create a more accurate and a faster volatility estimator than traditional ones. Financial applications are tremendous, enclosed are some examples:

- A smarter adjustment of the bid-ask spreads for brokers
- A better risk management
- A more accurate pricing and valuation of derivatives for traders

This project gathers at the same time Statistical and Mathematical researchers, Market-Making specialists and Big Data analytics professionals.

The combination of these three complementary expertises ensures that:

1. The data are the relevant
2. The real-time data-mining process is well-structured
3. Modelization of news diffusion process
4. Calibration of the models with market data
5. Application to bid-ask spread adjustments on the ETF market