Research Statement

Ze Shen

My research interest focuses on financial price analysis and risk management. My Ph.D. dissertation mainly studies the cryptocurrency market price and fluctuations, from cryptocurrency price formation to its price return volatility forecasting and risk management. I am also interested in the application of machine learning methods to financial and economic issues. For the next three to five years, I plan to: 1. continue the cryptocurrency market analysis to draw a new market efficient frontier; 2. study other financial market sectors, for example, commodity markets, by applying both economic and machine learning techniques.

Cryptocurrency has made a big impact on financial market in last decade. It is an emerging financial market with market capitalization reached over 768 billion USD at the beginning of 2018. The cryptocurrency market is popular among investors, with over 1,600 types of cryptocurrencies being traded as of March, 2018. There is an interesting puzzle associated with cryptocurrencies: unlike gold or other financial assets, cryptocurrency started from something that has no intrinsic value but derives its value from data on people's disks. Thus, how can it have value and how can its price increase over time? In other words, what is cryptocurrency price formation? Some early studies have identified that the cryptocurrency price has four determinants: economic driver, popularity among investors and macroeconomic and financial indicators. However, all previous studies on cryptocurrencies price formation considered factors, outside the cryptocurrency market. They did not take the competition within the market into consideration.

There are hundreds of different types of cryptocurrencies in the market, and the number is growing. The remarkable growth in the number of cryptocurrencies is a signal that the cryptocurrency market is attracting more investors and more capital. Intuitively, the competition within the market may have an impact on the cryptocurrencies prices formation. For example, bitcoin is the first and most popular cryptocurrency, which has been dominating the market from the beginning. However, its market share (the percentage of the cryptocurrency market capitalization that is attributed to bitcoin) was 94% in April, 2013, but is now only 33% as of January 2018. A number of altcoins (refers to the alternative cryptocurrencies that are not bitcoin) entered the market to compete with the market leader bitcoin. The altcoins share the same features with bitcoin, but some of them are even more advanced - such as Litecoin, which provides faster transaction confirmation times than bitcoin. There are important questions to consider. Will cryptocurrencies prices affect each other? Does competition within the cryptocurrency market matters for the price formation?

In answering these question, my first essay "How Cryptocurrencies Prices Affect Each Other" studies the cryptocurrencies prices formation by analyzing the relationships between cryptocurrencies prices rather than investigating the four determinants which were considered in previous studies. I selected four cryptocurrencies, bitcoin, Etheruem, XRP and Litecoin. These four cryptocurrencies were leading the market with large market capitalization, trading volume and long price history, ranges from August 2015 to August 2018. Vector autoregression model (VAR) is a good methodology to use because VAR is one of the most successful models for multivariate time series. VAR applies to the situation that all the time series variables are stationary without cointegrating relationships. However, we found the existence of cointegration between the four variables. Therefore, we applied another model, vector error correction model (VECM), which adds error correction terms to VAR model and allows us to analyze both long run equilibrium and short run adjustment of the variables.

The empirical results show that cryptocurrency's price affects each other. Its price is not only influenced by its past value, but other cryptocurrencies' prices as well. I also test for the causal relationship between the four variables by using Granger causality test and directed information theory. The causality tests tell the same story as VECM. Then I estimated the impulse response function (IRF) based on VECM to tract how each variable response to the impulses of system shocks. We found a shock in bitcoin has serious impact on other cryptocurrencies prices, and the impact lasts a long time.

The results of VECM and causality tests provide evidence that competition in the cryptocurrency market matters. Investors need to consider it when making investment decisions.

As financial asset, the most well-known feature of cryptocurrencies is their extreme fluctuations. I consider the bitcoin cryptocurrency market first because it dominants the market. One can see the violent fluctuations in bitcoin historical prices from 2009 to now. The most important thing that bitcoin holders are concerned about is how much risk exposure they have? Thus, the modeling and forecasting of bitcoin volatility is crucial for bitcoin investors' decision making analysis and risk management. Early studies of bitcoin volatility were founded on economic models, mainly using GARCH models. However, research on bitcoin volatility forecasting using machine learning algorithms is still void. Machine learning is an advanced and promising approach that may have a big impact and make future contributions in financial markets. Unlike economic models, where researcher picks a specific model based on economic principles and estimates the parameters, a machine learning algorithm is usually a data driven modeling focused on the selection process. Thus, a model of a machine learning algorithm is not fixed or predetermined but will be refined during a training process. Applying machine learning methods to solve for economic issues can potentially make a difference in the economic and financial field. It seems that machine learning methods are more advanced and more efficient than economic models. However, I plan to determine if this is true in bitcoin volatility forecasting.

