MATH 640 Project Proposal

Jeff Gould & Sergio Leon

Background

Bitcoin was launched in 2009 as a method to transfer value via a peer-to-peer decentralized network, essentially by-passing traditional financial intermediaries such as banks. Since its inception Bitcoin has grown into the most popular cryptocurrency, reaching a current market capitalization of over $1 trillion USD, which is higher than all but four companies on the NYSE. Although this growth has been rapid, the cryptocurrency market is still small compared to more traditional markets such as the global stock market. For that reason and several others, bitcoin prices tend to be extremely volatile. Bitcoin values in late 2017 rose to nearly $20,000 only to crash down to $3,000 a year later. Even thousand-dollar hourly swings in price are common.

For the final project we want to predict the volatility of bitcoin over a subsequent 24-hour period by using historical price volatility. Traditionally, more common econometric modeling techniques such as generalized autoregressive conditional heteroskedasticity (GARCH) models have been used to predict conditional volatility in financial assets. In the following section we will outline at a high level, a Bayesian approach to forecasting the dynamic nature of bitcoin price volatility.

Methods

Our data sample is the 5-minute intraday OHLCV-Candles for Bitcoin trading on the Coinbase exchange, dating from January 1, 2016-April 14, 2021. With these candles we calculate the 24-hour intraday Realized Volatility (RV), where $RV = \sqrt{\sum_{t=1}^{288} r_t^2}$, where $r_t = log(p_t) - log(p_{t-1})$ denotes the return for time period t, and 288 comes from 288 5-minute candles in a day. Our goal is to use a Gibbs Sampler or MH-Algorithm to estimate the parameters for 2-3 Stochastic Volatility models on the data sample ranging from 2016-2019, and then test the results on the data sample ranging from 2020-2021 for applicability. The general form of a stochastic volatilty model is $d\nu_t = \alpha_{\nu,t} dt + \beta_{\nu,t} dB_t$, where dB_t is standard Brownian Motion.

An example of a specifc model that we may use is the Heston Model, which parametrizes the differential equation above as the following:

$$d\nu_t = \theta(\omega - \nu_t)dt + \xi\sqrt{\nu_t}dB_t$$

In this example, we would build a Gibbs Sampler to draw θ, the rate at which volatility tends back to the mean, ω, the long run mean for RV, and ξ, the long run variance of volatility (sometimes called "Vol of Vol"). After we have sampled these parameters using our train data, we will then test our results on the out of sample data from January 2020-April 2021. This sample also includes the highly volatile trading days of March 2020, when all markets were highly volatile due to global uncertainty amid the Covid-19 outbreak. This makes it a good test sample to see how well the models perform under times of extreme outliers. Our goal is to have a model with accurate parameters that not just accurately predict the point estimate, but also capture the tail distribution of Realized Volatility.