The second essay of my dissertation, which is also my job market paper, uses conventional economic models-simple moving average and GARCH model-and machine learning model-recurrent neural network model- to forecast the volatility of bitcoin return, and calculate the forecasted value at risk of bitcoin return. We compare the out-of-sample performance of the three models by using RMSE and MAE. We also plot the figures that compare the forecasted volatility and the realized one with three different forecasting horizons: one day, five day and ten day ahead. We found the recurrent neural network model does better in capturing the bitcoin return volatility trends and clustering than GARCH model and simple moving average model. However, after calculating value at risk and comparing the out of sample coverage, we found recurrent neural network model performs poorly in value at risk forecasting, even worse than the benchmark simple moving average model. This results demonstrate recurrent neural network model outperforms the economic GARCH model in terms of statistical forecasting accuracy, but is less efficient in risk management in the framework of value at risk. People widely believed that machine learning method is more advanced than economic models in financial and economic field, but our finding shows something different. Machine learning method has an advantage in preserving more temporal information of a time series during training, however, it doesn't involve any economic intuition nor market information, which is provided in economic models.

Our study proposed an alternative way of volatility analysis. It illuminates the feasibility and potential to apply machine learning methods to economic time series forecasting. If the bitcoin market investors need information on how volatile the market would be in the future and the data amount is large enough, recurrent neural network is a good method as it captures the trend and volatility clustering better than other models; however, if additional market information is available and investors are interested in bitcoin market risk management, the economic models can be more efficient. This study also gives me an idea of applying recurrent neural network model to forecast commodity prices and fluctuations.

Let's go back to the question that cryptocurrency holders are concerned about: how much risk exposure they have? How much money will they potentially lose tomorrow? The most commonly used measurement for risk exposure is value at risk and its related measure, conditional value at risk, which is also known as expected shortfall. Value at risk, is essentially a quantile of the asset return distribution over a given period of time. Then another concern with risk measurement was raised that what is the cryptocurrency conditional return distribution, and whether different assumptions provide different risk information. Thus the main challenge for calculating value at risk and conditional value at risk is to estimate the asset return distribution. It is observed that the time series of cryptocurrency return is serially correlated and presents volatility clustering, which implies the existence of conditional heteroscedasticity in volatility. The GARCH model is the most common model used for conditional volatility modelling. Then the task of asset return distribution estimation can be transferred to the residuals distribution estimation constructed from GARCH model.

There are three categories of methods to estimate the residuals distribution, nonparametric methods, parametric methods and semiparametric method. The most popular nonparametric method is historical simulation, which is a model free method and assumes that tomorrow's asset return distribution can be well estimated by the empirical distribution of past asset returns. The parametric methods have a specific assumption of the residual distribution, normal distribution or non-normal distribution, such as my second essay, which assumes bitcoin return follows student t distribution. The semiparametric method is to estimate the residual distribution by using extreme value theory. My research goal is to estimate the distribution of residuals constructed from GARCH model with three approaches, filtered historical simulation method, standard t distribution (benchmark) and extreme value theory. We use cryptocurrency index (CRIX) proposed by Simon and Wolfgang (2018) to see which approach yields better cryptocurrency risk measurement estimation by calculating value at risk and conditional value at risk.

In reference to risk management, think of the classic portfolio theory, which tells investors to avoid asset specific risk by holding different assets. Financial market investors will want to know if they can reduce risk exposure by diversifying their holdings that includes cryptocurrencies. Early studies have found that there is no significant correlation between bitcoin and other financial assets, including S&P 500, gold, US bonds and oil. Some researchers allocated bitcoin and some other cryptocurrencies into investment portfolio to increase portfolio effectiveness. My next research is to involve cryptocurrency index – CRIX into investment portfolio, with traditional asset classes, to establish a new efficient frontier by using modern portfolio theory.

Besides cryptocurrency market analysis, I am also interested in stock market. Previous studies show that the decision makers are both risk averse and ambiguity averse when they make investment decisions. Ambiguity aversion refers a preference for known risks instead of unknown risks. The finding of ambiguity aversion is important since it is helpful to explain the equity premium puzzle. Many studies have addressed risk aversion, but the ambiguity aversion is still a new area. Jeong, Kim and Park (2014) took many popular asset pricing models into consideration and developed a fundamental asset pricing equation of multiple-priors recursive utility model to estimate ambiguity aversion parameter. They used Standard and Poor’s (S&P) 500 index to calculate the market return and found the ambiguity aversion is economically and statistically significant. My idea is to follow their work, and apply their model to different stock market sectors rather than the whole market. My goal is to examine whether the ambiguity aversion is significant in specific stock market sectors and try to figure out how the risk aversion, the elasticity of intertemporal substitution and ambiguity aversion differs between sectors. I will start with energy sector and consumer staples sector.